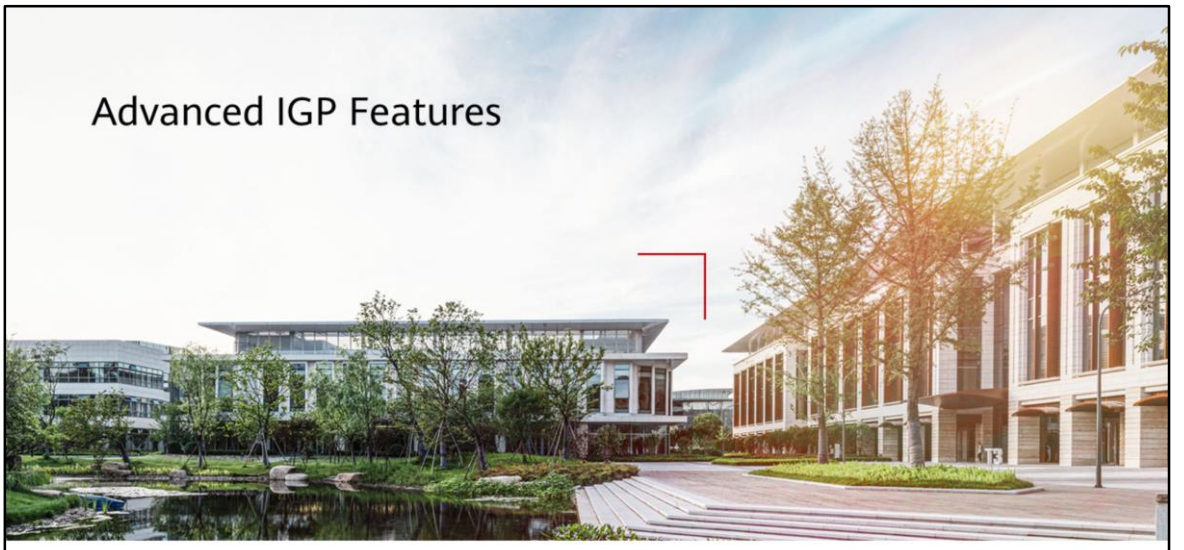
Advanced IGP Features



HUAWEI

Foreword

- Both Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS) are link-state-based Interior Gateway Protocols (IGPs). Routers that run them synchronize link state databases (LSDBs) and use the shortest path first (SPF) algorithm to calculate optimal routes.
- In response to network topology changes, OSPF and IS-IS support multiple fast convergence and protection mechanisms, which minimize traffic loss caused by network faults.
- To control the size of a routing table, OSPF and IS-IS support route selection and routing information control.
- This course describes the advanced features of OSPF and IS-IS, including fast convergence, routing control, and other features.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe various fast convergence techniques of OSPF and IS-IS.
  - Perform configurations for OSPF and IS-IS equal-cost routes.
  - Configure OSPF and IS-IS to advertise default routes.
  - Describe the application scenarios of OSPF and IS-IS multi-process.
  - Describe graceful restart (GR) and non-stop routing (NSR) fundamentals of OSPF and IS-IS.
  - Describe the application scenarios of OSPF forwarding addresses (FAs).
  - Describe the fundamentals of IS-IS LSP fragment extension.

HUAWEI

Contents

1. **OSPF Fast Convergence**
2. OSPF Route Control
3. Other OSPF Features
4. Advanced IS-IS Features

HUAWEI

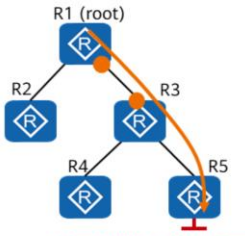
Overview of OSPF Fast Convergence

- OSPF fast convergence is an extended feature of OSPF to speed up route convergence. It features partial route calculation (PRC) and the intelligent timer.
- In addition, OSPF supports fast convergence upon fault rectification. For example, OSPF IP fast reroute (FRR) can be used to implement fast traffic switchover to a backup link, and OSPF can also be associated with BFD to implement fast fault detection.

HUAWEI

PRC

- PRC only calculates routes that have been changed on a network.
- PRC does not calculate nodes. Instead, it updates routes based on the shortest path tree (SPT) calculated using the SPF algorithm.



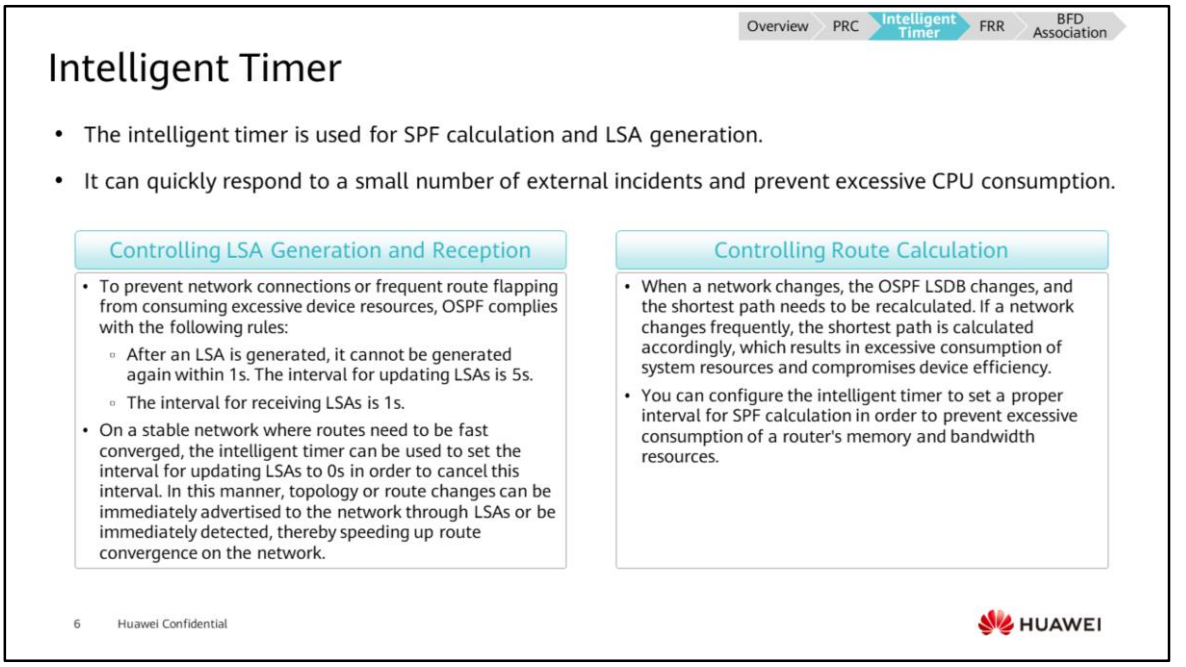
Loopback0: A directly connected network segment is added.

In route calculation, a node represents a router, and a leaf represents a route. PRC processes only the changed leaf information.

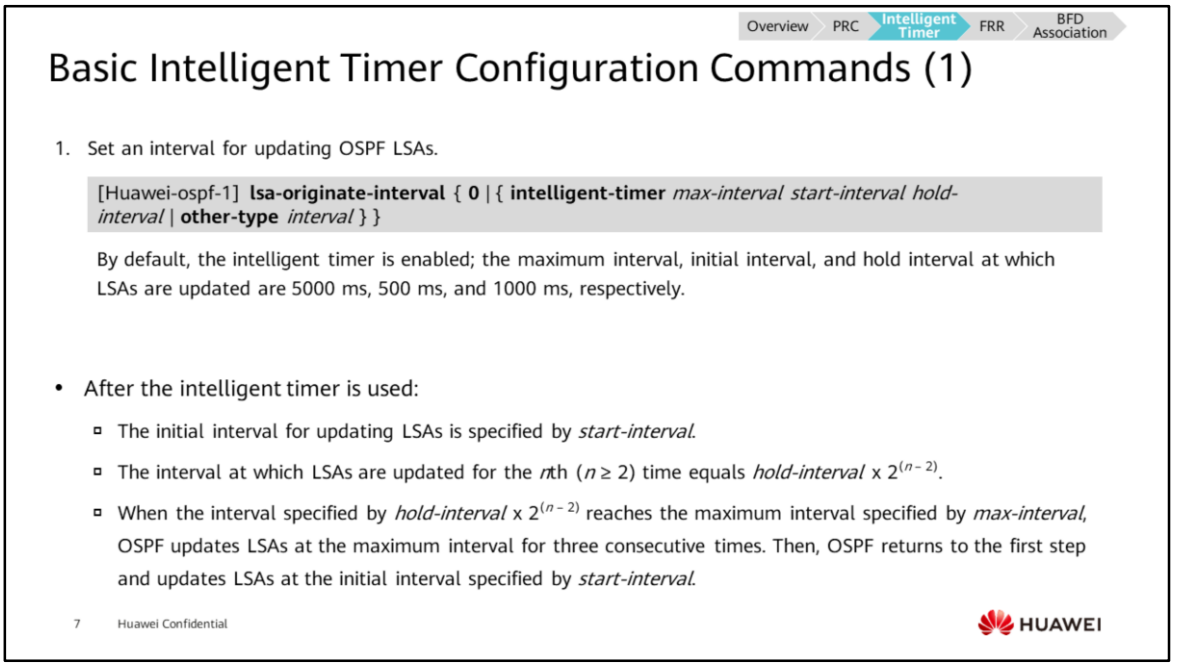
- Scenario description:
  - OSPF runs on a network. The figure shows the SPT with R1 as the root after network convergence. When R1 accesses R5, traffic is sent to the destination based on [outbound interface of R1's downlink, IP address of R3's uplink interface].
  - OSPF is enabled on Loopback0 of R5. This means that a new network segment is added to the OSPF network.
- PRC:
  - R5 floods a new LSA on the entire network.
  - After receiving the LSA, R1 creates a new route that inherits the original path and next hop used when R1 accesses R5. In this case, the SPT remains unchanged, and only a leaf is added to R5.
  - Therefore, when R1 accesses Loopback0 of R5, traffic is sent to the destination based on [outbound interface of R1's downlink, IP address of R3's uplink interface].
- Benefits:
  - PRC focuses only on routes that are changed due to the addition of network segments to an OSPF network, thereby speeding up route calculation.

HUAWEI

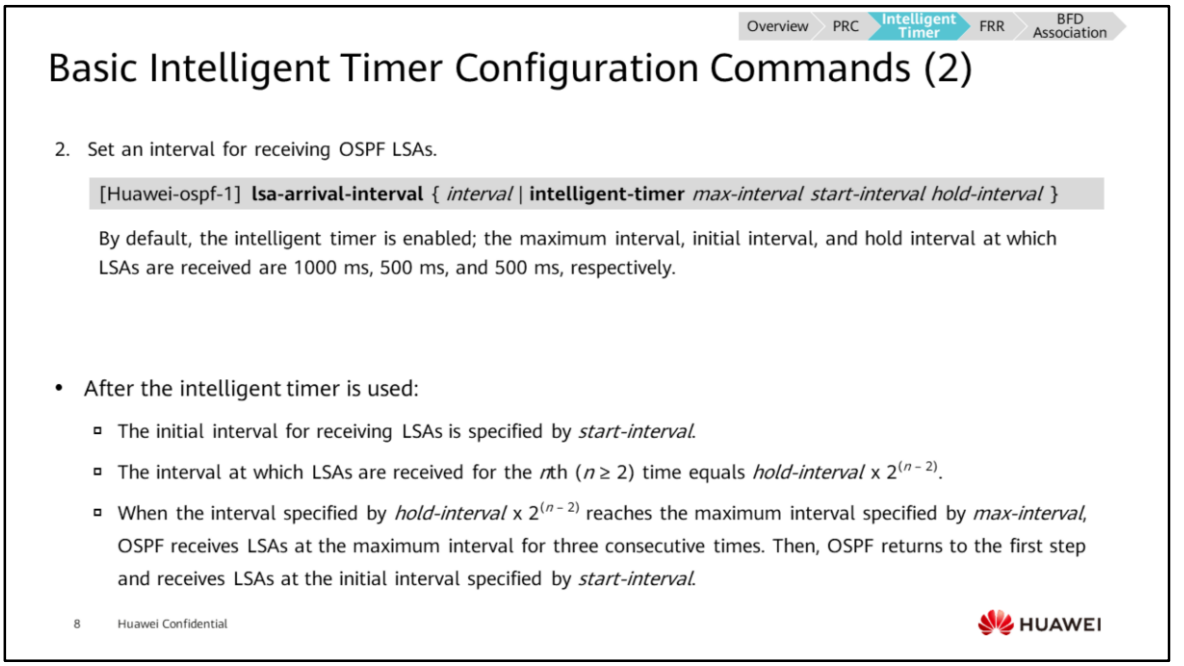
- Note: On Huawei devices, OSPF PRC is enabled by default.



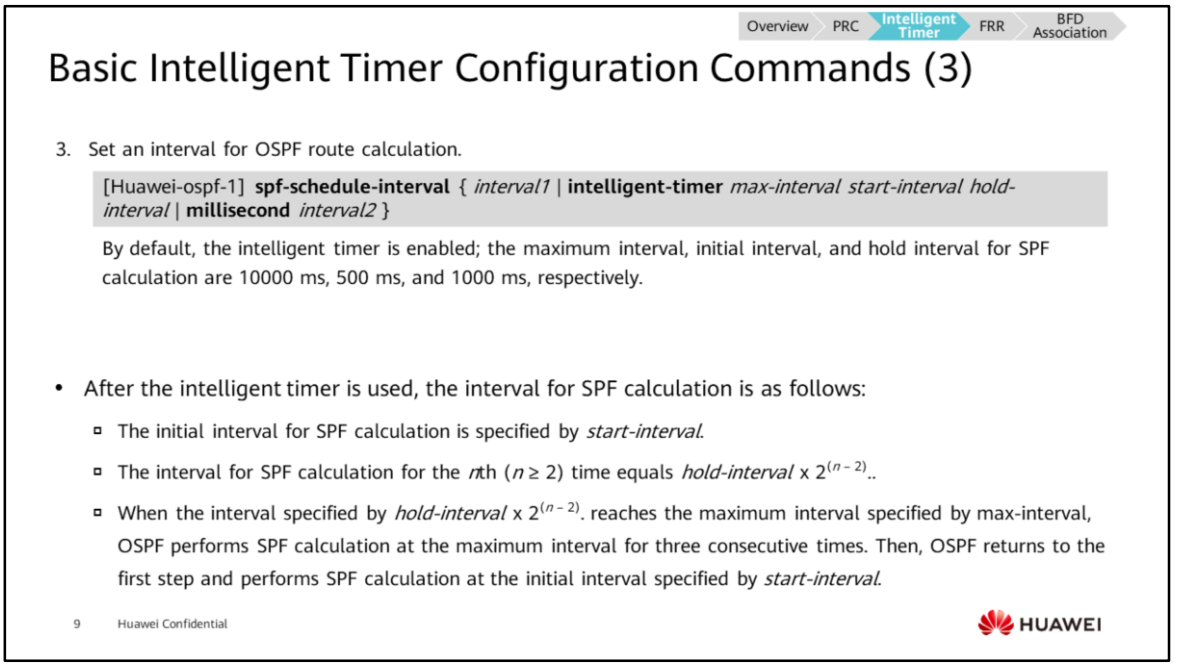
- If the interval for triggering route calculation is long, the network convergence speed is affected.
- The first timeout period of the intelligent timer is fixed. Before the intelligent timer expires, if an event that triggers the timer occurs, the next timeout period of the intelligent timer becomes longer.



- Command: [Huawei-ospf] **lsa-originate-interval { 0 | { intelligent-timer** *max-interval start-interval hold-interval* **| other-type** *interval* **} }**
  - **0**: sets the interval for updating LSAs to 0s, that is, cancels the interval of 5s for updating LSAs.
  - **intelligent-timer**: uses the intelligent timer to set the update interval for router-LSAs and network-LSAs.
  - *max-interval:* specifies the maximum interval for updating OSPF LSAs. The value is an integer ranging from 1 to 120000, in milliseconds. The default value is 5000.
  - *start-interval:* specifies the initial interval for updating OSPF LSAs. The value is an integer ranging from 0 to 60000, in milliseconds. The default value is 500.
  - *hold-interval:* specifies the hold interval for updating OSPF LSAs. The value is an integer ranging from 1 to 60000, in milliseconds. The default value is 1000.
  - **other-type:** sets an update interval for OSPF LSAs except router-LSAs and network-LSAs.
  - *interval:* specifies the interval for updating LSAs. The value is an integer ranging from 0 to 10, in seconds. The default value is 5.



- Command: [Huawei-ospf-1] **lsa-arrival-interval {** *interval* **| intelligent-timer** *max-interval start-interval hold-interval* **}**
  - *interval*: specifies the interval for receiving LSAs. The value is an integer ranging from 0 to 10000, in milliseconds.
  - **intelligent-timer**: uses the intelligent timer to set the receive interval for LSAs.
  - *max-interval:* specifies the maximum interval for receiving OSPF LSAs. The value is an integer ranging from 1 to 120000, in milliseconds. The default value is 1000.
  - *start-interval:* specifies the initial interval for receiving OSPF LSAs. The value is an integer ranging from 0 to 60000, in milliseconds. The default value is 500.
  - *hold-interval:* specifies the hold interval for receiving OSPF LSAs. The value is an integer ranging from 1 to 60000, in milliseconds. The default value is 500.

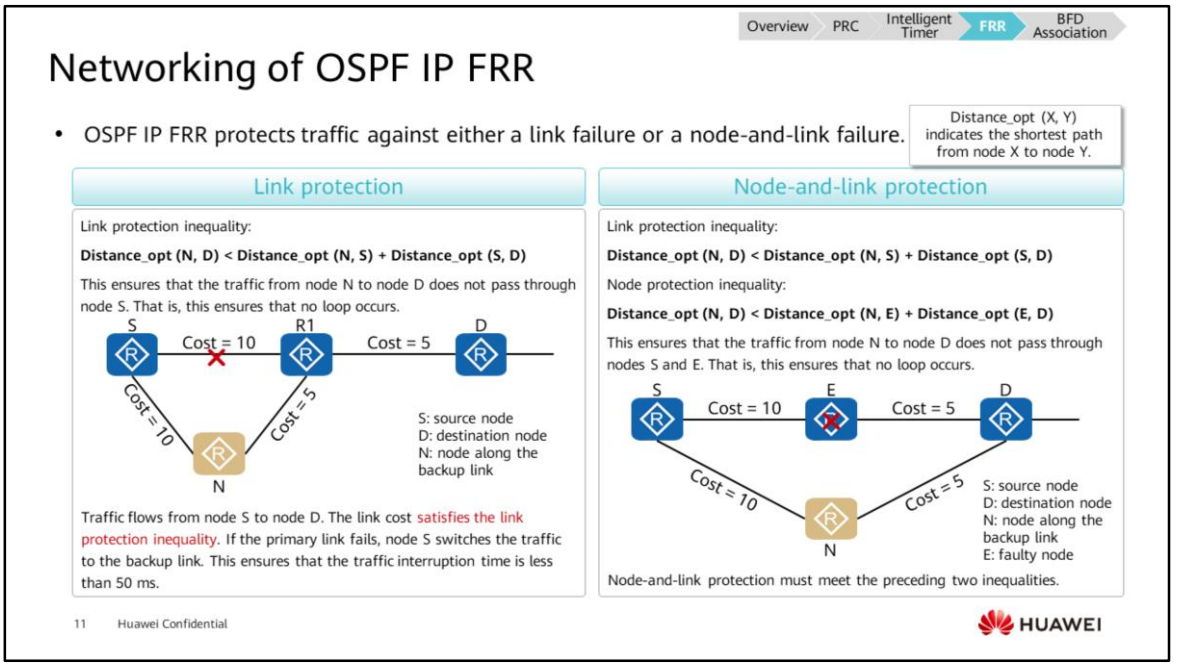


- Command: [Huawei-ospf-1] **spf-schedule-interval** { *interval1* | **intelligent-timer** *max-interval start-interval hold-interval* | **millisecond** *interval2* }
  - *interval1*: specifies an interval for OSPF SPF calculation. The value is an integer ranging from 1 to 10, in seconds.
  - **intelligent-timer**: uses the intelligent timer to set the interval for OSPF SPF calculation.
  - *max-interval:* specifies the maximum interval for OSPF SPF calculation. The value is an integer ranging from 1 to 120000, in milliseconds. The default value is 10000.
  - *start-interval:* specifies the initial interval for OSPF SPF calculation. The value is an integer ranging from 1 to 60000, in milliseconds. The default value is 500.
  - *hold-interval:* specifies the hold interval for OSPF SPF calculation. The value is an integer ranging from 1 to 60000, in milliseconds. The default value is 1000.
  - **millisecond** *interval2:* specifies an interval for OSPF SPF calculation. The value is an integer ranging from 1 to 10000, in milliseconds.

OSPF IP FRR

- OSPF IP fast reroute (FRR) is a dynamic IP FRR technology that uses the loop-free alternate (LFA) algorithm to pre-calculate a backup path and saves it in the forwarding table. If the primary link fails, traffic is rapidly switched to the backup link, ensuring traffic continuity and achieving traffic protection. OSPF IP FRR can reduce the fault recovery time to less than 50 ms.
- The LFA algorithm calculates a backup link based on the following principles:
  - A device uses the SPF algorithm to calculate shortest paths to the destination, with each neighbor that provides a backup link as a root node. The device then uses the inequality to calculate a loop-free backup link with the minimum cost.

HUAWEI



- Node-and-link protection:
  - As shown in the right figure, traffic flows from node S to node D. The link cost satisfies the node-and-link protection inequality. If the primary link fails, node S switches the traffic to the backup link. This ensures that the traffic interruption time is less than 50 ms.
- OSPF IP FRR protects traffic against either a link failure or a node-and-link failure.
  - Link protection takes effect when the traffic to be protected flows along a specified link.
  - Node-and-link protection takes effect when the traffic to be protected flows along a specified device. Node-and-link protection takes precedence over link protection.

Basic OSPF IP FRR Configuration Commands

1. Enable OSPF IP FRR.

```
[Huawei-ospf-1] frr
[Huawei-ospf-1-frr]
```

Create OSPF FRR and enter its view.

```
[Huawei-ospf-1-frr] loop-free-alternate
```

After OSPF IP FRR is enabled, the device uses the LFA algorithm to calculate the next hop and outbound interface for a backup link.

2. (Optional) Disable OSPF IP FRR on an interface.

```
[Huawei-GigabitEthernet0/0/1] ospf frr block
```

OSPF IP FRR can be disabled on an interface of a specific device that is running important services and resides on an FRR backup link. This setting prevents the device connected to this interface from being a part of a backup link and being burdened after FRR switches traffic to the backup link.

HUAWEI

Example for Configuring OSPF IP FRR

Device	Router ID	Interface	IP Address
R1	10.1.1.1	GE0/0/0	10.1.13.1/24
		GE0/0/1	10.1.12.1/24
R2	10.1.2.2	GE0/0/1	10.1.12.2/24
		GE0/0/2	10.1.23.2/24
R3	10.1.3.3	GE0/0/0	10.1.13.3/24
		GE0/0/1	10.1.34.3/24
		GE0/0/2	10.1.23.3/24
R4	10.1.4.4	GE0/0/1	10.1.34.4/24

If the link between R1 and R3 fails, traffic forwarded by R1 can be quickly switched to the backup link and forwarded by R2.

1. Assign an IP address to each interface and configure OSPF on each device. (The configuration details are not provided here.)
2. Configure an OSPF cost for each device. The following example uses the command output on R1.

```
[R1] interface GigabitEthernet 0/0/0
[R1-GigabitEthernet 0/0/0] ospf cost 10
[R1-GigabitEthernet 0/0/0] quit
[R1] interface GigabitEthernet 0/0/1
[R1-GigabitEthernet 0/0/1] ospf cost 10
[R1-GigabitEthernet 0/0/1] quit
```

3. Enable OSPF IP FRR on R1.

```
[R1] ospf
[R1-ospf-1] frr
[R1-ospf-1-frr] loop-free-alternate
[R1-ospf-1-frr] quit
[R1-ospf-1] quit
```

The cost configurations of R2, R3, and R4 are similar to the configuration of R1.

HUAWEI

Checking the OSPF IP FRR Configuration

Device	Router ID	Interface	IP Address
R1	10.1.1.1	GE0/0/0	10.1.13.1/24
		GE0/0/1	10.1.12.1/24
R2	10.1.2.2	GE0/0/1	10.1.12.2/24
		GE0/0/2	10.1.23.2/24
R3	10.1.3.3	GE0/0/0	10.1.13.3/24
		GE0/0/1	10.1.34.3/24
		GE0/0/2	10.1.23.3/24
R4	10.1.4.4	GE0/0/1	10.1.34.4/24

Check information about the route from R1 to GE0/0/1 of R4.

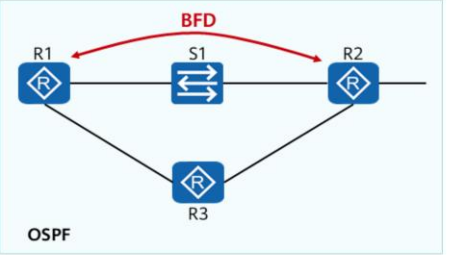
```
[R1]display ospf routing 10.1.34.4
          OSPF Process 1 with Router ID 10.1.1.1
 Destination :  10.1.34.0/24
 AdverRouter :  10.1.4.4      Area :      0.0.0.0
 Cost :         15            Type :      Transit
 NextHop :      10.1.13.3     Interface : GigabitEthernet0/0/0
 Priority :     Low           Age :       00h01m59s
 Backup Nexthop :  10.1.12.2  Backup Interface:GigabitEthernet0/0/1
 Backup Type :     LFA LINK
```

You can find that OSPF generates a backup link after OSPF IP FRR is enabled on R1.

HUAWEI

BFD for OSPF

- A link fault or a topology change causes devices to recalculate routes. Fast and efficient routing protocol convergence is necessary to improve network availability.
- BFD for OSPF associates BFD with OSPF. If a fault occurs on the link between a device and its neighbor, BFD can rapidly detect the link fault to speed up OSPF's response to network topology changes.

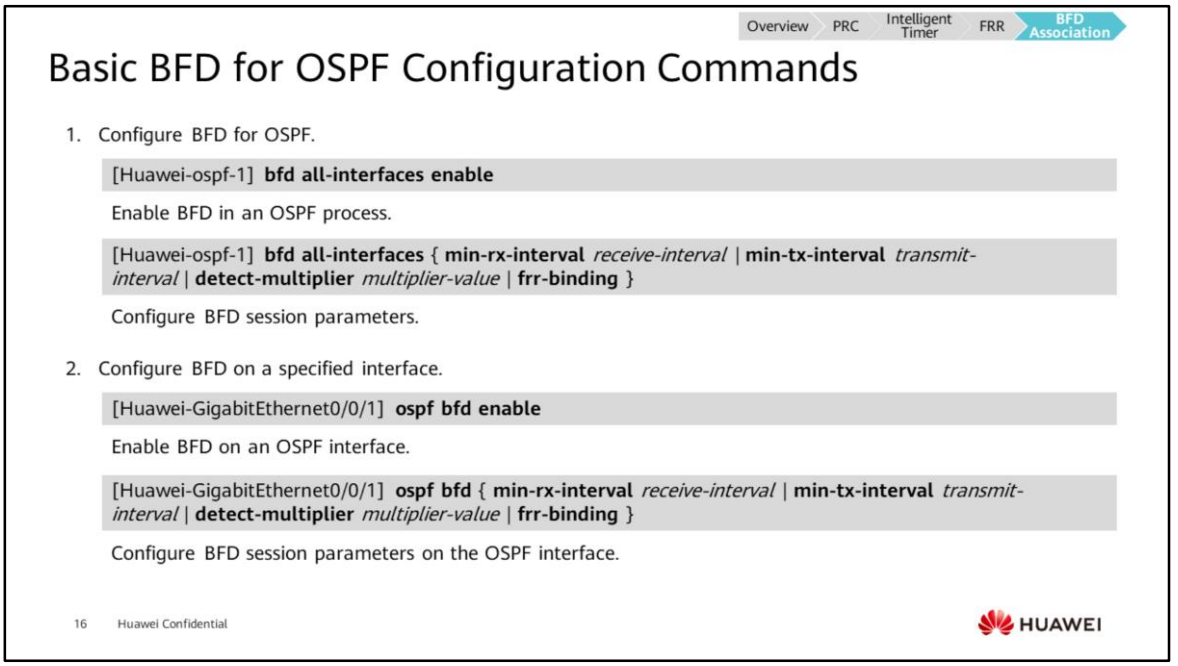


The working principle of BFD for OSPF is as follows:

- OSPF neighbor relationships are established between R1, R2, and R3. When the neighbor relationships enter the Full state, BFD is instructed to set up a BFD session.
- If a fault occurs on the link between R1 and R2, BFD detects the fault and notifies R1. R1 processes the BFD session down event and recalculates the route. The new path is R1-R3-R2.

HUAWEI

- OSPF periodically sends Hello packets to neighbors to detect faults. It takes more than 1s to detect a fault. By default, when the OSPF Dead timer expires, the neighbor is considered invalid. The default value of the OSPF Dead timer is 40s. With the development of technologies, voice, video, and video on demand (VOD) services are widely used. These services are sensitive to the packet loss rate and delay. When the traffic rate reaches gigabit per second (Gbit/s), long-time fault detection causes a large number of packets to be lost. This cannot meet high reliability requirements of the carrier-class network.
- BFD for OSPF is introduced to resolve this problem. After BFD for OSPF is configured in a specified process or on a specified interface, the link status can be rapidly detected and fault detection can be completed in milliseconds. This speeds up OSPF convergence when the link status changes.

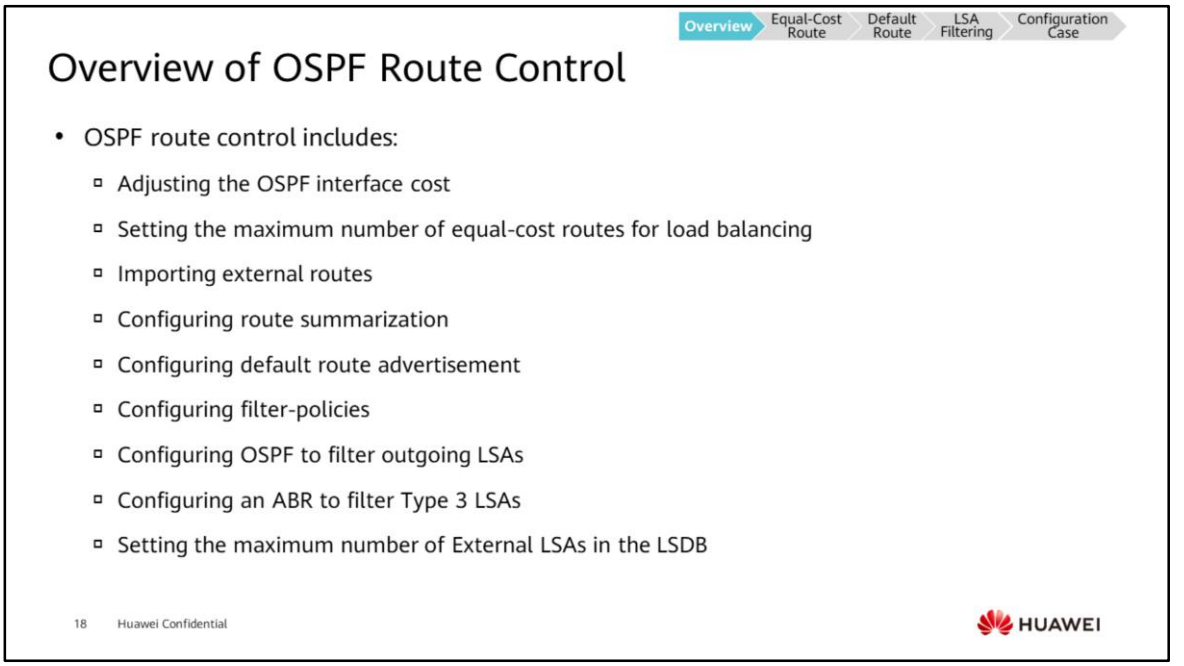


- Prerequisites:
  - Before using BFD to quickly detect link faults, run the bfd command in the system view to enable BFD globally.
- The BFD configuration on an interface takes precedence over that in a process. If BFD is enabled on an interface, the BFD parameters on the interface are used to establish BFD sessions.
- OSPF IP FRR can be associated with BFD.
  - During the OSPF IP FRR configuration, the underlying layer needs to fast respond to a link status change so that traffic can be switched to the backup link immediately.
  - OSPF IP FRR and BFD can be bound to rapidly detect link faults. This ensures that traffic is rapidly switched to the backup link in the case of link failures.
- Command: [Huawei-ospf-1] **bfd all-interfaces** { **min-rx-interval** *receive-interval* | **min-tx-interval** *transmit-interval* | **detect-multiplier** *multiplier-value* | **frr-binding** }
  - **min-rx-interval** *receive-interval*: specifies an expected minimum interval for receiving BFD packets from the peer. The value is an integer ranging from 10 to 2000, in milliseconds. The default value is 1000.
  - **min-tx-interval** *transmit-interval*: specifies a minimum interval for sending BFD packets to the peer. The value is an integer ranging from 10 to 2000, in milliseconds. The default value is 1000.
  - **detect-multiplier** *multiplier-value:* specifies a local detection multiplier. The value is an integer ranging from 3 to 50. The default value is 3.
  - **frr-binding:** binds the BFD session status to the link status of an interface. If a BFD session goes down, the physical link of the bound interface also goes down, triggering traffic to be switched to the backup link.

Contents

1. OSPF Fast Convergence
2. **OSPF Route Control**
3. Other OSPF Features
4. Advanced IS-IS Features

HUAWEI



- This course describes only equal-cost routes, default routes, and LSA filtering. For other information, see HCIP-Datacom-Core Technology.

Equal-Cost Route

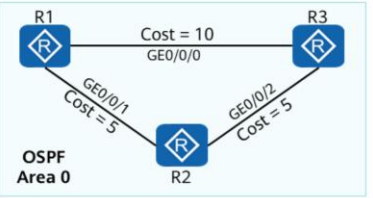
- If the destinations and costs of the multiple routes discovered by one routing protocol are the same, these routes are equal-cost routes and can participate in load balancing.
- The device sends packets to the same destination address through multiple equal-cost routes in load balancing mode.
- Set the maximum number of equal-cost routes for load balancing.

```
[Huawei-ospf-1] maximum load-balancing number
```


HUAWEI

- Command: [Huawei-ospf-1] **maximum load-balancing** *number*
  - *number*: specifies the maximum number of equal-cost routes for load balancing. The value range varies according to the device model. For details, see the product documentation of the corresponding device.

Example for Configuring the Number of Equal-Cost Routes for Load Balancing



Device	Interface	IP Address
R1	GE 0/0/0	10.1.13.1/24
	GE 0/0/1	10.1.12.1/24
R2	GE 0/0/1	10.1.12.2/24
	GE 0/0/2	10.1.23.2/24
R3	Loopback0	10.1.3.3/32
	GE 0/0/0	10.1.13.3/24
	GE 0/0/2	10.1.23.3/24

It is required that R1 can access the loopback interface address of R3 through the path R1 -> R3 or the path R1 -> R2 -> R3.

1. Assign an IP address to each interface and configure OSPF on each device. (Details are not provided here.)

2. Configure the maximum number of OSPF equal-cost routes for load balancing on R1.

```
[R1] ospf
[R1-ospf-1] maximum load-balancing 2
```

3. Verify the configuration.

```
[R1]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
10.1.3.3/32         OSPF    10   10          D   10.1.13.3       GigabitEthernet0/0/0
                    OSPF    10   10          D   10.1.12.2       GigabitEthernet0/0/1
```


HUAWEI

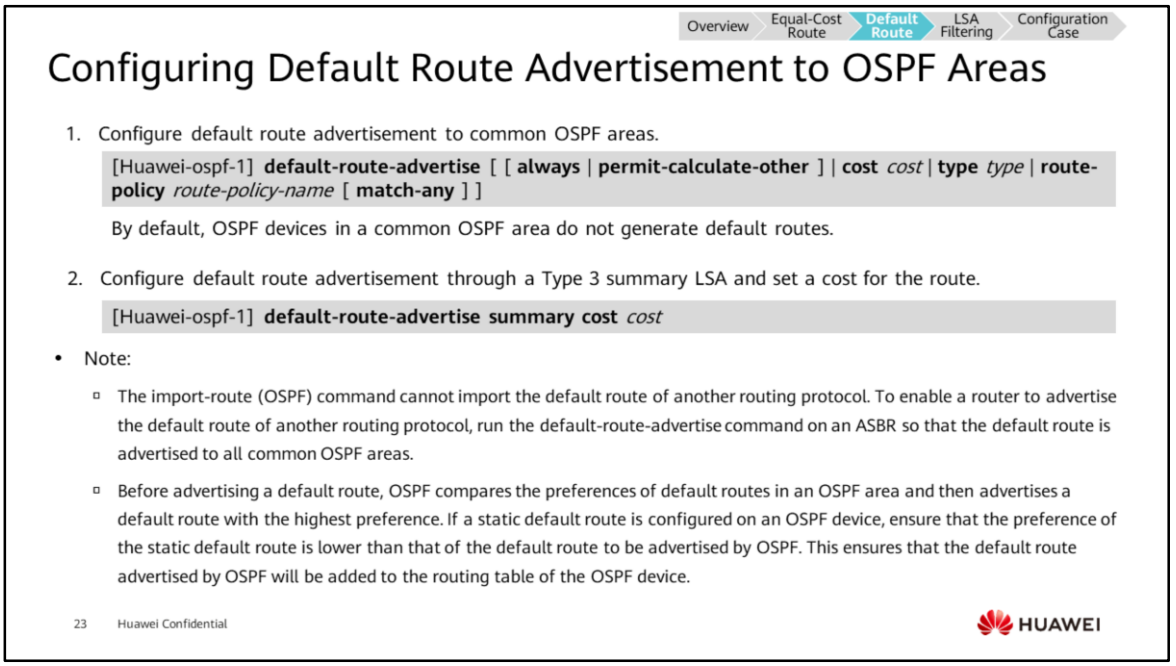
Default Route

- On the area border and AS border of an OSPF network generally reside multiple routers for egress backup or traffic load balancing. In this case, a default route can be configured to reduce the number of routing entries in the routing table and ensure high availability of the network.
- OSPF default routes are generally applied to the following scenarios:
  - An ABR advertises Type 3 LSAs carrying the default route so that routers in an area forward inter-area packets accordingly.
  - An ASBR advertises Type 5 or Type 7 LSAs carrying the default route so that routers in an AS forward AS-external packets.

Area Type	Trigger Condition	Advertised by	LSA Type	Flooding Scope
Common area	The **default-route-advertise** command is run.	ASBR	Type 5 LSA	Common area
Stub area and totally stubby area	Automatically generated	ABR	Type 3 LSA	Stub area
NSSA	The **nssa [default-route-advertise]** command is run.	ASBR	Type 7 LSA	NSSA
Totally NSSA	Automatically generated	ABR	Type 3 LSA	NSSA

HUAWEI

- Default routes have all 0s as the destination address and mask. A device uses a default route to forward packets when no matching route is available. Hierarchical management of OSPF routes prioritizes the default route carried in Type 3 LSAs over the default route carried in Type 5 or Type 7 LSAs.
- Common area:
  - By default, routers in a common OSPF area do not generate default routes. To enable a router in a common OSPF area to advertise a default route to OSPF, run the **default-route-advertise** command on the router. After the command is run, the router generates a default ASE LSA (Type 5 LSA) and advertises it to the entire OSPF AS.
- Stub area:
  - Type 5 LSAs cannot be advertised within a stub area. All routers within a stub area can learn AS external routes only through an ABR.
  - The ABR in a stub area automatically generates a default Type 3 LSA and advertises it to the entire stub area. The ABR uses the default route to divert traffic destined for a destination outside the AS to itself and then forwards the traffic.



- Command: [Huawei-ospf-1] **default-route-advertise [[always | permit-calculate-other] | cost** *cost* **| type** *type* **| route-policy** *route-policy-name* **[match-any]]**
  - **always:** An LSA that describes the default route is generated and advertised regardless of whether the local device has an active default route that does not belong to the current OSPF process.
    - If **always** is configured, the device does not calculate the default routes from other devices.
    - If **always** is not configured, an LSA that describes the default route can be generated only if an active default route that does not belong to the current OSPF process exists in the routing table of the local device.
  - **permit-calculate-other:** An LSA that describes the default route is generated and advertised only if the device has an active default route that does not belong to the current OSPF process, and the device still calculates the default routes from other devices.
  - **type** *type:* specifies the type of an external route. The value is 1 or 2. The default value is 2.
    - 1: Type 1 external route
    - 2: Type 2 external route
  - **route-policy** *route-policy-name:* specifies the name of a route-policy. The device advertises default routes according to the configuration of the route-policy when the routing table of the device contains a default route that matches the route-policy but does not belong to the current OSPF process. The value is a string of 1 to 40 case-sensitive characters. If spaces are used, the string must start and end with double quotation marks (").

Configuring OSPF to Filter Outgoing LSAs

- When multiple links exist between two routers, you can enable the function to filter outgoing LSAs, preventing them from being sent to particular links. This function can help reduce unnecessary retransmission of LSAs and reduce bandwidth consumption.
- Filtering the outgoing LSAs on a specified OSPF interface can prevent unwanted LSAs from being sent to neighbors, thus reducing the LSDB sizes of the neighbors and speeding up network convergence.
- To filter outgoing LSAs on an OSPF interface, run the following command:

[Huawei-GigabitEthernet0/0/1] **ospf filter-lsa-out** { **all** | { **summary** [**acl** { *acl-number* | *acl-name* }] | **ase** [**acl** { *acl-number* | *acl-name* }] | **nssa** [**acl** { *acl-number* | *acl-name* }] } }

The command configuration does not take effect for the LSAs that have been sent out before the command is run. The aging time of such an LSA is still 3600 seconds.

HUAWEI

- Command: [Huawei-GigabitEthernet0/0/1] **ospf filter-lsa-out** { **all | { summary [acl** { *acl-number* | *acl-name* }] | **ase [acl** { *acl-number* | *acl-name* }] | **nssa** [**acl** { *acl-number* | *acl-name* }] } }
  - **all**: filters all outgoing LSAs, except grace LSAs.
  - **summary**: filters outgoing network-summary-LSAs (Type 3).
  - **ase:** filters outgoing AS-external-LSAs (Type 5).
  - **nssa:** filters outgoing NSSA-LSAs (Type 7).
  - **acl** *acl-number:* specifies the number of a basic ACL. The value is an integer ranging from 2000 to 2999.
  - **acl** *acl-name*: specifies the name of an ACL. The value is a string of 1 to 32 case-sensitive characters. It cannot contain spaces and must start with a letter (a to z or A to Z).

Configuring OSPF to Filter ABR Type 3 LSAs

- Configure filtering policies for incoming and outgoing ABR Type 3 LSAs (network-summary-LSAs) in an area, allowing only those that pass the filtering to be sent and accepted.
- Filtering Type 3 LSAs in a specified OSPF area can prevent unwanted LSAs from being sent to neighbors, thus reducing the LSDB sizes and speeding up network convergence.
- To filter outgoing Type 3 LSAs in an OSPF area, run the following command:

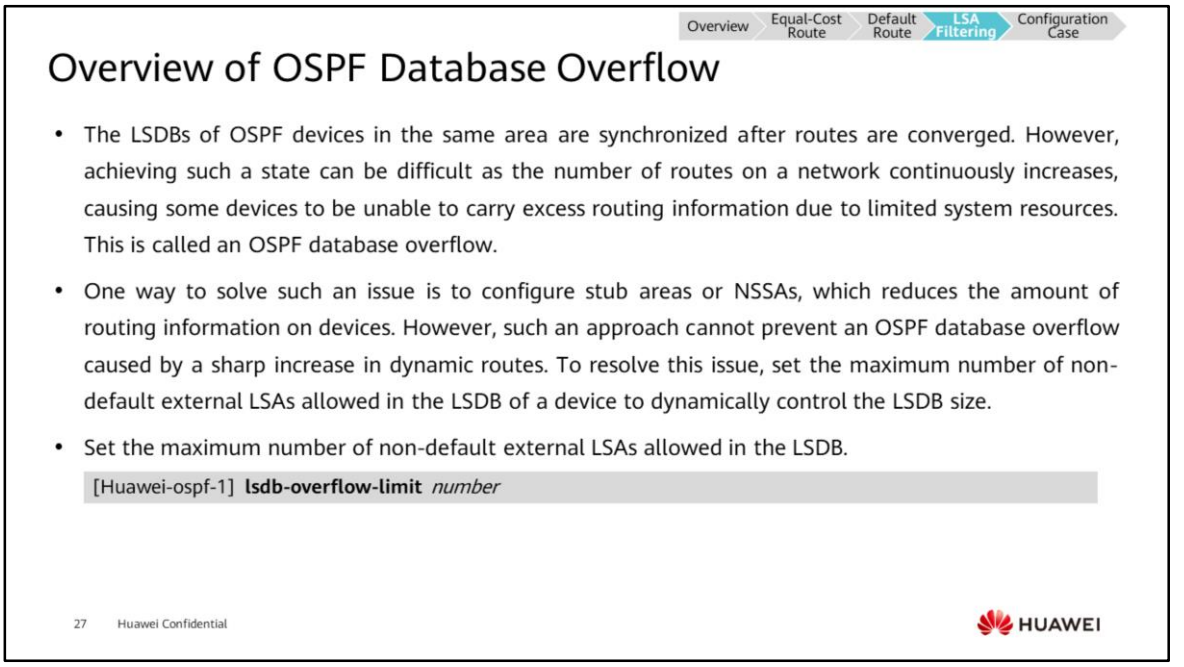
  [Huawei-ospf-1-area-0.0.0.1] **filter** { *acl-number* | **acl-name** *acl-name* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **export**

- To filter incoming Type 3 LSAs in an OSPF area, run the following command:

  [Huawei-ospf-1-area-0.0.0.1] **filter** { *acl-number* | **acl-name** *acl-name* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **import**

HUAWEI

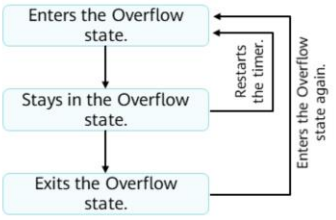
- Command: [Huawei-ospf-1-area-0.0.0.1] **filter** { *acl-number* | **acl-name** *acl-name* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **export**
  - *acl-number:* specifies the number of a basic ACL. The value is an integer ranging from 2000 to 2999.



- When the number of external LSAs (Type 5 and Type 7) imported by OSPF exceeds the maximum number supported, excessive external LSAs cannot be processed properly and are discarded. To address this issue, you can set a proper upper limit for the number of non-default external LSAs in the LSDB, so as to adjust and optimize the OSPF network.
- Command: [Huawei-ospf-1] **lsdb-overflow-limit** *number*
  - *number:* specifies the maximum number of non-default external LSAs allowed in the LSDB. The value is an integer ranging from 1 to 1000000.

Preventive Measures for OSPF Database Overflows

- Setting the maximum number of non-default external routes on a router can prevent OSPF database overflows.
- Set the same maximum number for all routers on an OSPF network. When the number of non-default external routes on a router reaches the maximum number, the router enters the Overflow state and starts the Overflow state timer (default timeout period: 5s). After the timer expires, the router automatically exits the Overflow state.



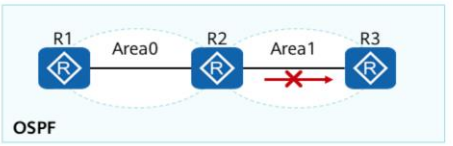
Overflow Phase	OSPF Processing
Enters the Overflow state.	• Deletes all non-default external routes generated by the router itself. • Starts the Overflow state timer.

Stays in the Overflow state.	• Does not generate non-default external routes.
	• Discards newly received non-default external routes and does not reply with LSAck packets.
	• Checks whether the number of external routes still exceeds the maximum number when the overflow state timer expires.
	  ▪ If not, the router exits the Overflow state.
	  ▪ If so, the router restarts the Overflow state timer.

Exits the Overflow state.	• Deletes the Overflow state timer. • Generates non-default external routes. • Accepts newly received non-default external routes and replies with LSAck packets. • Gets ready to enter the Overflow state again.

HUAWEI

Example for Configuring OSPF Route Control (1)



Type 3 LSA

- To reduce the number of LSAs on R3 and ensure that R3 can properly communicate with routers in other areas, check that the following requirements are met:
  - R2 does not inject Type 3 LSAs into Area 1.
  - R2 advertises default routes.

1. Assign an IP address to each interface and configure OSPF on each device. (omitted).
2. Configure R2 to filter Type 3 LSAs.

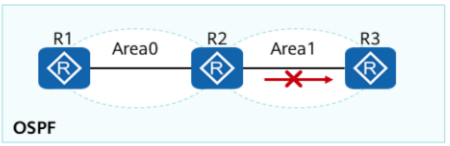
```
[R2] acl 2000
[R2-acl-basic-2000] rule deny
[R2-acl-basic-2000] quit
[R2] ospf
[R2-ospf-1] area 1
[R2-ospf-1-area-0.0.0.1] filter 2000 import
```

3. Configure R2 to advertise default routes.

```
[R2] ospf
[R2-ospf-1] default-route-advertise always
```


HUAWEI

Example for Configuring OSPF Route Control (2)



Type 3 LSA

- The LSDB of R3 in Area1 does not contain Type 3 LSAs but contains a default Type 5 LSA.
- R3 can access devices in other areas through the default route.

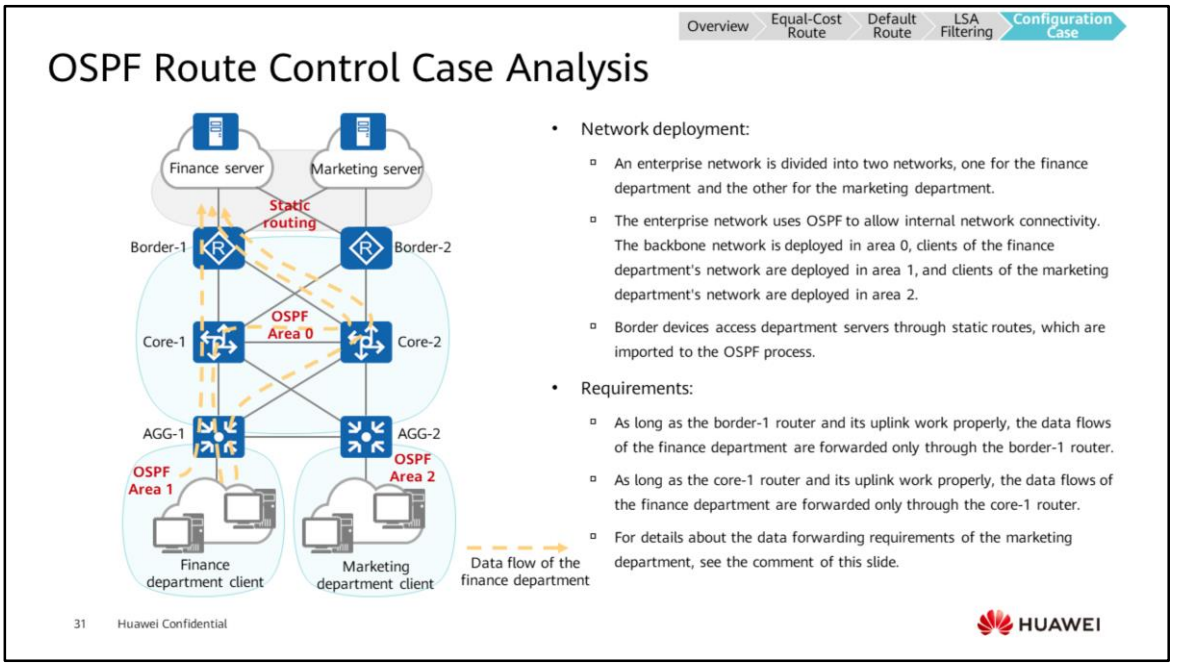
1. Check the LSDB of R3.

```
[R3]display ospf lsdb
         OSPF Process 1 with Router ID 10.1.23.3
                   Link State Database
                   Area: 0.0.0.1
Type       LinkState ID    AdvRouter       Age  Len  Sequence   Metric
Router     10.1.23.3       10.1.23.3       731  48   80000004   1
Router     10.1.12.2       10.1.12.2       406  36   80000008   1
Network    10.1.23.2       10.1.12.2       730  32   80000002   0

                   AS External Database
Type       LinkState ID    AdvRouter       Age  Len  Sequence   Metric
External   0.0.0.0         10.1.12.2       406  36   80000001   1
```

2. Check the routing table of R3.

```
[R3]dis ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: Public
         Destinations : 9        Routes : 9
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
0.0.0.0/0           O_ASE   150  1         D     10.1.23.2       GigabitEthernet0/0/1
......
```

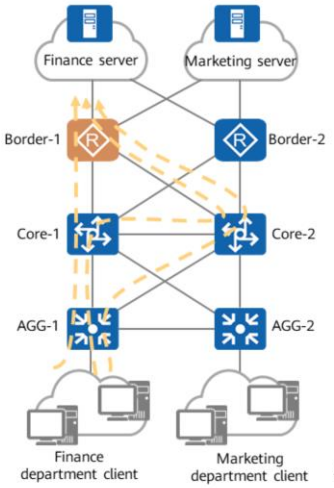

HUAWEI



- Data forwarding requirements of the marketing department:
  - As long as the border-2 router and its uplink work properly, the data flows of the marketing department are forwarded only through the border-2 router.
  - As long as the core-2 router and its uplink work properly, the data flows of the marketing department are forwarded only through the core-2 router.
- This case uses the data forwarding path of the finance department as an example. The data forwarding path of the marketing department is not described here.

Overview | Equal-Cost Route | Default Route | LSA Filtering | Configuration Case

Requirement Analysis



1. Controlling the network egress for data forwarding:
   - Data of the finance department is always forwarded through border-1.
   - Data of the marketing department is always forwarded through border-2.

To ensure that fixed ASBRs are used to forward data, the internal network changes must be ignored. That is, the internal route cost is not calculated.

—Use Type 2 external routes of OSPF.

2. Controlling the precise internal path of data flows:
   - Load-balancing path should not exist.

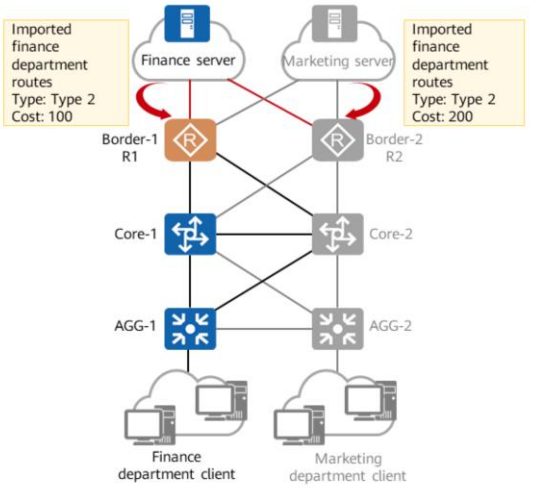
Data needs to be sent to a specific ASBR along the planned path on the network.

—Adjust the internal path cost.

HUAWEI

- Type 2 external route:
  - Because a Type 2 external route offers low reliability, its cost is considered to be much greater than the cost of any internal route to an ASBR.
  - Cost of a Type 2 external route = Cost of the route from an ASBR to the destination

Traffic Egress Control

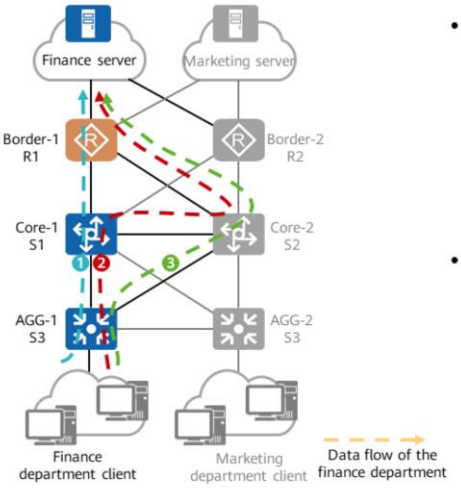


- Implementation:
  - Import static routes destined for the finance department server to the OSPF processes of R1 (border-1) and R2 (border-2) to implement egress backup through route-policies.
  - Set the type of the imported external route to Type 2.
  - On R1, set the cost of the external route to 100; on R2, set the cost of the external route to 200.
- Configuration result:
  - When there are two Type 2 external routes with different costs on the same network segment, the network device prefers the route with a smaller cost. In this case, each network device preferentially selects R1 as the egress.

HUAWEI

- The internal path cost to each ASBR is not considered during traffic egress control.

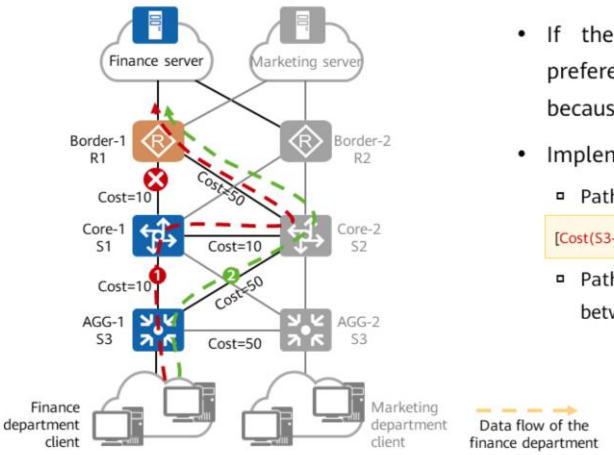
Controlling Internal Paths



- Network requirement analysis:
  - If the network is running properly, S3 (AGG-1) selects path 1.
  - If the link between S1 (core-1) and R1 fails, S3 selects path 2.
  - If S1 fails, S3 selects path 3.
- Implementation:
  - Path 1-cost < Path 2-cost < Path 3-cost

HUAWEI

Controlling Internal Paths: Adjusting the Cost Between Aggregation Devices and Core Devices

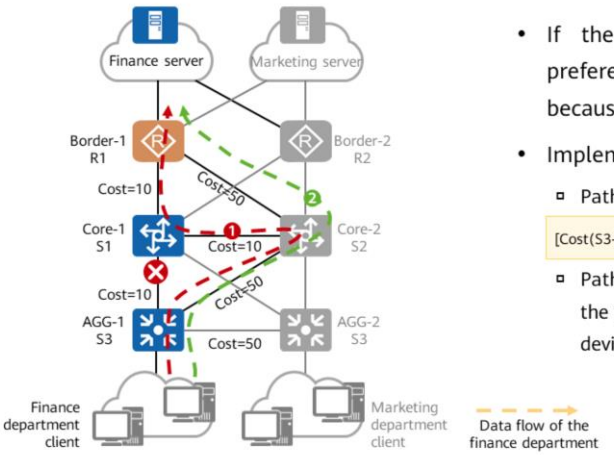


- If the link between S1 and R1 fails, S3 preferentially selects path 1 and then path 2 because S1 is working properly.
- Implementation:
  - Path 1-cost < Path 2-cost, that is:

    [Cost(S3-S1) + Cost(S1-S2) + Cost(S2-R1)] < [Cost(S3-S2) + Cost(S2-R1)]
  - Path 1 can be achieved by adjusting the path cost between aggregation devices and core devices.

HUAWEI

Controlling the Internal Path: Adjusting the Cost Between Core Devices and Boundary Devices



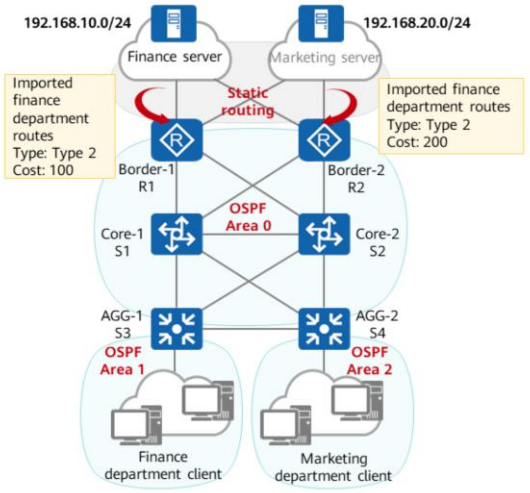
- If the link between S3 and S1 fails, S3 preferentially selects path 1 and then path 2 because S1 is working properly.
- Implementation:
  - Path 1-cost < Path 2-cost, that is:

[Cost(S3-S2) + Cost(S2-S1) + Cost(S1-R1)] < [Cost(S3-S2) + Cost(S2-R1)]

  - Path 1 can be preferentially selected by adjusting the path cost between core devices and boundary devices.

HUAWEI

Example for Configuring OSPF Route Control (1)



Configure a route-policy on R1 and set the cost to 100.

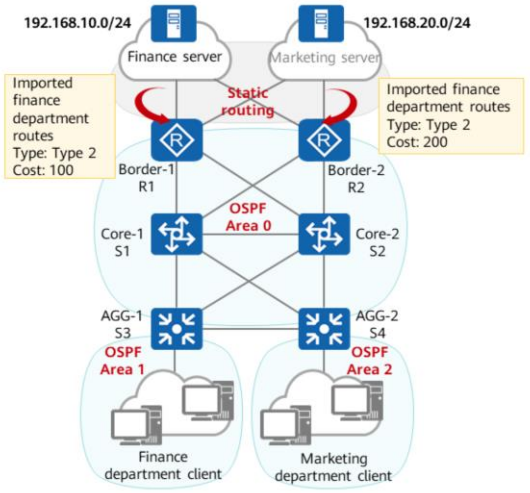
```
[R1] acl 2000
[R1-acl-basic-2000] rule permit source 192.168.10.0 0.0.0.255
[R1-acl-basic-2000] quit

[R1] route-policy static2ospf permit node 10
[R1-route-policy] if-match acl 2000
[R1-route-policy] apply cost 100
[R1-route-policy] quit
[R1] route-policy static2ospf permit node 20
[R1-route-policy] quit
```

Import static routes to the OSPF process on R1.

```
[R1] ospf
[R1-ospf-1] import-route static route-policy static2ospf type 2
```


HUAWEI

Example for Configuring OSPF Route Control (2)



Configure a route-policy on R2 and set the cost to 200.

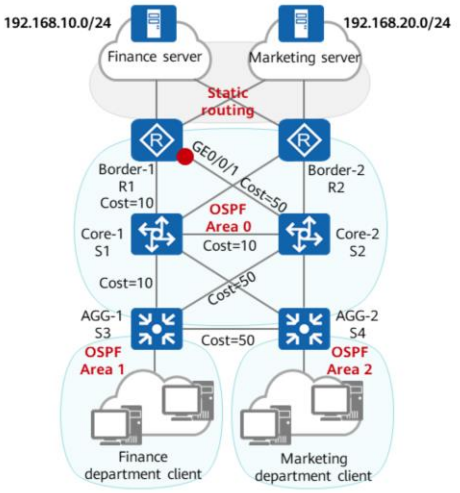
```
[R2] acl 2000
[R2-acl-basic-2000] rule permit source 192.168.10.0 0.0.0.255
[R2-acl-basic-2000] quit

[R2] route-policy static2ospf permit node 10
[R2-route-policy] if-match acl 2000
[R2-route-policy] apply cost 200
[R2-route-policy] quit
[R2] route-policy static2ospf permit node 20
[R2-route-policy] quit
```

Import static routes to the OSPF process on R2.

```
[R2] ospf
[R2-ospf-1] import-route static route-policy static2ospf type 2
```


HUAWEI

Example for Configuring OSPF Route Control (3)



Set the OSPF cost of an interface (GE0/0/1 of R1 is used as an example).

```
[R1] interface GigabitEthernet 0/0/1
[R1-GigabitEthernet0/0/1] ospf cost 50
```


HUAWEI

Contents

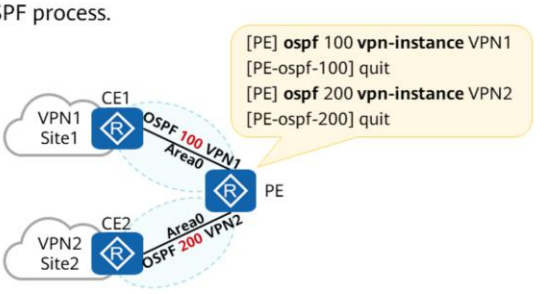
1. OSPF Fast Convergence
2. OSPF Route Control
3. **Other OSPF Features**
4. Advanced IS-IS Features

HUAWEI

Multi-Process | Association with BGP | Forwarding Address | GR | NSR

OSPF Multi-Process

- OSPF supports multiple processes that can separately run on the same device and do not affect each other. Route exchange between different OSPF processes is similar to route exchange between different routing protocols.
- An interface on a router can belong to only one OSPF process.
- Usage scenario:
  - A typical application of OSPF multi-process is in the VPN scenario.
  - As shown in the figure, a PE connects to two different VPN customers, and OSPF is deployed between the PE and CEs. Therefore, multiple processes can be deployed on the PE to isolate the VPN customers.

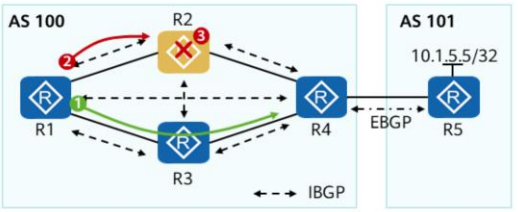


HUAWEI

- VPN: virtual private network
- If a VPN instance is specified for an OSPF process that is to be created, the OSPF process belongs to this instance. Otherwise, the OSPF process belongs to the global instance.

Association Between OSPF and BGP (1)

- When a new device is added or a device is restarted, network traffic may be lost during BGP convergence. This is because IGP route convergence is faster than BGP route convergence.



OSPF runs on R1, R2, R3, and R4 and full-mesh IBGP connections are established. R3 is the backup device of R2.

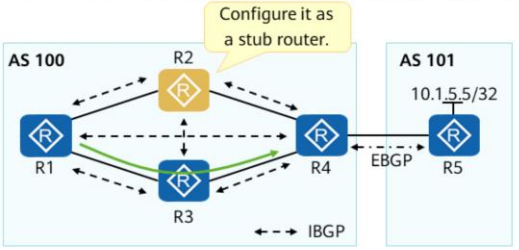
When the network is stable, the traffic from R1 to 10.1.5.5/32 passes through the path R1 -> R2 -> R4 -> R5.

1. If R2 fails, traffic is switched to the path R1 -> R3 -> R4 -> R5.
2. After R2 recovers, OSPF converges first because IGP route convergence is faster than BGP route convergence. If R1 needs to access 10.1.5.5/32, it searches for a BGP route with the next hop being R5. R2 then searches for the IGP route and sends traffic to R2 over the route.
3. After receiving the traffic, R2 searches for the BGP route. Because BGP route convergence is not complete, R2 does not find a route to 10.1.5.5/32 and therefore does not forward traffic. As a result, traffic is lost.

HUAWEI

Association Between OSPF and BGP (2)

- OSPF-BGP synchronization can be enabled to prevent traffic loss.
- After OSPF-BGP synchronization is enabled on a device, the device remains as a stub router within the set synchronization period. That is, the link metric in the LSA advertised by the device is the maximum value 65535. Therefore, the device instructs other OSPF devices not to use it for data forwarding.



Enable BGP association on R2. In this way, R1 continues to forward traffic through R3, but does not forward traffic to R2 until BGP route convergence on R2 is complete.

Configure the stub router.

[Huawei-ospf-1] **stub-router** [**on-startup** [*interval*]]

- Configuring a stub router is a special route selection method. The path with a stub router configured is not preferred.
- The implementation is to set the metric to the maximum value (65535) to prevent data from being forwarded through the router. It is used to protect the link of the router and is usually used in maintenance scenarios, such as upgrade.

HUAWEI

- Command: [Huawei-ospf-1] **stub-router [on-startup** [*interval]*]
  - **on-startup** [*interval*]: specifies the interval for a device to remain as a stub router when the device restarts or fails. The value is an integer ranging from 5 to 65535, in seconds. The default value is 500s.
    - If on-startup is not configured, the device is always a stub router. That is, the cost of all routes sent by this device is 65535.
    - If on-startup is specified, the device remains as a stub router only when it restarts or fails. The duration is determined by interval. If interval is not specified, the default value of 500s is used.

OSPF Forwarding Address

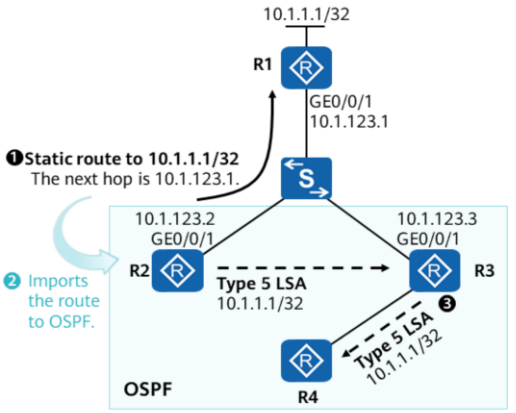
- Forwarding address (FA):
  - The FA is the address to which a data packet is forwarded before the packet reaches the advertised destination address. If the forwarding address is 0.0.0.0, the packet is forwarded to the originating ASBR.
  - Type 5 AS-External LSAs and Type 7 NSSA LSAs:

LS Age		Options	LS Type
Link State ID			
Advertising Router			
LS Sequence Number			
LS Checksum		Length	
Network Mask			
E	0	Metric	
Forwarding Address			
External Route Tag			
.....			

OSPF Type 5 and Type 7 LSAs contain a special FA field. The introduction of FA enables OSPF to avoid the sub-optimal path problem in some special scenarios.

HUAWEI

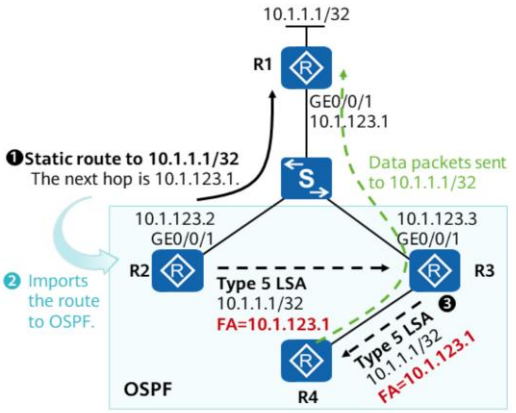
Problem Occurring When No FA Is Used



- OSPF runs on R2, R3, and R4 that are deployed in Area 0. OSPF is enabled on GE0/0/1 of R2 and GE0/0/1 of R3, and OSPF adjacencies are established between the two routers. However, no OSPF adjacencies are established between R1 and the two routers.
  - Configure a static route destined for 10.1.1.1/32 on R2, with 10.1.123.1 as the next hop.
  - R2 imports the static route to OSPF and generates a Type 5 LSA to be flooded in the area.
  - After R3 receives the Type 5 LSA from R2, it calculates an external route to 10.1.1.1/32 and sets the next-hop address of the route to R2 (10.1.123.2).
- The path from a router (for example, R4) in the OSPF domain to 10.1.1.1/32 is R4 -> R3 -> R2 -> R1, which is the sub-optimal path.

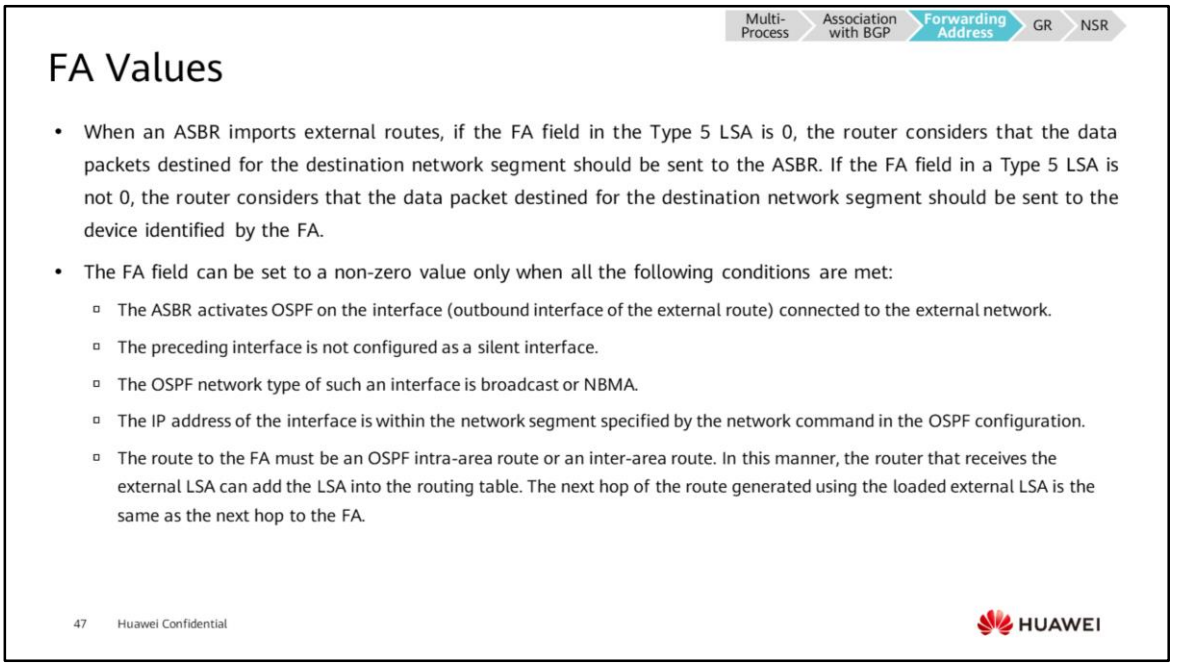
HUAWEI

Using FA to Solve the Suboptimal Path Problem



- When advertising an external route destined for 10.1.1.1/32 in the OSPF area, R2 sets the FA of the corresponding Type 5 LSA to 10.1.123.1, which is the next hop of the external route.
- When R3 receives the LSA, it calculates the route to 10.1.1.1/32 and finds that the FA is not 0. Therefore, R3 considers that the next hop to 10.1.1.1/32 is the address specified by the FA, that is, 10.1.123.1.

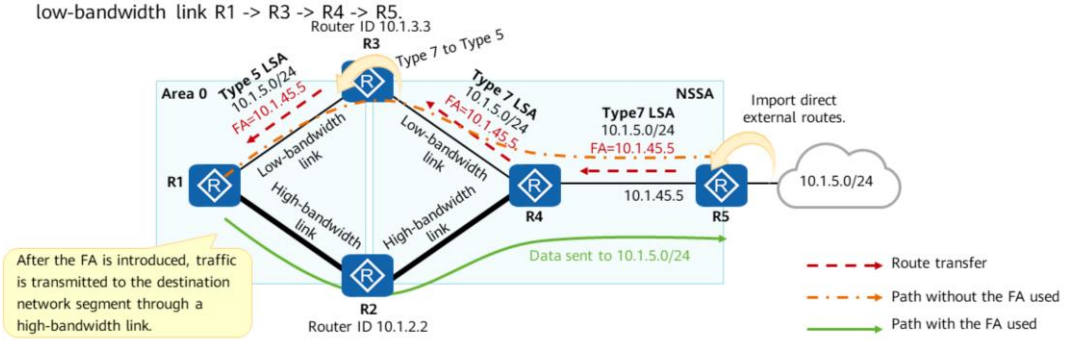
HUAWEI



- Type 7 LSAs in an NSSA are translated into Type 5 LSAs.
  - To advertise external routes imported by an NSSA to other areas, Type 7 LSAs must be translated into Type 5 LSAs. By default, the translator is the ABR with the largest router ID in an NSSA.
  - The propagate bit (P-bit) in the Options field of an LSA header is used to notify a translator whether the Type 7 LSA needs to be translated into a Type 5 LSA. A Type 7 LSA can be translated into a Type 5 LSA only when the P-bit is set to 1 and the FA is not 0.
  - The P-bit is not set for Type 7 LSAs generated by an ABR.
- Note: All OSPF LSAs have the same LSA header, and the P-bit is in the Options field of the LSA header.

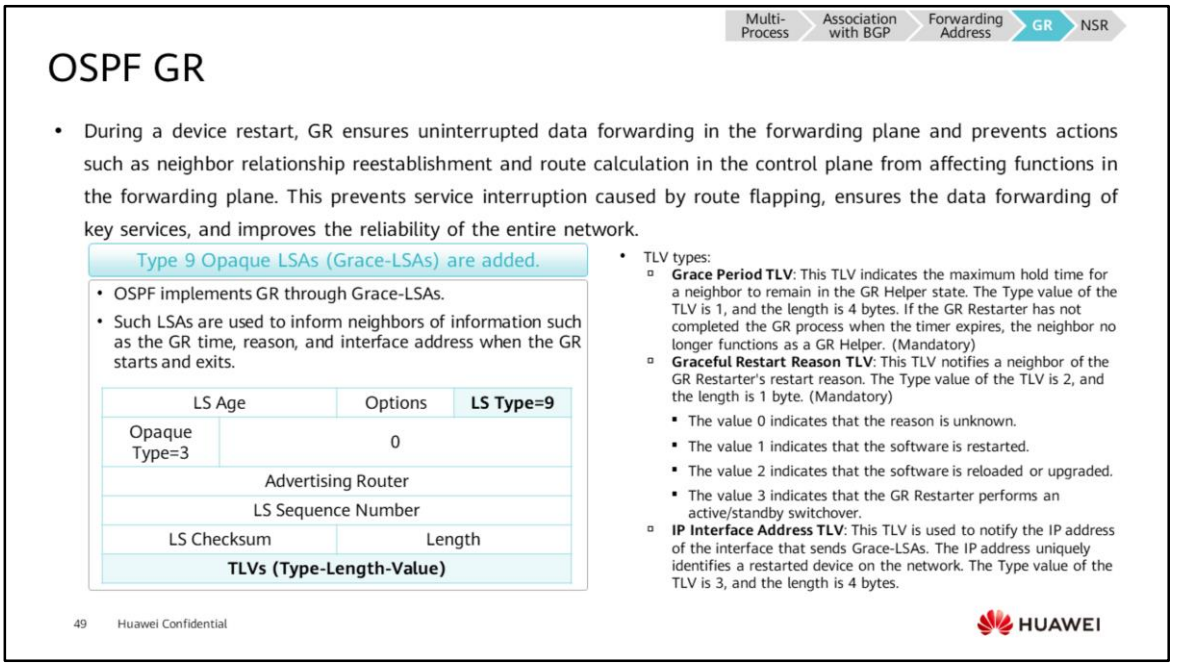
Case: Typical FA Application in NSSA Scenarios

- When multiple ABRs exist in an NSSA, the system automatically selects an ABR as a translator to translate Type 7 LSAs into Type 5 LSAs. Other ABRs do not perform LSA translation.
- As shown in the figure, if the FA is not considered, R1 considers that the packet must pass through the ABR (R3) to reach the destination because the router ID of R3 is greater than that of R2. In this way, traffic is diverted to the low-bandwidth link R1 -> R3 -> R4 -> R5.

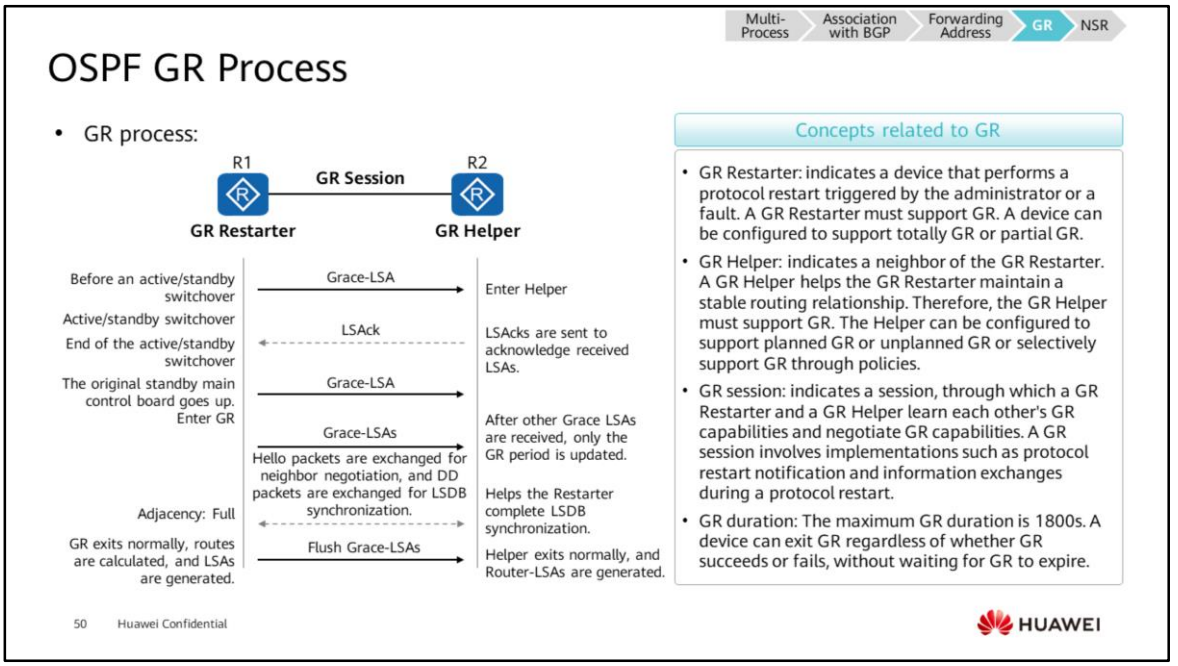


HUAWEI

- As shown in the figure:
  - Configure R5 to import direct external routes and set the IP address of the FA to 10.1.45.5, which is used by R5 to access the destination network segment 10.1.5.0/24.
  - R3 translates Type 7 LSAs into Type 5 LSAs and the LSAs continue to carry the FA 10.1.45.5.
  - Upon receipt, R1 searches its OSPF routing table for a route to the FA and uses the next hop address of the route as the next hop address of the external route.
  - Therefore, R1 will finally access the destination network segment 10.1.5.0/24 through the path R1 -> R2 -> R4 -> R5.



- GR is a fault-tolerant redundancy technique and has been widely used to ensure non-stop forwarding of key services during an active/standby switchover or system upgrade.
- GR reasons:
  - Unknown: GR is triggered for an unknown reason.
  - Software restart: GR is triggered by a command.
  - Software reload/upgrade: GR is triggered by a software reloading or upgrade.
  - Switch to redundant control processor: GR is triggered by an unexpected active/standby switchover.
- GR classification:
  - Totally GR: When a neighbor of a router does not support GR, the router exits the GR state.
  - Partial GR: When a neighbor does not support GR, only the interface associated with this neighbor exits GR, whereas the other interfaces perform GR processing normally.
  - Planned GR: Triggered by a command, a device is restarted or performs an active/standby switchover. Before the device restarts or performs an active/standby switchover, the Restarter sends Grace-LSAs.
  - Unplanned GR: Different from the implementation of a planned GR, the implementation of an unplanned GR is as follows: A router restarts or performs an active/standby switchover due to a fault or other reasons, without sending Grace-LSAs first, and enters GR after the original standby main control board goes up. The following procedure is the same as that in a planned GR.



- On the Restarter:
  - In planned GR mode, after a command is run to trigger an active/standby switchover on the Restarter, the Restarter sends a Grace-LSA to each neighbor to notify them of the GR period and reason, and then performs an active/standby switchover. This prevents an interruption of the neighbor relationships due to the long time taken for the original standby main control board to go up, which would otherwise cause a GR failure.
  - After the original standby main control board goes up, the router notifies the routing management (RM) module that the GR starts. After the interfaces previously enabled with OSPF on the original standby main control board recover, a Grace-LSA is sent immediately to notify neighbors that the local device has started a GR. Then, another five Grace-LSAs are sent consecutively to each neighbor to ensure reception. This implementation is defined by vendors, but not in the protocol. The router does not age FIB entries in this case. Therefore, the communication is normal. If GR is not supported, the router directly ages FIB entries. As a result, routing is interrupted. In this case, the router restarts normally (non-GR process). The Grace-LSAs sent by the Restarter are used to notify neighbors that the Restarter enters GR. During GR, the neighbors retain neighbor relationships with the Restarter so that other routers are unaware of the switchover on the Restarter.
  - The Restarter negotiates with its neighbors to enter the standard adjacency establishment process. The neighbor states change as follows: Down → Init → Exstart. Then, DD packets are exchanged to complete LSDB synchronization. During the update of LSAs, if the LSAs (Type 1 LSAs and Type 2 LSAs) received from a Helper do not contain the link to the local device, it indicates that the network topology has changed or the neighbor does not support the Helper mode. In this case, GR fails, and the local router exits GR and restarts normally.

OSPF GR Exit

- Reasons why a device exits GR:

GR Execution Result	GR Restarter	GR Helper
GR succeeds.	• Before GR expires, the Restarter re-establishes neighbor relationships with all its pre-active/standby switchover neighbors.	• When the GR Helper receives the Grace-LSA with the age being 3600s from the Restarter, the neighbor relationship with the Restarter is Full.
GR fails.	• GR expires, and neighbor relationships have not recovered completely. • The Router-LSA or Network-LSA sent by the Helper causes a failure in the bidirectional check on the Restarter. • The status of the GR Restarter's interface changes. • The Restarter receives a Grace-LSA generated by another router on the same network segment. Only one router can perform a GR on the same network segment at the same time. • The topology of neighbors on the same network segment with the Restarter changes.	• The GR Helper has not received any Grace-LSA from the Restarter when the neighbor relationship expires. • The status of the Helper's interface changes. • The GR Helper receives LSAs, which are different from the LSAs in its own LSDB, from other routers. You can configure the GR Helper not to perform a strict LSA check to avoid this issue. • On the same network segment, the GR Helper receives Grace-LSAs from two routers at the same time. • The neighbor relationships between the GR Helper and other routers change.

HUAWEI

Configuring OSPF GR (1)

1. Enable the opaque LSA capability.

   [Huawei-ospf-1] **opaque-capability enable**

   This command enables the opaque LSA capability so that the OSPF process can generate opaque LSAs and receive opaque LSAs from neighbors. To use the OSPF GR function, which is implemented through Type 9 LSAs, the opaque LSA capability must be enabled first.

2. Enable OSPF GR and configure GR session parameters on the Restarter.

   [Huawei-ospf-1] **graceful-restart** [**period** *period* | **planned-only** | **partial**]

   period *period*: specifies a GR period. The value is an integer ranging from 1 to 1800, in seconds. The default value is 120.
   planned-only: indicates that the router supports only planned GR. By default, a router supports both planned GR and unplanned GR.
   partial: indicates that the router supports partial GR. By default, a router supports totally GR.

HUAWEI

- Before configuring OSPF GR, complete the following tasks:
  - Configure an IP address for each involved interface to ensure that neighboring devices can communicate with each other at the network layer.
  - Configure basic OSPF functions.

Configuring OSPF GR (2)

3. (Optional) Configure the GR Helper mode and set GR session parameters on the Helper.

```
[Huawei-ospf-1] graceful-restart [ period period | partial | planned-only ] * helper-role { { [ ip-prefix ip-prefix-name | acl-number acl-number | acl-name acl-name ] | ignore-external-lsa | planned-only } * | never }
```

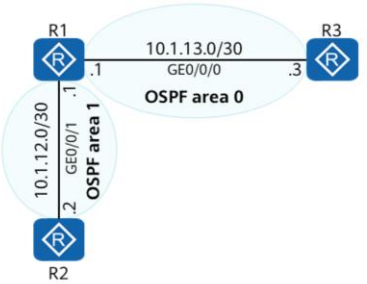
period *period*: specifies a GR period. The value is an integer ranging from 1 to 1800, in seconds. The default value is 120.
planned-only: indicates that the device supports only planned GR. By default, a device supports both planned GR and unplanned GR.
partial: indicates that the device supports partial GR. By default, the device supports totally GR.
ip-prefix *ip-prefix-name*: specifies the name of an IP prefix list.
acl-number *acl-number*: specifies the number of a basic ACL. An ACL can be used to configure a filtering policy, and the device enters the Helper mode only after received packets match the filtering policy.
acl-name *acl-name*: specifies the name of an ACL.
ignore-external-lsa: configures the Helper not to check Type 5 LSAs or Type 7 LSAs. By default, a Helper checks external LSAs.
never: indicates that the device does not support the Helper mode.

4. Check OSPF GR configurations and OSPF GR information.

```
[Huawei] display ospf [ process-id ] graceful-restart [ verbose ]
```


HUAWEI

Example for Configuring OSPF GR (1)



- Routers R1, R2, and R3 each are equipped with two main control boards, which back up each other. The routers interwork through OSPF and support GR.
- It is required that traffic forwarding be not interrupted when R1 restarts an OSPF process or performs an active/standby switchover in GR mode.

1. Assign an IP address to each involved interface and configure OSPF on each device. The configuration details are omitted.

2. Enable OSPF GR on R1, R2, and R3.

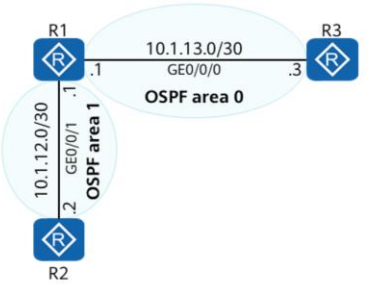
```
[R1] ospf 1
[R1-ospf-1] opaque-capability enable
[R1-ospf-1] graceful-restart
```

```
[R2] ospf 1
[R2-ospf-1] opaque-capability enable
[R2-ospf-1] graceful-restart
```

```
[R3] ospf 1
[R3-ospf-1] opaque-capability enable
[R3-ospf-1] graceful-restart
```


HUAWEI

Example for Configuring OSPF GR (2)



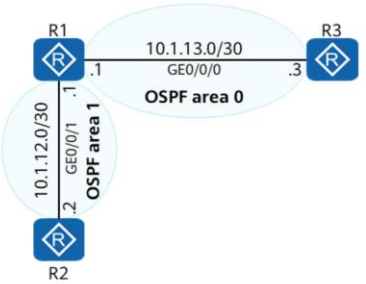
- Routers R1, R2, and R3 each are equipped with two main control boards, which back up each other. The routers interwork through OSPF and support GR.
- It is required that traffic forwarding be not interrupted when R1 restarts an OSPF process or performs an active/standby switchover in GR mode.

3. Check the OSPF GR status on each router. The following example uses the command output on R1.

```
<R1> display ospf 1 graceful-restart

                OSPF Process 1 with Router ID 10.1.13.1
Graceful-restart capability     : enabled
Graceful-restart support        : planned and un-planned, totally
Helper-policy support  : planned and un-planned, strict lsa check
Current GR state                : normal
Graceful-restart period         : 120 seconds

Number of neighbors under helper:
 Normal neighbors     : 0
 Virtual neighbors    : 0
 Sham-link neighbors : 0
 Total neighbors      : 0

Number of restarting neighbors : 0

Last exit reason:
 On graceful restart : none
 On Helper           : none
```


HUAWEI

Example for Configuring OSPF GR (3)



- Routers R1, R2, and R3 each are equipped with two main control boards, which back up each other. The routers interwork through OSPF and support GR.
- It is required that traffic forwarding be not interrupted when R1 restarts an OSPF process or performs an active/standby switchover in GR mode.

4. Run the **reset ospf process graceful-restart** command in the user view of R1 to restart OSPF process 1, and run the **display ospf peer** command on R2 to check the OSPF neighbor relationship with R1.

```
<R1> reset ospf 1 process graceful-restart
```

```
<R2> display ospf 1 peer

         OSPF Process 1 with Router ID 10.1.12.2
                  Neighbors

 Area 0.0.0.1 interface 10.1.12.2(GigabitEthernet0/0/0)'s neighbors
 Router ID: 10.1.13.1    Address: 10.1.12.1      GR State: Doing GR
   State: Full           Mode: Nbr is  Master    Priority: 1
   DR: 10.1.12.1         BDR: 10.1.12.2          MTU: 0
   Dead timer due in 32  sec
   Retrans timer interval: 5
   Neighbor is up for 00:00:13
   Authentication Sequence: [ 0 ]
```

The OSPF neighbor state on R2 is Full, indicating that the neighbor relationship between R1 and R2 is not interrupted when R1 restarts the OSPF process in GR mode.

HUAWEI

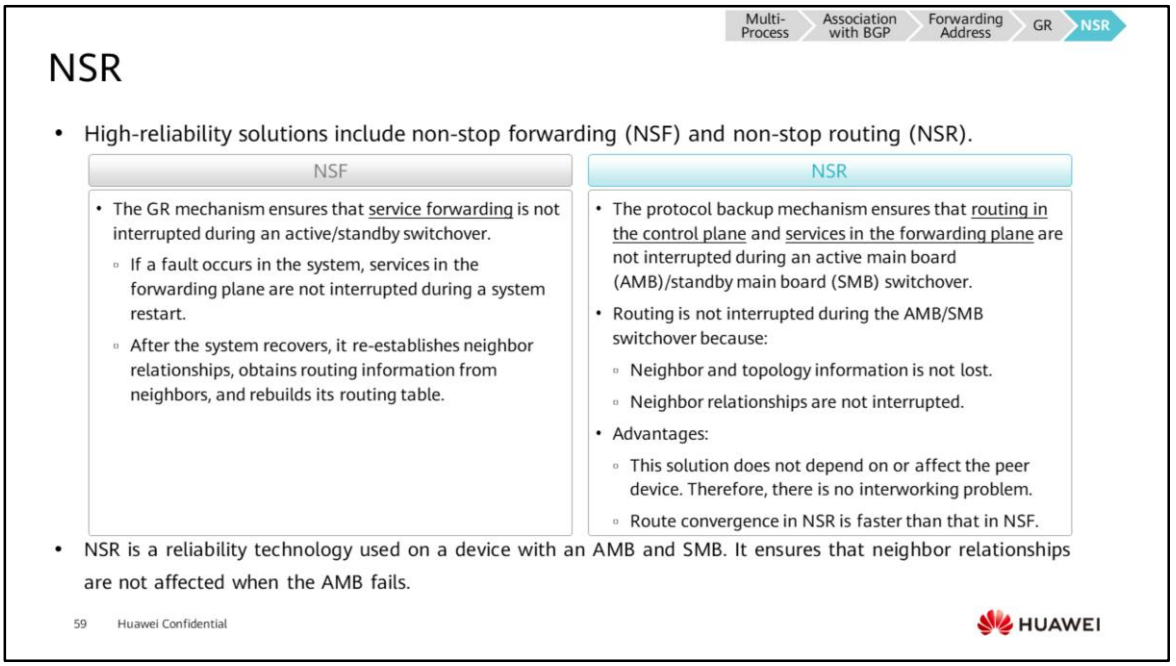
Example for Configuring OSPF GR (4)

5. Check the OSPF GR state on the routers. According to the following command outputs, R1 has exited the GR state normally, and R2 has exited from the Helper state normally.

```
<R1> display ospf 1 graceful-restart

                OSPF Process 1 with Router ID 10.1.13.1
Graceful-restart capability     : enabled
Graceful-restart support        : planned and un-planned, totally
Helper-policy support           : planned and un-planned, strict lsa check
Current GR state                : normal
Graceful-restart period         : 120 seconds

Number of neighbors under helper:
 Normal neighbors     : 0
 Virtual neighbors    : 0
 Sham-link neighbors  : 0
 Total neighbors      : 0

Number of restarting neighbors : 0

Last exit reason:
 On graceful restart : successful exit
 On Helper           : none
```

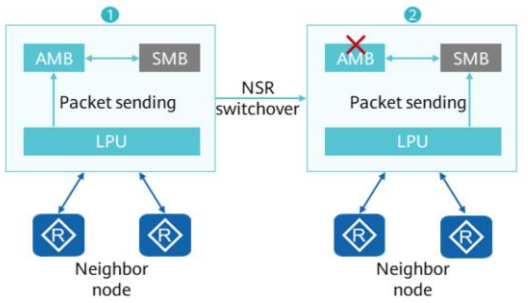
```
<R2> display ospf 1 graceful-restart

                OSPF Process 1 with Router ID 10.1.12.2
Graceful-restart capability     : enabled
Graceful-restart support        : planned and un-planned, totally
Helper-policy support           : planned and un-planned, strict lsa check
Current GR state                : normal
Graceful-restart period         : 120 seconds

Number of neighbors under helper:
 Normal neighbors     : 0
 Virtual neighbors    : 0
 Sham-link neighbors  : 0
 Total neighbors      : 0

Number of restarting neighbors : 0

Last exit reason:
 On graceful restart : none
 On Helper           : successful exit
```


HUAWEI



- NSR and GR:
  - The system in which an AMB/SMB switchover is performed supports two types of high-reliability protection: NSR and GR.
  - NSR and GR, however, are mutually exclusive. That is, for a specific protocol, only one of them can be used after a switchover.
  - When NSR is deployed on a device, the device can still function as a GR Helper to support the GR process of its neighbor, ensuring high reliability of services on all nodes on the network to the greatest extent.
- Application scenario:
  - NSF can be used if a network has low requirements for the packet loss rate and route convergence.
  - NSR can be used if a network has high requirements for the packet loss rate and route convergence.
- System-level NSR is triggered in the following situations:
  - A system fault triggers an AMB/SMB switchover.
  - The network administrator manually triggers an AMB/SMB switchover during a software upgrade or maintenance.
- Note: NSR fundamentals are the same in OSPF, IS-IS, and BGP. This course uses OSPF as an example.

NSR Fundamentals



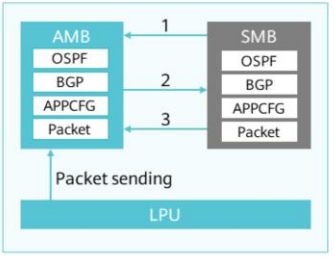
1. The routing protocols in the control plane back up routing information in real time.
2. The hardware channel detects the exception on the AMB, the SMB is instructed to become the new AMB, and LPUs are instructed to send packets to the new AMB.

- NSR is implemented in three phases:
  - Batch backup: After NSR is enabled and the SMB restarts, the AMB sends routing and forwarding information in batches to the SMB for backup. Batch backup is performed before real-time backup. An AMB/SMB switchover cannot be performed during batch backup in NSR.
  - Real-time backup: Real-time backup starts after batch backup is complete. All updates in the control and forwarding planes are backed up from the AMB to the SMB in real time. In this phase, the SMB can take over services from the AMB at any time.
  - AMB/SMB switchover: If the AMB fails in the NSR system in which backup has been completed, the SMB detects the failure through hardware status detection and becomes the new AMB. The new AMB instructs LPUs to send packets to itself. The AMB/SMB switchover is completed rapidly, with routing between the local node and its neighbor nodes uninterrupted.

HUAWEI

- High availability (HA): implements data backup from the AMB to the SMB.
- AMB and SMB: implement control plane processes.
- Line processing unit (LPU): implements forwarding plane processes.

NSR Switchover Phase: Batch Backup

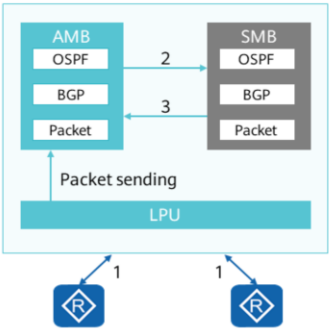


1. The SMB completes startup.
2. Service processes start backing up service data such as routes in batches through the HA channel. The system enters the batch backup phase.
3. After service processes complete batch backup, the system enters the redundancy protection state.

- After NSR is enabled and the SMB restarts, the service process on the AMB receives a message indicating that the SMB goes online. Service processes then start to back up internal data in batches.
  - After batch backup is complete, the device enters the redundancy protection state. If the AMB fails in this case, the SMB becomes the new AMB based on the data backed up from the original AMB and restores services.
  - If the AMB fails before batch backup is complete, the SMB may fail to become the new AMB due to incomplete service data. As a result, an NSR switchover cannot be performed; in this case, the device restarts and restores the pre-fault status.

HUAWEI

NSR Switchover Phase: Real-Time Backup

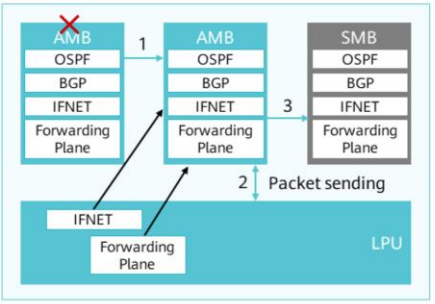


1. The neighbor status or routing information changes.
2. The AMB backs up updated information to the SMB through the HA channel.
3. The SMB sends an acknowledgement to the AMB.

- After batch backup is complete, the device enters the real-time backup phase. If the neighbor status or routing information changes in this phase, the AMB backs up the updated information to the SMB in real time.

HUAWEI

NSR Switchover Phase: AMB/SMB Switchover



1. The hardware channel detects the exception on the AMB, the SMB is instructed to become the new AMB, and LPUs are instructed to send packets to the new AMB.
2. After the original SMB becomes the new AMB, LPUs send the information updated during the switchover to the new AMB.

- After the batch backup is complete, the system enters the redundancy protection state. If a software or hardware fault occurs on the AMB, the SMB detects the fault on the AMB through underlying hardware detection and automatically becomes the new AMB.
- The new AMB uses the data backed up from the original AMB to process services. At the same time, LPUs send the information updated during the switchover to the new AMB. Routing and traffic forwarding are not interrupted in this case.

3. After the switchover is complete, the new AMB performs NSR backup to the new SMB.

HUAWEI

Configuring NSR

1. Enable NSR.

   [Huawei] **switchover mode** { **nonstop-routing** | **nonstop-forwarding** }

   This command is used to set an HA working mode. The default HA working mode is NSF.
   nonstop-routing: sets the HA working mode to NSR. In NSR mode, the system can process routes and forward services without interruption.
   nonstop-forwarding: sets the HA working mode to NSF. The NSF mode greatly reduces the service interruption time.

2. Verify the configuration.

   [Huawei] **display switchover mode**

   Check the HA working mode of the current system.

   [Huawei] **display ip routing-table** [**vpn-instance** *vpn-instance-name*] [**verbose**]

   Check post-AMB/SMB switchover routing information.

   [Huawei] **display fib** [*slot-id*] [**vpn-instance** *vpn-instance-name*] [**verbose**]

   Check post-AMB/SMB switchover forwarding information.

HUAWEI

Contents

1. OSPF Fast Convergence
2. OSPF Route Control
3. Other OSPF Features
4. **Advanced IS-IS Features**

HUAWEI

Overview of IS-IS Fast Convergence

- IS-IS fast convergence is an extended feature designed to speed up route convergence. It provides a series of functions, covering incremental SPF (I-SPF), PRC, intelligent timer, and LSP fast flooding.
- IS-IS supports fast convergence after a fault is rectified. For example, IS-IS auto FRR can be used to quickly switch traffic to a backup link, and IS-IS can be associated with BFD to quickly detect faults.

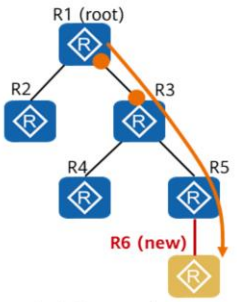
HUAWEI

- The functions, including PRC, intelligent timer, and FRR of IS-IS are similar to those of OSPF and therefore not detailed here.

Fast Convergence | Routing Control | Other Features

I-SPF

- I-SPF implementation: When the network topology changes, I-SPF recalculates routes only for affected nodes instead of all nodes, speeding up route calculation.

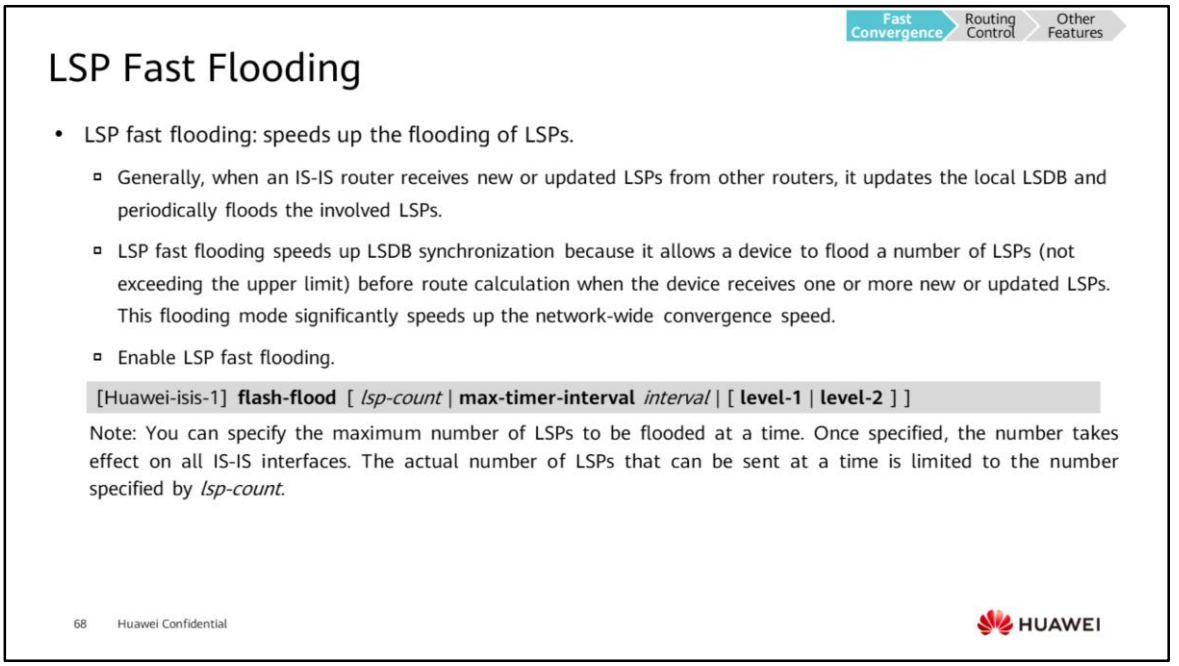


In route calculation, a node represents a router, and a leaf represents a route. I-SPF processes the information of only changed nodes.

- Scenario description:
  - On an IS-IS network, an SPT with R1 as the root is calculated after route convergence, as shown in the left figure. When R1 accesses R5, the traffic of R1 is forwarded to R5 based on [the outbound interface of R1's downlink and the IP address of the inbound interface of R3's uplink].
  - R6 is added as a downstream device of R5. IS-IS is enabled on R6, meaning that there is a new network node on the IS-IS network.
- I-SPF calculation:
  - R5 and R6 both flood LSPs carrying information about their neighbor relationship on the entire network.
  - After receiving such an LSP, R1 performs I-SPF calculation only for R5 and R6 to generate a new SPT.
  - Therefore, when R1 accesses R5 and R6, the traffic of R1 is forwarded to R5 and R6 based on [the outbound interface of R1's downlink and the IP address of the inbound interface of R3's uplink].

HUAWEI

- SPF for route calculation: If a node on a network changes, SPF recalculates routes for all the nodes on the network, which takes a long time, consumes a large number of CPU resources, and consequently reduces the network-wide convergence speed.
- I-SPF is an improvement of SPF. Unlike SPF that calculates all nodes, I-SPF calculates only affected nodes. The SPT generated using I-SPF is the same as that generated using SPF. This significantly decreases CPU usage and speeds up network convergence.
- I-SPF and PRC are used together on an IS-IS network.
  - If the SPT calculated by I-SPF changes, PRC processes all the leaves (routes) of only the changed node.
  - If the SPT calculated by I-SPF does not change, PRC processes only the changed leaves (routes). For example, if IS-IS is newly enabled on an interface of a node, the SPT on the network remains unchanged. In this case, PRC updates only the routes of this interface, which consumes less CPU resources.



- Command: [Huawei-isis-1] **flash-flood** [*lsp-count* **| max-timer-interval** *interval* | [**level-1 | level-2**]]
  - *lsp-count:* specifies the maximum number of LSPs that can be flooded on each interface at a time. The value is an integer ranging from 1 to 15. The default value is 5.
  - **max-timer-interval** *interval:* specifies the maximum interval at which LSPs are flooded. The value is an integer ranging from 10 to 50000, in milliseconds. The default value is 10.
  - **level-1**: enables the LSP flash-flood function in the Level-1 area. If no level is specified in the command, this function is enabled in both Level-1 and Level-2 areas.
  - **level-2**: enables the LSP flash-flood function in the Level-2 area. If no level is specified in the command, this function is enabled in both Level-1 and Level-2 areas.

Fast Convergence | Routing Control | Other Features

Overview of IS-IS Route Control

- In real-world applications, IS-IS routes calculated using SPF sometimes cannot meet network planning and traffic management requirements, which may lead to various problems, such as slow table lookup due to a large number of routing entries in routing tables and unbalanced link usage on a network. To optimize IS-IS networks and facilitate traffic management, more precise route control is required. You can use any of the following methods to implement the control:
  - Adjusting the IS-IS preference
  - Adjusting the IS-IS interface cost
  - Configuring equal-cost routes
  - Configuring IS-IS route leaking
  - Configuring default route advertisement
  - Importing external routes
  - Configuring filter-policies

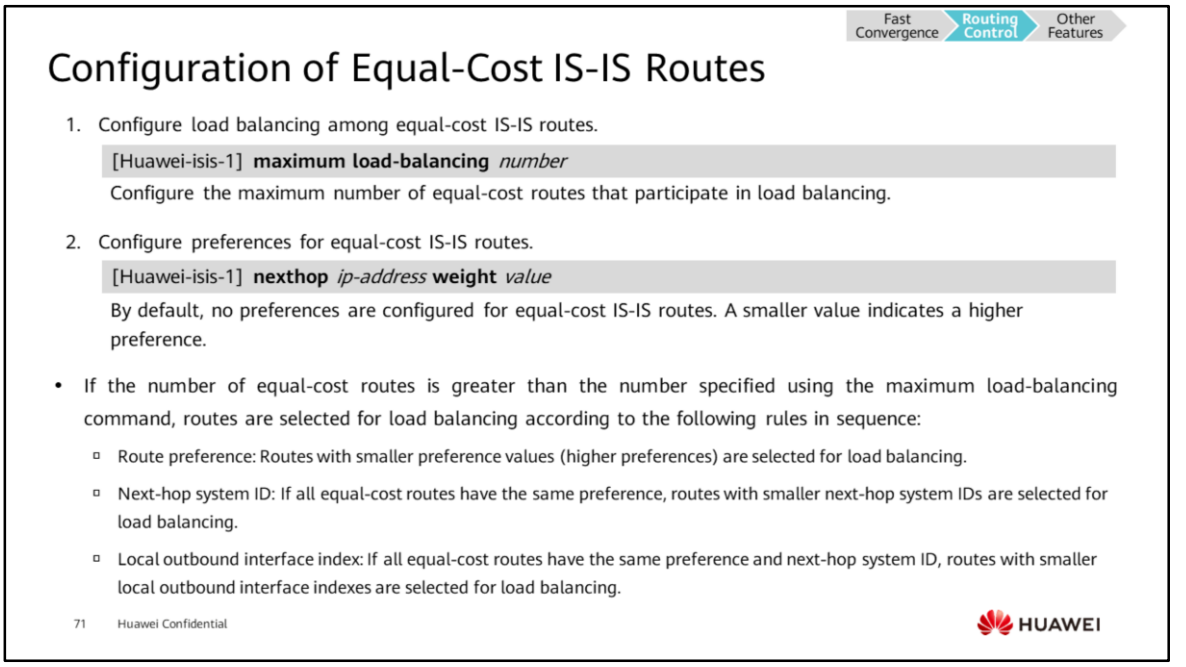
HUAWEI

- This course involves only equal-cost and default routes. For details about other control methods, see *HCIP-Datacom-Core Technology*.

Equal-Cost Route

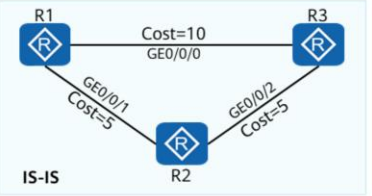
- If there are multiple redundant links on an IS-IS network, multiple equal-cost routes to the same destination may exist. In this case, you can use either of the following methods to configure equal-cost routes:
  - Configure load balancing so that traffic is evenly distributed to relevant links.
    - This method improves link utilization and reduces the possibility of congestion caused by overloaded links. However, because traffic will be randomly forwarded, this method may make traffic management difficult.
  - Configure a preference for each equal-cost route so that the route with the highest preference is preferentially selected and the others function as backups.
    - This method is used to specify the preferred route among multiple equal-cost routes, without the need to modify original configurations. It facilitates traffic management and improves network reliability.
    - Note: After preferences are configured for equal-cost routes, IS-IS devices forward traffic to the next hop with the highest preference, instead of forwarding traffic in load balancing mode.

HUAWEI



- Command: [Huawei-isis-1] **maximum load-balancing** *number*
  - *number:* specifies the maximum number of equal-cost routes that participate in load balancing. The value varies according to the device model.
- Command: [Huawei-isis-1] **nexthop** *ip-address* **weight** *value*
  - *ip-address*: specifies the IP address of the next hop. The value is in dotted decimal notation.
  - **weight** *value:* specifies the weight of the next hop. A smaller value indicates a higher preference. The value is an integer ranging from 1 to 254.

Example for Configuring Equal-Cost Routes (1)



Device	Interface	IP Address
R1	GE0/0/0	10.1.13.1/24
	GE0/0/1	10.1.12.1/24
R2	GE0/0/1	10.1.12.2/24
	GE0/0/2	10.1.23.2/24
R3	Loopback0	10.1.3.3/32
	GE0/0/0	10.1.13.3/24
	GE0/0/2	10.1.23.3/24

- On the network shown in the figure, IS-IS runs on R1, R2, and R3. It is required that R1 be able to access the loopback interface address of R3 through path R1 -> R3 or R1 -> R2 -> R3.

1. Assign IP addresses to device interfaces and configure IS-IS on each device. (Omitted)

2. Configure the number of IS-IS equal-cost routes for load balancing on R1.

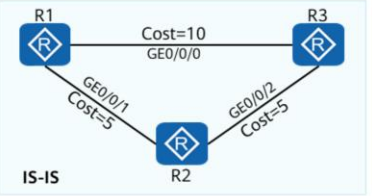
```
[R1] isis
[R1-isis-1] maximum load-balancing 2
```

3. Verify the configuration.

```
[R1]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
10.1.3.3/32         ISIS-L2 15   10          D   10.1.13.3       GigabitEthernet0/0/0
                    ISIS-L2 15   10          D   10.1.12.2       GigabitEthernet0/0/1
```


HUAWEI

Example for Configuring Equal-Cost Routes (2)



Device	Interface	IP Address
R1	GE0/0/0	10.1.13.1/24
	GE0/0/1	10.1.12.1/24
R2	GE0/0/1	10.1.12.2/24
	GE0/0/2	10.1.23.2/24
R3	Loopback0	10.1.3.3/32
	GE0/0/0	10.1.13.3/24
	GE0/0/2	10.1.23.3/24

- On the network shown in the figure, IS-IS runs on R1, R2, and R3. It is required that R1 be able to access the loopback interface address of R3 through path R1 -> R3 or R1 -> R2 -> R3.

1. Assign IP addresses to device interfaces and configure IS-IS on each device. (Omitted)

2. (Optional) Configure preference for equal-cost routes on R1.

```
[R1] isis
[R1-isis-1] nexthop 10.1.13.3 weight 1
[R1-isis-1] nexthop 10.1.12.2 weight 2
```

3. Verify the configuration.

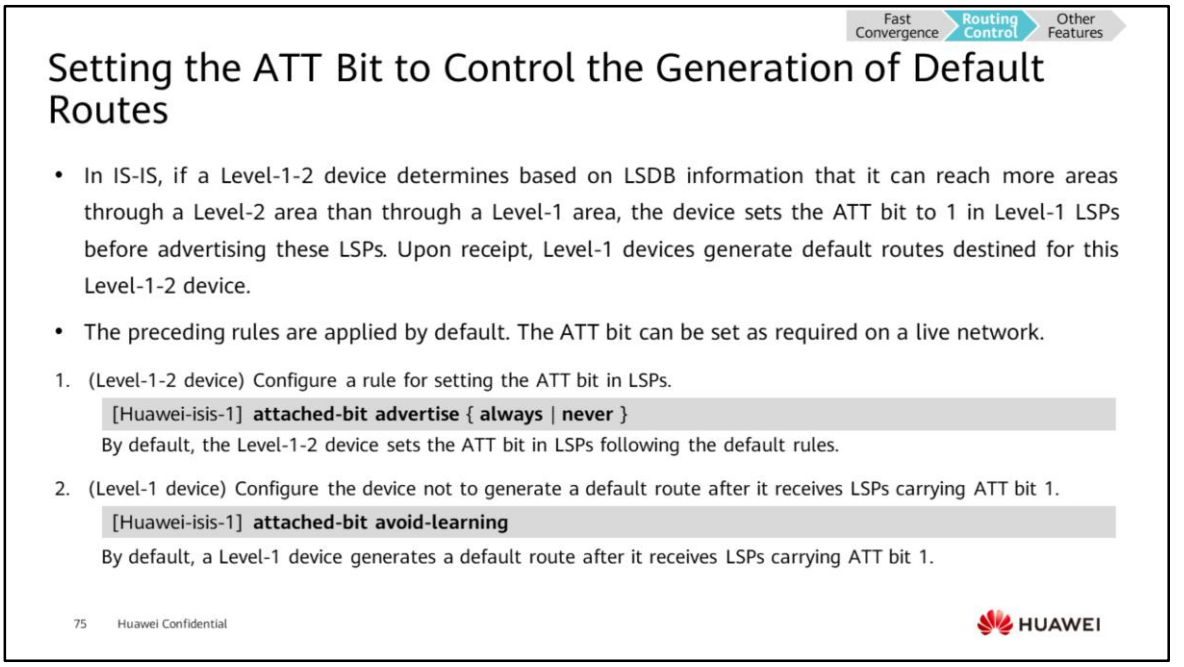
```
[R1]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
10.1.3.3/32         ISIS-L2 15   10          D   10.1.13.3       GigabitEthernet0/0/0
```


HUAWEI

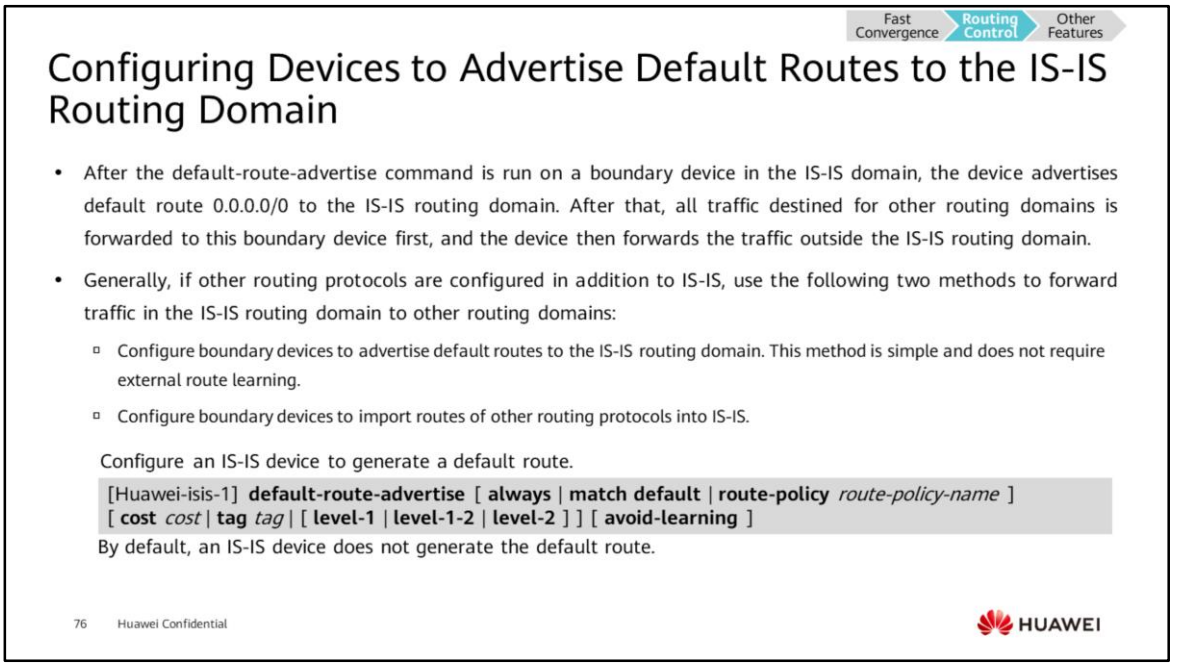
Default Route

- IS-IS allows you to control the generation and advertisement of default routes using the following methods:
  - On Level-1-2 devices, configure a rule for setting the attached (ATT) bit in Level-1 LSPs.
  - Configure Level-1 devices not to automatically generate default routes even if they receive Level-1 LSPs with the ATT bit set to 1.
  - Configure devices to advertise default routes to the IS-IS routing domain.
- LSP packet format:
  - The ATT bit is generated by a Level-1-2 router to identify whether the originating router connects to other areas. Note that the ATT field has four bits, and Huawei datacom products use only one of the four bits.

HUAWEI

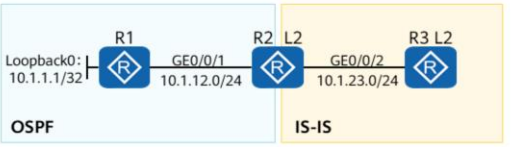


- Command: [Huawei-isis-1] **attached-bit advertise { always | never }**
  - **always**: indicates that the ATT bit is set to 1. After receiving an LSP with ATT bit 1, a Level-1 device generates a default route.
  - **never**: indicates that the ATT bit is set to 0. This prevents the Level-1 device from generating default routes and reduces the size of the routing table.
- Although the ATT bit is defined in both Level-1 and Level-2 LSPs, it is set to 1 only in Level-1 LSPs advertised by Level-1-2 devices. Therefore, this command takes effect only on Level-1-2 devices.
- To prevent Level-1 devices from advertising default routes to their routing tables, perform either of the following operations:
  - Run the **attached-bit advertise never** command on Level-1-2 devices to disable them from advertising LSPs with ATT bit 1.
  - Run the **attached-bit avoid-learning** command on Level-1 devices that connect to Level-1-2 devices.
- The difference between the preceding commands lies in that the attached-bit avoid-learning command applies to specified Level-1 devices.



- Command: [Huawei-isis-1] **default-route-advertise [always | match default | route-policy** *route-policy-name* **] [cost** *cost* **| tag** *tag* **| [level-1 | level-1-2 | level-2]] [avoid-learning]**
  - **always:** configures an IS-IS device to unconditionally advertise the default route and set itself as the next hop in the route.
  - **match default:** advertises the default route generated by another routing protocol or IS-IS process through LSPs if such a route already exists in the routing table.
  - **route-policy** *route-policy-name*: specifies the name of the route-policy. A Level-1-2 device advertises the default route to the IS-IS routing domain only when external routes matching the route-policy exist in the routing table of the device. This prevents routing blackholes caused by the advertisement of the default route when link faults make some important external routes unavailable. This route-policy does not affect the import of external routes into IS-IS. 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 (").
  - **cost** *cost:* specifies the cost of the default route. The value is an integer. The value range depends on cost-style. When cost-style is narrow, narrow-compatible, or compatible, the value ranges from 0 to 63. When cost-style is wide or wide-compatible, the value ranges from 0 to 4261412864.

Example for Configuring Default Routes



- As shown in the figure, OSPF runs on GE0/0/1 and Loopback0 of R1 and GE0/0/1 of R2, and IS-IS runs on GE0/0/2 of R2 and R3.
- To enable R3 to access the OSPF network, configure R2 to advertise the default route to the IS-IS routing domain.
- Similarly, R2 also needs to advertise the default route to the OSPF routing domain.

1. Assign an IP address to each interface and configure OSPF/IS-IS on each device.
2. Configure R2 to advertise the default route to the IS-IS routing domain.

```
[R2] isis
[R2-isis-1] default-route-advertise always
```

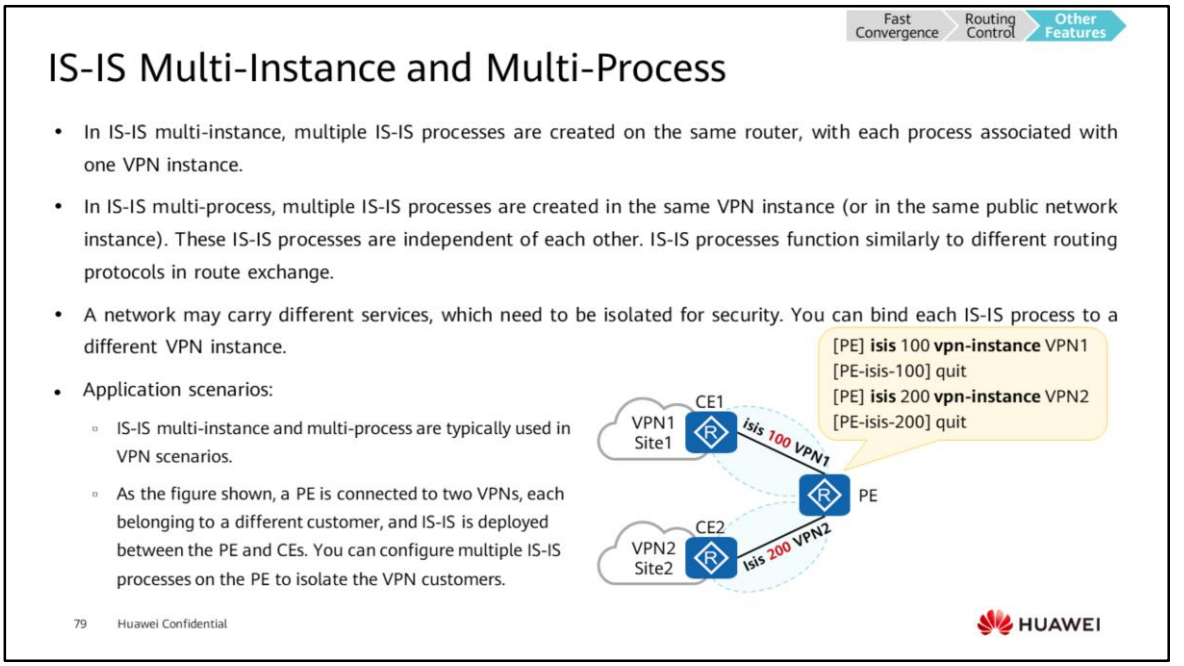
3. Configure R2 to advertise the default route to the OSPF routing domain.

```
[R2] ospf
[R2-ospf-1] default-route-advertise always
```

4. Check routing entries on R3.

```
[R3]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: Public
         Destinations : 8        Routes : 8
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
0.0.0.0/0           ISIS-L2 15   10          D   10.1.23.2       GigabitEthernet0/0/2
......
```

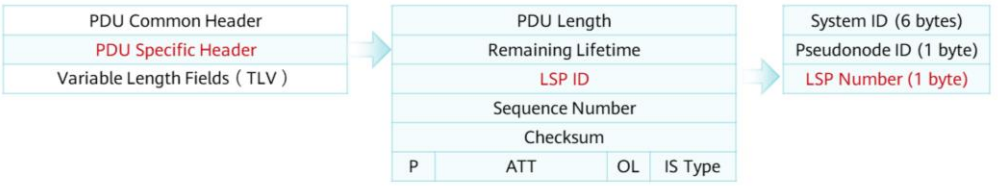

HUAWEI



- IS-IS multi-process and multi-instance have the following characteristics:
  - In IS-IS multi-process, processes share the same global routing table. IS-IS multi-instance, however, uses the routing tables of VPNs, with each VPN having a separate routing table.
  - IS-IS multi-process allows a set of interfaces to be associated with a specified IS-IS process. This ensures that the protocol operations in the specified IS-IS process are confined only to this set of interfaces. In this way, multiple IS-IS processes can work on a single router, with each process responsible for a unique set of interfaces.
  - When creating an IS-IS process, you can bind it to a VPN instance. The IS-IS process then accepts and processes only the events related to the VPN instance. When the bound VPN instance is deleted, the IS-IS process is also deleted.
- Command: [Huawei] **isis** [*process-id*] [**vpn-instance** *vpn-instance-name*]
  - *process-id*: specifies the ID of an IS-IS process. If no IS-IS process is specified, IS-IS process 1 is started. The value is an integer ranging from 1 to 65535. The default value is 1.
  - **vpn-instance** *vpn-instance-name*: specifies the name of a VPN instance. If this parameter is not specified, no VPN instance is associated with the IS-IS process. The value is a string of 1 to 31 case-sensitive characters. If spaces are used, the string must start and end with double quotation marks (").

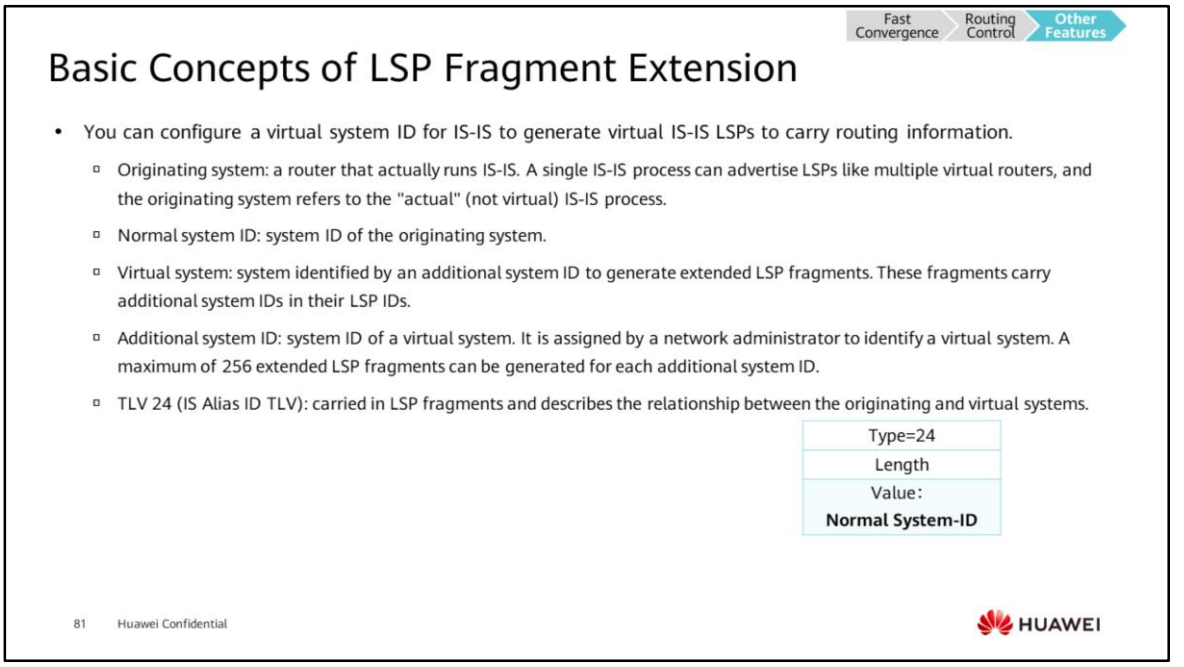
LSP Fragment

- When a PDU to be advertised by IS-IS contains too much information, an IS-IS router generates LSP fragments to carry the information.
- LSP packet format

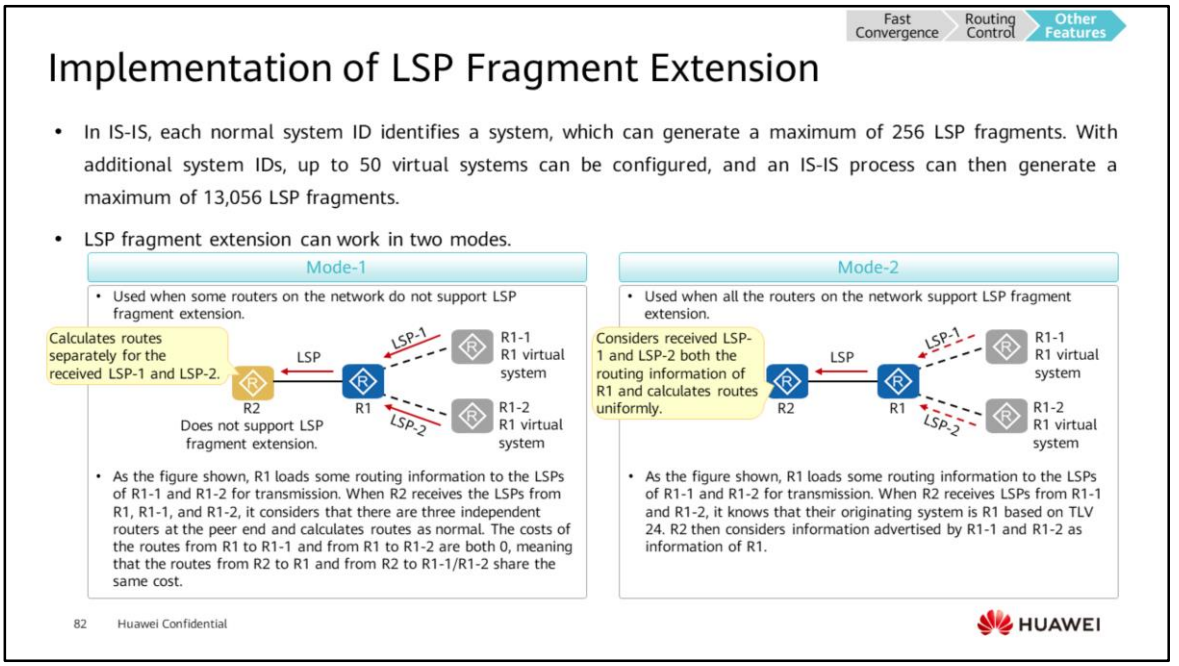


- An IS-IS LSP fragment is identified by the 1-byte LSP Number field in LSP ID. So, an IS-IS process can generate a maximum of 256 LSP fragments, which means only limited information can be carried.

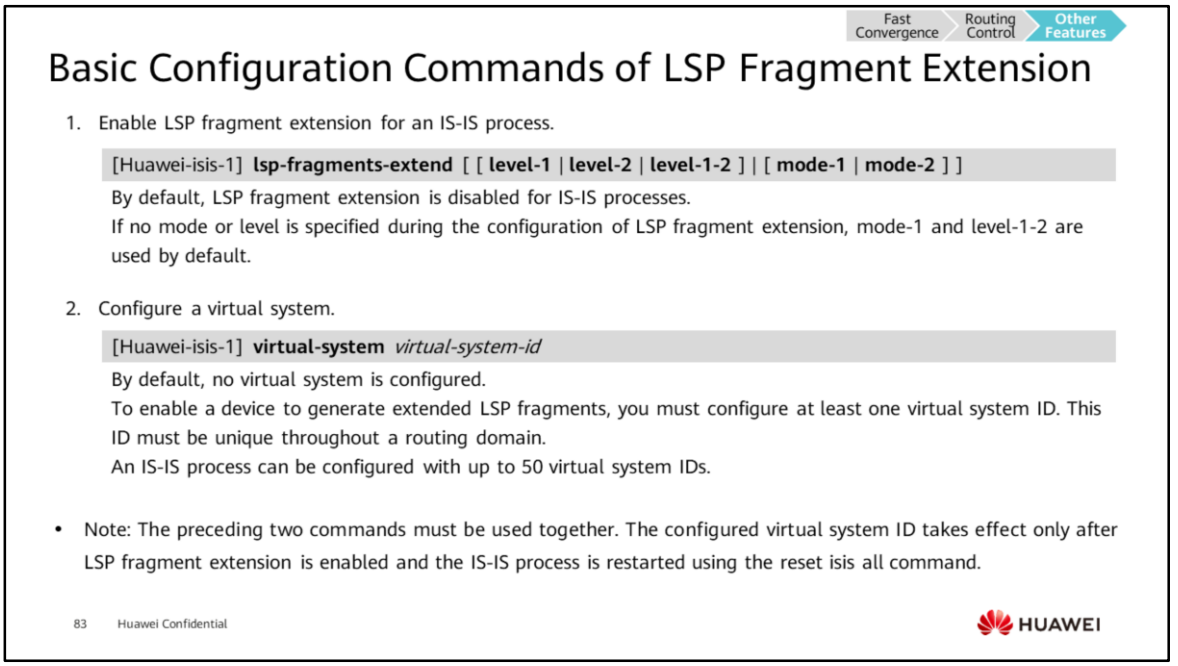
HUAWEI



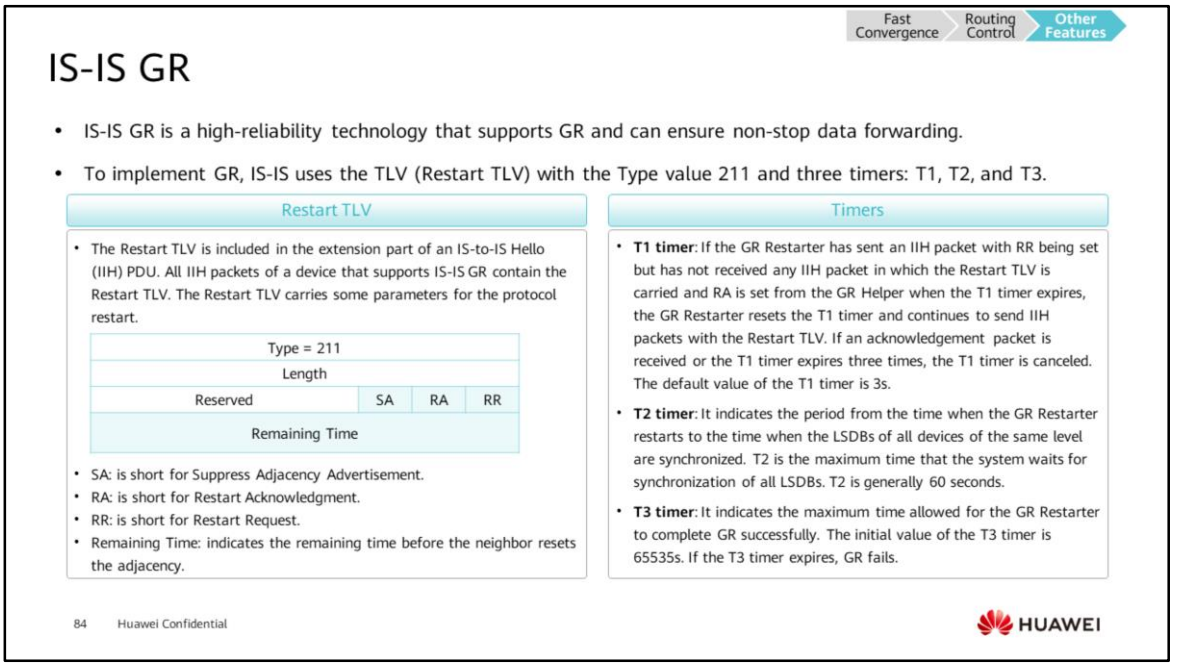
- After LSP fragment extension is configured, the system prompts you to restart the IS-IS process if information is lost because LSPs overflow. After the IS-IS process is restarted, the originating system loads as much routing information as possible to its LSPs. The information that cannot be loaded is placed in the LSPs of virtual systems for transmission. The originating system then notifies other routers of its relationship with the virtual systems through TLV 24.
- Note: the additional and normal system IDs must be unique throughout a routing domain.



- Mode-1 implementation:
  - Virtual systems participate in SPF calculation. The LSPs advertised by the originating system contain information about links to each virtual system. Similarly, the LSPs advertised by each virtual system contain information about links to the originating system. In this way, virtual systems function like physical routers that connect to the originating system.
  - Mode-1 is a transitional mode used to support earlier versions that are incapable of LSP fragment extension. In these earlier versions, IS-IS cannot identify TLV 24. As a result, the LSPs sent by a virtual system must look like LSPs sent by an originating system.
  - Precautions:
    - The LSPs sent by a virtual system must contain the same area address and overload bit as those in LSPs sent by an originating system. Other TLVs must also be the same.
    - The neighbor of a virtual system must point to an originating system, and the metric is the maximum value minus 1. The neighbor of the originating system must point to the virtual system, and the metric must be 0. This ensures that the virtual system is the downstream node of the originating system when other routers calculate routes.
- Mode-2 implementation:
  - Virtual systems do not participate in SPF calculation. All the routers on the network know that the LSPs generated by the virtual systems actually belong to the originating system.
  - IS-IS working in mode-2 can identify TLV 24, which is used as the basis for calculating an SPT and routes.
- Note: In both modes, the originating system and virtual systems must include TLV 24 in their LSPs whose LSP Number is 0 to indicate which is the originating system.



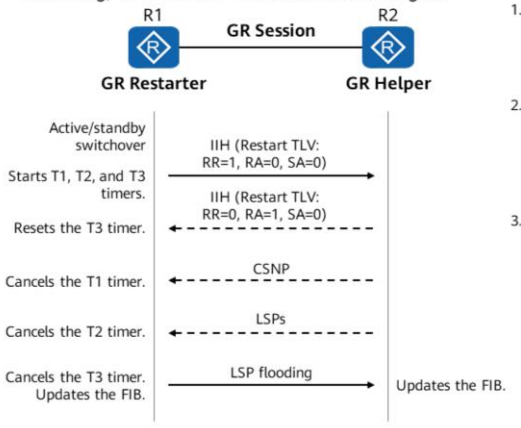
- Command: [Huawei-isis-1] **lsp-fragments-extend [[level-1 | level-2 | level-1-2] | [mode-1 | mode-2]]**
  - **level-1:** enables LSP fragment extension in Level-1.
  - **level-2:** enables LSP fragment extension in Level-2.
  - **level-1-2:** enables LSP fragment extension in Level-1-2.
  - **mode-1:** allows routers to be compatible with other routers running earlier versions that are incapable of LSP fragment extension.
  - **mode-2:** requires all routers to support LSP fragment extension.
- Command: [Huawei-isis-1] **virtual-system** *virtual-system-id*
  - *virtual-system-id:* specifies a virtual system ID of an IS-IS process. The length is 6 bytes (48 bits), and the format is XXXX.XXXX.XXXX.



- Background:
  - After an active/standby switchover is performed on a device, because the device has not stored any information about the neighbor relationships established before the restart, the initial Hello packets sent by the device carry an empty neighbor list. After receiving the Hello packets, a neighbor (Helper device) performs the two-way neighbor relationship check. After the neighbor detects that it is not in the neighbor list in the Hello packets sent by the Restarter, the neighbor tears down the neighbor relationship. The neighbor then generates new LSPs and floods the topology changes to all other devices in the area. The other devices in the area then calculate routes based on their new LSDBs, leading to routing interruptions or routing loops.
  - The IETF defined the GR standard (RFC 3847) for IS-IS. The protocol restarts in which the FIB is retained and the protocol restarts in which the FIB is not retained are both processed, preventing route flapping and traffic interruptions caused by protocol restarts.

IS-IS GR Process (Restarting) (1)

- The GR process triggered by an active/standby switchover or the restart of an IS-IS process is referred to as restarting, in which the FIB remains unchanged.



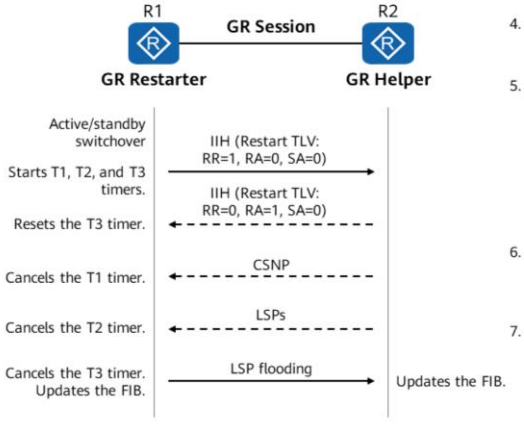
1. After performing a protocol restart, the GR Restarter performs the following actions:
   - Starts T1, T2, and T3 timers.
   - Sends IIH packets that contain the Restart TLV through all interfaces. In the packets, RR is set, and RA and SA are cleared.
2. After receiving an IIH packet, the GR Helper performs the following actions:
   - Maintains the neighbor relationship and updates the current Holdtime.
   - Replies with an IIH packet containing the Restart TLV. In the packet, RR is cleared, RA is set, and Remaining Time indicates the period from the current moment to the timeout of the Holdtime.
   - Sends CSNPs and all LSPs to the GR restarter.
3. After receiving the IIH packet, in which RR is 0 and RA is 1, from the neighbor, the GR Restarter performs the following actions:
   - Compares the current value of the T3 timer with the value of Remaining Time in the packet. The smaller value is used as the value of the T3 timer.
   - Cancels the T1 timer maintained by the interface when the interface receives an acknowledgement and CSNPs.
   - Resets the T1 timer and resends the IIH packet that contains the Restart TLV when the T1 timer expires but the interface has not received any acknowledgement or CSNPs. The reset may be performed multiple times. If the expiry count of the T1 timer exceeds the threshold value, the GR Restarter forcibly cancels the T1 timer and starts normal IS-IS processing.

HUAWEI

- Notes:
  - In Step 2, if the neighbor does not have the GR Helper capability, it ignores the Restart TLV and resets the adjacency with the GR Restarter according to normal IS-IS processing.
  - In Step 3, the Restarter sets the value of the T3 timer to the Holdtime of the neighbor, preventing neighbor disconnection during the GR, which would otherwise cause routes to be recalculated on the whole network.

IS-IS GR Process (Restarting) (2)

- The GR process triggered by an active/standby switchover or the restart of an IS-IS process is referred to as restarting, in which the FIB remains unchanged.



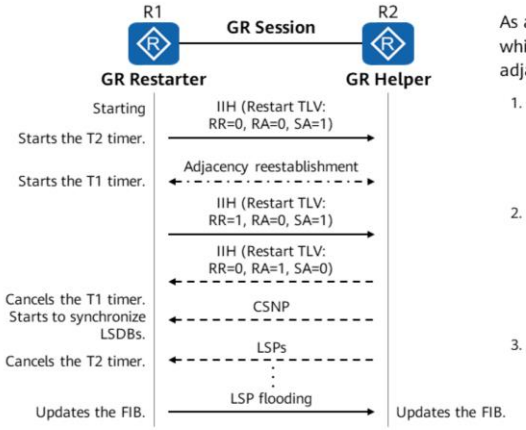
4. After the GR Restarter cancels the T1 timers on all interfaces, clears the CSNP list, and collects all LSPs, it considers the synchronization with all neighbors complete and cancels the T2 timer.
5. The cancellation of the T2 timer indicates that LSDBs of the corresponding level are synchronous.
   - In the case of a single-level device, the SPF calculation is triggered directly.
   - In the case of a Level-1-2 device, the device checks whether the T2 timer of the other level is also canceled. If the T2 timers of both levels are canceled, SPF calculation is triggered. Otherwise, the device waits for the T2 timer of the other level to expire.
6. After the T2 timers of both levels are canceled, the GR Restarter cancels the T3 timer and updates its FIB. The GR Restarter re-generates the LSPs of each level and floods them. During synchronization, if the GR Restarter receives the LSPs that were generated by itself before the restarting, it deletes them.
7. At this point, the IS-IS restarting of the GR Restarter is complete.

HUAWEI

- During the restarting, the Restarter starts the T1, T2, and T3 timers at the same time after the protocol restart. The value of the T1 timer indicates the longest time during which the GR Restarter waits for the Hello packet used for GR acknowledgement from the GR Helper. The value of the T2 timer indicates the longest time during which the system waits for the LSDB synchronization. The value of the T3 timer indicates the longest time allowed for a GR. The device cancels the T3 timer after synchronization of Level-1 and Level-2 LSDBs completes during the GR. If LSDB synchronization has not completed when the T3 timer expires, the GR fails.

IS-IS GR Process (Starting) (1)

- The GR process triggered by a device restart is referred to as starting, in which the FIB is updated.



As a starting device does not retain its FIB, it wants the neighbors with which the adjacencies are up before the device starts, to reset the adjacencies and suppress advertisement of the adjacencies in a period.

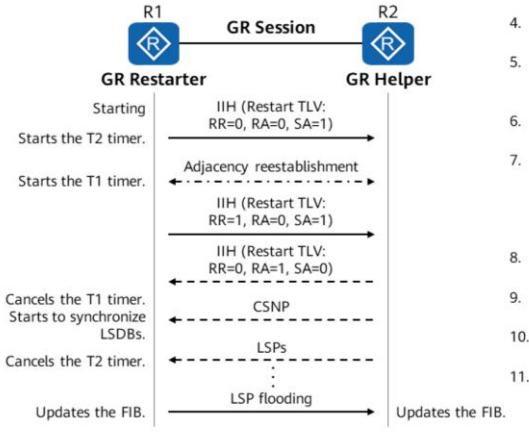
1. After the GR Restarter starts, it performs the following actions:
   - Starts one T2 timer for synchronization of LSDBs at each level.
   - Sends IIH packets that contain the Restart TLV through all interfaces. In the packets, RR is cleared, and SA is set.
2. After receiving an IIH packet that carries the Restart TLV, a neighbor performs the following actions based on whether GR is supported:
   - **If GR is supported**, the neighbor re-initiates the adjacency.
   - **If GR is not supported**, the neighbor ignores the Restart TLV and resets the adjacency with the GR Restarter.
3. After adjacencies are re-initiated, the GR Restarter re-establishes them with neighbors on each interface. When an adjacency goes up, the GR Restarter starts the T1 timer for the interface.

HUAWEI

- Additional remarks for Step 1:
  - If RR is cleared, starting has completed.
  - The IIH packet in which SA is set indicates that the Restarter requests its neighbor to suppress the advertisement of their adjacency until the neighbor receives an IIH packet in which SA is cleared from the Restarter.
- Additional remarks for Step 2:
  - **If GR is supported**, the neighbor re-initiates the adjacency with the GR Restarter and deletes the description of the adjacency from the LSPs to be sent. In addition, the neighbor ignores the link connected to the GR Restarter when performing SPF calculation until it receives an IIH packet in which SA is cleared from the GR Restarter.
  - **If GR is not supported**, the neighbor ignores the Restart TLV, resets the adjacency with the GR Restarter, replies with an IIH packet that does not contain the Restart TLV, and returns to normal IS-IS processing. In this case, the neighbor does not suppress the advertisement of the adjacency with the GR Restarter. In the case of a P2P link, the neighbor also sends a CSNP.

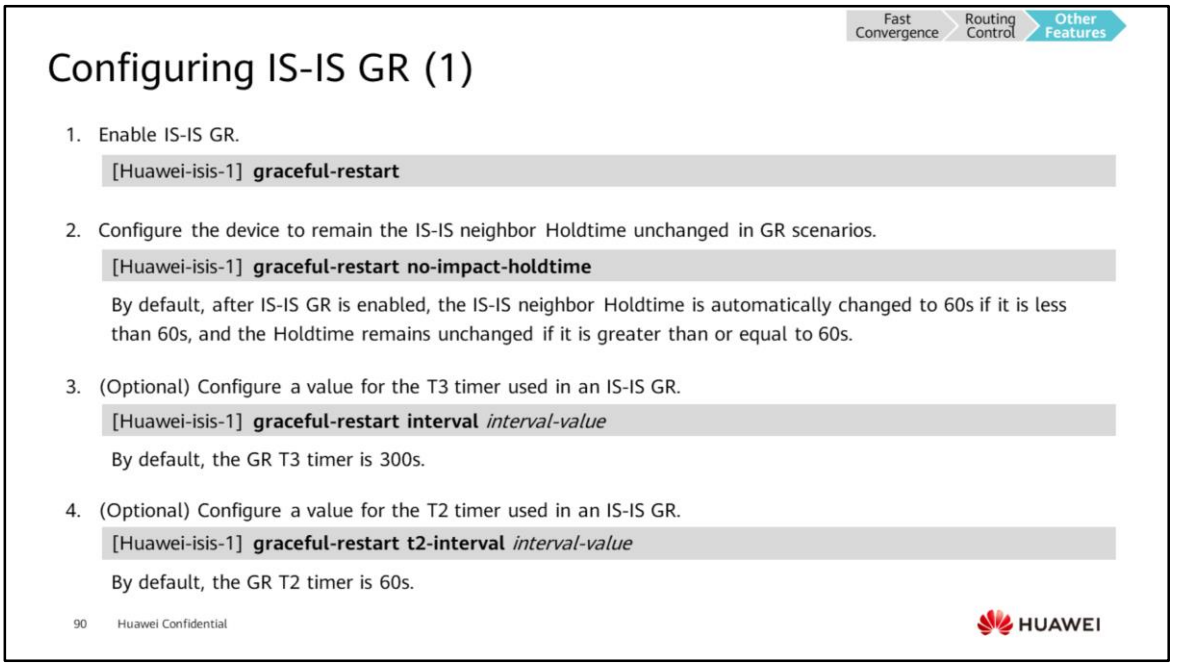
IS-IS GR Process (Starting) (2)

- The GR process triggered by a device restart is referred to as starting, in which the FIB is updated.

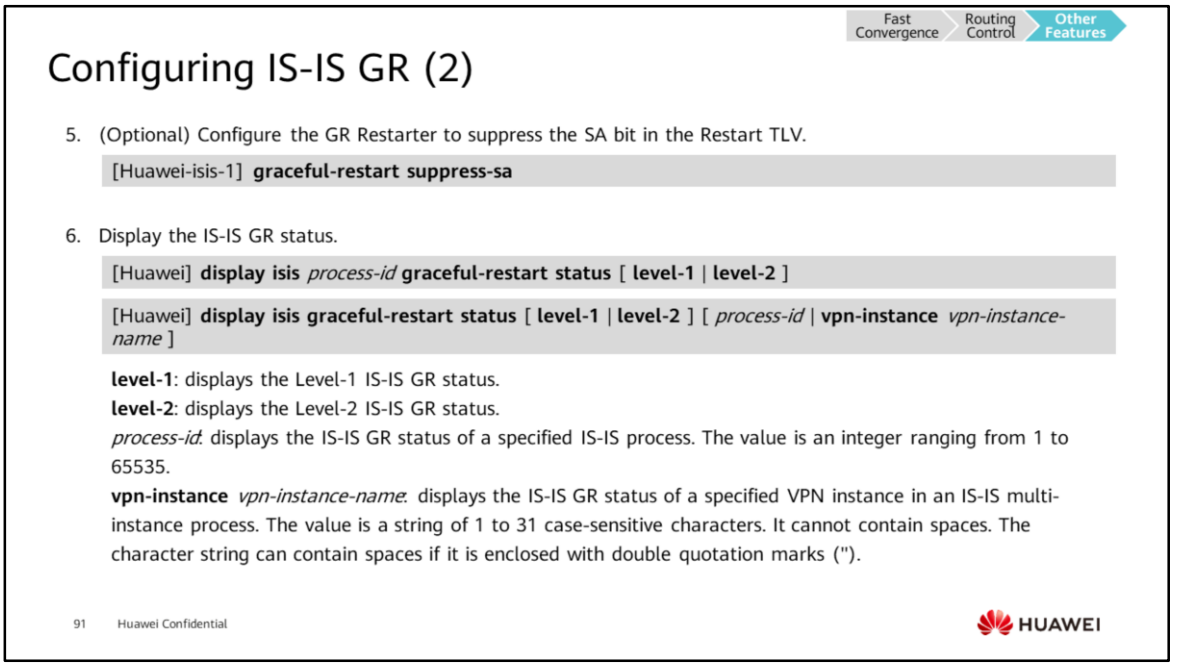


4. When the T1 timer expires, the GR Restarter sends an IIH packet in which both RR and SA are set.
5. After the neighbor receives the IIH packet in which both RR and SA are set, it replies with an IIH packet, in which RR is cleared and RA is set, and sends a CSNP.
6. After receiving the IIH packet used for acknowledgement and CSNP from the neighbor, the GR Restarter cancels the T1 timer.
7. If the GR Restarter does not receive an IIH packet or CSNP, it constantly resets the T1 timer and resends the IIH packet in which RR and SA are set. If the expiry count of the T1 timer exceeds the threshold value, the GR Restarter forcibly cancels the T1 timer and starts normal IS-IS processing to perform LSDB synchronization.
8. After receiving the CSNP from the Helper, the GR Restarter starts LSDB synchronization.
9. After LSDBs of a level are synchronized, the GR Restarter cancels the corresponding T2 timer.
10. After all T2 timers are canceled, SPF calculation is started and LSPs are regenerated and flooded.
11. At this point, the IS-IS starting of the GR Restarter is complete.

HUAWEI

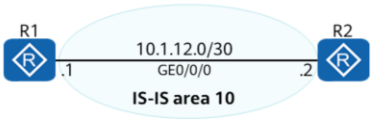


- Usage scenario of the **graceful-restart no-impact-holdtime** command:
  - On an IS-IS network, if GR is configured on a device, its IS-IS neighbors may automatically update corresponding neighbor Holdtimes. Specifically, the Holdtimes are changed to 60s if they are less than 60s, and remain unchanged if they are greater than or equal to 60s. Consequently, the minimum Holdtime value is 60s after the update. If one of the neighbors fails in non-GR scenarios, it takes at least 60s for the other end to detect the failure. As a result, a large number of packets may be discarded within this period, reducing network security and reliability.
  - To resolve this problem, you can run the **graceful-restart no-impact-holdtime** command so that the Holdtimes of neighbors remain unchanged in an IS-IS GR scenario. In this way, neighbor status can be fast detected, implementing rapid network convergence.
- **graceful-restart interval** *interval-value* and **graceful-restart t2-interval** *interval-value* commands
  - **interval** *interval-value*: specifies a value of the GR T3 or T2 timer. The value is an integer that ranges from 30 to 1800, in seconds.
  - It is recommended that the value of the T3 timer is greater than that of the T2 timer. Otherwise, the GR may fail.



- Usage scenario of the **graceful-restart suppress-sa** command:
  - A router does not maintain the forwarding status when it starts for the first time (excluding the post-GR cases). If it is not the first time that a router has started, the LSPs generated when the router ran last time may still exist in the LSP database of other routers on the network.
  - Because the sequence numbers of LSP fragments are also reinitialized when the router starts, the LSP copies stored on other routers seem to be newer than the LSPs generated after the local router starts. This leads to a temporary "blackhole" on the network, and the blackhole persists until the router regenerates its own LSPs and advertises them with the largest sequence number.
  - If the neighbors of the router are suppressed from advertising adjacencies to the router during the startup of the router until the router advertises updated LSPs, the preceding problem can be prevented. Therefore, you can run the **graceful-restart suppress-sa** command to suppress the SA bit in the Restart TLV.

Example for Configuring IS-IS GR (1)



- Routers R1 and R2 each are equipped with two main control boards, which back up each other. The routers interwork through IS-IS and support GR.
- It is required that traffic forwarding be not interrupted when R1 restarts an IS-IS process or performs an active/standby switchover in GR mode.

1. Assign an IP address to each involved interface and configure IS-IS on each device. The configuration details are omitted.

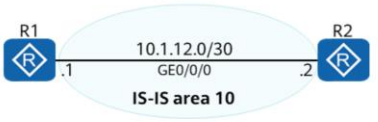
2. Enable IS-IS GR on R1 and R2.

```
[R1] isis 1
[R1-isis-1] graceful-restart
[R1-isis-1] graceful-restart no-impact-holdtime
```

```
[R2] isis 1
[R2-isis-1] graceful-restart
[R2-isis-1] graceful-restart no-impact-holdtime
```


HUAWEI

Example for Configuring IS-IS GR (2)



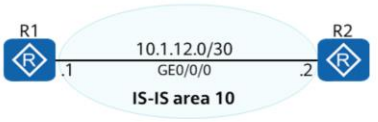
- Routers R1 and R2 each are equipped with two main control boards, which back up each other. The routers interwork through IS-IS and support GR.
- It is required that traffic forwarding be not interrupted when R1 restarts an IS-IS process or performs an active/standby switchover in GR mode.

3. Check the IS-IS GR status on R2.

```
<R2> display isis 1 graceful-restart status

                    Restart information for ISIS(1)
                    --------------------------------

IS-IS(1) Level-1 Restart Status
Restart Interval: 300
SA Bit Supported
  Total Number of Interfaces = 1
  Restart Status: RESTART COMPLETE

IS-IS(1) Level-2 Restart Status
Restart Interval: 300
SA Bit Supported
  Total Number of Interfaces = 1
  Restart Status: RESTART COMPLETE
```


HUAWEI

Example for Configuring IS-IS GR (3)



- Routers R1 and R2 each are equipped with two main control boards, which back up each other. The routers interwork through IS-IS and support GR.
- It is required that traffic forwarding be not interrupted when R1 restarts an IS-IS process or performs an active/standby switchover in GR mode.

When R1 restarts the IS-IS process in GR mode, the FIB remains unchanged, ensuring uninterrupted data forwarding.

4. Run the **reset isis all graceful-restart** command in the user view of R1 to restart IS-IS process 1, and run the **display isis peer** command on R2 to check the IS-IS adjacency with R1.

```
<R1> reset isis 1 all graceful-restart
```

```
<R2> display fib
Route Flags: G - Gateway Route, H - Host Route,    U - Up Route
             S - Static Route,  D - Dynamic Route, B - Black Hole Route
             L - Vlink Route
------------------------------------------------------------------------------
FIB Table:
Total number of Routes : 7

Destination/Mask    Nexthop      Flag  TimeStamp  Interface     TunnelID
10.1.12.3/32        127.0.0.1    HU    t[15]      InLoop0       0x0
10.1.12.2/32        127.0.0.1    HU    t[15]      InLoop0       0x0
255.255.255.255/32  127.0.0.1    HU    t[3]       InLoop0       0x0
127.255.255.255/32  127.0.0.1    HU    t[3]       InLoop0       0x0
127.0.0.1/32        127.0.0.1    HU    t[3]       InLoop0       0x0
127.0.0.0/8         127.0.0.1    U     t[3]       InLoop0       0x0
10.1.12.0/30        10.1.12.2    U     t[15]      GE0/0/0       0x0
```


HUAWEI

Quiz

1. (Multiple-answer question) Which of the following fast convergence mechanisms are supported by OSPF? ( )
   A. PRC
   B. LSP fast flooding
   C. Intelligent timer
   D. OSPF IP FRR
2. (True or false) The FA field in Type 5 LSAs of OSPF must be 0.0.0.0. ( )
   A. True
   B. False

HUAWEI

1. ACD
2. B

Quiz

3. (True or false) On an IS-IS network, if a device runs mode-2 LSP fragment extension, virtual systems do not participate in SPF calculation. All routers on the network know that the LSPs generated by the virtual systems actually belong to the originating system. ( )

   A. True

   B. False

HUAWEI

3. A

Summary

- To better adapt to network topology changes, OSPF and IS-IS support multiple fast convergence modes. I-SPF and PRC algorithms speed up route calculation; FRR enables fast traffic switching to the backup link; and intelligent timers allow you to control the speeds at which link state information is generated and routes are calculated.
- To control the size of routing tables and improve network performance, OSPF and IS-IS support route filtering, equal-cost routes, and delivery of default routes. To isolate protocol routing tables, OSPF and IS-IS support multi-process deployment on the same device. Routing tables of different processes are independent of each other.
- To prevent a suboptimal route from being selected when external routes are imported, OSPF uses the FA in Type 5 or Type 7 LSAs to guide data packet forwarding. To carry more routing information, IS-IS supports LSP fragment extension.
- To ensure that key services are not interrupted during a device restart, OSPF and IS-IS also support GR and NSR.
- Thanks to the preceding features, OSPF and IS-IS are widely and flexibly used on live networks.

HUAWEI

Thank you.

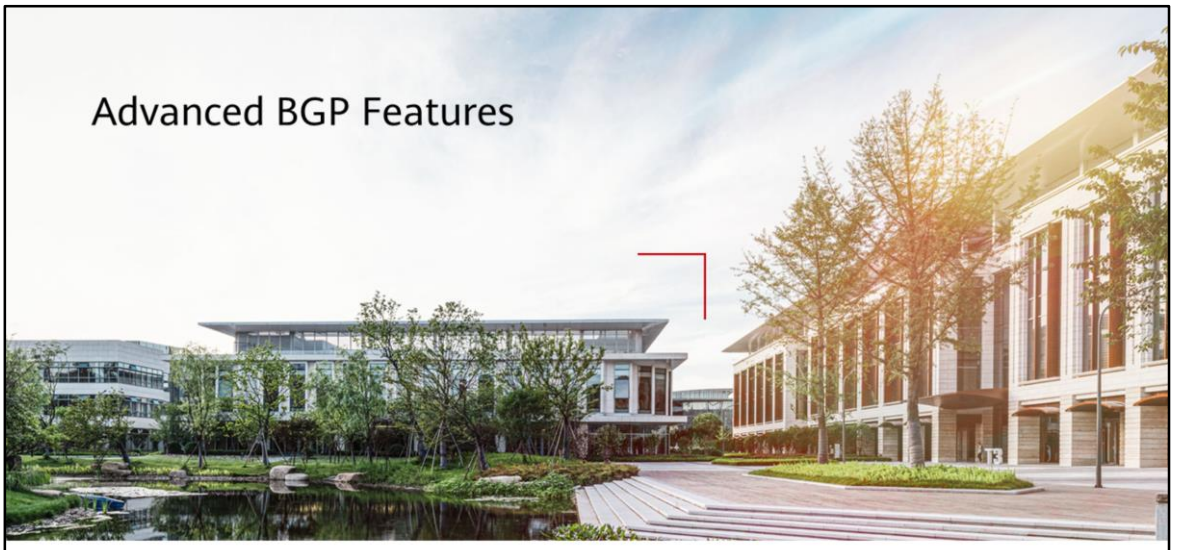
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Advanced BGP Features



HUAWEI

Foreword

- Border Gateway Protocol (BGP) is usually deployed on large-scale networks. Compared with an Interior Gateway Protocol (IGP), BGP has more flexible route control capabilities. Each BGP route can carry multiple path attributes. Special route matching tools, such as AS_Path and community filters, are available for filtering routes based on the route attributes. Routing policies can be used to control route advertisement and acceptance according to the actual networking requirements.
- In addition, BGP provides various advanced features and networking deployment solutions to improve network performance.
- This course describes the fundamentals and configurations of BGP route control, common advanced BGP features, such as ORF, peer group, security features, and 4-byte AS number, as well as the networking modes of BGP route reflectors (RRs).

HUAWEI

Objectives

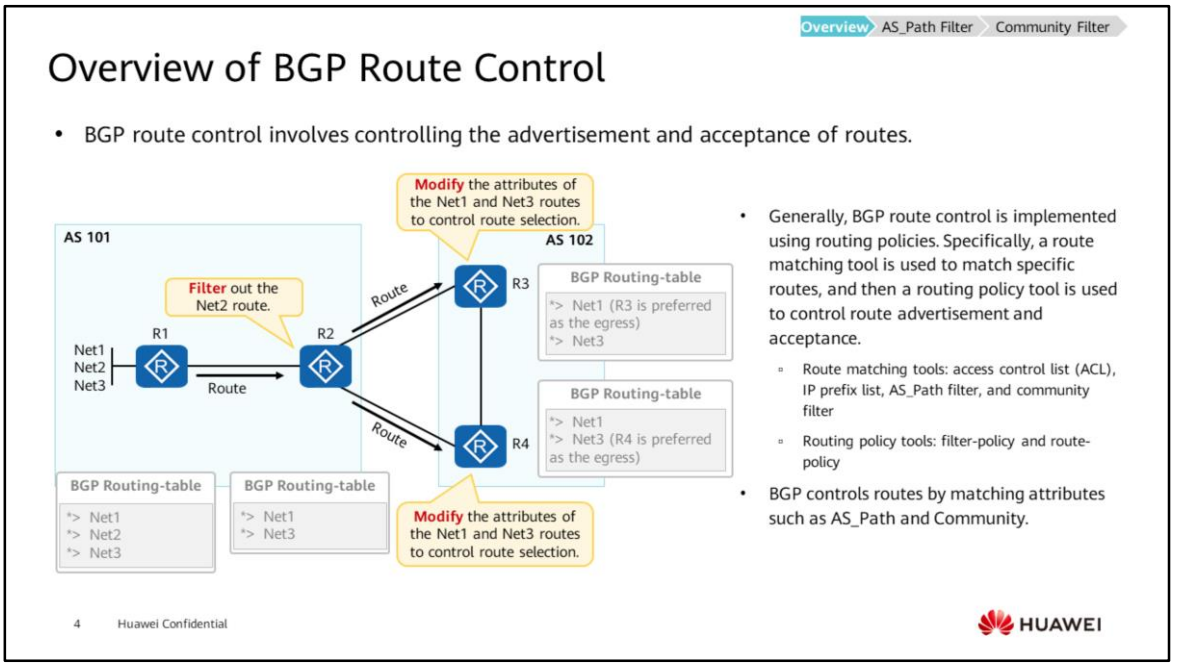
- On completion of this course, you will be able to:
  - Use regular expressions when configuring AS_Path and community filters.
  - Use AS_Path and community filters to control BGP routes.
  - Apply the ORF and peer group functions of BGP.
  - Perform basic configurations of BGP security.
  - Describe the concepts related to 4-byte AS numbers and usage.
  - Describe the networking modes of BGP RRs.

HUAWEI

Contents

1. **BGP Route Control**
2. Introduction to BGP Features
3. Networking Modes of BGP RRs

HUAWEI



- As shown in the figure:
  - R1 and R2 reside in AS 101, and establish an IBGP peer relationship with each other. R3 and R4 reside in AS 102, and each establishes an EBGP peer relationship with R2.
  - R1 is directly connected to three subnets: Net1, Net2, and Net3. R1 advertises routes to the three subnets to its BGP routing table.
  - R2 can filter out the Net2 route through BGP route control so that R2's BGP routing table does not contain the Net2 route.
  - R3 and R4 can implement BGP route control by modifying the attributes of the Net1 and Net3 routes, respectively. In this way, when a device in AS 102 accesses Net1, R3 is preferentially selected as the egress device; when the device accesses Net3, R4 is preferentially selected as the egress device.
- Note: For details about the ACL, IP prefix list, filter-policy, route-policy, and BGP path attributes, see the "HCIP-Datacom-Core Technology" course.

Regular Expression

- A regular expression (regex) is a formula used to match character strings based on a specific template. It consists of common characters (for example, letters from a to z) and special characters.
- Common characters are used to match themselves in a string:
  - Include all upper-case and lower-case letters, digits, punctuations, and special symbols.
  - For example, the regex a matches the letter "a" in "abc", the regex 10 matches the digits "10" in "10.113.25.155", and the regex @ matches the symbol "@" in "xxx@xxx.com".
- Special characters, together with common characters, are used to match complex or special character strings:
  - A special character that precedes or follows a common character is used to restrict or extend the independent control character or placeholder of the common character.
  - A special character describes how the character that precedes the special character is reused.
  - A special character specifies a complete range.

HUAWEI

- A regex has the following functions:
  - Checks and obtains the sub-character string that matches a specific rule in the character string.
  - Replaces the character string based on matching rules.

Example of Special Characters (1)

- Type 1

.	Matches any single character, including a space.	0.0 matches 0x0, 020, and so on.
^	Matches the start position of a character string.	^10 matches 10.1.1.1 but not 20.1.1.1.
$	Matches the end position of a character string.	1$ matches 10.1.1.1 but not 10.1.1.2.
_	Underscore, which matches any separator. Matches a comma (,), left brace ({), right brace (}), left parenthesis ((), or right parenthesis ()). (Same as ^) Matches the start position of an input character string. (Same as $) Matches the end position of an input character string. Matches a space.	_10 matches (10, {10, space10, and so on. 10_ matches 10), 10}, 10space, and so on.
|	Pipeline character, which is a logical OR operator. x|y matches x or y.	100|200 matches 100 or 200.
\	Defines an escape character, which is used to mark the next character (common or special character) as a common character.	* matches *.

- Type 2

*	Matches a sub-regular expression that it follows zero or multiple times.	10* matches 1, 10, 100, 1000, and so on.	(10) * matches null, 10, 1010, 101010, and so on.
+	Matches a sub-regular expression that it follows once or multiple times.	10+ matches 10, 100, 1000, and so on.	(10)+ matches null, 10, 1010, 101010, and so on.
?	Matches a sub-regular expression that it follows zero times or once.	10? matches 1 or 10.	(10)? matches null or 10.

HUAWEI

- Note: The parentheses () can be used to define the scope and priority of an operator. For example, gr(a|e)y is equivalent to gray|grey.

Example of Special Characters (2)

- Type 3

[xyz]	Matches any character contained in a regex.	[123] matches the character 2 in 255.
[^xyz]	Matches any character that is not contained in a regex.	[^123] matches any character except 1, 2, and 3.
[a-z]	Matches any character within a specified range in a regex.	[0-9] matches all digits in the range from 0 to 9.
[^a-z]	Matches any character beyond the range specified in a regex.	[^0-9] matches all non-digit characters (matching any characters except digits 0 to 9).

- Think about character strings that can be matched against the regexes of the following types.

Type 1

```
^a.$
^100_
^100$
100$|400$
^\(65000\)$
```

Type 2

```
abc*d
abc+d
abc?d
a(bc)?d
```

Type 3

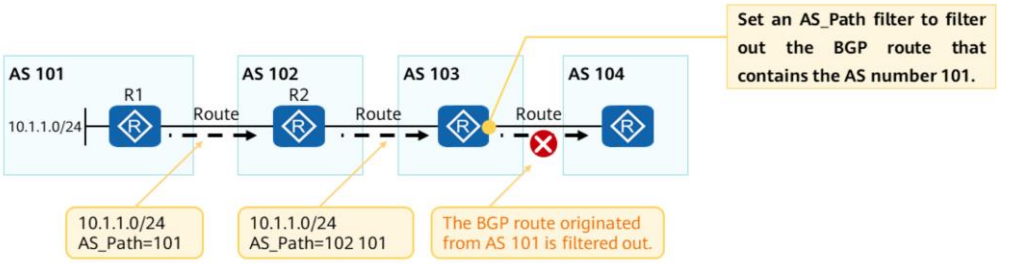
```
[abcd]
[a-c 1-2]$
[^act]$
[123].[7-9]
```


HUAWEI

- Type 1:
  - ^a.$: matches a character string that starts with the character a and ends with any single character, for example, a0, a!, ax, and so on.
  - ^100_: matches a character string starting with 100, for example, 100, 100 200, 100 300 400, and so on.
  - ^100$: matches only 100.
  - 100$|400$: matches a character string ending with 100 or 400, for example, 100, 1400, 300 400, and so on.
  - ^\(65000\)$: matches (65000) only.
- Type 2:
  - abc*d: matches the character c zero or multiple times, for example, abd, abcd, abccd, abcccd, abcccdef, and so on.
  - abc+d: matches the character c once or multiple times, for example, abcd, abccd, abcccd, abcccdef, and so on.
  - abc?d: matches the character c zero times or once, for example, abd, abcd, abcdef, and so on.
  - a(bc)?d: matches the character string bc zero times or once, for example, ad, abcd, aaabcdef, and so on.

Route Matching Tool: AS_Path Filter

- An AS_Path filter uses the AS_Path attribute of BGP routes as a matching condition to filter BGP routes.
- If you do not want routes from certain ASs, you can use an AS_Path filter to filter out the routes that carry the associated AS numbers.



HUAWEI

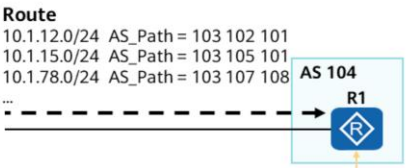
- The AS_Path attribute is a well-known mandatory attribute of BGP. All BGP routes must carry this attribute. This attribute records the numbers of all the ASs that a BGP route traversed during transmission.
- The value of the AS_Path attribute can be 0, 1, or a set of multiple AS numbers.

Using a Regex to Match the AS_Path Attribute

- A regex can be used to match the AS_Path attribute of routes.
  - For example, a regex can match the AS number 103 in AS_Path = 103 102 101.

- Other examples:

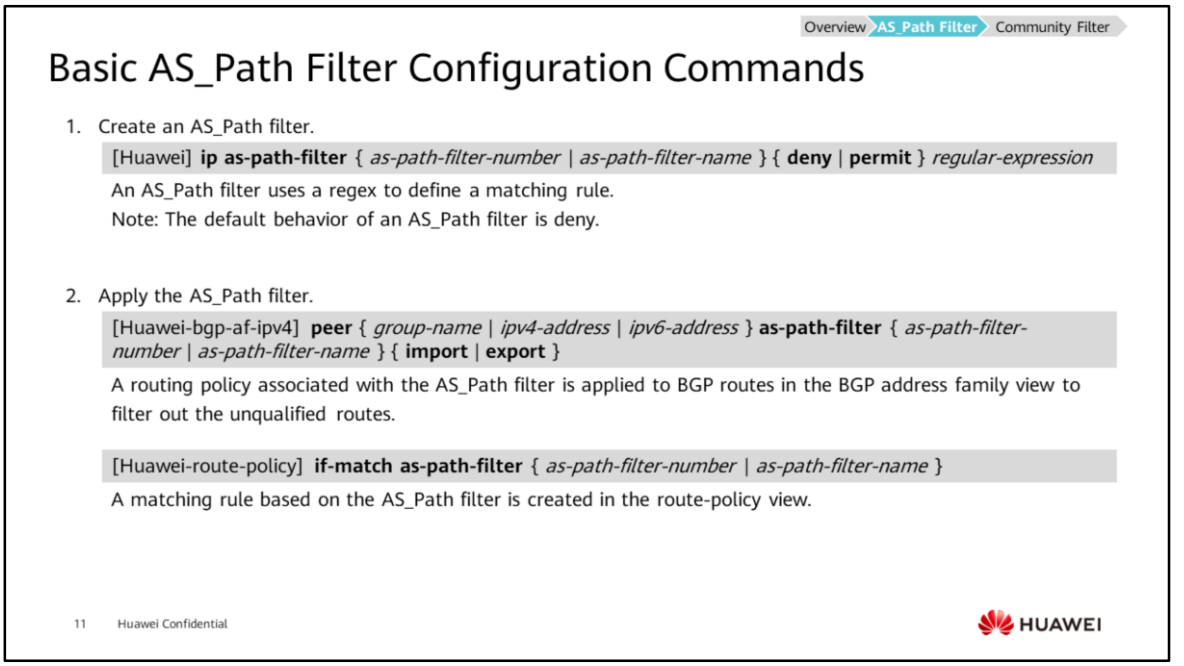
^$	Matches the AS_Path list that does not contain any AS number, that is, matching routes in the local AS.
.*	Matches all or any routes.
^10[012349]$	Matches the AS numbers 100, 101, 102, 103, 104, and 109.
^10[^0-6]$	Matches the AS_Path lists that exclude the AS numbers 100 to 106.
^10.	Matches the AS numbers 100 to 109 and 10.
^12(_34)?_56$	Matches the AS_Path lists 12 56 and 12 34 56.



R1 receives many BGP routes, each of which has its own AS_Path value.
To meet a specific requirement, R1 needs to apply a policy to the routes carrying the AS number 101 in their AS_Path attribute. In this case, you can associate an AS_Path filter with a regex to match routes, regardless of the route prefixes.

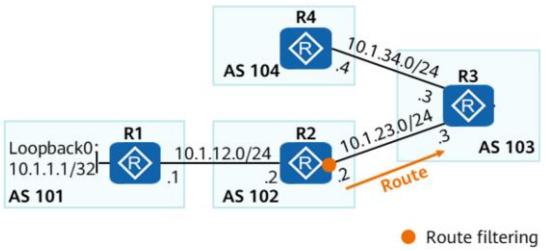
You can use different regexes to meet different requirements, for example, using ^101$ to match AS 101, and _101$ to match the AS_Path lists ending with AS 101.

HUAWEI



- Multiple matching rules (each in permit or deny mode) can be specified in an AS_Path filter. These rules are in the OR relationship, which means that if a route matches one of the matching rules, the route is considered to match the AS_Path filter.
- Command: [Huawei] **ip as-path-filter** {*as-path-filter-number* | *as-path-filter-name*} {**deny** | **permit**} *regular-expression*
    - *as-path-filter-number*: specifies the number of an AS_Path filter. The value is an integer ranging from 1 to 256.
    - *as-path-filter-name*: specifies the name of an AS_Path filter. The value is a string of 1 to 51 case-sensitive characters. It cannot be comprised of only digits. If spaces are used, the string must start and end with double quotation marks (").
    - **deny:** sets the matching mode of the AS_Path filter to deny.
    - **permit**: sets the matching mode of the AS_Path filter to permit.
    - *regular-expression*: specifies a regex for the AS_Path filter. The value is a string of 1 to 255 characters and can contain spaces.
- The default behavior of an AS_Path filter is deny. That is, if a route is not permitted in a filtering, the route fails to match the AS_Path filter. If all matching rules in an AS_Path filter work in deny mode, all BGP routes are denied by the filter. To prevent this problem, configure a matching rule in permit mode after one or more matching rules in deny mode so that the routes except for those denied by the preceding matching rules can be permitted by the filter.

Examples for Configuring AS_Path Filters (1)



R2 transmits EBGP routes to R3. Among these routes, some are locally originated by R2, and the others are transmitted from AS 101 to R2 and then updated by R2 to R3.

A routing policy is configured on R2 to deny the routes originated from AS 101.

1. Create an AS_Path filter.

```
[R2] ip as-path-filter 1 deny _101$
[R2] ip as-path-filter 1 permit .*
```

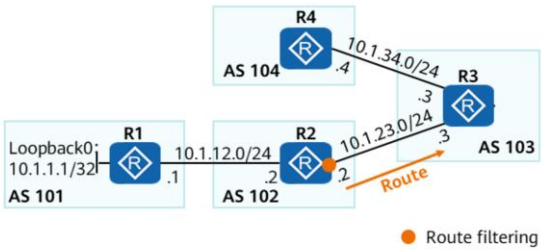
The AS_Path filter is configured to filter out the routes originated from AS 101 and permit the other routes.

2. (Direct reference) Apply the AS_Path filter.

```
[R2] bgp 102
[R2-bgp] peer 10.1.23.3 as-number 103
[R2-bgp] ipv4-family unicast
[R2-bgp-af-ipv4] peer 10.1.23.3 as-path-filter 1 export
```


HUAWEI

Examples for Configuring AS_Path Filters (2)



R2 transmits EBGP routes to R3. Among these routes, some are locally originated by R2, and the others are transmitted from AS 101 to R2 and then updated by R2 to R3.

A routing policy is configured on R2 to deny the routes originated from AS 101.

1. Create an AS_Path filter.

```
[R2] ip as-path-filter 1 deny _101$
[R2] ip as-path-filter 1 permit .*
```

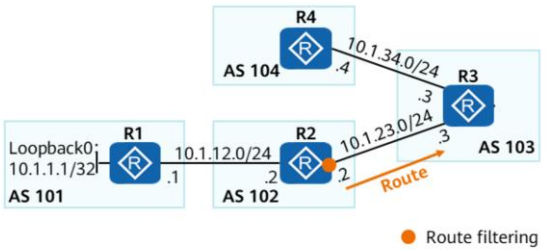
The AS_Path filter is configured to filter out the routes originated from AS 101 and permit the other routes.

2. (Through a route-policy) Apply the AS_Path filter.

```
[R2] route-policy AS_Path permit node 10
[R2-route-policy] if-match as-path-filter 1
[R2-route-policy] quit
[R2] bgp 102
[R2-bgp] peer 10.1.23.3 as-number 103
[R2-bgp] ipv4-family unicast
[R2-bgp-af-ipv4] peer 10.1.23.3 route-policy AS_Path export
```


HUAWEI

Checking the AS_Path Filter Information



1. Check the AS_Path filter.

```
[R2]display ip as-path-filter 1
As path filter number: 2
            deny        _101$
            permit      .*
```

2. Check all the routes whose AS_Path lists match the specified regex in the BGP routing table.

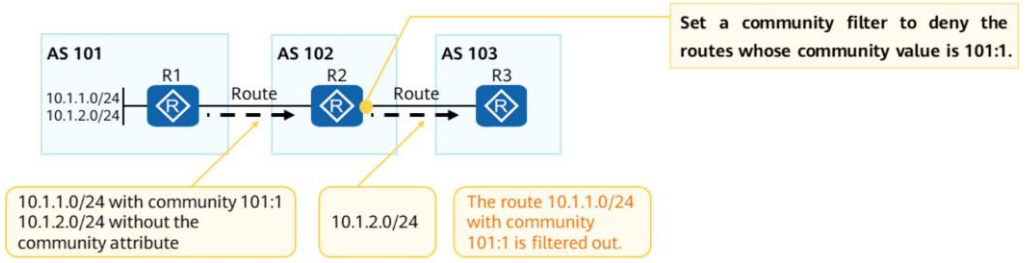
```
[R2]display bgp routing-table regular-expression _101$

 Total Number of Routes: 1

 BGP Local router ID is 10.1.12.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
    Network         NextHop    MED  LocPrf  PrefVal  Path/Ogn
 *> 10.1.1.1/32     10.1.12.1  0                0        101i
```

The route 10.1.1.1/32 originated from AS 101 is filtered by the AS_Path filter.

HUAWEI

Route Matching Tool: Community Filter

- The community filter and community attribute can be used together to simplify route management when the IP prefix list and AS_Path filter are not available.
- There are two types of community filters:
  - Basic community filter, which matches the community number or well-known community attribute.
  - Advanced community filter, which uses a regex to match the community number.



HUAWEI

- The community attribute is an optional transitive attribute. It can identify the routes with the same characteristics, regardless of the scattered route prefixes and various AS numbers. That is, a specific community value can be assigned to some routes so that these routes can be matched against the community value instead of the network number or mask. Then, a corresponding routing policy can be applied to the matched routes.

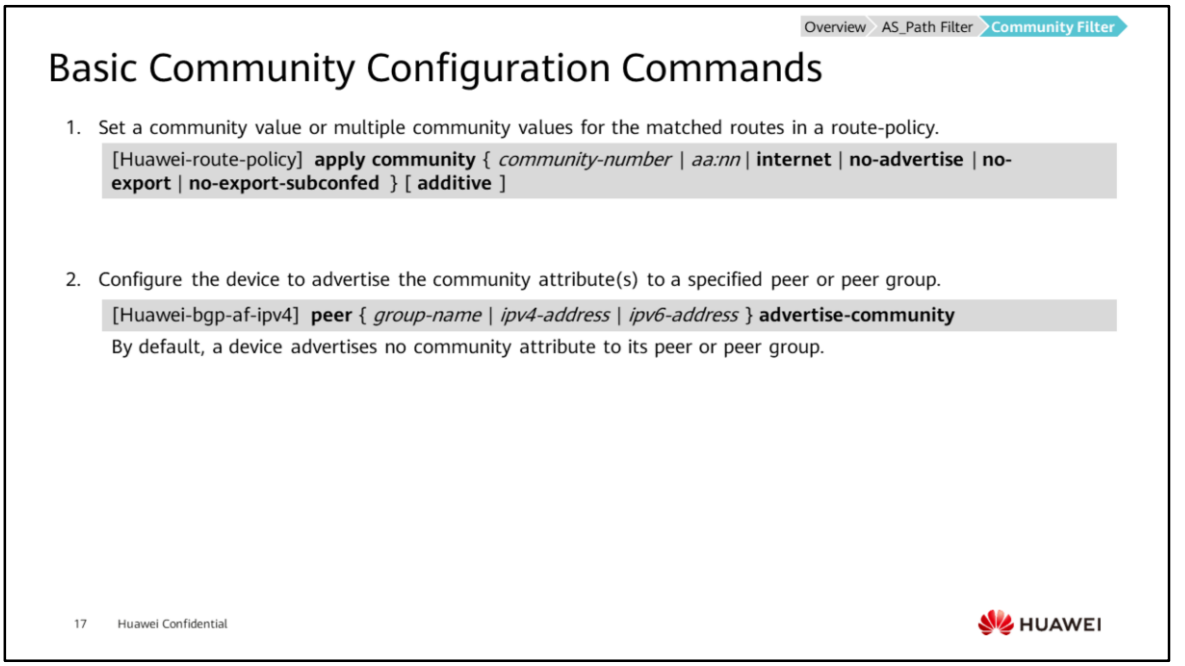
Community Attributes

- Well-known community attributes

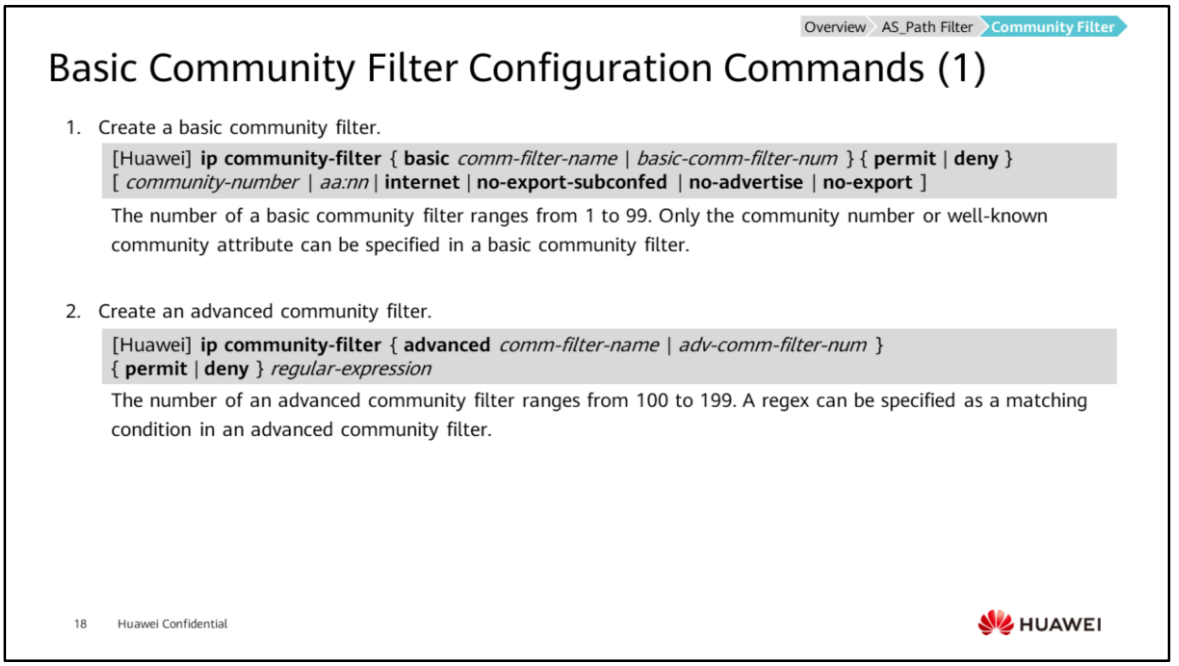
Name	Number	Description
Internet	0 (0x00000000)	After a device receives a route with this attribute, the device can advertise this route to any BGP peer. By default, all routes belong to the Internet community.
No_Advertise	4294967042 (0xFFFFFF02)	After a device receives a route with this attribute, the device does not advertise this route to any BGP peer.
No_Export	4294967041 (0xFFFFFF01)	After a device receives a route with this attribute, the device does not advertise this route outside the local AS.
No_Export_Subconfed	4294967043 (0xFFFFFF03)	After a device receives a route with this attribute, the device does not advertise this route outside the local AS. If a confederation is used, the device does not advertise this route to the other sub-ASs in the confederation.

- Community attribute format:
  - The length of a community attribute is 32 bits, which can be presented in either of the following formats:
    - Decimal integer
    - In the *AA*:*NN* format, *AA* indicates an AS number, and *NN* is a user-defined number.

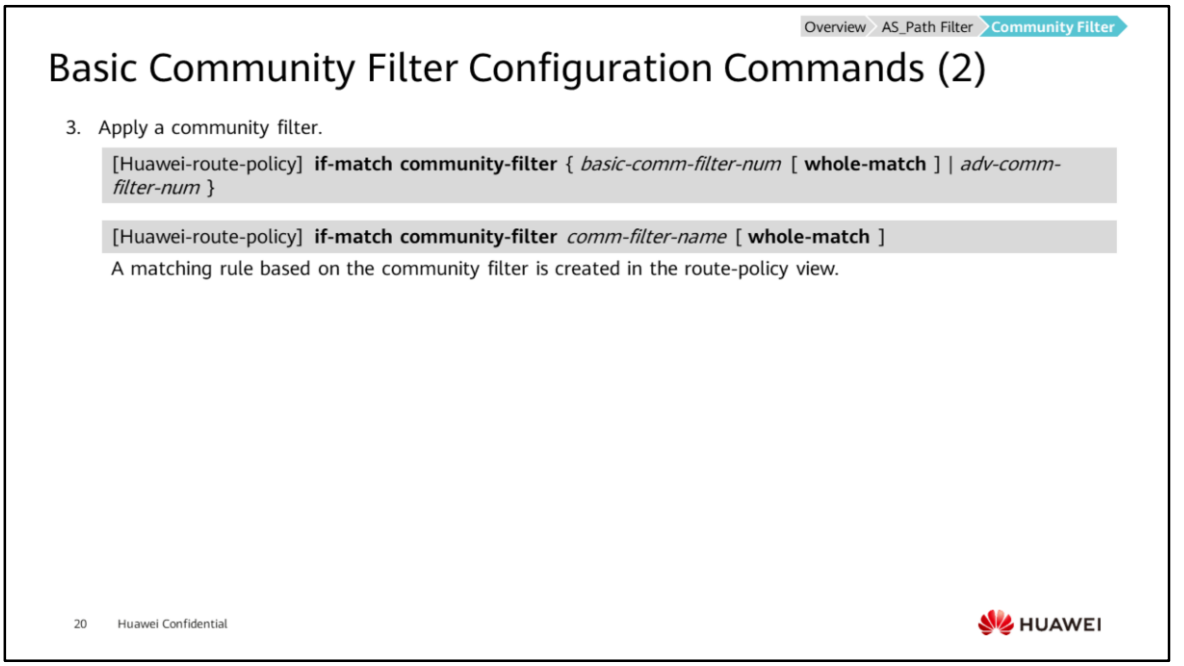
HUAWEI



- Command: [Huawei-route-policy] **apply community** { *community-number* | *aa:nn* | **internet** | **no-advertise** | **no-export** | **no-export-subconfed** } [**additive**]
  - *community-number* | *aa:nn*: specifies a community number for a community attribute. A maximum of 32 community numbers can be specified at a time using this command. The value of *community-number* is an integer ranging from 0 to 4294967295. The values of *aa* and *nn* are also integers ranging from 0 to 65535.
  - **internet**: allows the matched routes to be advertised to any peers. By default, all routes belong to the Internet community.
  - **no-advertise**: prevents the matched routes from being advertised to any peer. After a device receives a route with this attribute, it cannot advertise this route to any other BGP peers.
  - **no-export**: prevents the matched routes from being advertised outside the local AS but allows them to be advertised to other sub-ASs in the local AS. After a device receives a route with this attribute, it cannot advertise this route outside the local AS.
  - **no-export-subconfed**: prevents the matched routes from being advertised outside the local AS or to other sub-ASs in the local AS. After a device receives a route with this attribute, it cannot advertise this route to any other sub-ASs.
  - **additive:** adds community attributes to the routes that match the filtering conditions.



- Command: [Huawei**] ip community-filter { basic** *comm-filter-name* | *basic-comm-filter-num* **} { permit | deny } [** *community-number* | *aa:nn* | **internet | no-export-subconfed | no-advertise | no-export]**
  - **basic** *comm-filter-name:* specifies the name of a basic community filter. The value is a string of 1 to 51 case-sensitive characters. It cannot be comprised of only digits.
  - *basic-comm-filter-num:* specifies the number of a basic community filter. The value is an integer ranging from 1 to 99.
  - **deny:** sets the matching mode of the community filter to deny.
  - **permit:** sets the matching mode of the community filter to permit.
  - *community-number:* specifies a community number. The value is an integer ranging from 0 to 4294967295.
  - *aa:nn:* specifies a community number. A maximum of 20 community numbers can be specified at a time using this command. The values of aa and nn are integers ranging from 0 to 65535.
  - **internet:** allows the matched routes to be advertised to any peers.
  - **no-export-subconfed:** prevents the matched routes from being advertised outside the local AS. If a confederation is used, the matched routes will not be advertised to the other sub-ASs in the confederation.
  - **no-advertise:** prevents the matched routes from being advertised to any other peers.
  - **no-export:** prevents the matched routes from being advertised outside the local AS. If a confederation is used, the matched routes will not be advertised outside the confederation but will be advertised to the other sub-ASs in the confederation.



- Command: [Huawei-route-policy] **if-match community-filter** { *basic-comm-filter-num* [**whole-match**] | *adv-comm-filter-num* }
- Command: [Huawei-route-policy] **if-match community-filter** *comm-filter-name* [**whole-match**]
  - *basic-comm-filter-num*: specifies the number of a basic community filter. The value is an integer ranging from 1 to 99.
  - *adv-comm-filter-num*: specifies the number of an advanced community filter. The value is an integer ranging from 100 to 199.
  - *comm-filter-name*: specifies the name of a community filter. The value is a string of 1 to 51 case-sensitive characters. It cannot be comprised of only digits. If spaces are used, the string must start and end with double quotation marks (").
  - **whole-match**: indicates complete matching. That is, all the community attributes in the specified community filter must be matched. This parameter applies only to basic community filters.

Examples for Configuring Community Filters

1. Example for configuring a basic community filter.

   Match the routes carrying community values [100:1, 200:1, 300:1]. (The relationship between multiple community values in a command configuration is AND.)

   ```
   ip community-filter 1 permit 100:1 200:1 300:1
   ```

   Match the routes carrying community value [100:1] or community values [200:1, 300:1]. (The relationship between different groups of community values is OR. A set of community values in a command configuration is called a group.)

   ```
   ip community-filter 1 permit 100:1
   ip community-filter 1 permit 200:1 300:1
   ```

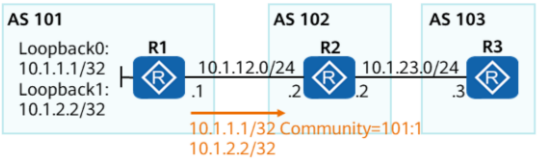
2. Example for configuring an advanced community filter.

   Match the routes carrying community values starting with 10.

   ```
   ip community-filter 100 permit ^10
   ```


HUAWEI

Configuring Community Attributes (1)



Configure a routing policy on R1 to allow the community value 101:1 to be carried only in the BGP route 10.1.1.1/32 when R1 advertises this route.

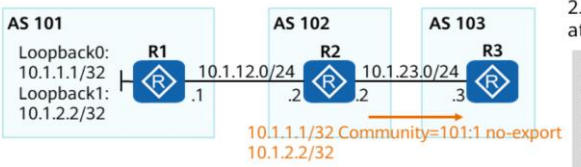
1. Configure a routing policy on R1 to add a community attribute to matched routes, and enable R1 to advertise the community attribute to the EBGP peer R2.

```
[R1] ip ip-prefix 1 permit 10.1.1.1 32
[R1] route-policy Community permit node 10
[R1-route-policy] if-match ip-prefix 1
[R1-route-policy] apply community 101:1
[R1-route-policy] quit
[R1] route-policy Community permit node 20
[R1-route-policy] quit
[R1] bgp 101
[R1-bgp] peer 10.1.12.2 as-number 102
[R1-bgp] peer 10.1.12.2 route-policy Community export
[R1-bgp] peer 10.1.12.2 advertise-community
[R1-bgp] network 10.1.1.1 32
[R1-bgp] network 10.1.2.2 32
```


HUAWEI

- Command: [R1] **route-policy** Community **permit** node 20
- Run this command to allow the route 10.1.2.2/32 to be advertised properly.

Configuring Community Attributes (2)



Configure a routing policy on R1 to allow the community value 101:1 to be carried only in the BGP route 10.1.1.1/32 when R1 advertises this route.

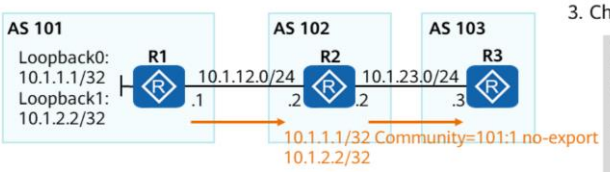
Configure a routing policy on R2 to allow the community attribute no-expert to be added to the BGP route 10.1.1.1/32 when R2 advertises this route.

2. Configure R2 to advertise the community attribute to the EBGP peer R3.

```
[R2] ip ip-prefix 1 permit 10.1.1.1 32
[R2] route-policy Community permit node 10
[R2-route-policy] if-match ip-prefix 1
[R2-route-policy] apply community no-expert additive
[R2-route-policy] quit
[R2] route-policy Community permit node 20
[R2-route-policy] quit
[R2] bgp 102
[R2-bgp] peer 10.1.12.1 as-number 101
[R2-bgp] peer 10.1.23.3 as-number 102
[R2-bgp] peer 10.1.23.3 advertise-community
[R2-bgp] peer 10.1.23.3 route-policy Community export
```


HUAWEI

Configuring Community Attributes (3)



Configure a routing policy on R1 to allow the community value 101:1 to be carried only in the BGP route 10.1.1.1/32 when R1 advertises this route.

Configure a routing policy on R2 to allow the community attribute no-expert to be added to the BGP route 10.1.1.1/32 when R2 advertises this route.

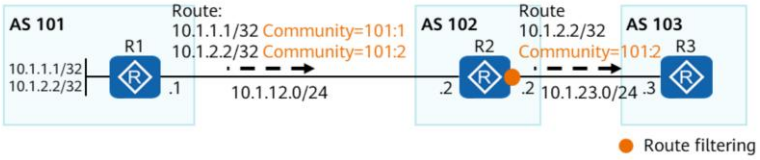
3. Check the BGP routing information on R3.

```
[R3] bgp 103
[R3-bgp] peer 10.1.23.2 as-number 102
[R3-bgp] quit

[R3] display bgp routing-table 10.1.1.1
 BGP local router ID : 10.1.23.3
 Local AS number : 103
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 10.1.1.1/32:
 From: 10.1.23.2 (10.1.12.2)
 Route Duration: 00h00m21s
 Direct Out-interface: GigabitEthernet0/0/2
 Original nexthop: 10.1.23.2
 Qos information : 0x0
 Community:<101:1>, no-export
 AS-path 102 101, origin igp, pref-val 0, valid, external,
best, select, active, pre 255
 Not advertised to any peer yet
```

Check the route 10.1.1.1/32 on R3. The command output shows that the route carries two community values: 101:1 and no-export.

HUAWEI

Configuring a Community Filter (1)



R2 transmits routes to the EBGP peer R3. Configure a routing policy on R2 to filter out the route carrying the community value 101:1.

1. Configure a community filter to match the route carrying the community value 101:1.

```
[R2] ip community-filter 1 permit 101:1
```

2. Configure a matching rule that is based on the community filter.

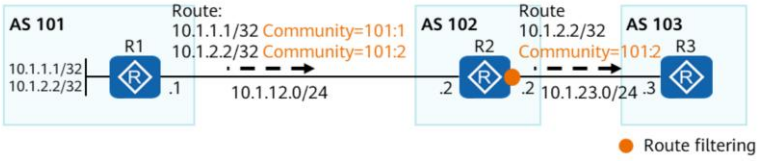
```
[R2] route-policy Community deny node 10
[R2-route-policy] if-match community-filter 1
[R2-route-policy] quit
[R2] route-policy Community permit node 20
[R2-route-policy] quit
[R2] bgp 102
[R2-bgp] peer 10.1.23.3 route-policy Community export
```

3. Check the community filter information on R2.

```
[R2]display ip community-filter 1
Community filter Number: 1
        permit 101:1
```


HUAWEI

Configuring a Community Filter (2)



R2 transmits routes to the EBGP peer R3. Configure a routing policy on R2 to filter out the route carrying the community value 101:1.

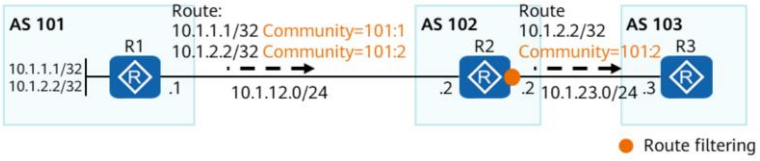
4. Check the BGP routing information on R2.

```
[R2]dis bgp routing-table 10.1.1.1
 BGP local router ID : 10.1.12.2
 Local AS number : 102
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 10.1.1.1/32:
 From: 10.1.12.1 (10.1.1.1)
 Route Duration: 00h13m39s
 Direct Out-interface: GigabitEthernet0/0/1
 Original nexthop: 10.1.12.1
 Qos information : 0x0
 Community:<101:1>
 AS-path 101, origin igp, MED 0, pref-val 0, valid, external, best, select, active, pre 255
 Not advertised to any peer yet
```

Check the route 10.1.1.1/32 on R2. The command output shows that the route carries the community value 101:1.
Use the community filter to filter out this route.

HUAWEI

Configuring a Community Filter (3)



R2 transmits routes to the EBGP peer R3. Configure a routing policy on R2 to filter out the route carrying the community value 101:1.

5. Check the BGP routing table on R3.

```
[R3]display bgp routing-table
 BGP Local router ID is 10.1.23.3
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
 Total Number of Routes: 1
      Network      NextHop      MED        LocPrf    PrefVal  Path/Ogn
 *>   10.1.2.2/32  10.1.23.2                         0        102 101i
```

R3 does not receive the BGP route 10.1.1.1/32.

HUAWEI

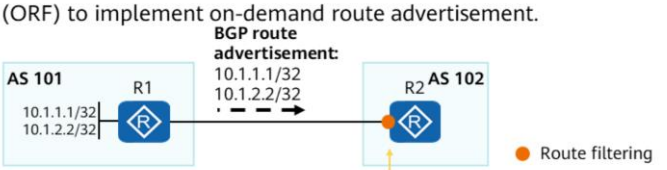
Contents

1. BGP Route Control
2. **Introduction to BGP Features**
3. Networking Modes of BGP RRs

HUAWEI

On-Demand Route Advertisement by BGP Peers

- If a device expects to receive only required routes from a remote device, and the remote device cannot maintain a separate outbound routing policy for each connected peer, you can configure a prefix-based outbound route filter (ORF) to implement on-demand route advertisement.

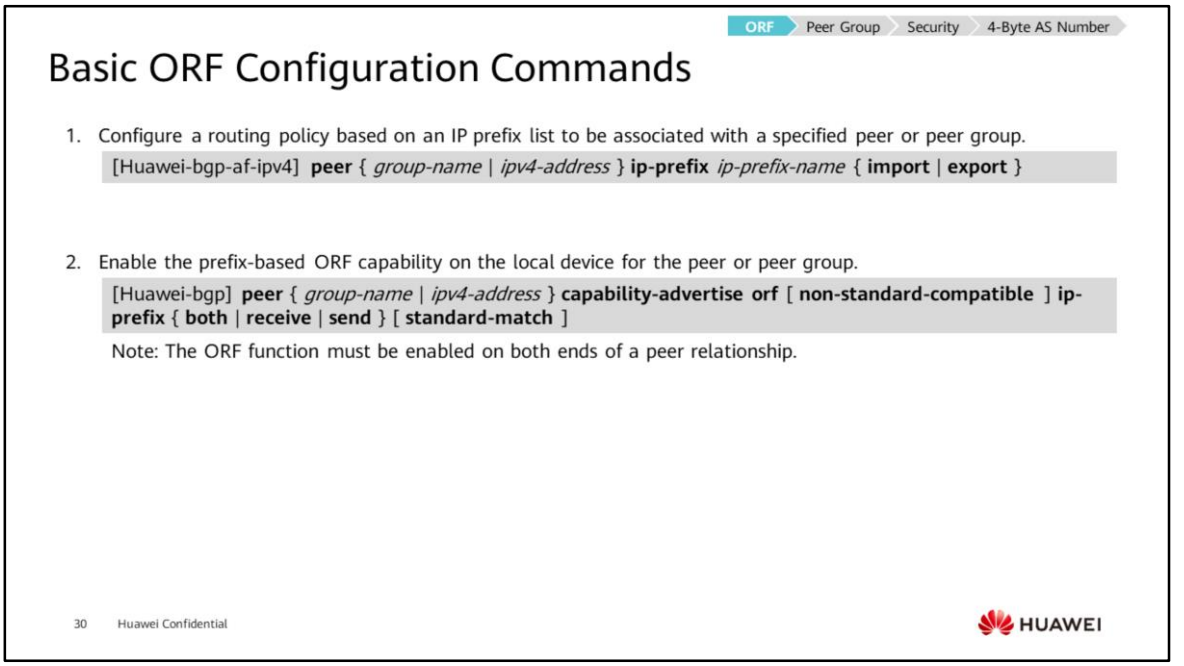


- Prefix-based ORF enables a device to send locally configured prefix-based import policies to its BGP peers through Route-refresh messages. The BGP peers construct export policies based on the received policies (in Route-refresh messages) to filter routes to be advertised.
- This prevents the local device from receiving a large number of unwanted routes, lowers the CPU usage of the local device, reduces the configuration workload on BGP peers, and lowers the link bandwidth usage.

R2 filters routes on the inbound interface and accepts only the route 10.1.1.1/32.

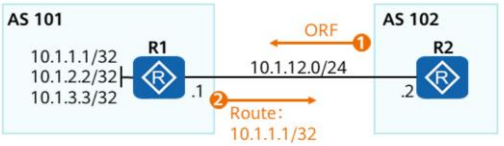
For the routes that are filtered out (for example, route 10.1.2.2/32), actually R1 does not need to advertise them to R2.

HUAWEI



- Command: [Huawei-bgp-af-ipv4] **peer** { *group-name* | *ipv4-address* } **ip-prefix** *ip-prefix-name* { **import** | **export** }
  - **import**: applies the routing policy to the routes received from the peer or peer group.
  - **export**: applies the routing policy to the routes to be advertised to the peer or peer group.
- Command: [Huawei-bgp] **peer** { *group-name* | *ipv4-address* } **capability-advertise orf** [**non-standard-compatible**] **ip-prefix** { **both** | **receive** | **send** } [**standard-match**]
  - **non-standard-compatible**: indicates that the ORF capability supported by the Huawei device is compatible with that supported by a non-Huawei device.
  - **both**: enables the local device to both send and accept ORF packets.
  - **receive**: enables the local device to only accept ORF packets.
  - **send**: enables the local device to only advertise ORF packets.
  - **standard-match**: matches routes according to the prefix matching rules defined in an RFC standard.

Example for Configuring ORF



R2 expects R1 to only advertise the route 10.1.1.1/32, and sends an ORF packet to R1 to meet this expectation.

1. Configure a routing policy on R2 to filter the route 10.1.1.1/32, and enable ORF on R2 to allow R2 to send ORF packets.

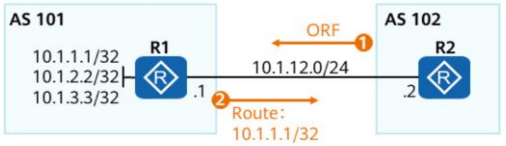
```
[R2] ip ip-prefix 1 permit 10.1.1.1 32
[R2] bgp 102
[R2-bgp] peer 10.1.12.1 as-number 101
[R2-bgp] peer 10.1.12.1 ip-prefix 1 import
[R2-bgp] peer 10.1.12.1 capability-advertise orf ip-prefix send
```

2. Enable ORF on R1 to allow R1 to accept ORF packets.

```
[R1] bgp 101
[R1-bgp] peer 10.1.12.2 as-number 102
[R1-bgp] peer 10.1.12.2 capability-advertise orf ip-prefix receive
[R1-bgp] network 10.1.1.1 32
[R1-bgp] network 10.1.2.2 32
[R1-bgp] network 10.1.3.3 32
```


HUAWEI

Verifying the ORF Configuration



1. On R1, check the prefix-based ORF information sent by R2.

```
[R1]display bgp peer 10.1.12.2 orf ip-prefix
 Total number of ip-prefix received: 1
 Index     Action     Prefix          MaskLen  MinLen  MaxLen
 10        Permit     10.1.1.1        32
```

2. Check the BGP routing information on R2.

```
[R2]display bgp routing-table peer 10.1.12.1 received-routes

BGP Local router ID is 10.1.12.2
Status codes: * - valid, > - best, d - damped,
              h - history,  i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete

Total Number of Routes: 1
     Network      NextHop      MED        LocPrf    PrefVal  Path/Ogn
*>   10.1.1.1/32  10.1.12.1    0                    0        101i
```

Check the ORF information on R1. The command output shows that the route 10.1.1.1/32 is required.

Check the BGP routing information on R2. The command output shows that R2 has received only the route 10.1.1.1/32.

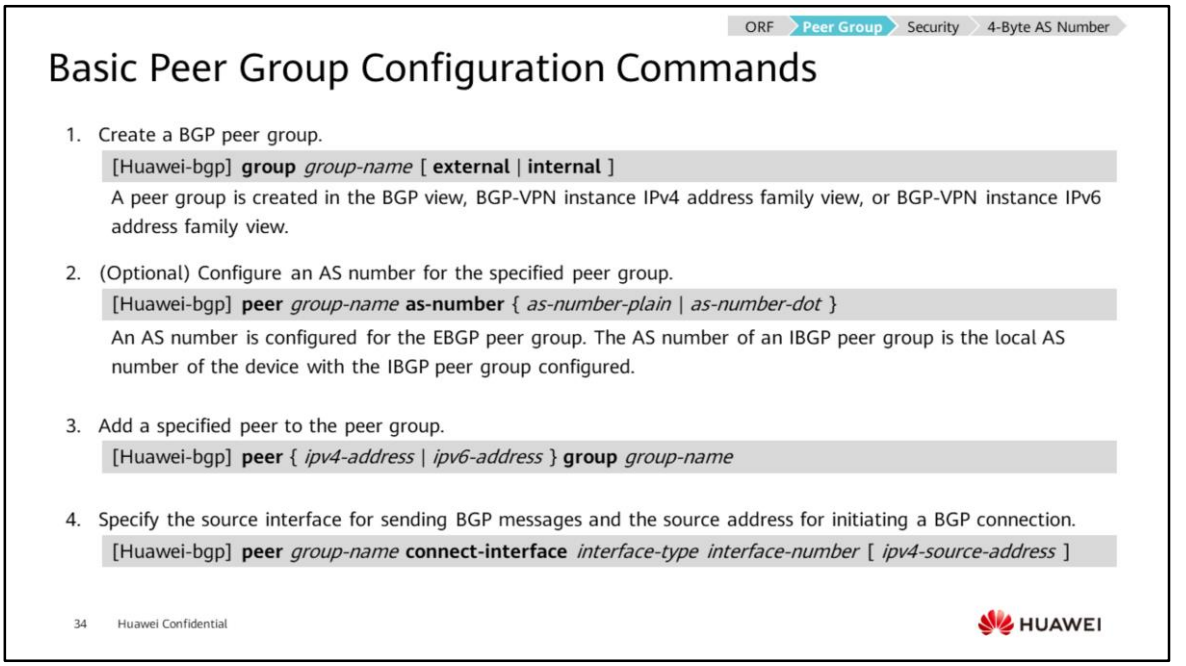
HUAWEI

BGP Peer Group

- A peer group is a set of peers with the same policies. When a peer is added to a peer group, it inherits the configurations of the peer group. If the configurations of the peer group change, the configurations of all the peers in the group change accordingly.
- A large number of BGP peers may exist on a large-scale BGP network, many of which need the same policies, you can configure a peer group to simplify configuration.

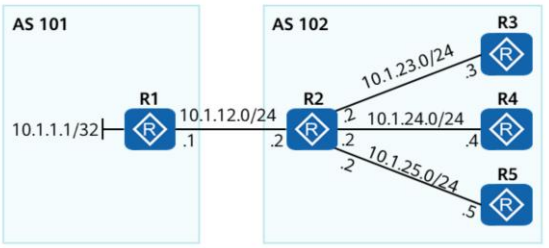
HUAWEI

- Each peer in a peer group can be configured with its own policies for route advertisement and acceptance.



- Command: [Huawei-bgp] **group** *group-name* [**external | internal**]
  - *group-name:* specifies the name of a peer group. The value is a string of 1 to 47 case-sensitive characters. If spaces are used, the string must start and end with double quotation marks (").
  - **external:** creates an EBGP peer group.
  - **internal:** creates an IBGP peer group.

Example for Configuring a BGP Peer Group



After R2 (ASBR) receives a route from its EBGP peer (R1), it advertises the route to all its IBGP peers (R3, R4, and R5). If R2 supports the BGP peer group function, it saves memory and improves route forwarding efficiency.

Device	Interface	Interface Address	Device	Interface	Interface Address
R1	Loopback0	10.1.1.1/32	R4	Loopback0	10.1.4.4/32
R2	Loopback0	10.1.2.2/32	R5	Loopback0	10.1.5.5/32
R3	Loopback0	10.1.3.3/32	/		

1. Complete the basic BGP configuration on R1.

```
[R1] bgp 101
[R1-bgp] peer 10.1.12.2 as-number 102
[R1-bgp] network 10.1.1.1 32
```

2. Complete the basic EBGP configuration and basic IBGP peer group configuration on R2.

```
[R2] bgp 102
[R2-bgp] peer 10.1.12.1 as-number 101
[R2-bgp] group in internal
[R2-bgp] peer 10.1.3.3 group in
[R2-bgp] peer 10.1.4.4 group in
[R2-bgp] peer 10.1.5.5 group in
[R2-bgp] peer in connect-interface Loopback 0
```

3. Complete the basic IBGP configuration on R3.

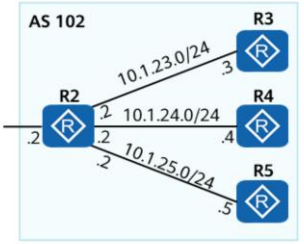
```
[R3] bgp 102
[R3-bgp] peer 10.1.2.2 as-number 102
[R3-bgp] peer 10.1.2.2 connect-interface Loopback 0
```

The configurations of R4 and R5 are similar to the configuration of R3, and are not provided here.

HUAWEI

- As shown in the figure, assume that static routes are used or OSPF is used to ensure internal network reachability in AS 102. The configuration details are not provided here.

Verifying the BGP Peer Group Configuration



1. Check the BGP peer group information on R2.

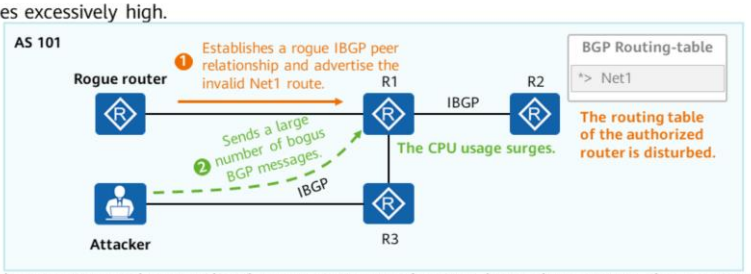
```
[R2]display bgp group in
 BGP peer-group: in
 Remote AS: 102
 Authentication type configured: None
 Type : internal
 Configured hold timer value: 180
 Keepalive timer value: 60
 Connect-retry timer value: 32
 Minimum route advertisement interval is 15 seconds
 Connect-interface has been configured
 PeerSession Members:
   10.1.3.3        10.1.4.4        10.1.5.5

 Peer Preferred Value: 0
 No routing policy is configured
 Peer Members:
  Peer        V    AS  MsgRcvd  MsgSent  OutQ  Up/Down      State       PrefRcv
  10.1.3.3    4   102       5        7     0   00:03:33    Established   0
  10.1.4.4    4   102       5        6     0   00:03:11    Established   0
  10.1.5.5    4   102       4        6     0   00:02:52    Established   0
```

The command output shows that the peer group created on R2 is named in and has three IBGP peer members: 10.1.3.3, 10.1.4.4, and 10.1.5.5. All these members have established peer relationships with R2.

HUAWEI

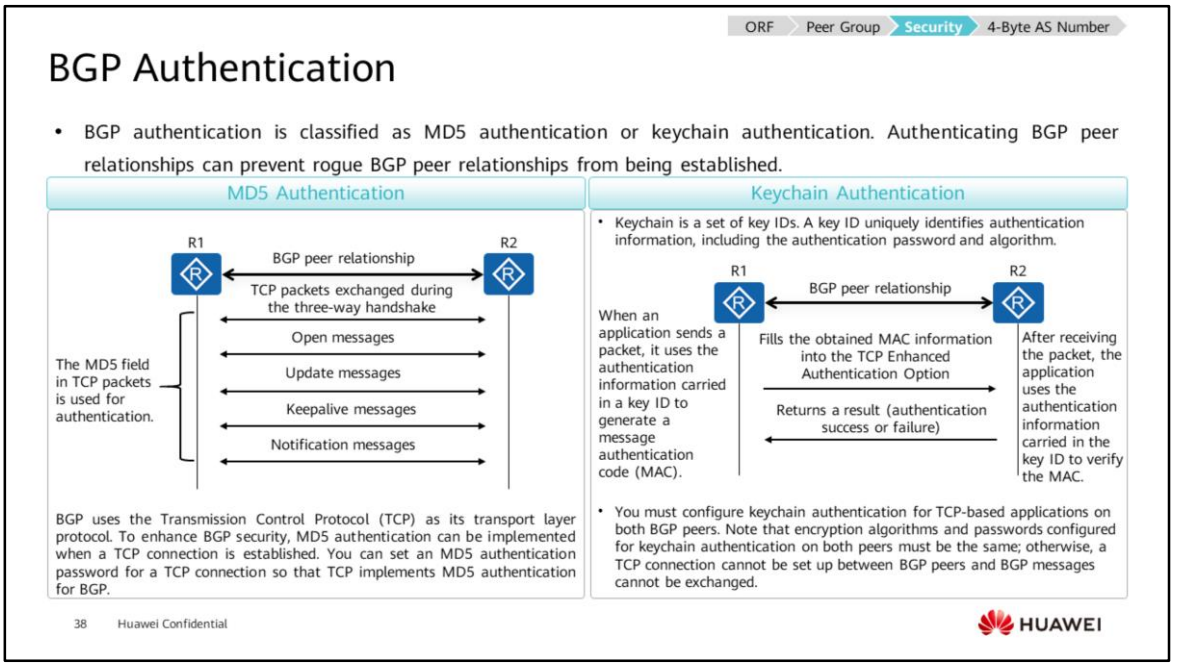
BGP Security

- Common BGP attacks are as follows:
  - A rogue BGP peer relationship is established, and invalid routes are advertised to disturb the normal routing table.
  - A router receives a large number of bogus BGP messages and sends them to the CPU. As a result, the CPU usage goes excessively high.



- BGP uses authentication and Generalized TTL Security Mechanism (GTSM) to ensure the security of message exchange between BGP peers.

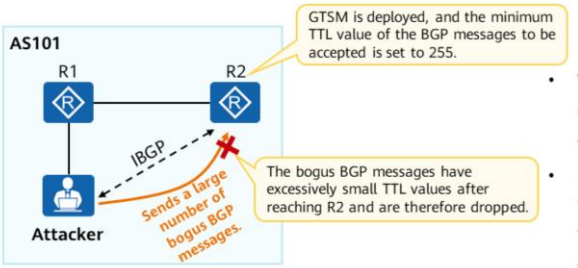
HUAWEI



- BGP uses TCP as its transport layer protocol and considers a TCP packet valid only if the source IP address, destination IP address, source port number, destination port number, and TCP sequence number in the packet are correct. Most of the preceding parameters in a TCP packet can be obtained by attackers without much difficulty. To protect BGP from attacks, use MD5 authentication or keychain authentication between BGP peers to reduce the possibility of attacks.
    - The MD5 algorithm is easy to configure and generates a single password, which can only be manually changed.
    - The keychain algorithm is complex to configure and generates a set of passwords. Keychain authentication allows passwords to be changed automatically based on configurations. Therefore, keychain authentication is applicable to networks requiring high security.
- Note: BGP MD5 authentication and BGP keychain authentication are mutually exclusive.

BGP GTSM

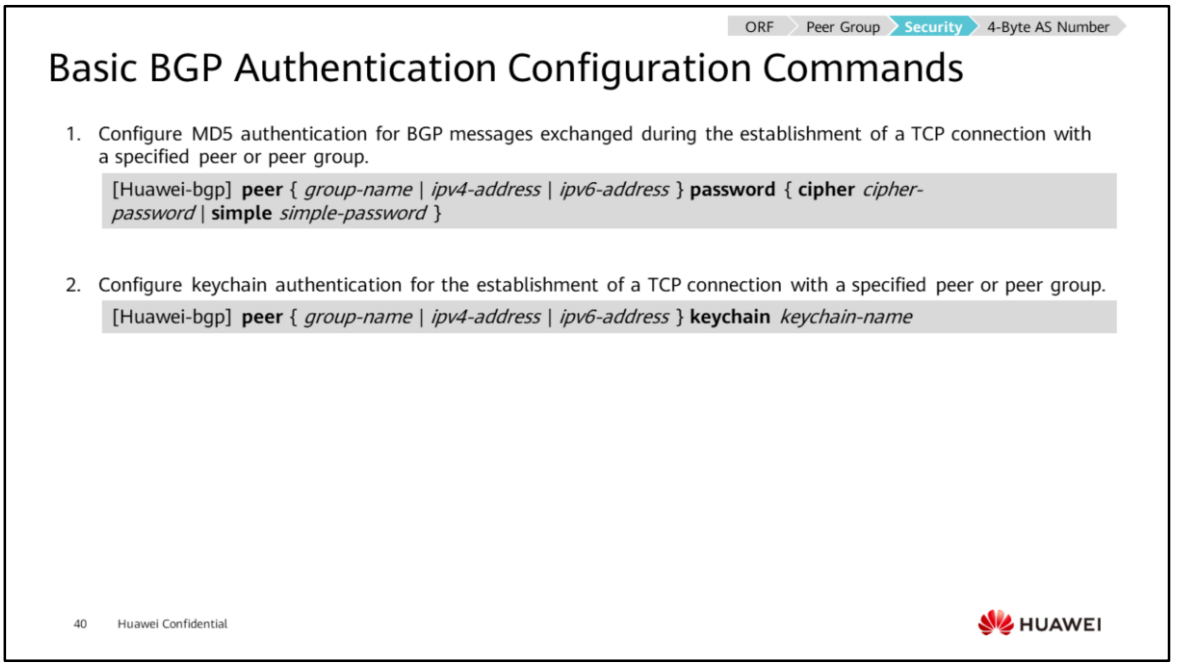
- BGP GTSM can check whether the time to live (TTL) value in the IP packet header is within a preset range, and drop the packets whose TTL values are not within the preset range. This prevents bogus BGP messages from attacking the device.



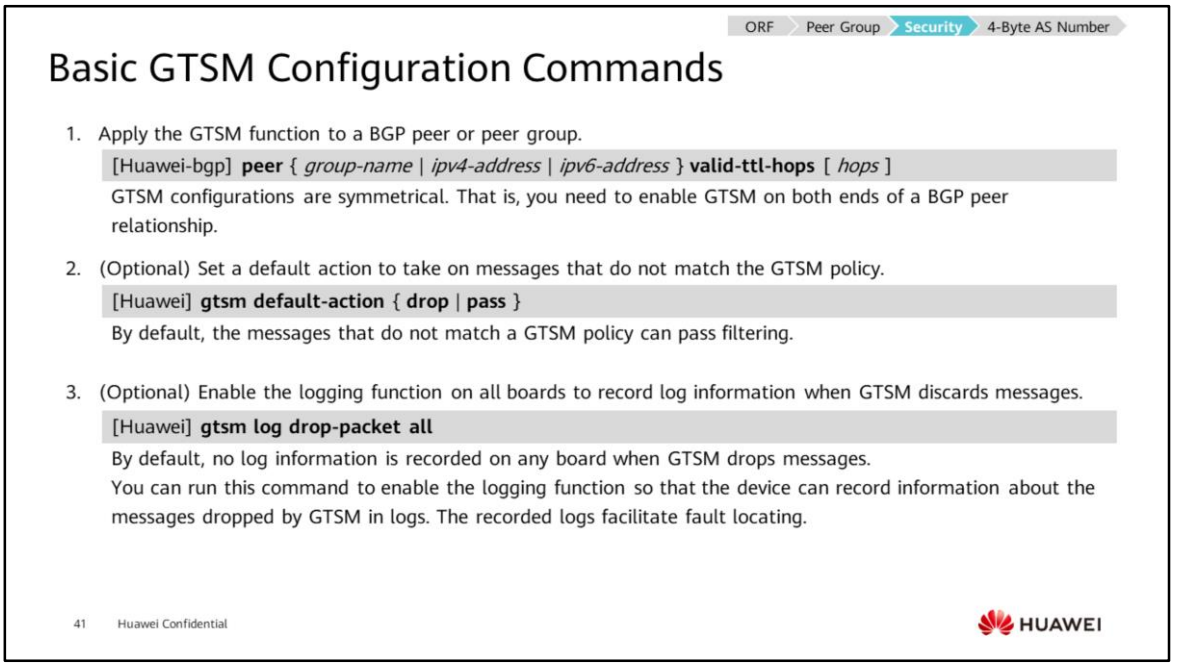
- When an attacker continuously sends bogus BGP messages to attack R2, the TTL values of such messages are smaller than 255.
- If BGP GTSM is enabled on R2 and the valid TTL range for messages sent by the IBGP peer is set to [255, 255], the system checks the TTL values of all BGP messages and drops the bogus messages whose TTL values are smaller than 255. This prevents the bogus messages from consuming CPU resources.

HUAWEI

- As shown in the figure, if BGP GTSM is not enabled, the device finds that the received numerous bogus BGP messages are destined for itself, and directly sends them to the control plane for processing. As a result, the control plane has to process a large number of bogus messages, causing the CPU usage to go excessively high and the system to be unexpectedly busy.

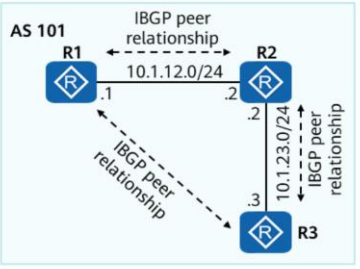


- Command: [Huawei-bgp] **peer** { *group-name* | *ipv4-address* | *ipv6-address* } **keychain** *keychain-name*
  - *keychain-name:* specifies the name of a keychain. The value is a string of 1 to 47 case-insensitive characters. It cannot contain question marks (?). If spaces are used, the string must start and end with double quotation marks (").



- Command: [Huawei-bgp] **peer** { *group-name* | *ipv4-address* | *ipv6-address* } **valid-ttl-hops** [*hops*]
  - *hops:* specifies the number of TTL hops to be checked. The value is an integer ranging from 1 to 255. The default value is 255. If you specify hops, the valid range of TTL values in the messages to be checked is [255 – hops + 1, 255].
- Command: [Huawei] **gtsm default-action** { **drop** | **pass** }
  - **drop:** indicates that the messages that do not match the GTSM policy cannot pass filtering and are dropped.
  - **pass:** indicates that the messages that do not match the GTSM policy can pass filtering.
- Command: [Huawei] **gtsm log drop-packet all**
  - **all:** indicates all boards.

Example for Configuring GTSM (1)



R1, R2, and R3 all belong to AS 101 and use loopback0 interfaces to establish full-mesh IBGP connections. GTSM needs to be enabled on them to prevent CPU attacks.

Device	Interface	Interface Address
R1	Loopback0	10.1.1.1/32
R2	Loopback0	10.1.2.2/32
R3	Loopback0	10.1.3.3/32

1. Establish full-mesh IBGP connections.

```
[R1] bgp 101
[R1-bgp] peer 10.1.2.2 as-number 101
[R1-bgp] peer 10.1.2.2 connect-interface Loopback 0
[R1-bgp] peer 10.1.3.3 as-number 101
[R1-bgp] peer 10.1.3.3 connect-interface Loopback 0
[R1-bgp] network 10.1.1.1 32
```

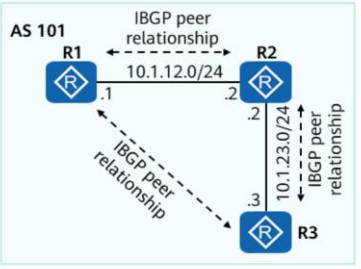
```
[R2] bgp 101
[R2-bgp] peer 10.1.1.1 as-number 101
[R2-bgp] peer 10.1.1.1 connect-interface Loopback 0
[R2-bgp] peer 10.1.3.3 as-number 101
[R2-bgp] peer 10.1.3.3 connect-interface Loopback 0
```

```
[R3] bgp 101
[R3-bgp] peer 10.1.1.1 as-number 101
[R3-bgp] peer 10.1.1.1 connect-interface Loopback 0
[R3-bgp] peer 10.1.2.2 as-number 101
[R3-bgp] peer 10.1.2.2 connect-interface Loopback 0
```


HUAWEI

- As shown in the figure:
  - Assume that static routes are used or OSPF is used to ensure internal network reachability in AS 101. The configuration details are not provided here.
  - R1 advertises the route destined for the IP address of its loopback0 interface to the BGP routing table.

Example for Configuring GTSM (2)



R1, R2, and R3 all belong to AS 101 and use loopback0 interfaces to establish full-mesh IBGP connections. GTSM needs to be enabled on them to prevent CPU attacks.

Device	Interface	Interface Address
R1	Loopback0	10.1.1.1/32
R2	Loopback0	10.1.2.2/32
R3	Loopback0	10.1.3.3/32

2. Enable GTSM between R1 and R2. As the two routers are directly connected, the valid TTL range of the messages from one router to the other is [255, 255]. In this case, the value of **valid-ttl-hops** is 1.

```
[R1-bgp] peer 10.1.2.2 valid-ttl-hops 1
```

```
[R2-bgp] peer 10.1.1.1 valid-ttl-hops 1
```

3. Enable GTSM between R2 and R3. As the two routers are directly connected, the valid TTL range of the messages from one router to the other is [255, 255]. In this case, the value of **valid-ttl-hops** is 1.

```
[R2-bgp] peer 10.1.3.3 valid-ttl-hops 1
```

```
[R3-bgp] peer 10.1.2.2 valid-ttl-hops 1
```

4. Enable GTSM between R1 and R3. As the two routers are connected through R2, the valid TTL range of the messages from one end to the other is [254, 255]. In this case, the value of **valid-ttl-hops** is 2.

```
[R1-bgp] peer 10.1.3.3 valid-ttl-hops 2
```

```
[R3-bgp] peer 10.1.1.1 valid-ttl-hops 2
```


HUAWEI

Verifying the GTSM Configuration

```
[R1]display bgp peer 10.1.3.3 verbose
                BGP Peer is 10.1.3.3,  remote AS 101
                Type: IBGP link
                BGP version 4, Remote router ID 10.1.3.3
                Update-group ID: 1
                BGP current state: Established, Up for 00h02m17s
                BGP current event: KATimerExpired
                BGP last state: OpenConfirm
                BGP Peer Up count: 1
                Received total routes: 0
                Received active routes total: 0
                Advertised total routes: 1
                Port:  Local - 179           Remote - 51077
                Configured: Connect-retry Time: 32 sec
                Configured: Active Hold Time: 180 sec  Keepalive Time:60 sec
                Received  : Active Hold Time: 180 sec
                Negotiated: Active Hold Time: 180 sec  Keepalive Time:60 sec
                Peer optional capabilities:
                Peer supports bgp multi-protocol extension
                Peer supports bgp route refresh capability
                Peer supports bgp 4-byte-as capability
```

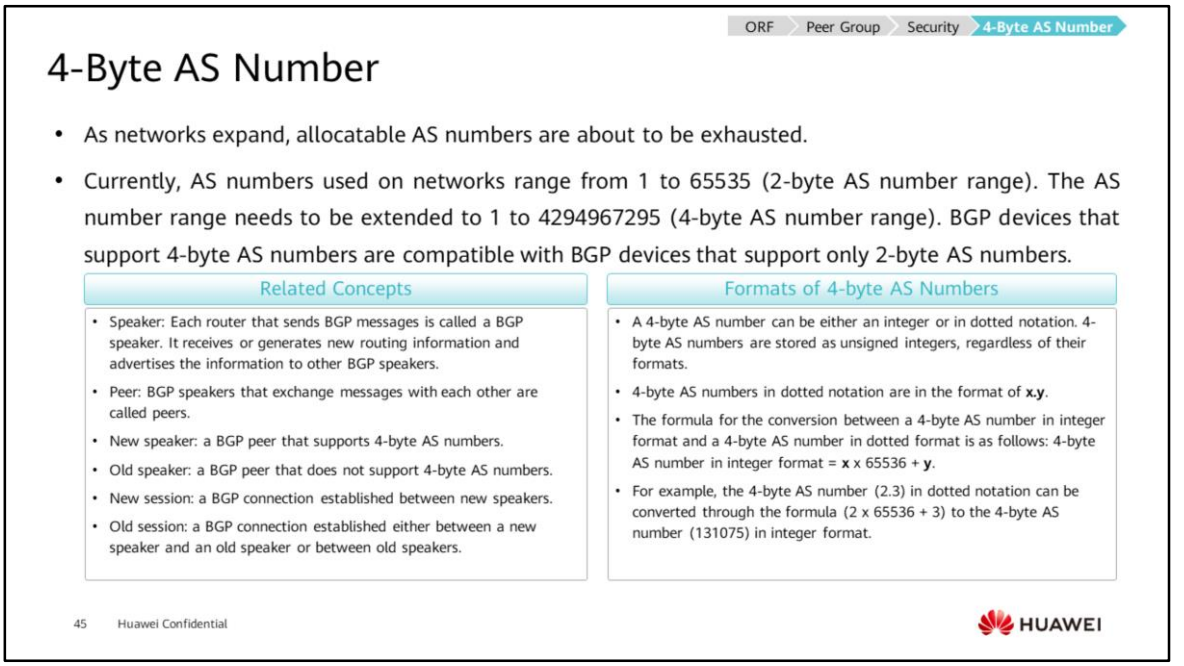
To be continued on the right

```
Received: Total 5 messages
                        Update messages     0
                        Open messages       1
                        KeepAlive messages  3
                        Notification messages 0
                        Refresh messages    1
Sent: Total 8 messages
                        Update messages     2
                        Open messages       2
                        KeepAlive messages  3
                        Notification messages 0
                        Refresh messages    1
Authentication type configured: None
Last keepalive received: yy/mm/dd 17:34:13 UTC-08:00
Last keepalive sent   : yy/mm/dd 17:34:13 UTC-08:00
Last update   sent    : yy/mm/dd 17:34:02 UTC-08:00
Minimum route advertisement interval is 15 seconds
Optional capabilities:
Route refresh capability has been enabled
4-byte-as capability has been enabled
Connect-interface has been configured
GTSM has been enabled, valid-ttl-hops: 2
Peer Preferred Value: 0
Routing policy configured:
No routing policy is configured
```

GTSM is enabled on R1, and the number of valid hops between R1 and its IBGP peer 10.1.3.3 (R3) is 2.

Run the **display bgp peer** command to check BGP peer information.
By specifying the **verbose** parameter, you can check information about the BGP timer, numbers of received and sent routes, capabilities supported by the peer, and enabled functions.

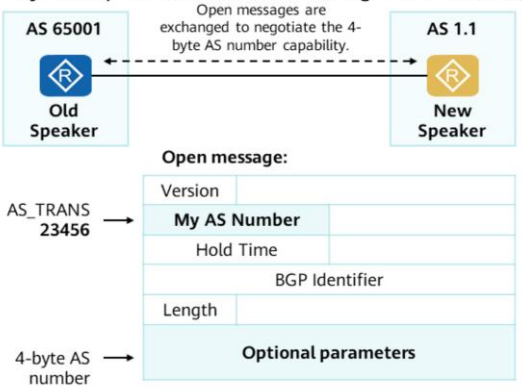
HUAWEI



- 4-byte AS numbers are extended from 2-byte AS numbers. BGP peers use a new capability code and new optional transitive attributes to negotiate the 4-byte AS number capability and transmit 4-byte AS numbers. This mechanism enables communication between new speakers as well as between old speakers and new speakers.
    - To support 4-byte AS numbers, an open capability code 0x41 is defined in a standard protocol for BGP connection capability negotiation. 0x41 indicates that the BGP speaker supports 4-byte AS numbers.
    - In addition, two new optional transitive attributes AS4_Path and AS4_Aggregator are defined in the standard protocol to transmit 4-byte AS information in old sessions.
    - To set up a BGP connection between a new speaker with an AS number greater than 65535 and an old speaker, the peer AS number on the old speaker must be set to AS_TRANS, which is a reserved AS number with the value being 23456.
- NetEngine 8000 series routers support configuration of 4-byte AS numbers in either format. The formats of 4-byte AS numbers displayed in the configuration file are the same as those used by users during configuration.

Open Messages Carrying 4-Byte AS Numbers

- BGP speakers negotiate capabilities by exchanging Open messages. The format of Open messages sent by new speakers is shown in the figure at the bottom of this slide.

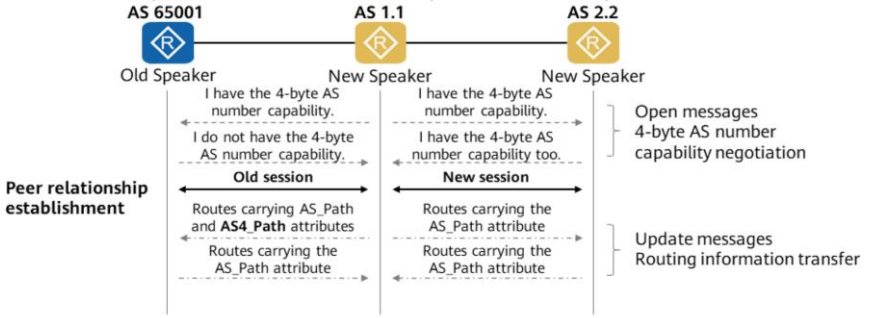


- The header of a BGP Open message is fixed, in which **My AS Number** is supposed to be the local AS number. However, **My AS Number** has only 2 bytes, unable to support 4-byte AS numbers.
- Therefore, a new speaker places the AS_TRANS 23456 in **My AS Number** and its AS number in **Optional parameters** before it sends an Open message to a peer. After the peer receives the message, it can determine whether the other end supports 4-byte AS numbers by checking **Optional parameters** in the message.

HUAWEI

Fundamentals of 4-Byte AS Numbers

- The following figure shows the process of establishing peer relationships and transmitting routing information between new speakers and between new speakers and old speakers.



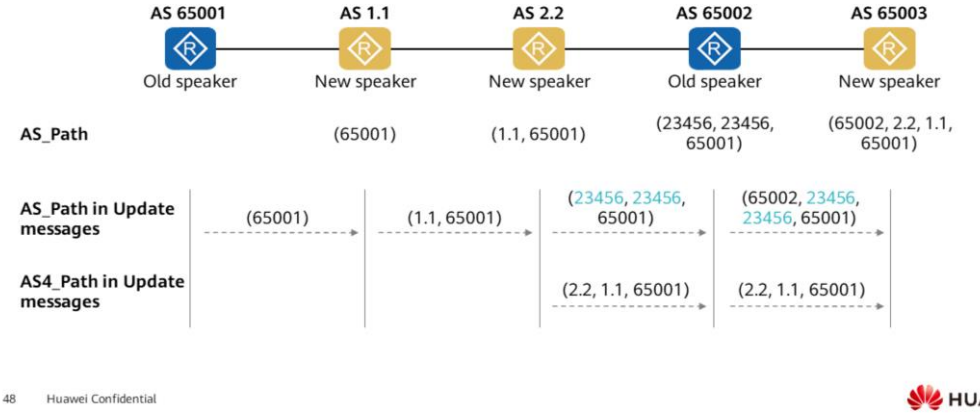
- BGP speakers notify each other of whether they have the 4-byte AS number capability by exchanging Open messages. After the capability negotiation, new sessions are established between new speakers, and old sessions are established between a new speaker and an old speaker.

HUAWEI

- Note: When transmitting routing information to an old speaker, a new speaker encapsulates both the AS_Path and AS4_Path attributes to help transmit 4-byte AS numbers. The AS4_Path attribute is transitive. Therefore, when the old speaker receives the AS4_Path attribute, it transmits the AS4_Path attribute to other speakers.

Transmission of 4-Byte AS Numbers

- In the Update messages exchanged between new speakers, AS numbers in the AS_Path attribute are encoded in 4-byte mode. In the Update messages sent by old speakers, AS numbers in the AS_Path attribute are encoded in 2-byte mode.



- When a new speaker sends an Update message carrying an AS number greater than 65535 to an old speaker, the new speaker uses the AS4_Path attribute to assist the AS_Path attribute in transferring 4-byte AS numbers. The AS4_Path attribute is transparent to the old speaker.
  - On the network shown in this figure, before the new speaker in AS 2.2 sends an Update message to the old speaker in AS 65002, the new speaker replaces each 4-byte AS number (1.1 and 2.2) with 23456 (AS_TRANS) in AS_Path; therefore, the AS_Path carried in the Update message is (23456, 23456, 65001), and the carried AS4_Path is (2.2, 1.1, 65001). Upon receiving the Update message, the old speaker in AS 65002 transparently transmits AS4_Path (2.2, 1.1, 65001) to another AS.
- When a new speaker receives an Update message carrying the AS_Path and AS4_Path attributes from an old speaker, the new speaker obtains the actual AS_Path attribute based on the reconstruction algorithm.
  - In the figure, after the new speaker in AS 65003 receives an Update message carrying AS_Path (65002, 23456, 23456, 65001) and AS4_Path (2.2, 1.1, 65001) from the old speaker in AS 65002, the new speaker obtains the actual AS_Path (65002, 2.2, 1.1, 65001) through reconstruction.

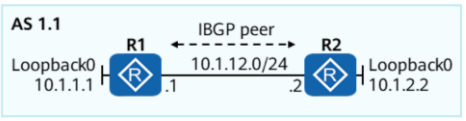
4-Byte AS Number Configuration

1. Enable BGP and enter the BGP view, or directly enter the BGP view.

   [Huawei] **bgp** *as-number*

   as-number: specifies an AS number.
   For an AS number in integer format, the value ranges from 1 to 4294967295.
   For an AS number in dotted notation, it is in the format of x.y, in which x and y are integers, with x ranging from 1 to 65535 and y ranging from 0 to 65535.

- Configuration example:
  - R1 and R2 are in AS 1.1 and use Loopback0 interfaces to establish an IBGP peer relationship.



Establish an IBGP connection.

```
[R1] bgp 1.1
[R1-bgp] peer 10.1.2.2 as-number 1.1
[R1-bgp] peer 10.1.2.2 connect-interface Loopback 0
```

```
[R2] bgp 1.1
[R2-bgp] peer 10.1.1.1 as-number 1.1
[R2-bgp] peer 10.1.1.1 connect-interface Loopback 0
```


HUAWEI

- In the figure:
  - Assume that static routes are configured or OSPF is used to ensure internal network reachability in AS 1.1. The configuration details are not provided here.
  - Configure R1 to advertise the route to Loopback0's address to BGP. The configuration details are not provided here.
- Notes:
  - This slide uses NetEngine 8000 series routers as an example. For the 4-byte AS number configuration on any other type of product, see the corresponding product documentation.
  - If you adjust the display format of 4-byte AS numbers, the matching results in the case of filtering using AS_Path regular expressions or extended community filters are affected. Specifically, after the display format of 4-byte AS numbers is changed when an AS_Path regular expression or extended community filter has been used in an export or import policy, the AS_Path regular expression or extended community filter needs to be reconfigured. If reconfiguration is not performed, routes may fail to match the export or import policy, leading to a network fault.

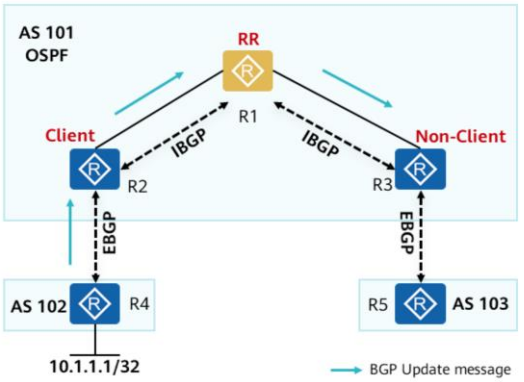
Contents

1. BGP Route Control
2. Introduction to BGP Features
3. **Networking Modes of BGP RRs**

HUAWEI

RR

- Using RRs can avoid the need for full-mesh IBGP connections and reduce the burden on the network and CPU.



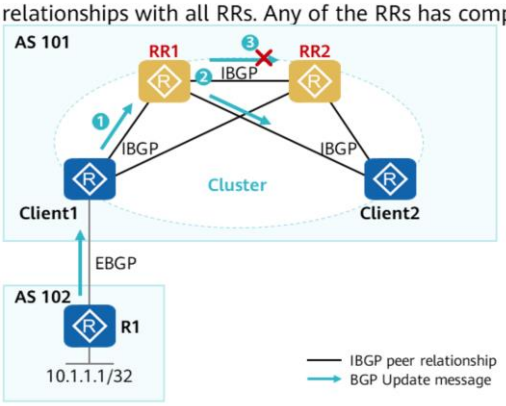
- After an RR is introduced, three roles exist:
  - Route reflector
  - Client
  - Non-client
- The RR reflects the learned routes so that IBGP routes are advertised in the AS, avoiding the need to establish full-mesh IBGP connections.
- After the RR receives routes from its peers, it selects the optimal route based on BGP route selection rules and advertises the optimal route to IBGP peers based on certain rules.

HUAWEI

- RR-related roles:
  - RR: BGP device that reflects the routes learned from an IBGP peer to other IBGP peers. An RR is similar to the designated router (DR) on an OSPF network.
  - Client: IBGP peer whose routes are reflected by the RR to other IBGP peers. In an AS, clients only need to be directly connected to the RR.
  - Non-client: IBGP device that is neither an RR nor a client. In an AS, full-mesh connections still must be established between non-clients and RRs, and between all non-clients.
  - Originator: device that originates routes in an AS. The Originator_ID attribute is used to prevent routing loops in a cluster.
  - Cluster: a set of RRs and their clients. The Cluster_List attribute is used to prevent routing loops between clusters.
- When configuring a BGP router as an RR, you also need to specify a client of the RR. A client does not need to be configured because it is not aware that an RR exists on the network.
- Rules for an RR to advertise routes:
  - After learning routes from non-clients, the RR selects and advertises the optimal route to all its clients.
  - After learning routes from clients, the RR selects and advertises the optimal route to all its non-clients and clients (except the originating client).
  - After learning routes learned from EBGP peers, the RR selects and advertises the optimal route to all its clients and non-clients.

Backup RR Networking

- To improve network reliability and prevent a single point of failure (SPOF) from affecting the network, multiple RRs need to be configured in a cluster. Routers on a forwarding path establish IBGP peer relationships with all RRs. Any of the RRs has complete BGP routes.

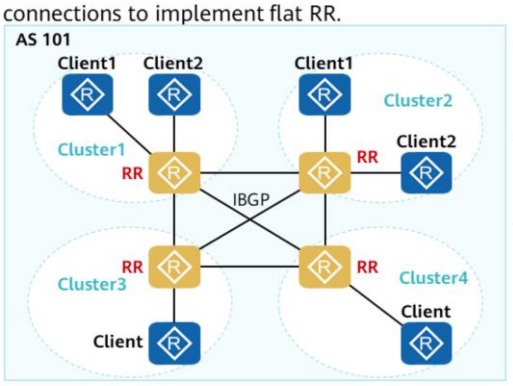


- RR1 and RR2 are in the same cluster and configured with the same cluster ID.
- Route reflection in single-cluster RR networking (RR1 is used as an example):
  1. When client 1 receives an updated route from an EBGP peer, it advertises this route to RR1 and RR2 through IBGP peer relationships.
  2. After RR1 and RR2 receive this route, they add the local cluster ID to the top of the cluster list of the route, and then reflect the route to client 2 and to each other.
  3. After RR1 and RR2 receive the reflected route, they check the cluster list of the route, finding that the cluster list contains their local cluster ID. RR1 and RR2 discard this route to reduce BGP memory usage.

HUAWEI

Multi-cluster RR Networking (1)

- If multiple clusters exist in an AS, RRs of the clusters establish IBGP peer relationships with each other.
- When RRs reside at the same network layer, RRs of different clusters can establish full-mesh connections to implement flat RR.



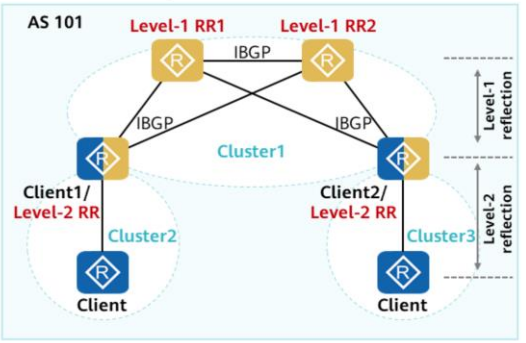
- An AS on a backbone network may be divided into multiple clusters. RRs of the clusters are non-clients of each other and establish full-mesh IBGP connections with each other.
- Although each client in a cluster establishes an IBGP connection only with the RR in the same cluster, each RR and client can receive all the routing information.
- As shown in the figure, one RR resides in each cluster: 1, 2, 3, and 4. An IBGP connection is established between any two of the RRs. Each client in a cluster establishes an IBGP connection only with the RR in the same cluster.

—— IBGP peer relationship

HUAWEI

Multi-cluster RR Networking (2)

- If multiple clusters exist in an AS, RRs of the clusters establish IBGP peer relationships with each other.
- When the RRs reside at different network layers, the RRs at the lower network layer can be configured as clients to implement hierarchical RR.



- In practice, hierarchical RR deployment is used more widely.
- As shown in the figure, AS 101 is divided into three clusters:
  - The four devices in cluster 1 are core routers and work in master/backup mode to ensure high reliability. Two Level-1 RRs are deployed in cluster 1, and the other two routers function as clients of the RRs.
  - A Level-2 RR is deployed in each of cluster 2 and cluster 3. The Level-2 RRs are also clients of the Level-1 RRs. An IBGP connection does not need to be established between the Level-2 RRs.

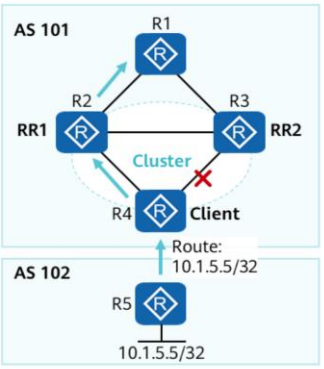
IBGP peer relationship

HUAWEI

- The route advertisement rules for hierarchical RR networking are the same as those for single-cluster RR networking.
- The following factors need to be considered for hierarchical RR design:
  - Size of the top-layer full-mesh topology: If the number of full-mesh IBGP connections has exceeded the management capacity, hierarchical RR networking can be deployed.
  - Number of alternate paths: This factor affects load balancing and resource consumption. More layers reduce the number of links for load balancing but require fewer router resources.

Single-Cluster Problem

- To provide the desired redundancy in an RR-based architecture, it is important to properly divide an AS into clusters.

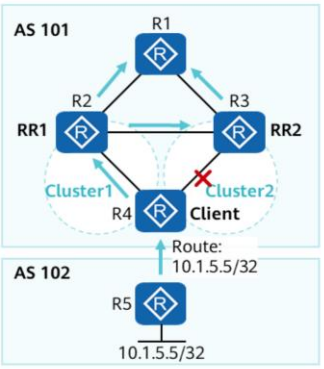


- Scenario description:
  - As shown in the figure, AS 101 uses the backup RR networking. RR1 and RR2 use the same cluster ID, and provide a redundant link for traffic from R1 to access 10.1.5.5/32.
  - After R4 advertises the route 10.1.5.5/32, the two RRs advertise the route to R1 and to each other. Because RR1 and RR2 have the same cluster ID, the Update messages exchanged between the RRs are discarded.
- An IBGP session failure causes the redundancy failure:
  - Assume that the IBGP session between R3 and R4 fails (for example, due to incorrect configurations). As R3 ignores the route 10.1.5.5/32 advertised by R2, no redundant link is available for traffic from R1 to access 10.1.5.5/32.

HUAWEI

Multi-cluster Design

- Multi-cluster design not only provides physical redundancy in response to link failures, but also provides logical redundancy in response to IBGP session failures between clients and RRs.



- As shown in the figure, R2 and R4 are added to cluster 1, and R3 and R4 are added to cluster 2.
- If the IBGP session between R3 and R4 fails, R3 can still forward traffic because R3 learns the route 10.1.5.5/32 advertised by R2.

HUAWEI

Quiz

1. (Single-answer question) Which of the following AS_Path value is matched against the ip as-path-filter 1 permit ^[100|200]$ command? ( )
   A. AS_Path 100
   B. AS_Path 200
   C. AS_Path 100 200
   D. AS_Path 100 or AS_Path 200

HUAWEI

1. D

Quiz

2. (True or false) Authenticate BGP peer relationships can prevent unauthorized BGP peer relationships from being established. ( )
   A. True
   B. False
3. (True or false) In the backup RR networking of BGP, the master and backup RRs discard the routes reflected by each other to prevent routing loops. ( )
   A. True
   B. False

HUAWEI

2. A
3. A

Summary

- As dedicated route matching tools of BGP, AS_Path and community filters can be used to match BGP routes based on the AS_Path and community attributes, respectively, and can be used with route-policies to implement BGP route control.
- BGP also supports various advanced features, including ORF to achieve on-demand route advertisement, peer groups to simplify configurations, BGP security features to prevent the establishment of rogue peer relationships as well as attacks through bogus BGP messages, and 4-byte AS numbers to meet future network expansion requirements.
- BGP uses RRs to break the split horizon rule of IBGP, avoid the need for full-mesh IBGP connections, and reduce the burden on the network and CPU. Common RR networking modes include backup-RR networking and multi-cluster RR networking.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Network Security Technologies

HUAWEI

Foreword

- Currently, Ethernet technologies are widely used on networks. Network attacks often occur, for example, attacks based on the Address Resolution Protocol (ARP) and Dynamic Host Configuration Protocol (DHCP). Such attacks cause authorized users' failure to access network resources and threaten network information security. In this situation, Ethernet switching security becomes increasingly important.
- On networks of large- and medium-sized enterprises, firewalls are usually deployed in hot standby mode to ensure that the standby firewall can smoothly take over services of the active firewall when the active firewall fails, ensuring service continuity. The virtual system feature allows a physical firewall to be logically divided into multiple independent virtual systems. Each virtual system functions as a real device.
- This course describes common Ethernet switching security technologies, including port isolation, port security, MAC address flapping detection, storm control, interface rate limiting, MAC address table security, DHCP snooping, and IP source guard; and advanced firewall features, including hot standby and virtual system.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe types and configurations of port isolation.
  - Illustrate the working mechanism of port security.
  - Describe MAC address flapping detection.
  - Explain traffic suppression and storm control functions of switches.
  - Describe application scenarios of DHCP snooping.
  - Illustrate the working mechanism of IP source guard.
  - Describe the working scenarios of firewall hot backup.

HUAWEI

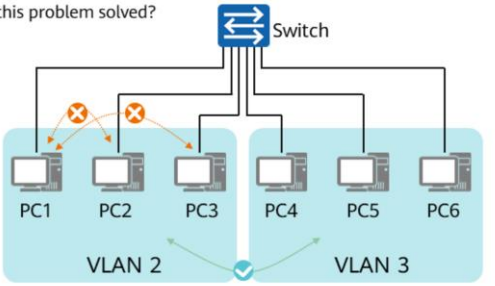
Contents

1. **Ethernet Switching Security**
   - Port Isolation
   - MAC Address Table Security
   - Port Security
   - MAC Address Flapping Prevention and Detection
   - MACsec
   - Traffic Control
   - DHCP Snooping
   - IP Source Guard
2. Advanced Firewall Features

HUAWEI

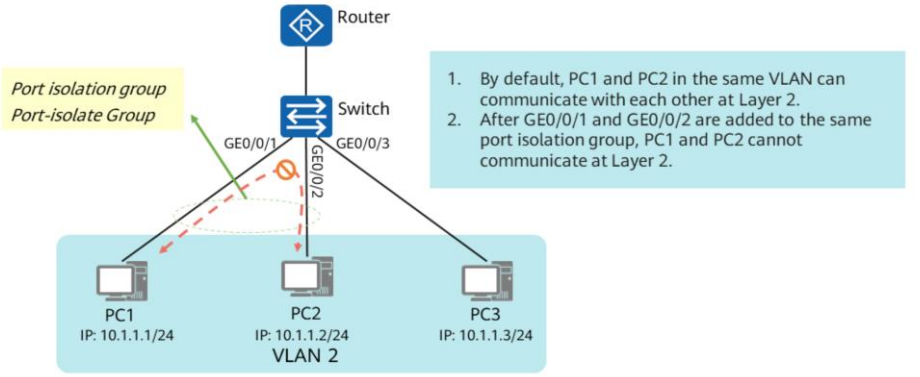
Background of Port Isolation

- On an Ethernet switching network, to implement Layer 2 isolation between packets, users usually add different interfaces to different VLANs to isolate Layer 2 broadcast domains.
- On a large-scale network, there are various service requirements. Simply using VLANs to implement Layer 2 isolation of packets wastes limited VLAN resources.
- As shown in the following figure, although PC1 and PC2 belong to the same VLAN, they cannot communicate with each other at Layer 2 but can communicate with each other at Layer 3. PC1 and PC3 cannot communicate with each other in any case, but hosts in VLAN 3 can access hosts in VLAN 2. How is this problem solved?

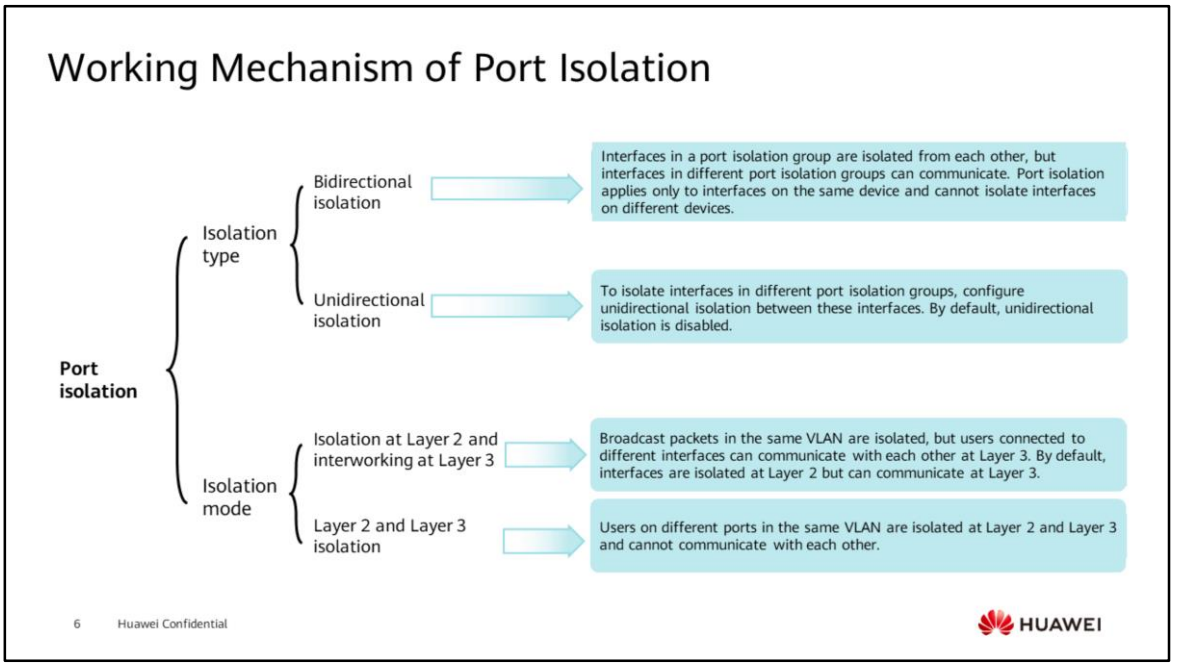


Overview of Port Isolation

- Port isolation can isolate interfaces in a VLAN. That is, you only need to add interfaces to a port isolation group to implement Layer 2 isolation between these interfaces. Port isolation provides secure and flexible networking schemes for customers.



HUAWEI



- When Layer 2 isolation and Layer 3 interworking are used, you can enable intra-VLAN proxy ARP on the VLANIF interface and configure **arp-proxy inner-sub-vlan-proxy enable** to implement communication between hosts in the same VLAN.

Port Isolation Configuration Commands

1. Enable port isolation.

   [Huawei-GigabitEthernet0/0/1] **port-isolate enable** [**group** *group-id*]

   By default, port isolation is disabled on an interface. If *group-id* is not specified, an interface is added to port isolation group 1 by default.

2. (Optional) Configure a port isolation mode.

   [Huawei] **port-isolate mode** { **l2** | **all** }

   By default, the port isolation mode is Layer 2 isolation and Layer 3 interworking.

   l2: Layer 2 isolation and Layer 3 interworking

   all: Layer 2 and Layer 3 isolation

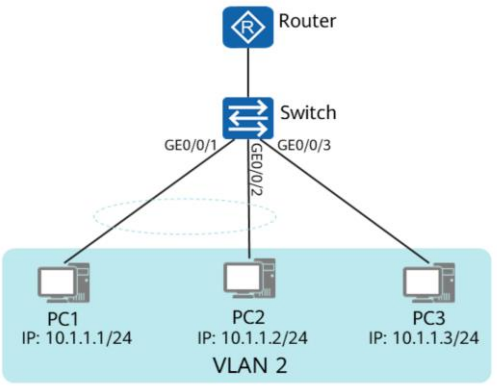
3. Configure unidirectional isolation.

   [Huawei-GigabitEthernet0/0/1] **am isolate** {*interface-type interface-number*}&<1-8>

   This command is used to unidirectionally isolate the current interface from a specified interface. If interface A is isolated from interface B unidirectionally, packets sent from interface A cannot reach interface B, but packets sent from interface B can reach interface A. By default, unidirectional isolation is disabled.

HUAWEI

Example for Configuring Port Isolation



Switch configuration:

```
[Switch] vlan 2
[Switch] port-isolate mode all
[Switch] interface GigabitEthernet 0/0/1
[Switch-GigabitEthernet0/0/1] port link-type access
[Switch-GigabitEthernet0/0/1] port default vlan 2
[Switch-GigabitEthernet0/0/1] port-isolate enable group 2
[Switch] interface GigabitEthernet 0/0/2
[Switch-GigabitEthernet0/0/2] port link-type access
[Switch-GigabitEthernet0/0/2] port default vlan 2
[Switch-GigabitEthernet0/0/2] port-isolate enable group 2
[Switch] interface GigabitEthernet 0/0/3
[Switch-GigabitEthernet0/0/3] port link-type access
[Switch-GigabitEthernet0/0/3] port default vlan 2
```

- As shown in the figure, PC1, PC2, and PC3 belong to VLAN 2. After port isolation is configured, PC3 can communicate with PC1 and PC2, but PC1 and PC2 cannot communicate with each other.

HUAWEI

Verifying the Configuration

1. Run **the display port-isolate group group-number** command to check interfaces in the port isolation group.

```
[SW]display port-isolate group 2
  The ports in isolate group 2:
GigabitEthernet0/0/1     GigabitEthernet0/0/2
```

2. Verify that hosts in the same port isolation group cannot communicate with each other.

```
PC1>ping 10.1.1.2
Ping 10.1.1.2: 32 data bytes, Press Ctrl_C to break
From 10.1.1.1: Destination host unreachable
From 10.1.1.1: Destination host unreachable
From 10.1.1.1: Destination host unreachable
From 10.1.1.1: Destination host unreachable
From 10.1.1.1: Destination host unreachable
--- 10.1.1.2 ping statistics ---
  5 packet(s) transmitted
  0 packet(s) received
  100.00% packet loss
```

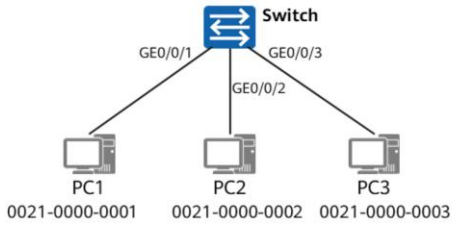

HUAWEI

Contents

1. **Ethernet Switching Security**
   - Port Isolation
   - MAC Address Table Security
   - Port Security
   - MAC Address Flapping Prevention and Detection
   - MACsec
   - Traffic Control
   - DHCP Snooping
   - IP Source Guard
2. Advanced Firewall Features

HUAWEI

Types of MAC Address Entries

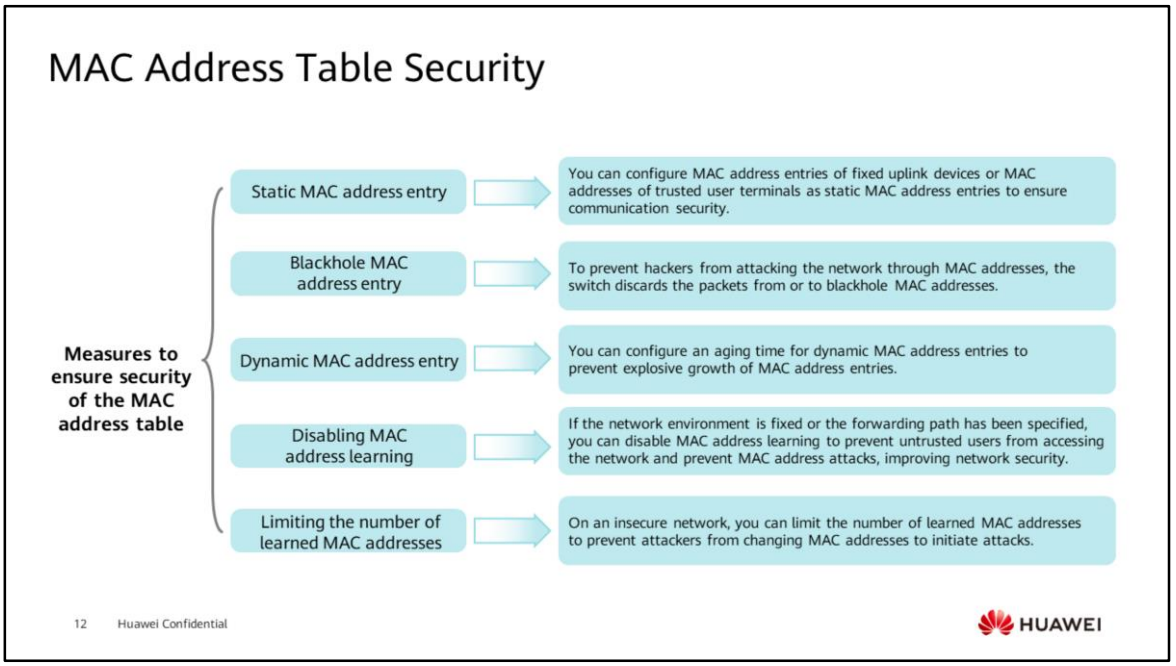


MAC Address	VLAN	Interface	Type
0021-0000-0001	10	GE0/0/1	Static
0021-0000-0002	10	GE0/0/2	Blackhole
0021-0000-0003	10	GE0/0/3	Dynamic

- MAC address entries fall into the following types:
  - **Dynamic MAC address entries** are obtained by learning source MAC addresses of packets received on an interface, and will age out. After a device resets or an interface board is hot swapped or resets, dynamic MAC address entries on the device or interface board are lost.
  - **Static MAC address entries** that are manually configured by users and delivered to each interface card. Static MAC address entries will never age out. After a device resets or an interface board is hot swapped or resets, the static MAC address entries saved on the device or interface board are not lost. After an interface is statically bound to a MAC address, other interfaces discard packets originating from that source MAC address.
  - **Blackhole MAC address entries** that are manually configured by users and delivered to each interface card. Blackhole MAC address entries will never age out. After blackhole MAC address entries are configured, packets with the source or destination MAC addresses being the blackhole MAC addresses are discarded.

HUAWEI

- A MAC address table is used by the switch to record the mappings between learned MAC addresses of other devices and interfaces on which MAC addresses are learned, as well as VLANs to which the interfaces belong.
- When performing Layer 2 switching, the device searches the MAC address table according to the destination MAC address of the packet. If the MAC address table contains the entry corresponding to the destination MAC address of the packet and the interface that receives the packet is different from the interface corresponding to the entry, the packet is directly forwarded through the outbound interface in the entry. If they are the same, the packet is discarded.
- If the MAC address table does not contain the entry matching the destination MAC address of the packet, the device broadcasts the packet through all the interfaces in the VLAN except the interface that receives the packet.



- To prevent unauthorized users from modifying MAC address entries of some key devices (such as servers or uplink devices), you can configure the MAC address entries of these devices as static MAC address entries. Static MAC address entries take precedence over dynamic MAC address entries and can hardly be modified by unauthorized users.
- To prevent useless MAC address entries from occupying the MAC address table and prevent hackers from attacking user devices or networks using MAC addresses, you can configure untrusted MAC addresses as blackhole MAC addresses. In this way, when the device receives a packet with the destination or source MAC address as the blackhole MAC address, the device discards the packet without modifying the original MAC address entry or adding a MAC address entry.
- To reduce manual configuration of static MAC address entries, Huawei S series switches are enabled with dynamic MAC address learning by default. The aging time needs to be set properly for dynamic MAC address entries so that the switch can delete unneeded MAC address entries.
- To improve network security and prevent the device from learning invalid MAC addresses and incorrectly modifying the original MAC address entries in the MAC address table, you can disable MAC address learning on a specified interface or all interfaces in a specified VLAN so that the device does not learn new MAC addresses from these interfaces.
- You can limit the number of MAC address entries that can be learned on the device. When the number of learned MAC address entries reaches the limit, the device does not learn new MAC address entries. You can also configure an action to take when the number of learned MAC address entries reaches the limit. This prevents MAC address entries from being exhausted and improves network security.

Configuring MAC Address Entries

1. Configure a static MAC address entry.

   [Huawei] **mac-address static** *mac-address interface-type interface-number* **vlan** *vlan-id*

   The specified VLAN must have been created and added to the bound interface. The specified MAC address must be a unicast MAC address and cannot be a multicast or broadcast MAC address.

2. Configure a blackhole MAC address entry.

   [Huawei] **mac-address blackhole** *mac-address* **[vlan** *vlan-id* **]**

   The device discards the received packets originating from or destined for blackhole MAC addresses.

3. Set an aging time for a dynamic MAC address entry.

   [Huawei] **mac-address aging-time** *aging-time*

HUAWEI

Disabling MAC Address Learning

1. Disable MAC address limiting on an interface.

   [Huawei-GigabitEthernet0/0/1] **mac-address learning disable [action { discard | forward }]**

   By default, MAC address learning is enabled on an interface.
   - By default, the device takes the forward action after MAC address learning is disabled. That is, the device forwards packets according to the MAC address table.
   - When the action is set to discard, the device looks up the source MAC address of the packet in the MAC address table. If the interface and MAC address match the MAC address entry, the device forwards the packet according to the destination MAC address. If the interface and MAC address do not match the MAC address entry, the device discards the packet.

2. Disable MAC address learning in a VLAN.

   [Huawei-vlan2] **mac-address learning disable**

   By default, MAC address learning is enabled in a VLAN.
   If both interface-based and VLAN-based MAC address learning is disabled, the latter takes effect.

HUAWEI

Limiting the Number of Learned MAC Address Entries

1. Limit the number of MAC address entries learned on an interface.

   [Huawei-GigabitEthernet0/0/1] **mac-limit maximum** *max-num*

   By default, the number of MAC address entries learned on an interface is not limited.

2. Configure an action for the device to take when the number of learned MAC addresses reaches the limit.

   [Huawei-GigabitEthernet0/0/1] **mac-limit action { discard | forward }**

   By default, the device discards packets with new MAC addresses when the number of learned MAC address entries reaches the limit on the interface.

3. Configure the device whether to generate an alarm when the number of learned MAC addresses reaches the limit.

   [Huawei-GigabitEthernet0/0/1] **mac-limit alarm { disable | enable }**

   By default, the device generates an alarm when the number of learned MAC address entries reaches the limit.

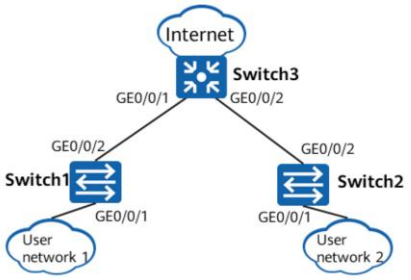
4. Limit the number of MAC address entries learned in a VLAN.

   [Huawei-vlan2] **mac-limit maximum** *max-num*

   By default, the number of MAC address entries learned in a VLAN is not limited.

HUAWEI

Example for Configuring a MAC Address Table



- Requirements:
  - The basic configuration of the network topology is complete. User network 1 belongs to VLAN 10, and user network 2 belongs to VLAN 20.
  - Switch3 is disabled from learning MAC addresses from user network 1.
  - The maximum number of MAC addresses on user network 2 learned by Switch3 is set.

Switch3 configuration:

Method 1: Interface view

```
# Disable MAC address learning on GE0/0/1.
[Switch3-GigabitEthernet0/0/1] mac-address learning disable action discard
# Set the maximum number of MAC address entries learned on GE0/0/2, and configure the device to generate an alarm and set the action to discard when the number of learned MAC address entries reaches the limit.
[Switch3-GigabitEthernet0/0/2] mac-limit maximum 100
[Switch3-GigabitEthernet0/0/2] mac-limit alarm enable
[Switch3-GigabitEthernet0/0/2] mac-limit action discard
```

Method 2: VLAN view

```
# Disable MAC address learning in VLAN 10.
[Switch3-vlan10] mac-address learning disable
# Set the maximum number of MAC address entries learned in VLAN 20, and configure the device to generate an alarm when the number of learned MAC address entries reaches the limit.
[Switch3-vlan20] mac-limit maximum 100 alarm enable
```


HUAWEI

Verifying the Configuration

- Run the **display mac-limit command** in any view to check whether MAC address limiting rules are configured successfully.

```
[Switch3]display mac-limit
MAC Limit is enabled
Total MAC Limit rule count : 2

PORT            VLAN/VSI/SI     SLOT  Maximum  Rate(ms)      Action     Alarm
--------------------------------------------------------------------------------
GE0/0/2         -               -     100      -             forward    enable
-               20              -     100      -             forward    enable
```

VLAN-based MAC address limiting

Interface-based MAC address limiting

HUAWEI

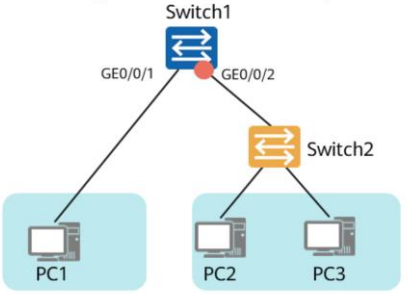
Contents

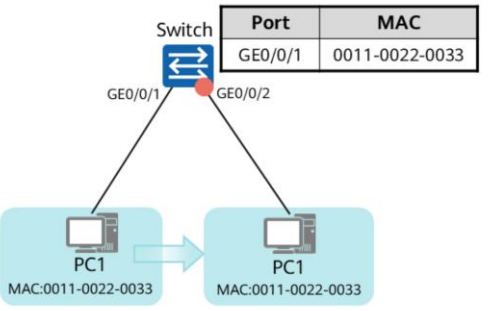
1. **Ethernet Switching Security**
   - Port Isolation
   - MAC Address Table Security
   - Port Security
   - MAC Address Flapping Prevention and Detection
   - MACsec
   - Traffic Control
   - DHCP Snooping
   - IP Source Guard
2. Advanced Firewall Features

HUAWEI

Background of Port Security

- An enterprise requires that each access switch interface connected to terminals allow only one PC to access the network (the number of MAC address entries is limited). If an employee attempts to connect a small switch or hub to an interface, this behavior should be detected or prohibited, as shown in the figure on the left.
- In addition, some enterprises may require that only data frames sent by terminals with trusted MAC addresses can be forwarded to the upper-layer network by the switch. Employees cannot change their locations (change access interfaces of the switch), as shown in the figure on the right.
- Port security of the switch can solve the problems.





HUAWEI

Introduction to Port Security

- You can configure port security on a specified interface of a switch to limit the number of MAC address entries learned by the interface and configure a punishment action when the number of learned MAC address entries exceeds the threshold.
- Port security converts dynamic MAC addresses learned on an interface into secure MAC addresses (including dynamic and static secure MAC addresses, and sticky MAC addresses). This function prevents unauthorized users from communicating with the switch using this interface and therefore enhances device security.

HUAWEI

Working Mechanism of Port Security

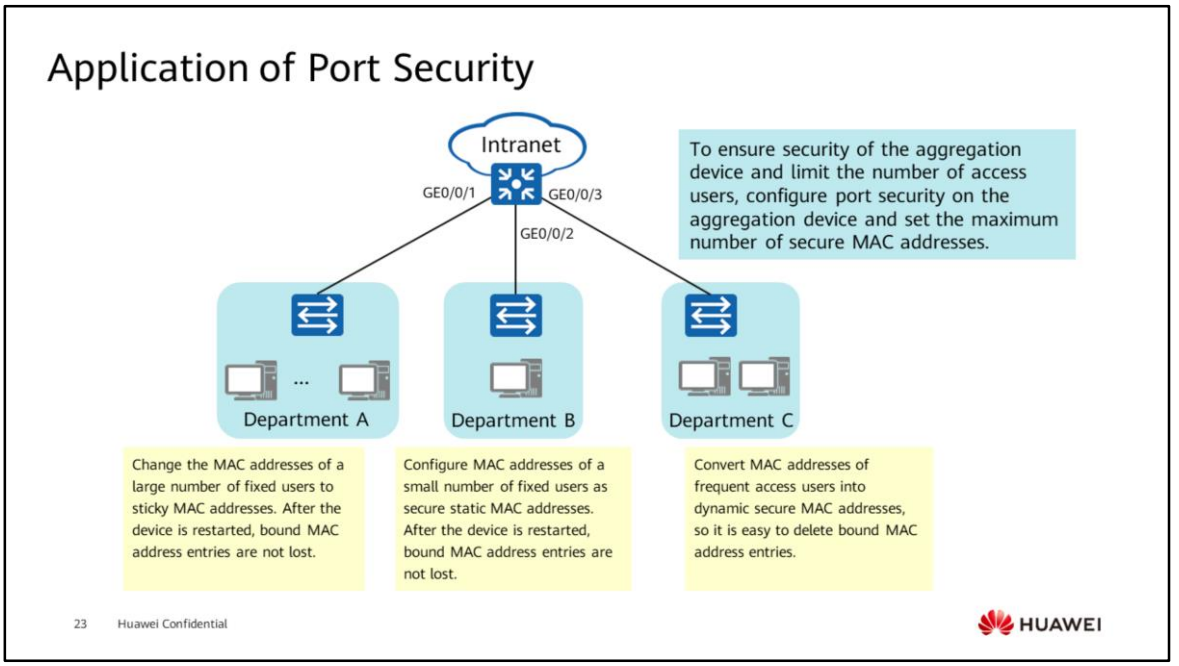
- Secure MAC addresses are classified into the following types.

Type	Definition	Characteristics
Secure dynamic MAC address	MAC address that is converted on an interface with port security enabled but sticky MAC disabled.	After a device restarts, dynamic secure MAC addresses are lost and need to be relearned. By default, dynamic secure MAC addresses are not aged out, and can be aged out only when the aging time is set.
Secure static MAC address	MAC address that is manually configured on an interface with port security enabled.	Secure static MAC addresses are not aged out and are not lost after a device restart.
Sticky MAC address	MAC address that is converted on an interface with both port security and sticky MAC enabled.	Sticky MAC addresses are not aged out and are not lost after a device restart.

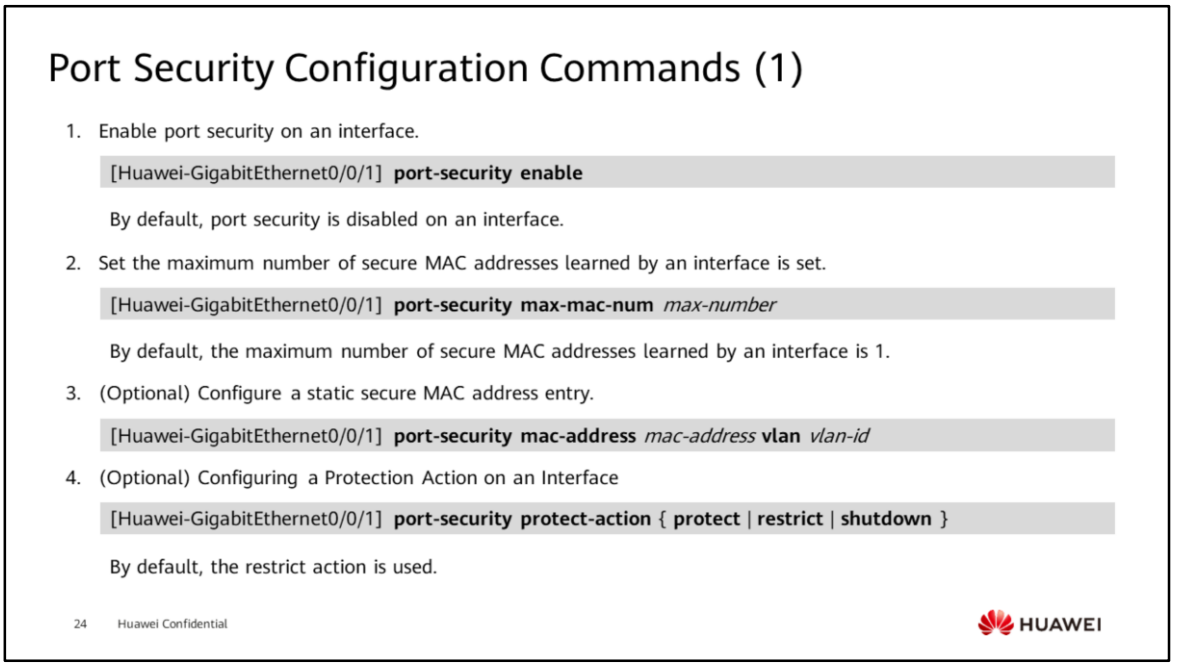
- Secure MAC addresses are usually used together with security protection actions. Common security protection actions are as follows:
  - Restrict: Discards packets with a nonexistent source MAC address and sends a trap.
  - Protect: Discards packets with a nonexistent source MAC address but does not send a trap.
  - Shutdown: Sets the interface state to error-down and generates an alarm.

HUAWEI

- Dynamic secure MAC addresses can be aged out using two modes: absolute aging and relative aging.
  - Absolute aging time: If the absolute aging time is set to 5 minutes, the system calculates the lifetime of each MAC address every minute. If the lifetime is larger than or equal to 5 minutes, the secure dynamic MAC address is aged immediately. If the lifetime is smaller than time minutes, the system determines whether to delete the secure dynamic MAC address after 1 minute.
  - Relative aging time: If the value is set to 5 minutes, the system checks whether there is traffic from a specified dynamic secure MAC address every 1 minute. If no traffic is received from the secure dynamic MAC address, this MAC address is aged out 5 minutes later.
- By default, an interface in error-down state can be restored only after the restart command is run in the interface view.
- To enable an interface in error-down state to automatically go Up after a period of time, run the **error-down auto-recovery cause port-security interval interval-value** command in the system view. In this command, interval-value specifies the period of time after which an interface in error-down state can automatically go Up.



- You can configure port security and set the maximum number of secure MAC addresses learned by an interface on networks demanding high access security. Port security enables the switch to convert MAC addresses learned by an interface into secure MAC addresses and to stop learning new MAC addresses after the maximum number of learned MAC addresses is reached. In this case, the switch can only communicate with devices with learned MAC addresses. If the switch receives packets with a nonexistent source MAC address after the number of secure MAC addresses reaches the limit, the switch considers that the packets are sent from an unauthorized user, regardless of whether the destination MAC address of packets is valid, and takes the configured action on the interface. This prevents untrusted users from accessing these interfaces, improving security of the switch and the network.
- Port security enables the switch to convert MAC addresses learned by an interface into secure MAC addresses and to stop learning new MAC addresses after the maximum number of learned MAC addresses is reached. In this case, the switch can only communicate with devices with learned MAC addresses. If the number of access users changes, you can restart the device or set the aging time of secure MAC address entries to update the MAC address entries. If you do not want to change the MAC address entries of stable access users, you can enable the sticky MAC function on the interface. After the configuration is saved, the MAC address entries will not be updated or lost.



- The port-security protect-action command configures the protection action to be used when the number of learned MAC addresses on an interface exceeds the upper limit or static MAC address flapping is detected.
  - protect
    - Discards packets with new source MAC addresses when the number of learned MAC addresses exceeds the limit.
    - When static MAC address flapping occurs, the interface discards the packets with this MAC address.
  - restrict
    - Discards packets with new source MAC addresses and sends a trap message when the number of learned MAC addresses exceeds the limit.
    - When static MAC address flapping occurs, the interface discards the packets with this MAC address and sends a trap.
  - shutdown
    - Sets the interface state to error-down and generates a trap when the number of learned MAC addresses exceeds the limit.
    - Sets the interface state to error-down and generates a trap when static MAC address flapping occurs.

Port Security Configuration Commands (2)

5. (Optional) Configure an aging time for dynamic secure MAC address entries learned by the interface.

   [Huawei-GigabitEthernet0/0/1] **port-security aging-time** *time* [**type** { **absolute** | **inactivity** }]

   By default, dynamic secure MAC address entries learned by an interface are not aged out.

6. Enable the sticky MAC address function on an interface.

   [Huawei-GigabitEthernet0/0/1] **port-security mac-address sticky**

   By default, the sticky MAC address function is disabled on an interface.

7. Set the maximum number of sticky MAC addresses that can be learned by an interface.

   [Huawei-GigabitEthernet0/0/1] **port-security max-mac-num** *max-number*

   By default, an interface enabled with the sticky MAC address function can learn only one sticky MAC address.

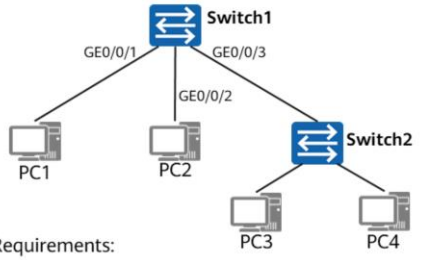
8. (Optional) Configure a sticky MAC address entry.

   [Huawei-GigabitEthernet0/0/1] **port-security mac-address sticky** *mac-address* **vlan** *vlan-id*

HUAWEI

- Check secure MAC addresses.
  - Run the **display mac-address security [vlan vlan-id | interface-type interface-number] * [verbose]** command to check dynamic secure MAC address entries.
  - Run the **display mac-address sec-config [vlan vlan-id | interface-type interface-number] * [verbose]** command to check static secure MAC address entries.
  - Run **the display mac-address sticky [vlan vlan-id | interface-type interface-number] * [verbose]** command to check sticky MAC address entries.

Example for Configuring Port Security - Secure Dynamic MAC Addresses



Switch1 configuration:

```
[Switch1] interface GigabitEthernet 0/0/1
[Switch1-GigabitEthernet 0/0/1] port-security enable
[Switch1-GigabitEthernet 0/0/1] port-security max-mac-num 1
[Switch1-GigabitEthernet 0/0/1] port-security protect-action restrict
[Switch1] interface GigabitEthernet 0/0/2
[Switch1-GigabitEthernet 0/0/2] port-security enable
[Switch1-GigabitEthernet 0/0/2] port-security max-mac-num 1
[Switch1-GigabitEthernet 0/0/2] port-security protect-action restrict
[Switch1] interface GigabitEthernet 0/0/3
[Switch1-GigabitEthernet 0/0/3] port-security enable
[Switch1-GigabitEthernet 0/0/3] port-security max-mac-num 2
[Switch1-GigabitEthernet 0/0/3] port-security protect-action shutdown
```

- Requirements:
  - Configure port security on Switch1.
  - Set the maximum number of MAC addresses learned by GE0/0/1 and GE0/0/2 to 1. When the interface is connected to multiple PCs, Switch1 needs to send an alarm. In addition, interfaces can still forward data frames from authorized PCs.
  - Set the maximum number of MAC addresses that can be learned by GE0/0/3 to 2. When the number of learned MAC addresses exceeds the maximum number, the switch generates an alarm and shuts down GE0/0/3.

HUAWEI

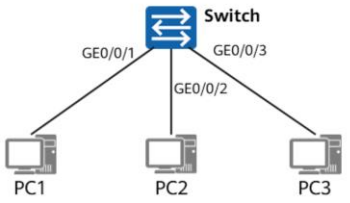
Verifying the Configuration

- Run the **display mac-address security** command to check dynamic secure MAC address entries.

```
[Switch1]display mac-address security
MAC address table of slot 0:
-------------------------------------------------------------------------------
MAC Address    VLAN/      PEVLAN CEVLAN Port          Type          LSP/LSR-ID
               VSI/SI                                               MAC-Tunnel
-------------------------------------------------------------------------------
5489-98ac-71a9 1          -      -      GE0/0/3       security      -
5489-98b1-7b30 1          -      -      GE0/0/1       security      -
5489-9815-662b 1          -      -      GE0/0/2       security      -
-------------------------------------------------------------------------------
Total matching items on slot 0 displayed = 3
```


HUAWEI

Configuring Port Security - Sticky MAC Addresses



Switch configuration:

```
[Switch] interface GigabitEthernet 0/0/1
[Switch-GigabitEthernet 0/0/1] port-security enable
[Switch-GigabitEthernet 0/0/1] port-security max-mac-num 1
[Switch-GigabitEthernet 0/0/1] port-security mac-address sticky
[Switch] interface GigabitEthernet 0/0/2
[Switch-GigabitEthernet 0/0/2] port-security enable
[Switch-GigabitEthernet 0/0/2] port-security max-mac-num 1
[Switch-GigabitEthernet 0/0/2] port-security mac-address sticky
[Switch] interface GigabitEthernet 0/0/3
[Switch-GigabitEthernet 0/0/3] port-security enable
[Switch-GigabitEthernet 0/0/3] port-security max-mac-num 1
[Switch-GigabitEthernet 0/0/3] port-security mac-address sticky
[Switch-GigabitEthernet 0/0/3] port-security mac-address sticky
5489-98ac-71a9 vlan 1
```

- Requirements:
  - Configure port security on the switch. Enable port security on GE0/0/1 through GE0/0/3.
  - Set the maximum number of MAC addresses that can be learned by GE0/0/1 and GE0/0/2 to 1 and convert secure dynamic MAC addresses learned by GE0/0/1 and GE0/0/2 to sticky MAC addresses.
  - On GE0/0/3, set the maximum number of MAC addresses that can be learned to 1, manually create a sticky MAC address entry for the interface, and bind the interface to MAC address 5489-98ac-71a9. Retain the default penalty on each interface.

HUAWEI

Verifying the Configuration

- Run **the display mac-address sticky** command to check sticky MAC address entries.

```
[Switch1]display mac-address sticky
MAC address table of slot 0:
-------------------------------------------------------------------------------
MAC Address    VLAN/      PEVLAN CEVLAN  Port          Type      LSP/LSR-ID
               VSI/SI                                             MAC-Tunnel
-------------------------------------------------------------------------------
5489-98ac-71a9 1          -      -       GE0/0/3       sticky    -
5489-98b1-7b30 1          -      -       GE0/0/1       sticky    -
5489-9815-662b 1          -      -       GE0/0/2       sticky    -
-------------------------------------------------------------------------------
Total matching items on slot 0 displayed = 3
```


HUAWEI

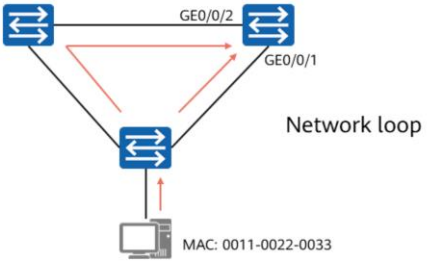
Contents

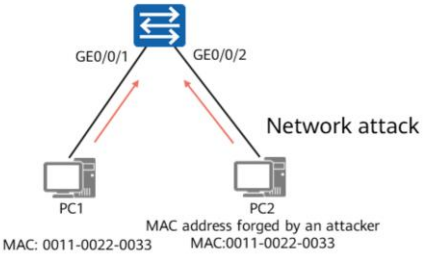
1. **Ethernet Switching Security**
   - Port Isolation
   - MAC Address Table Security
   - Port Security
   - MAC Address Flapping Prevention and Detection
   - MACsec
   - Traffic Control
   - DHCP Snooping
   - IP Source Guard
2. Advanced Firewall Features

HUAWEI

MAC Address Flapping Detection

- MAC address flapping occurs when a MAC address is learned by two interfaces in the same VLAN on a switch and the MAC address entry learned later overrides the earlier one.
- When a MAC address frequently switches between two interfaces, MAC address flapping occurs.
- MAC address flapping frequently occurs on networks where loops or network attacks occur.



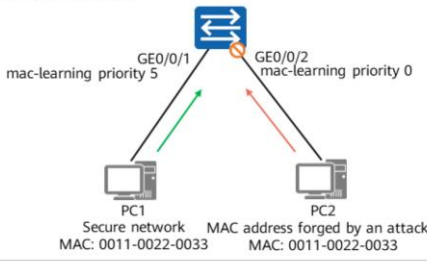


MAC Address Flapping Prevention

- If MAC address flapping is caused by loops, deploy loop prevention technologies, such as STP, to eliminate Layer 2 loops. If MAC address flapping is caused by network attacks or other reasons, you can use the following MAC address flapping prevention measures.

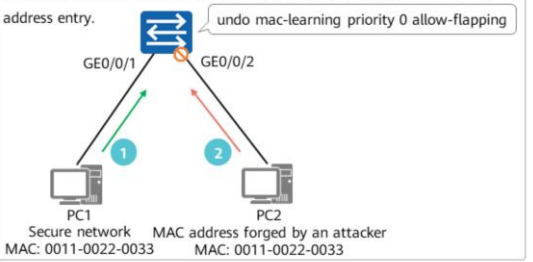
Configuring a MAC address learning priority for an interface

If the same MAC address is learned on interfaces that have different priorities, the MAC address entry on the interface with the highest priority overrides that on the other interfaces.



Preventing MAC address entries from being overridden on interfaces with the same priority

If the interface connected to a bogus network device has the same priority as the interface connected to an authorized device, the MAC address entry of the bogus device learned later does not override the original correct MAC address entry.



HUAWEI

- By default, the MAC address learning priority of an interface is 0. A larger priority value indicates a higher MAC address learning priority. If the same MAC address is learned on interfaces that have different priorities, the MAC address entry on the interface with the highest priority overrides that on the other interfaces.
- When the device is configured to prevent MAC address entries from being overridden on interfaces with the same priority, if the authorized device is powered off, the MAC address entry of the bogus device is learned. After the authorized device is powered on again, its MAC address cannot be learned. Exercise caution when using this feature. If a switch interface is connected to a server, when the server is powered off, other interfaces can learn the same MAC address as the server. When the server is powered on again, the switch cannot learn the correct MAC address.

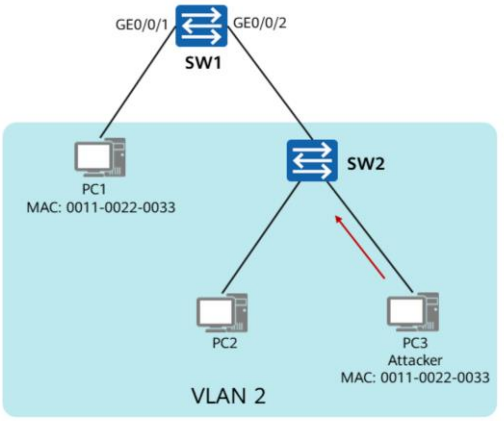
MAC Address Flapping Detection

- The switch supports the following MAC address flapping detection mechanisms.
  - VLAN-based MAC address flapping detection
    - After MAC address flapping detection is configured in a VLAN, the switch can detect MAC address flapping in a specified VLAN.
    - You can configure an action to take on an interface when MAC address flapping is detected on an interface, for example, sending a trap or blocking the interface or MAC address.
  - Global MAC address flapping detection
    - The global MAC address flapping detection function detects all MAC addresses on the device.
    - If MAC address flapping occurs, the device will send a trap to the NMS.
    - You can also specify an action to take when MAC address flapping is detected, for example, shutting down the interface or removing the interface from the VLAN.

HUAWEI

- Whether all Huawei switches support MAC address flapping detection depends on the switch model.

VLAN-based MAC Address Flapping Detection



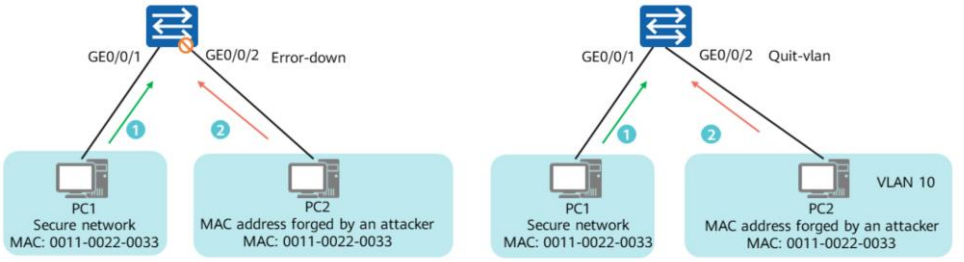
- When VLAN-based MAC address flapping detection is configured and detects MAC address flapping on an interface, you can configure one of the following actions:
  - **Trap sending:** The device only sends a trap to the NMS.
  - **Interface blocking:** The interface is blocked for a specified period of time and the interface is disabled from sending and receiving packets.
  - **MAC address blocking:** The device blocks only the current MAC address but not the physical interface. Communication of other MAC addresses on the current interface is not affected.

HUAWEI

- After MAC address flapping occurs, the following actions are performed: 1. A trap is generated and reported. 2. GE0/0/2 on SW1 is disabled from sending and receiving packets. 3. GE0/0/2 on SW1 is disabled from sending and receiving packets with a specified MAC address.
- When an interface is blocked:
  - When detecting MAC address flapping in VLAN 2, the device blocks the interface where MAC address flapping occurs.
  - The interface will be blocked for 10s (specified by the block-time keyword). The blocked interface cannot receive or send data.
  - After 10 seconds, the interface is unblocked and starts to send and receive data. If MAC address flapping is not detected within 20 seconds, the interface is unblocked. If MAC address flapping is detected again on the interface within 20 seconds, the switch blocks the interface again. If the switch still detects MAC address flapping on the interface, the switch permanently blocks the interface. The retry-times parameter specifies the number of times that MAC address flapping is detected.

Global MAC Address Flapping Detection

- When a switch detects MAC address flapping, it only reports a trap by default and does not take other actions. In practice, you can define the following actions after MAC address flapping is detected:
  - error-down
    - When an interface configured with MAC address flapping detection detects MAC address flapping, the interface is set to enter the Error-Down state and does not forward data.
  - quit-vlan
    - When an interface configured with MAC address flapping detection detects MAC address flapping, the interface is removed from the VLAN to which the interface belongs.



HUAWEI

- By default, global MAC address flapping detection is enabled on a Huawei switch. Therefore, the switch performs MAC address flapping detection in all VLANs by default.
- In some scenarios, MAC address flapping detection needs to be disabled in some VLANs. You can configure a VLAN whitelist for MAC address flapping detection.
- If an interface is set to enter the Error-Down state due to MAC address flapping, the interface does not automatically restore to the Up state by default.
- To enable an interface in Error-Down state to automatically go Up, run the **error-down auto-recovery cause mac-address-flapping interval** *time-value* command in the system view.
- If MAC address flapping occurs on an interface and the interface is removed from the VLAN, you can run the following command in the system view to implement automatic recovery of the interface:
  - **mac-address flapping quit-vlan recover-time** *time-value*

Configuration Commands of MAC Address Flapping Prevention and Detection (1)

1. Configure a MAC address learning priority for an interface.

   [Huawei-GigabitEthernet0/0/1] **mac-learning priority** *priority-id*

   By default, the MAC address learning priority of an interface is 0. A larger priority value indicates a higher MAC address learning priority.

2. Configure the device to discard packets when the device is configured to prohibit MAC address flapping.

   [Huawei-GigabitEthernet0/0/1] **mac-learning priority flapping-defend action discard**

   By default, the action is **forward** when the device is configured to prohibit MAC address flapping.

3. Configure the device to prevent MAC address entries from being overridden on interfaces with the same priority.

   [Huawei] **undo mac-learning priority** *priority-id* **allow-flapping**

   By default, MAC address flapping between interfaces with the same priority is allowed.

4. Configuring MAC Address Flapping Detection.

   [Huawei-vlan2] **mac-address flapping detection**

HUAWEI

Configuration Commands of MAC Address Flapping Prevention and Detection (2)

5. (Optional) Configure a VLAN whitelist for MAC address flapping detection.

   [Huawei] **mac-address flapping detection exclude vlan** { *vlan-id1* [*to vlan-id2*] } &<1-10>

   By default, the device performs MAC address flapping detection in all VLANs.

6. (Optional) Configure the action for the device to take after MAC address flapping is detected on an interface.

   [Huawei-GigabitEthernet0/0/1] **mac-address flapping action** { **quit-vlan** | **error-down** }

   By default, the device discards packets with new MAC addresses when the number of learned MAC address entries reaches the limit on the interface.

7. (Optional) Set an aging time for flapping MAC addresses.

   [Huawei] **mac-address flapping aging-time** *aging-time*

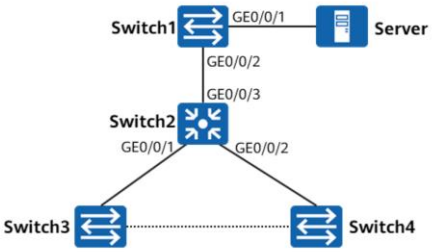
8. Configure MAC address flapping detection.

   [Huawei-vlan2] **loop-detect eth-loop** { [**block-mac**] **block-time** *block-time* **retry-times** *retry-times* | **alarm-only** }

HUAWEI

- Insecure networks are vulnerable to MAC address attacks. If attackers send large numbers of forged packets with different source MAC addresses to the switch, its MAC address table will be filled with unwanted address entries. As a result, the device is unable to learn the source MAC addresses of valid packets.
- You can limit the number of MAC address entries that can be learned on the device. When the number of learned MAC address entries reaches the limit, the device does not learn new MAC address entries. You can also configure an action to take when the number of learned MAC address entries reaches the limit. This prevents MAC address entries from being exhausted and improves network security.

Example for Configuring MAC Address Flapping Prevention and Detection



- Requirements:
  - Basic network configurations are complete, but an incorrect connection of a network cable between Switch3 and Switch4 causes a loop on the network.
  - Configure MAC address flapping prevention on GE0/0/1 of Switch1 to prevent attacks from unauthorized users.
  - Configure MAC address flapping detection on Switch2 to detect loops on the network and rectify the faults.

1. Set the MAC address learning priority of GE0/0/1 that connects Switch1 to the server to be higher than that of other interfaces. The default MAC address learning priority is 0.

```
[Switch1] interface GigabitEthernet 0/0/1
[Switch1-GigabitEthernet 0/0/1] mac-learning priority 3
```

2. Configure MAC address flapping detection on Switch2 and configure an action to taken when MAC address flapping is detected on an interface.

```
[Switch2] mac-address flapping detection
[Switch2] mac-address flapping aging-time 500
[Switch2-GigabitEthernet0/0/1] mac-address flapping action error-down
[Switch2-GigabitEthernet0/0/2] mac-address flapping action error-down
[Switch2] error-down auto-recovery cause mac-address-flapping interval 500
```


HUAWEI

- When Switch3 and Switch4 are incorrectly connected, the MAC address of GE0/0/1 on Switch2 is learned by GE0/0/2, causing GE0/0/2 to enter the Error-Down state.
- You can run the **display mac-address flapping record** command to check MAC address flapping records.

Verifying the Configuration

- When the MAC address of GE0/0/1 on Switch2 is learned by GE0/0/2, GE0/0/2 is shut down. You can run the **display mac-address flapping record** command to check MAC address flapping records.

```
[Switch2] display mac-address flapping record
 S  : start time
 E  : end time
(Q) : quit vlan
(D) : error down
--------------------------------------------------------------------------------
Move-Time               VLAN MAC-Address     Original-Port   Move-Ports   MoveNum
--------------------------------------------------------------------------------
S:2020-06-22 17:22:36    1   5489-9815-662b  GE0/0/1         GE0/0/2(D)   83
E:2020-06-22 17:22:44
--------------------------------------------------------------------------------
Total items on slot 0: 1
```

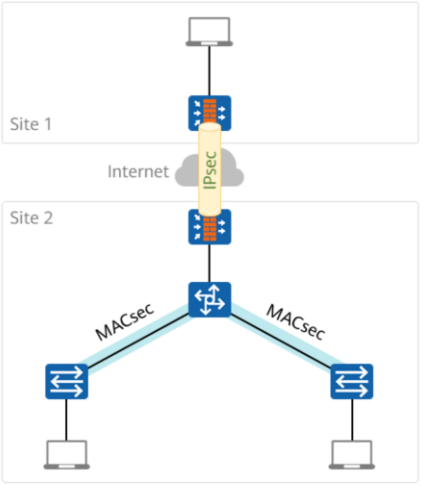

HUAWEI

Contents

1. **Ethernet Switching Security**
   - Port Isolation
   - MAC Address Table Security
   - Port Security
   - MAC Address Flapping Prevention and Detection
   - MACsec
   - Traffic Control
   - DHCP Snooping
   - IP Source Guard
2. Advanced Firewall Features

HUAWEI

MACsec Provides Secure Layer 2 Data Transmission



Background

Most data is transmitted in plain text on LAN links, posing security risks in scenarios demanding high security.

Overview of MACsec

Media Access Control Security (MACsec), in compliance with 802.1AE, defines an Ethernet-based data security communication method. It encrypts data hop by hop to ensure data transmission security.

- Data integrity check
- Data encryption
- Identity authentication
- Replay protection

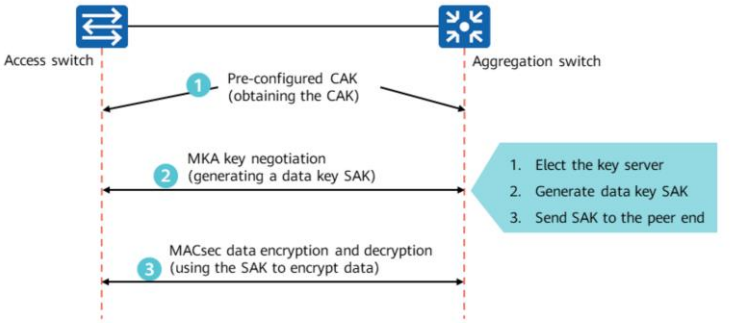
Typical Application Scenarios

- MACsec is deployed between switches to ensure data security. For example, MACsec is deployed between access switches and uplink aggregation or core switches.
- When transmission devices exist between switches, MACsec can be deployed to ensure data security.

HUAWEI

Working Mechanism of MACsec

- When the device runs point-to-point MACsec, a network administrator pre-configures the same Secure Connectivity Association Key (CAK) on the two devices using commands. The two devices use the MACsec Key Agreement (MKA) to elect a key server. The key server determines the encryption scheme, uses an encryption algorithm to generate a Secure Association Key (SAK) based on parameters such as the CAK, and distributes the SAK to the peer device. In this way, the two devices have the same SAK, which can be used to encrypt and decrypt MACsec data packets.



HUAWEI

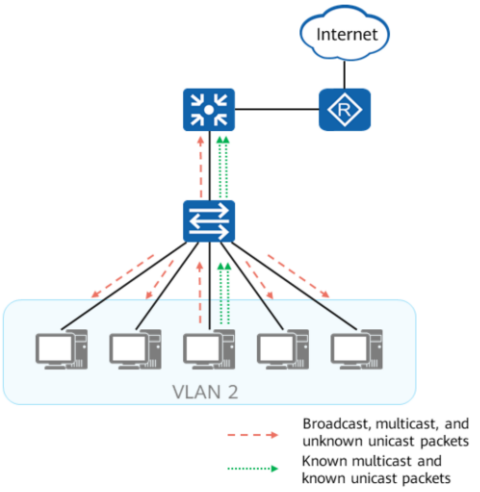
- A CAK is not directly used to encrypt data packets. Instead, the CAK and other parameters derive the encryption key of data packets. The CAK can be delivered during 802.1X authentication or statically configured.
- MKA is used for negotiation of MACsec data encryption keys.
- The SAK is derived based on the CAK using an algorithm and is used to encrypt data transmitted over secure channels. The MKA limits the number of packets that can be encrypted by each SAK. When the PNs encrypted by a SAK are exhausted, the SAK is updated. For example, on a 10 Gbit/s link, the SAK can be updated every 4.8 minutes.
- The key server determines the encryption scheme and the MKA entity that distributes the key.

Contents

1. **Ethernet Switching Security**
   - Port Isolation
   - MAC Address Table Security
   - Port Security
   - MAC Address Flapping Prevention and Detection
   - MACsec
   - Traffic Control
   - DHCP Snooping
   - IP Source Guard
2. Advanced Firewall Features

HUAWEI

Overview of Traffic Suppression

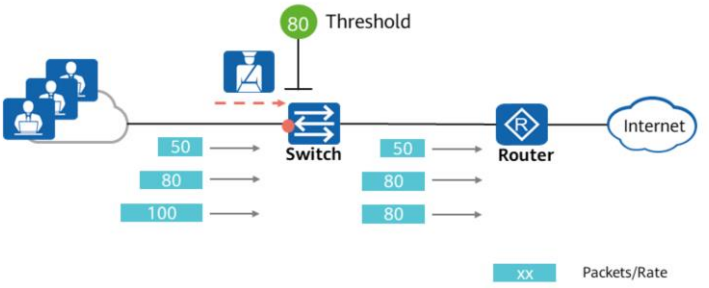


- **Issues on networks:**
  - In normal situations, when a Layer 2 Ethernet interface receives broadcast, unknown multicast, or unknown unicast packets, it forwards the packets to other Layer 2 Ethernet interfaces in the same VLAN. As a result, traffic flooding occurs and the forwarding performance of the device deteriorates.
  - When an Ethernet interface on the device receives known multicast or unicast packets, heavy traffic of a certain type of packets may affect the processing of other services on the switch.
- **Solution:**
  - Traffic suppression can rate-limit the broadcast, unknown multicast, unknown unicast, known multicast, and known unicast packets by setting thresholds. This prevents traffic flooding caused by broadcast, unknown multicast, and unknown unicast packets and the impact incurred by a large number of known multicast and known unicast packets.

HUAWEI

Working Mechanism of Traffic Suppression (1)

- In the inbound direction of an interface, the switch can suppress broadcast, unknown multicast, unknown unicast, known multicast, and known unicast packets based on the percentage, packet rate, and bit rate. The device monitors the rate of various types of packets on an interface and compares the rate with the threshold. When the traffic rate on the interface in the inbound direction reaches the threshold, the device discards excess traffic.

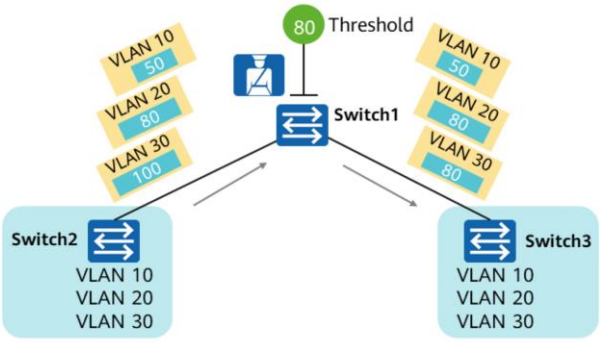


HUAWEI

- In the outbound direction of an interface, the device can block broadcast packets, unknown multicast packets, and unknown unicast packets.

Working Mechanism of Traffic Suppression (2)

- In the VLAN view, the device can rate-limit broadcast packets by the bit rate. The device monitors the rate of broadcast packets in the same VLAN and compares the rate with the threshold. When the traffic rate in the VLAN reaches the threshold, the device discards excess traffic.



HUAWEI

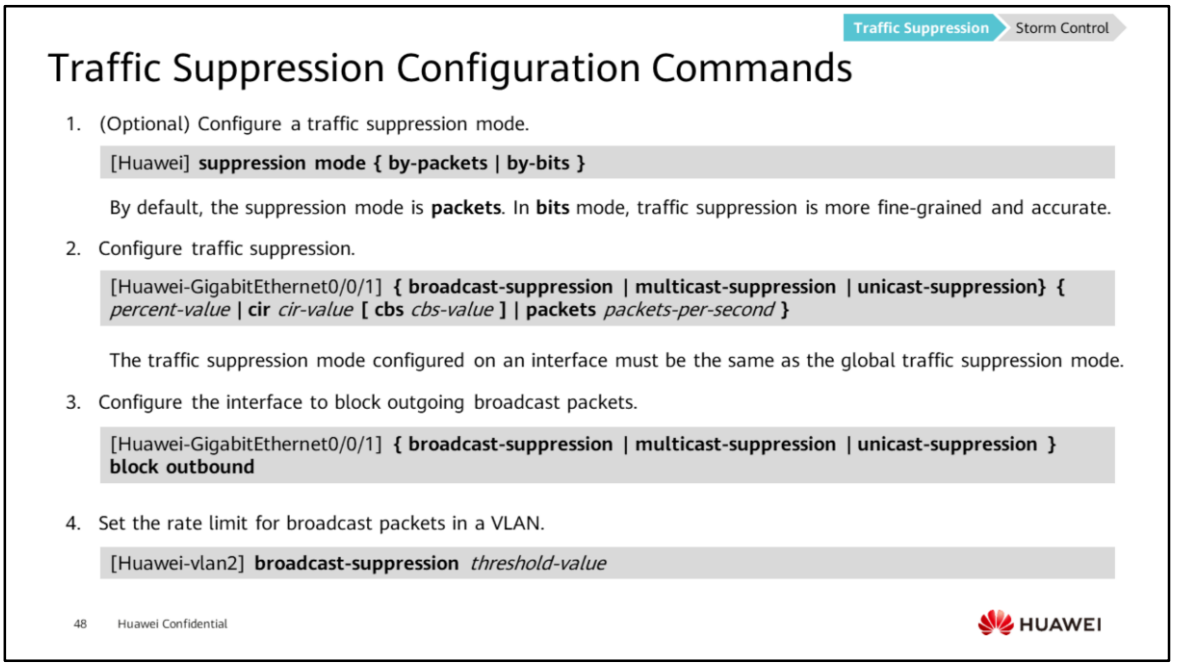
- Traffic suppression can also rate-limit ICMP packets by setting a threshold. A large number of ICMP packets may be sent to the CPU without traffic suppression. When this happens, other service functions may become abnormal.

Application of Traffic Suppression

- Traffic suppression limits the rate at which packets are sent by taking different measures for different types of packets. It involves the following situations:
  - In the inbound direction of a switch interface, for example, in the inbound direction of GE0/0/1 on SW1, traffic suppression can be used to limit the rate at which any packet is sent.
  - In the outbound direction of a switch interface, for example, in the outbound direction of GE0/0/1 on SW1, traffic suppression can be used to block broadcast, unknown multicast, and unknown unicast packets.
  - In the VLAN view of the switch, configure traffic suppression in a VLAN to limit the number of broadcast packets in the VLAN.

HUAWEI

- The threshold can be configured for incoming packets on interfaces. The system discards the traffic exceeding the threshold and forwards the traffic within the threshold. In this way, the system limits the traffic rate in an acceptable range.
- Note that traffic suppression can also block outgoing packets on interfaces.
- In storm control, rate thresholds are configured for incoming packets only on interfaces. When the traffic exceeds the threshold, the system rejects the packets of this particular type on the interface or shuts down the interface.



- Run the **display flow-suppression interface** *interface-type interface-number* command to check the traffic suppression configuration.
- When traffic suppression is configured in both the interface view and VLAN view, the configuration in the interface view takes precedence over the configuration in the VLAN view.

Example for Configuring Traffic Suppression

- Requirements:
  - Configure traffic suppression in the view of GE0/0/1 to limit the capability of forwarding broadcast, unknown multicast, and unknown unicast packets on the Layer 2 network.
  - Set the bandwidth percentage for broadcast packets to 60%.
  - Set the bandwidth percentage for unknown multicast packets to 70%.
  - Set the bandwidth percentage for unknown unicast packets to 80%.

Switch configuration:

```
[Switch]suppression mode by-packets
[Switch-GigabitEthernet0/0/1] unicast-suppression 80
[Switch-GigabitEthernet0/0/1] multicast-suppression 70
[Switch-GigabitEthernet0/0/1] broadcast-suppression 60
```


HUAWEI

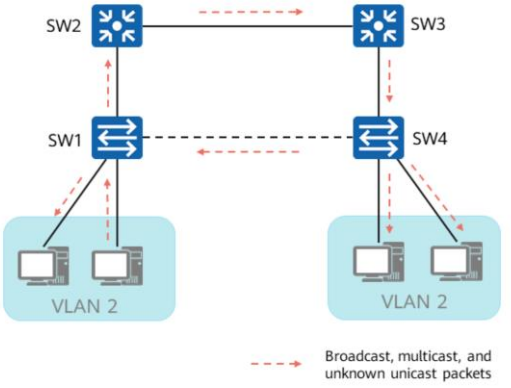
Verifying the Configuration

- Run the **display flow-suppression interface** command to check the traffic suppression configuration.

```
[Switch]dis flow-suppression interface GigabitEthernet 0/0/1
 storm type          rate mode           set rate value
-------------------------------------------------------------------------------
 unknown-unicast     percent             percent: 80%
 multicast           percent             percent: 70%
 broadcast           percent             percent: 60%
-------------------------------------------------------------------------------
```

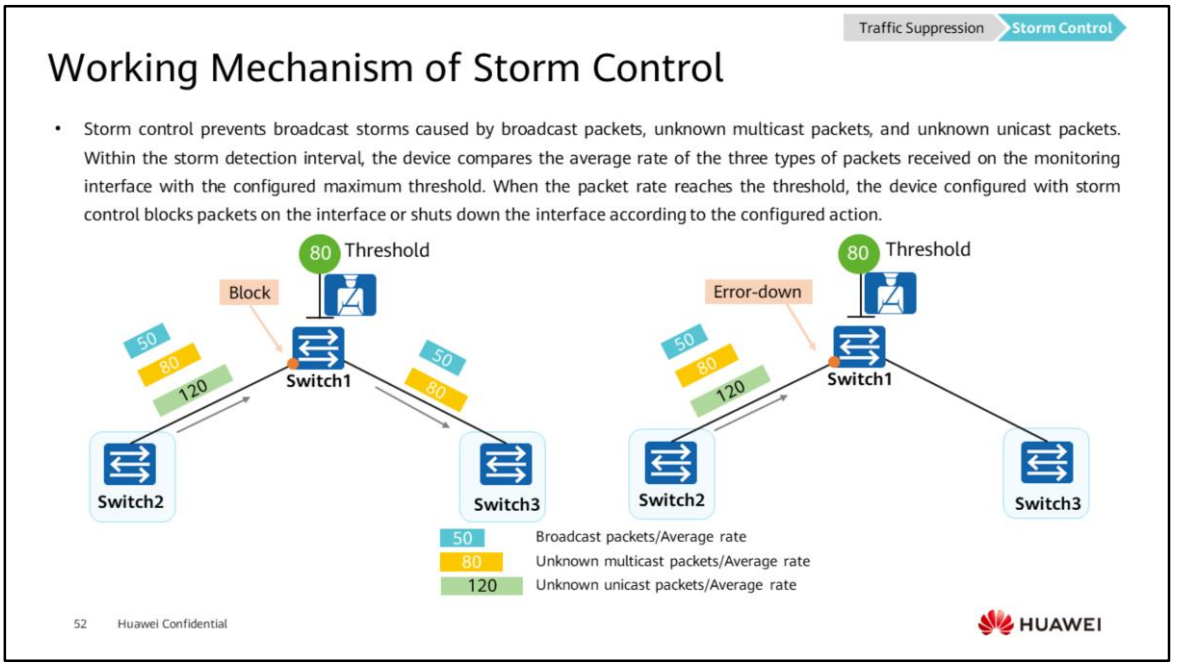

HUAWEI

Overview of Storm Control



- **Issues on networks:**
  - In normal situations, when a Layer 2 Ethernet interface receives broadcast, unknown multicast, or unknown unicast packets, it forwards the packets to other Layer 2 Ethernet interfaces in the same VLAN. As a result, traffic flooding occurs and the forwarding performance of the device deteriorates.
- **Solution:**
  - Storm control blocks broadcast, unknown multicast, and unknown unicast packets by disabling related interfaces.

HUAWEI



- The difference between traffic suppression and storm control is as follows: The storm control function can take the punishment action (block or shutdown) for an interface, whereas the traffic suppression function only limits the traffic on an interface.

Application of Storm Control

- Compared with traffic suppression, storm control monitors the average rates of broadcast packets, unknown multicast packets, and unknown unicast packets on an interface, and then blocks packets on the interface or shuts down the physical interface according to the threshold.
- As shown in the figure, the switch is connected to a Layer 2 network and a router. To limit the rates of broadcast packets, unknown multicast packets, and unknown unicast packets forwarded by the Layer 2 network, configure storm control on the Layer 2 Ethernet interface GE0/0/1 of the switch.

HUAWEI

- In traffic suppression, rate thresholds are configured for incoming packets on interfaces. When the traffic exceeds the threshold, the system discards excess traffic and allows the packets within the threshold to pass through. In this way, the traffic is limited within a proper range. Note that traffic suppression can also block outgoing packets on interfaces.
- In storm control, rate thresholds are configured for incoming packets only on interfaces. When the traffic exceeds the threshold, the system rejects the packets of this particular type on the interface or shuts down the interface.

Storm Control Configuration Commands

1. Configure storm control on an interface.

   [Huawei-GigabitEthernet0/0/1] **storm-control { broadcast | multicast | unicast } min-rate** *min-rate-value* **max-rate** *max-rate-value*

   Storm control is performed on broadcast packets, multicast packets, or unknown unicast packets on the interface.

2. Configure a storm control action.

   [Huawei-GigabitEthernet0/0/1] **storm-control action { block | error-down }**

3. Set the storm detection interval.

   [Huawei-GigabitEthernet0/0/1] **storm-control interval** *interval-value*

4. Enable automatic recovery of the interface status.

   [Huawei-GigabitEthernet0/0/1] **error-down auto-recovery cause storm-control interval** *interval-value*

   Enable the interface in Error-Down state to go Up and set the auto recovery delay.

5. (Optional) Add specified protocol packets to the traffic suppression and storm control whitelist.

   [Huawei] **storm-control whitelist protocol { arp-request | bpdu | dhcp | igmp | ospf }***

HUAWEI

Example for Configuring Storm Control

- Requirements
  - The switch is required to prevent broadcast storms caused by broadcast packets, unknown multicast packets, and unknown unicast packets forwarded on the Layer 2 network.
- Configuration roadmap:
  - Configure storm control on GE0/0/1 to prevent broadcast storms on the Layer 2 network.

Switch configuration:

```
[Switch] storm-control whitelist protocol arp-request
[Switch] interface gigabitethernet0/0/1
[Switch-GigabitEthernet0/0/1] storm-control broadcast min-rate 1000 max-rate 2000
[Switch-GigabitEthernet0/0/1] storm-control multicast min-rate 1000 max-rate 2000
[Switch-GigabitEthernet0/0/1] storm-control unicast min-rate 1000 max-rate 2000
[Switch-GigabitEthernet0/0/1] storm-control interval 90
[Switch-GigabitEthernet0/0/1] storm-control action block
[Switch-GigabitEthernet0/0/1] storm-control enable trap
# Enable the trap function for storm control.
```


HUAWEI

Verifying the Configuration

- Run the **display storm-control interface** command to check the storm control configuration on GE0/0/1.

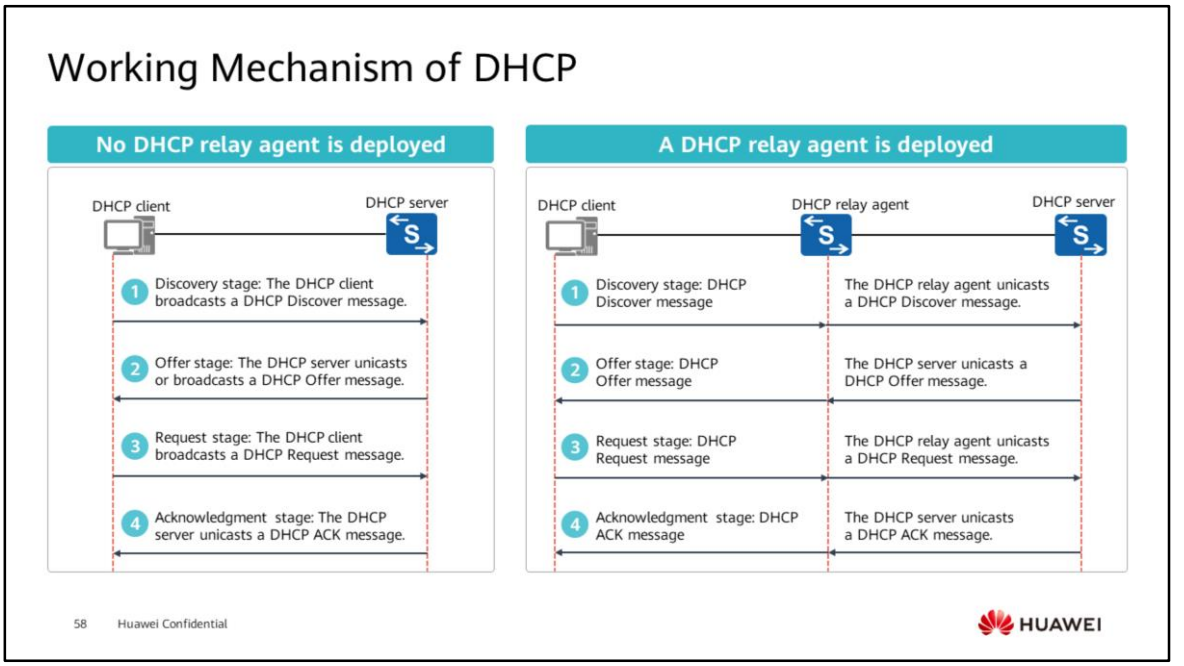
```
[Switch]display storm-control interface GigabitEthernet 0/0/1
PortName    Type       Rate (Min/Max)   Mode  Action  Punish-Status  Trap  Log  Int  Last-Punish-Time
------------------------------------------------------------------------------------------------------
GE0/0/1     Multicast  1000 /2000       Pps   Block   Normal         On    Off       90
GE0/0/1     Broadcast  1000 /2000       Pps   Block   Normal         On    Off       90
GE0/0/1     Unicast    1000 /2000       Pps   Block   Normal         On    Off       90
```


HUAWEI

Contents

1. **Ethernet Switching Security**
   - Port Isolation
   - MAC Address Table Security
   - Port Security
   - MAC Address Flapping Prevention and Detection
   - MACsec
   - Traffic Control
   - DHCP Snooping
   - IP Source Guard
2. Advanced Firewall Features

HUAWEI



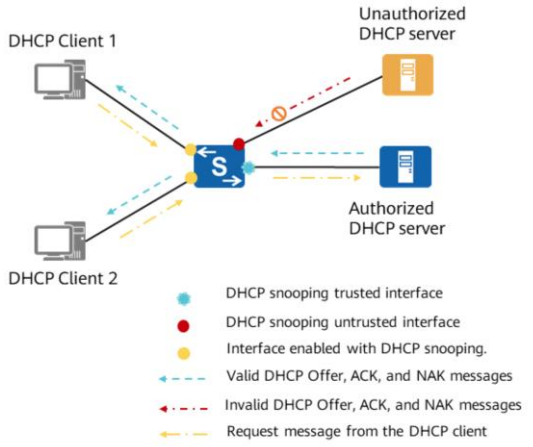
- No DHCP relay agent is deployed:
  - In the discovery stage, the DHCP client broadcasts a DHCP Discover message to discover DHCP servers. Information carried in a DHCP Discover message includes the client's MAC address (Chaddr field), parameter request list (Option 55), and broadcast flag (Flags field, determining whether the response should be sent in unicast or broadcast mode).
  - In the offer stage, 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.
  - In the request stage, 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 (Option 54) and IP address (Option 50, with the IP address specified in the Yiaddr field of the accepted DHCP Offer message). 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, after receiving the DHCP ACK message, the DHCP client broadcasts a gratuitous ARP packet to check whether any other terminal on the network segment uses the IP address allocated by the DHCP server.

Overview of DHCP Snooping

- DHCP snooping is equivalent to a firewall between DHCP clients and the DHCP server to defend against DHCP attacks on the network, ensuring security for communication services.
- DHCP snooping ensures that DHCP clients obtain IP addresses only from authorized DHCP servers and a DHCP snooping-enabled device records mappings between IP addresses and MAC addresses of DHCP clients, preventing DHCP attacks on the network.
- Some attacks are launched based on DHCP. These attacks include the bogus DHCP server attack, DHCP server DoS attack, and bogus DHCP message attack.
- DHCP snooping uses the DHCP snooping trusted interface and DHCP snooping binding table to ensure DHCP network security.

HUAWEI

DHCP Snooping Trust Function



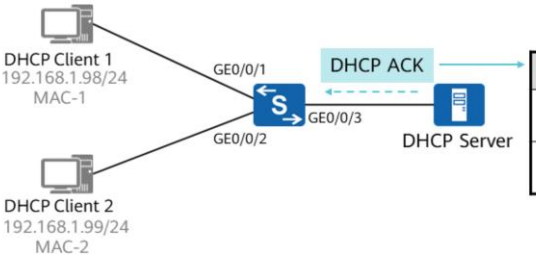
- The DHCP snooping trust function ensures that DHCP clients obtain IP addresses from authorized DHCP servers.
- DHCP snooping involves two interface roles: trusted interface and untrusted interface.
  - DHCP ACK messages, NAK messages, and Offer messages are received from the trusted interface.
  - In addition, the device only forwards DHCP Request messages from DHCP clients to the authorized DHCP server through the trusted interface.
  - DHCP ACK messages, NAK messages, and Offer messages are discarded on untrusted interfaces.

HUAWEI

- After the dhcp snooping enable command is run on an interface, the interface forwards received DHCP Request messages to all trusted interfaces and discards received DHCP Reply messages.
- After an interface on which the dhcp snooping trusted command is run receives a DHCP Request message, it forwards the message to all other trusted interfaces. If there are no other trusted interfaces, it discards the message. After receiving a DHCP Reply message, it forwards the message only to the interfaces that are connected to clients and have the dhcp snooping enable command configured. If such interfaces cannot be found, it discards the DHCP Reply message.

DHCP Snooping Binding Table

- The Layer 2 access device enabled with DHCP snooping obtains required information, such as the PC's MAC address, IP address, and address lease, from the DHCP ACK messages, learns information (interface number and VLAN ID) about the DHCP snooping-enabled interface connected to the PC, and generates a DHCP snooping binding entry for the PC.
- The DHCP snooping binding table records the mapping between IP addresses and MAC addresses of DHCP clients. The device can check DHCP messages against the DHCP snooping binding table to prevent attacks initiated by unauthorized users.



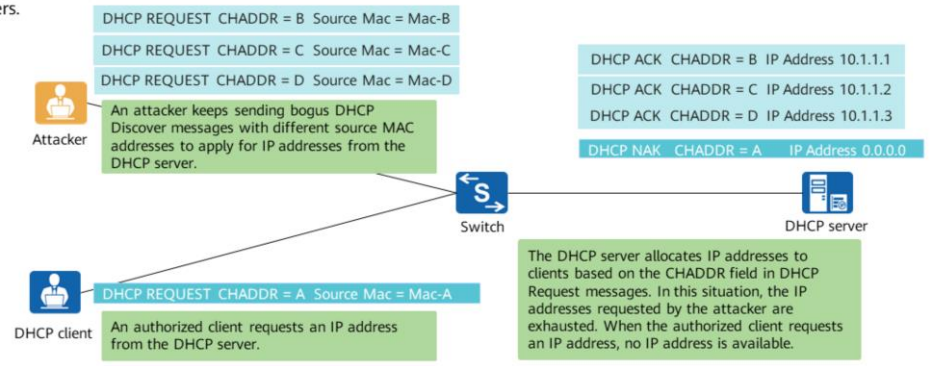
IP	MAC	VLAN	Interface	Lease
192.168.1.98	MAC-1	1	GE0/0/1	1
192.168.1.99	MAC-2	1	GE0/0/2	1

HUAWEI

- DHCP snooping binding entries are aged out when the DHCP release expires, or the entries are deleted when users send DHCP Release messages to release IP addresses.

DHCP Starvation Attacks

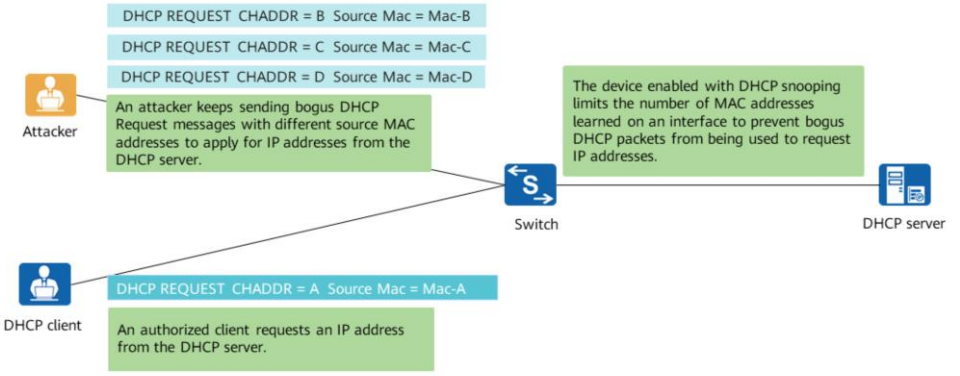
- An attacker continuously applies to the DHCP server for a large number of IP addresses until the IP addresses in the address pool of the DHCP server are exhausted. As a result, the DHCP server cannot allocate IP addresses to authorized users.
- Vulnerability analysis: When the DHCP server allocates IP addresses to clients, it cannot distinguish authorized and unauthorized users.



HUAWEI

Defense Against DHCP Starvation Attacks

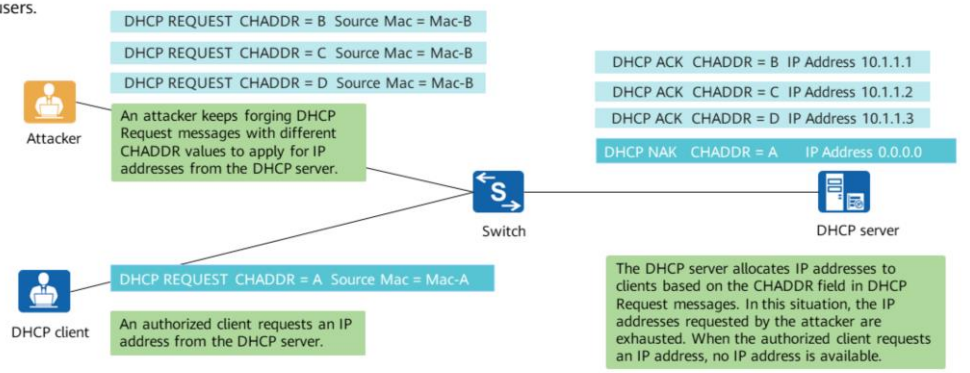
- Solution: Configure MAC address limiting of DHCP snooping to prevent starvation attacks. This function limits the maximum number of MAC addresses that can be learned on an interface of a switch to prevent a large number of DHCP Request messages with variable MAC addresses from being sent.



HUAWEI

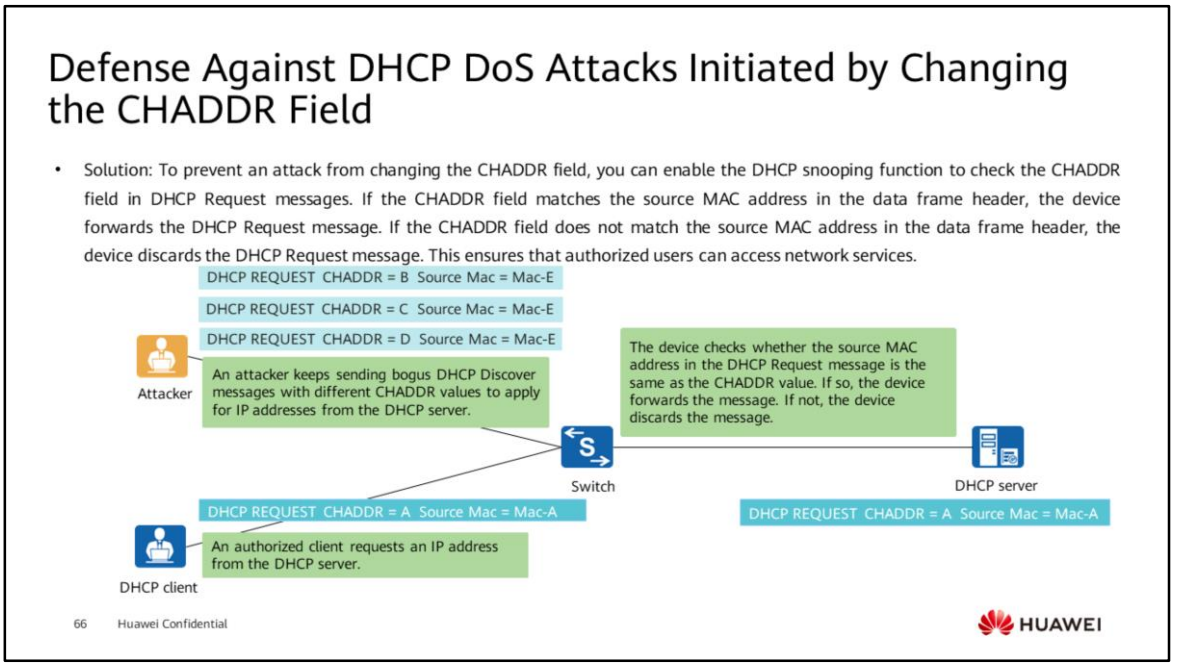
DoS Attacks by Changing the CHADDR Field

- An attacker continuously applies to the DHCP server for a large number of IP addresses until the IP addresses in the address pool of the DHCP server are exhausted. As a result, the DHCP server cannot allocate IP addresses to authorized users.
- Vulnerability analysis: When the DHCP server allocates IP addresses to clients, it cannot distinguish authorized and unauthorized users.



HUAWEI

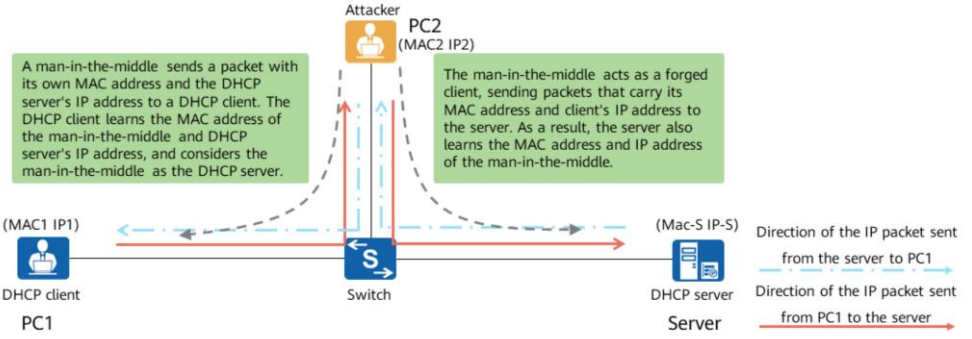
- In a DHCP starvation attack, an attacker continuously applies for a large number of IP addresses from the DHCP server to exhaust IP addresses in the address pool of the DHCP server. As a result, the DHCP server cannot allocate IP addresses to authorized users. The DHCP message contains the Client Hardware Address (CHADDR) field. This field is filled in by a DHCP client, indicating the hardware address of the client, that is, the MAC address of the client. The DHCP server assigns IP addresses based on the CHADDR field, and assigns different IP addresses if values of the CHADDR field are different. The DHCP server cannot distinguish valid CHADDR field. By exploiting this vulnerability, an attacker fills a different value in the CHADDR field of a DHCP message each time the attacker applies for an IP address. In this way, the attacker forges different users to request IP addresses.



- In a DHCP starvation attack, an attacker continuously applies for a large number of IP addresses from the DHCP server to exhaust IP addresses in the address pool of the DHCP server. As a result, the DHCP server cannot allocate IP addresses to authorized users. The DHCP message contains the Client Hardware Address (CHADDR) field. This field is filled in by a DHCP client, indicating the hardware address of the client, that is, the MAC address of the client. The DHCP server assigns IP addresses based on CHADDR values. The DHCP server cannot distinguish valid CHADDR values. By exploiting this vulnerability, an attacker fills a different value in the CHADDR field of a DHCP message each time the attacker applies for an IP address. In this way, the attacker forges different users to request IP addresses.
- To prevent starvation attacks, DHCP snooping checks whether the source MAC address of a DHCP Request message is the same as the CHADDR value on an interface. If they are the same, the interface forwards the DHCP Request message. If they are different, the interface discards the message. To check the consistency between the source MAC address and the CHADDR field on an interface, run the **dhcp snooping check dhcp-chaddr enable** command on the interface.
- An attacker may continuously change both the MAC address and CHADDR value simultaneously, and uses the same CHADDR value as the MAC address each time. In this way, the consistency check between the source MAC address and the CHADDR can be avoided.

Man-in-the-Middle Attacks

- An attacker uses the ARP mechanism to enable a client to learn the mapping between the DHCP server's IP address and attacker's MAC address, and enable the server to learn the mapping between the client's IP address and attacker's MAC address. In this way, all IP packets exchanged between the client and server traverse the attacker's device.
- Vulnerability analysis: The man-in-the-middle attack is a spoofing IP/MAC attack. This attack uses the mapping between the forged IP address and MAC address to deceive the DHCP client and server.

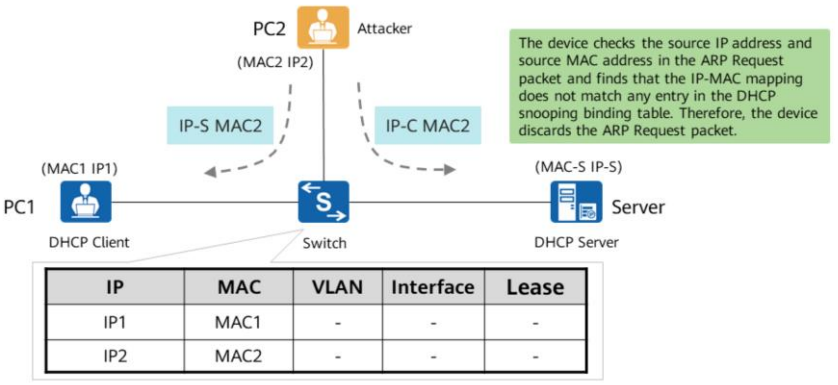


HUAWEI

- As shown in the figure, the attacker uses the ARP mechanism to enable PC1 to learn the mapping between IP-S and MAC2 and enable the server to learn the mapping between IP1 and MAC2. When PC1 sends an IP packet to the DHCP server, the destination IP address is IP-S and the source IP address is IP1. The destination MAC address of the frame in which the IP packet is encapsulated is MAC2 and the source MAC address is MAC1, so the frame reaches PC2 first. After receiving the frame, the attacker changes the destination MAC address to MAC-S and the source MAC address to MAC2, and then sends the frame to the server. When the DHCP server sends an IP packet to PC1, the destination IP address is IP1 and the source IP address is IP-S. The destination MAC address of the frame in which the IP packet is encapsulated is MAC2 and the source MAC address is MAC-S, so the frame reaches PC2 first. After receiving the frame, the attacker changes the destination MAC address to MAC1 and the source MAC address to MAC2, and then sends the frame to PC1.
- The IP packets transmitted between PC1 and the DHCP server traverse the attacker's device (man-in-the-middle). Therefore, the attacker can easily obtain some information in the IP packets and use the information to perform other damage operations. The attacker can easily tamper with the DHCP messages transmitted between PC1 and the DHCP server. These messages are encapsulated in UDP packets, and UDP packets are encapsulated in IP packets. In this way, the attacker can directly attack the DHCP server.

Defense Against DHCP Man-in-the-Middle Attacks

- Solution: To defend against man-in-the-middle attacks and IP/MAC spoofing attacks, configure the DHCP snooping binding table. When an interface receives an ARP or IP packet, the interface matches the source IP address and source MAC address in the ARP or IP packet against the DHCP snooping binding table. Packets that match entries are forwarded, whereas packets that do not match entries are discarded.



IP	MAC	VLAN	Interface	Lease
IP1	MAC1	-	-	-
IP2	MAC2	-	-	-

HUAWEI

- A DHCP man-in-the-middle attack is a spoofing IP/MAC attack. Preventing DHCP man-in-the-middle attacks is equivalent to preventing spoofing IP/MAC attacks.
- The switch running DHCP snooping listens to DHCP messages exchanged between the client and the DHCP server, and obtains the MAC address of the client from the DHCP messages. The MAC address (value of the CHADDR field in the DHCP messages, client's IP address (IP address allocated by the DHCP server to the corresponding CHADDR), and other information are stored in a database, which is also called the DHCP snooping binding table. The switch running DHCP snooping creates and dynamically maintains the DHCP snooping binding table. The binding table contains the MAC address, IP address, IP address lease, and VLAN ID of each client.
- As shown in the figure, if the DHCP server assigns IP address IP1 to PC1 and IP address IP2 to PC2, IP1 is bound to MAC1 and IP2 is bound to MAC2. These bindings are stored in the DHCP snooping binding table. To enable the server to learn the mapping between IP1 and MAC2, the attacker sends an ARP Request packet in which the source IP address is set to IP1 and the source MAC address is set to MAC2. After receiving the ARP Request packet, the switch checks the source IP address and source MAC address in the packet and finds that the IP-MAC (IP1-MAC2) mapping does not match any entry in the DHCP snooping binding table. Therefore, the switch discards the ARP Request packet, this effectively prevents spoofing IP/MAC attacks.
- To prevent IP/MAC spoofing attacks, run the **arp dhcp-snooping-detect enable** command in the system view of the switch.

DHCP Snooping Configuration Commands (1)

1. Enable DHCP snooping globally.

   [Huawei] **dhcp snooping enable** [**ipv4** | **ipv6**]

2. Enable DHCP snooping in the VLAN view.

   [Huawei-vlan2] **dhcp snooping enable**

   If you run this command in the VLAN view, the command takes effect for all DHCP messages in a specified VLAN received by all the interfaces on the device.

3. Configure an interface as the trusted interface in the VLAN view.

   [Huawei-vlan2] **dhcp snooping trusted interface** *interface-type interface-number*

   If you run this command in the VLAN view, the command takes effect only for the DHCP messages received by the interface in the VLAN that the interface belongs to.

HUAWEI

DHCP Snooping Configuration Commands (2)

4. Enable DHCP snooping in the interface view.

   [Huawei-GigabitEthernet0/0/1] **dhcp snooping enable**

5. Configure an interface as the trusted interface in the interface view.

   [Huawei-GigabitEthernet0/0/1] **dhcp snooping trusted**

   By default, all interfaces are untrusted interfaces.

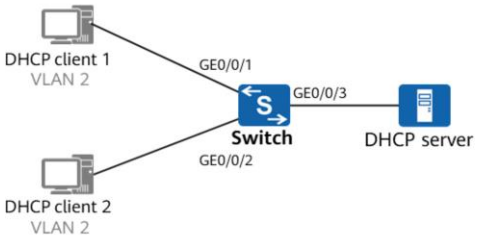
6. (Optional) Configure the device to discard DHCP Request messages with non-0 GIADDR field.

   [Huawei] **dhcp snooping check dhcp-giaddr enable vlan** { *vlan-id1* [**to** *vlan-id2*] }

   Enable the device to check whether the GIADDR field in a DHCP Request message is 0. This command can be run in both the VLAN view and interface view.
   If you run this command in the VLAN view, the command configuration takes effect for the DHCP messages received by all interfaces on the device from the specified VLAN. If you run this command in the interface view, the command configuration takes effect for all DHCP messages on the specified interface.

HUAWEI

Examples for Configuring DHCP Snooping



- As shown in the figure, basic DHCP and VLAN configurations are complete, and DHCP snooping is configured on the switch.

Method 1: Interface view

```
[Switch] dhcp snooping enable ipv4
[Switch] interface GigabitEthernet 0/0/1
[Switch-GigabitEthernet0/0/1] dhcp snooping enable
[Switch] interface GigabitEthernet 0/0/2
[Switch-GigabitEthernet0/0/2] dhcp snooping enable
[Switch] interface GigabitEthernet 0/0/3
[Switch-GigabitEthernet0/0/3] dhcp snooping enable
[Switch-GigabitEthernet0/0/3] dhcp snooping trusted
```

Method 2: VLAN view

```
[Switch] dhcp snooping enable ipv4
[Switch] vlan 2
[Switch-vlan2] dhcp snooping enable
[Switch] interface GigabitEthernet 0/0/3
[Switch-GigabitEthernet0/0/3] dhcp snooping trusted
```


HUAWEI

Verifying the Configuration

- Run the **display dhcp snooping interface** command to check DHCP snooping information on an interface.

```
[Switch]display dhcp snooping interface GigabitEthernet 0/0/3
 DHCP snooping running information for interface GigabitEthernet0/0/3 :
 DHCP snooping                    : Enable
 Trusted interface                : Yes
 Dhcp user max number             : 1024     (default)
 Current dhcp user number         : 0
 Check dhcp-giaddr                : Disable  (default)
 Check dhcp-chaddr                : Disable  (default)
 Alarm dhcp-chaddr                : Disable  (default)
 Check dhcp-request               : Disable  (default)
 Alarm dhcp-request               : Disable  (default)
----- more ------
```


HUAWEI

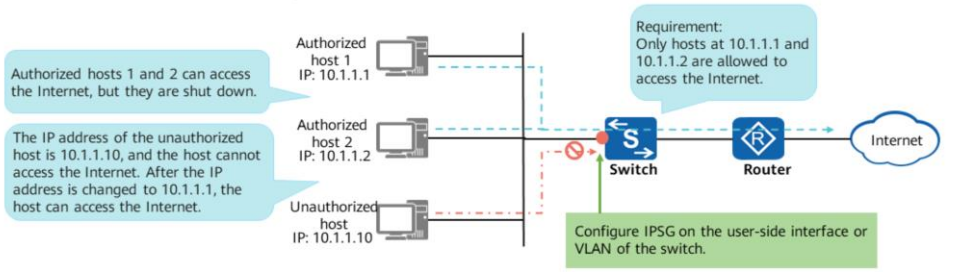
Contents

1. **Ethernet Switching Security**
   - Port Isolation
   - MAC Address Table Security
   - Port Security
   - MAC Address Flapping Prevention and Detection
   - MACsec
   - Traffic Control
   - DHCP Snooping
   - IP Source Guard
2. Advanced Firewall Features

HUAWEI

Overview of IPSG

- Some attackers forge IP addresses of authorized users to obtain network access rights and access networks. As a result, authorized users are unable to access networks or sensitive information may be intercepted. IP source guard (IPSG) provides a mechanism to effectively defend against IP address spoofing attacks.
- IPSG is a Layer 2 interface-based source IP address filtering technology. It prevents unauthorized hosts from using IP addresses of authorized hosts or specified IP addresses to access or attack a network.

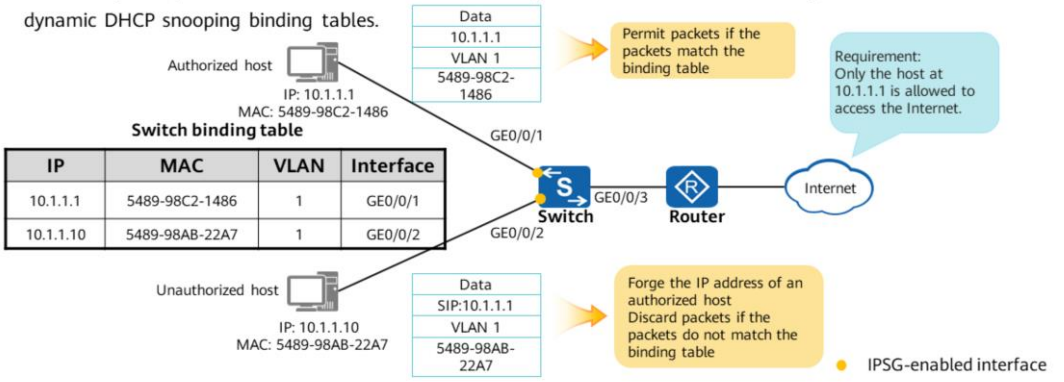


HUAWEI

- If the unauthorized host forges the IP address of an authorized host to obtain network access rights, configure IPSG on the switch's user-side interface or VLAN. The switch then checks the IP packets received by the interface and discards the packets from unauthorized hosts to prevent IP address spoofing attacks.
- Generally, IPSG is configured on the interfaces or VLANs of the access device connected to users.

Working Mechanism of IPSG

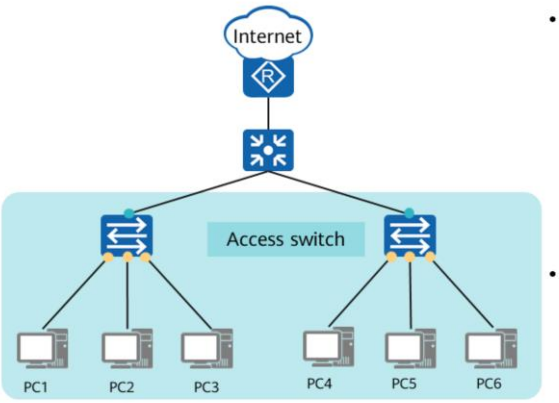
- IPSG checks IP packets on Layer 2 interfaces against a binding table that contains the bindings of source IP addresses, source MAC addresses, VLAN IDs, and inbound interfaces. Only packets that match the binding table are forwarded, and packets that do not match the binding table are discarded. Binding tables include static and dynamic DHCP snooping binding tables.



IP	MAC	VLAN	Interface
10.1.1.1	5489-98C2-1486	1	GE0/0/1
10.1.1.10	5489-98AB-22A7	1	GE0/0/2

- After the binding table is generated, the IPSG-enabled device delivers ACL rules to the specified interface or VLAN according to the binding table, and then checks all IP packets against the ACL rules. The switch forwards the packets from hosts only when the packets match binding entries, and discards the packets that do not match binding entries. When the binding table is modified, the IPSG-enabled device delivers the ACL rules again.
- By default, if IPSG is enabled in the scenario where no binding table is generated, the switch rejects all IP packets except DHCP Request messages.
- A static binding entry contains the MAC address, IP address, VLAN ID, and inbound interface. IPSG checks received packets against all options in a static binding entry.
- A dynamic binding entry contains the MAC address, IP address, VLAN ID, and inbound interface. You can specify the options to be checked, and IPSG filters the packets received by interfaces according to the specified options. By default, the IPSG-enabled device checks packets against all the four options.
    - Common check items:
        - Source IP address
        - Source MAC address
        - Source IP address + Source MAC address
        - Source IP address + Source MAC address + Interface
        - Source IP address + Source MAC address + Interface + VLAN

Application of IPSG



- IPSG prevents PCs from changing their own IP addresses.
  - PCs can only use the IP addresses allocated by the DHCP server or static IP addresses configured by an administrator to access the network. If a PC changes its IP address without permission, the PC cannot access the network. This prevents PCs from obtaining network rights without permission.
- If IP addresses on a small-scale network are statically allocated, IPSG can be used to prevent unauthorized PCs from accessing the network.
  - Users cannot access the intranet with their own computers. This prevents intranet resource leaks.

HUAWEI

IPSG Configuration Commands

1. Configure a static binding table.

   [Huawei] **user-bind static { { { ip-address | ipv6-address } { *start-ip* [to *end-ip*] } &<1-10> | ipv6-prefix** *prefix/prefix-length* **} | mac-address** *mac-address* **} * [interface** *interface-type interface-number* **] [vlan** *vlan-id* **[ce-vlan** *ce-vlan-id* **]]**

   The IPSG-enabled device matches packets against the static binding table.

2. Enable IPSG.

   [Huawei-GigabitEthernet0/0/1] **ip source check user-bind enable**

   The configuration of IP packet check in the VLAN view is the same as that in the interface view.

3. Enable the alarm function of IP packet check.

   [Huawei-GigabitEthernet0/0/1] **ip source check user-bind alarm enable**

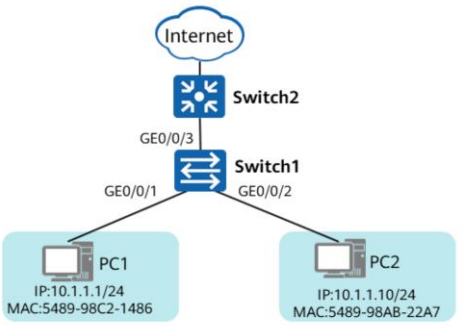
4. Set the alarm threshold for IP packet check.

   [Huawei-GigabitEthernet0/0/1] **ip source check user-bind alarm threshold** *threshold*

   After this alarm function is configured, the switch generates an alarm if the number of discarded IP packets exceeds the threshold.

HUAWEI

Example for Configuring IPSG



Switch1 configuration:

```
# Configure a static binding table on the access switch.
[Switch1] user-bind static ip-address 10.1.1.1 mac-address 5489-98C2-1486
[Switch1] user-bind static ip-address 10.1.1.10 mac-address 5489-98AB-22A7
# Enable IPSG and configure the alarm function of IP packet check on
GE0/0/1.
[Switch1] interface GigabitEthernet 0/0/1
[Switch1-GigabitEthernet0/0/1] ip source check user-bind enable
[Switch1-GigabitEthernet0/0/1] ip source check user-bind alarm enable
[Switch1-GigabitEthernet0/0/1] ip source check user-bind alarm threshold
100
# The configuration of GE0/0/2 is similar to that of GE0/0/1.
```

- As shown in the figure, PCs are configured with static IP addresses for unified management. IPSG is configured on the access switch to prevent hosts from changing their own IP addresses to access the network.
  - Configure a static binding table.
  - Enable IPSG and configure the alarm function.

HUAWEI

Verifying the Configuration

- Run the **display dhcp static user-bind all** command on the switch to check the static binding table.
- PC1 and PC2 can access the Internet using statically configured IP addresses, and cannot access the Internet after changing their IP addresses.

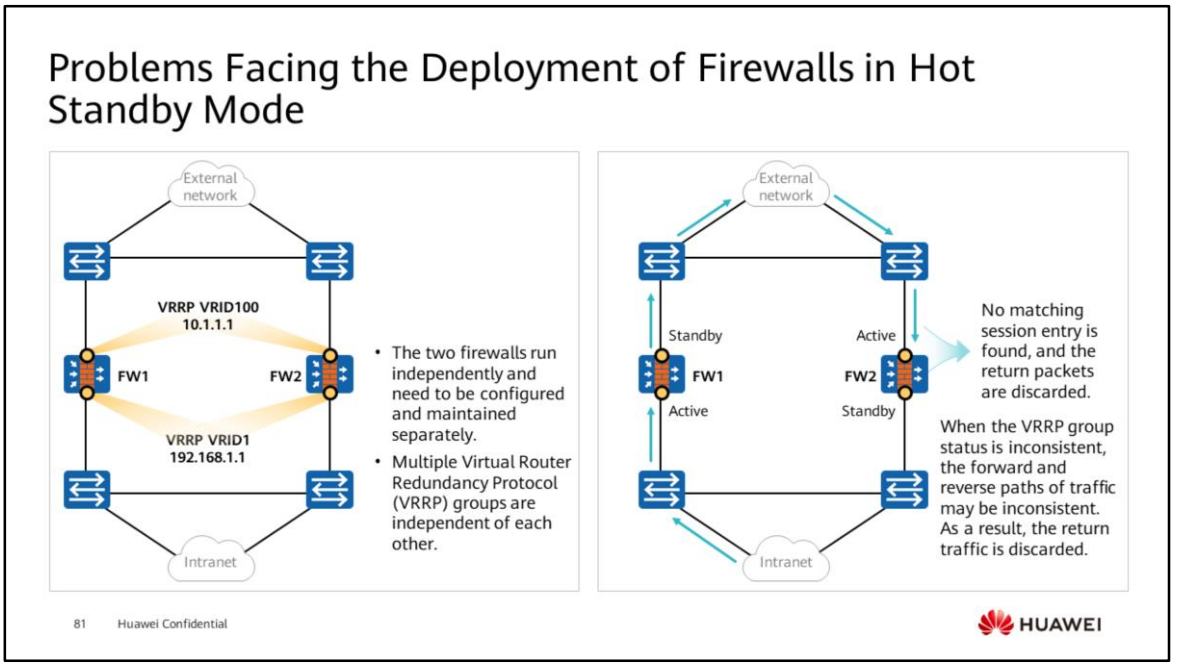
```
[Switch1] display dhcp static user-bind all
DHCP static Bind-table:
Flags:O - outer vlan ,I - inner vlan ,P - Vlan-mapping
IP Address                MAC Address                VSI/VLAN(O/I/P) Interface
--------------------------------------------------------------------------------
10.1.1.1                  5489-98C2-1486             --  /--  /--    --
10.1.1.10                 5489-98AB-22A7             --  /--  /--    --
--------------------------------------------------------------------------------
Print count:        2     Total count:        2
```


HUAWEI

Contents

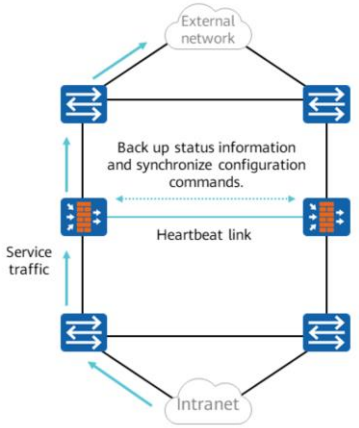
1. Ethernet Switching Security
2. **Advanced Firewall Features**
   - Hot Standby
   - Virtual System

HUAWEI



- With the rapid development of services such as mobile office, online shopping, instant messaging, Internet finance, and Internet education, the network needs to carry a growing number of services and the services carried are more and more important. As such, uninterrupted transmission of service data urgently needs to be fulfilled in the network development process.
- In this example, the firewall is deployed at the egress of the enterprise network, and services between the intranet and external network are forwarded by the firewall. Such services will be all interrupted if the firewall is faulty. Therefore, if only one device is deployed in a key position of the network, the network may be interrupted due to a single point of failure, regardless of the reliability of this single device. Therefore, in network architecture design, a key position on the network usually has two network devices planned for high availability.
- A firewall is a stateful inspection device. It inspects the first packet of a flow and establishes a session to record packet status information (including the source IP address, source port number, destination IP address, destination port number, and protocol). Subsequent packets of the flow are then forwarded according to the session entry. Only those matching this entry will be forwarded. Packets that do not match this entry will be discarded by the firewall.

Hot Standby Overview



Firewalls working in hot standby mode

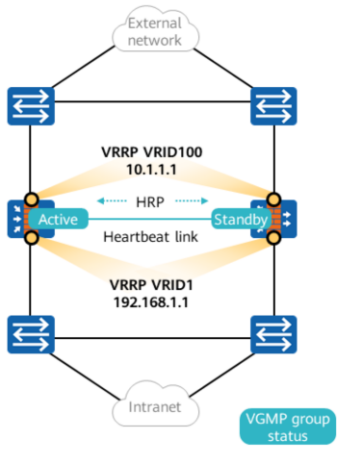
- Hot standby requires two firewalls with the same hardware and software configurations to form a hot standby system. The firewalls are connected through an independent link (heartbeat link) to learn about the health status of each other and back up configurations and entries (such as session entries and IPsec SA information) to each other.
- If a firewall fails, service traffic can be smoothly switched to the other firewall, preventing service interruption.

Deployment requirements

- Currently, hot standby can be implemented only between two devices.
- The active and standby devices must have the same product model and version.
- The active and standby devices must have the same numbers and types of cards installed in the same arrangement. Otherwise, the information synchronized from the active device does not match the physical configuration of the standby device. As a result, faults occur after an active/standby switchover.

HUAWEI

Key Components of Firewall Hot Standby



Virtual Router Redundancy Protocol (VRRP)

VRRP is a fault-tolerant protocol. It ensures that when the next-hop router (default gateway) of a host is faulty, the backup router can automatically replace the faulty router to forward packets, thus ensuring the continuity and reliability of network communication.

VRRP Group Management Protocol (VGMP)

All VRRP groups on a firewall are added to a VGMP group for centralized status monitoring and management. When detecting a status change in one VRRP group, the VGMP group forces all VRRP groups to perform status switchover, ensuring the status consistency among them.

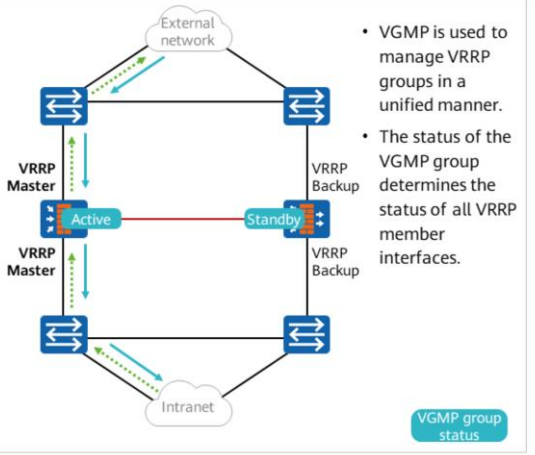
Huawei Redundancy Protocol (HRP)

HRP implements backup of dynamic status data and key configuration commands between the active and standby firewalls.

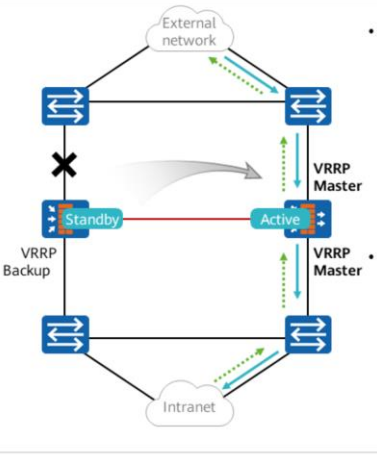
HUAWEI

- Each firewall has a VGMP group. A VGMP group can be in any of the following states:
  - Initialize: indicates the temporary initial status of a VGMP group after hot standby is enabled.
  - Load Balance: When the priority of the local VGMP group is the same as that of the peer VGMP group, the VGMP groups at both ends are in the Load Balance state.
  - Active: When the priority of the local VGMP group is higher than that of the peer VGMP group, the local VGMP group is in Active state.
  - Standby: When the priority of the local VGMP group is lower than that of the peer VGMP group, the local VGMP group is in Standby state.
- After two firewalls are deployed in hot standby mode, the VGMP groups on them have the same priority, and both are in Load Balance state. In this case, the two firewalls are in load balancing state.
- You can configure VRRP or manually specify a standby device to enable the two firewalls to work in active/standby mode. The VRRP configuration method applies to networks where the firewalls connect to Layer 2 switches, and the method of manually specifying a standby device applies to other hot standby networks.
- A firewall has an initial VGMP group priority. When an interface or a card on the firewall becomes faulty, the initial VGMP group priority is decreased by a specific value.
- This course uses USG6000 V500R001 as an example.

Key Components of Firewall Hot Standby: VRRP and VGMP



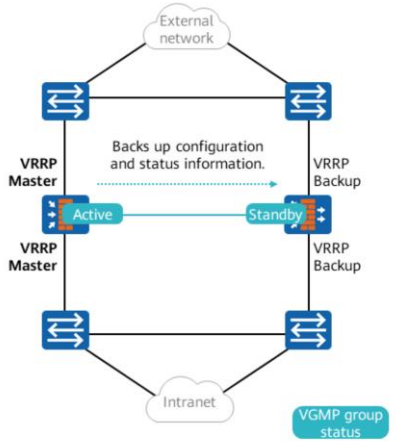
- VGMP is used to manage VRRP groups in a unified manner.
- The status of the VGMP group determines the status of all VRRP member interfaces.



- When the status of a VGMP group changes, the status of all VRRP member interfaces in the VGMP group is forcibly changed.
- Consistent status ensures that the incoming and outgoing service traffic is forwarded along the same path.

HUAWEI

Key Components of Firewall Hot Standby: HRP and Heartbeat Link



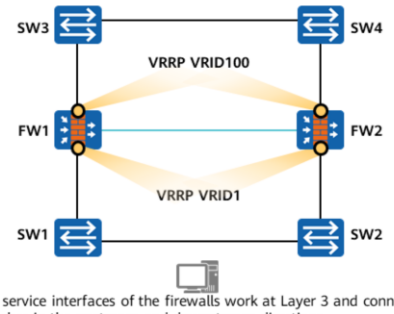
- To ensure successful switchover, key configuration commands and status information must be synchronized between the active and standby firewalls.
- Configurations that can be backed up are as follows:
  - **Policies**: security policy, NAT policy (including the NAT address pool), and NAT Server
  - **Objects**: address, region, service, application, and user
  - **Network**: security zone, DNS, IPsec, and SSL VPN
  - **System**: administrator, virtual system, and log configuration
- Status information that can be backed up is as follows:
  - Session table, server mapping table, blacklist/whitelist, port mapping table in PAT mode, address mapping table in NO-PAT mode, Layer 2 forwarding table (static MAC address backup), AAA user table (default user **admin** is not backed up), online user monitoring table, PKI certificate, and IPsec

HUAWEI

- To ensure successful switchover, key configuration commands and status information (such as session table information) must be synchronized between the active and standby firewalls.
- To achieve this, Huawei firewalls use HRP to back up dynamic status data and key configuration commands between the active and standby firewalls.
- In active/standby backup mode, the active firewall synchronizes configuration commands and status information to the standby firewall.
- In load balancing mode, both firewalls are active. Therefore, if both firewalls synchronize commands to each other, command overwrite or conflict problems may occur. To centrally manage the configurations of the two firewalls, you need to configure the designated active and standby devices.

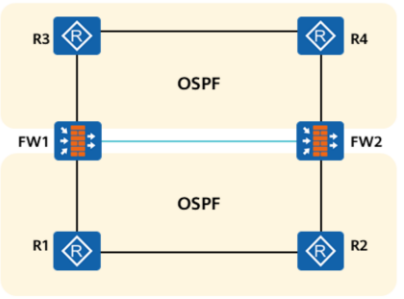
Typical Networking Scenarios of Firewall Hot Standby

Hot standby is deployed in in-line mode and connects to Layer 2 devices.



- The service interfaces of the firewalls work at Layer 3 and connect to switches in the upstream and downstream directions.
- The default gateway of the terminal can be set to the virtual IP address of VRRP VRID1. When configuring the return route on SW3/SW4, you can set the next hop to the virtual IP address of VRRP VRID100.

Hot standby is deployed in in-line mode and connects to Layer 3 devices.



- The service interfaces of the firewalls work at Layer 3 and connect to routers in the upstream and downstream directions.
- OSPF runs between firewalls and routers. When a service interface of FW1 is faulty, FW1 becomes the standby device, and FW2 becomes the active device. The cost of the routes advertised by FW1 is automatically changed to 65500. After the routes are re-converged, traffic is forwarded through FW2.

HUAWEI

Contents

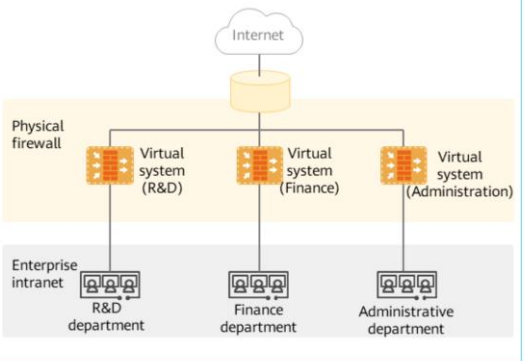
1. Ethernet Switching Security
2. **Advanced Firewall Features**
   - Hot Standby
   - Virtual System

HUAWEI

Application Scenarios of the Firewall Virtual System

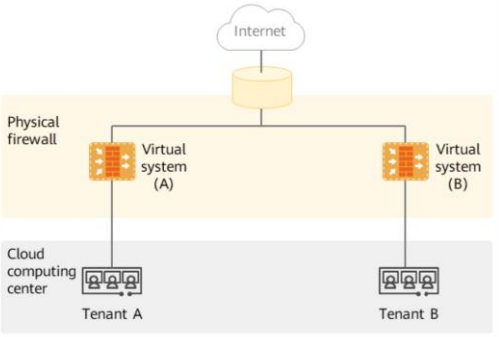
Network isolation for large and medium-sized enterprises

The virtual systems of the firewall isolate the R&D, finance, and administrative departments. The departments can access each other based on the permissions assigned to them, and their administrators have different permissions.

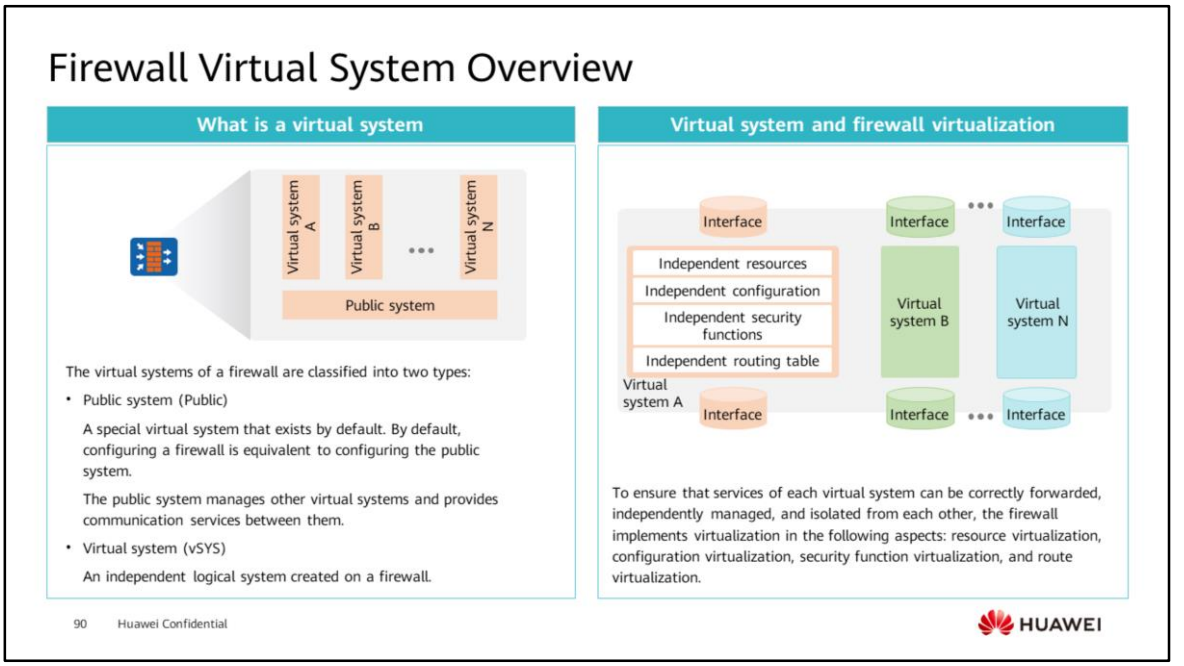


Security gateway for cloud computing centers

With virtual system technology, you can deploy a firewall at the egress of a cloud computing center and create virtual systems for each customer to isolate and protect the traffic of different tenants. Such a firewall functions as the security gateway.

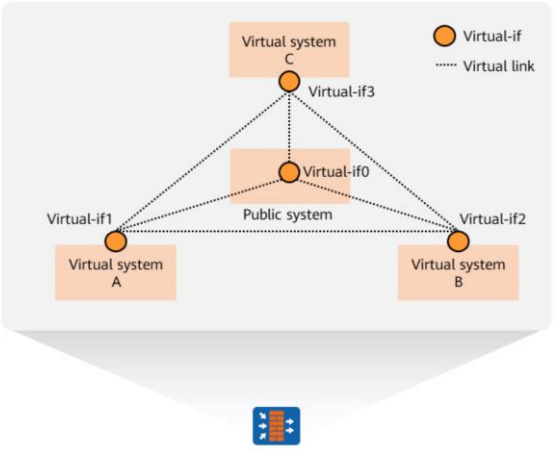


HUAWEI



- The firewall implements virtualization in the following aspects:
  - Resource virtualization: Each virtual system has dedicated resources, including interfaces, VLANs, policies, and sessions. The resources are assigned by the public system administrator and managed by virtual system administrators.
  - Configuration virtualization: Each virtual system has its own virtual system administrator and configuration interface. Virtual system administrators can only manage their own virtual systems.
  - Security function virtualization: Each virtual system has independent security policies and other security functions, which apply only to packets of the virtual system.
  - Route virtualization: Virtual systems maintain separate routing tables, independent and isolated from each other. Currently, only static routes can be virtualized.
- With the preceding virtualization techniques, each virtual system can function as a dedicated firewall that is exclusively managed by its administrator.

Virtual Interface



- Virtual systems communicate with each other through virtual interfaces.
- Virtual interfaces are logical interfaces used for inter-virtual system communication. After a virtual system is created, the device automatically creates a virtual interface for the virtual system.
- A virtual interface can work properly only after it is assigned an IP address and added to a security zone.
- Virtual interfaces are named in the format of **Virtual-if**+ *Interface number*, with the virtual interface of the public system numbered 0 (Virtual-if0). Other virtual interfaces are automatically numbered from 1.

HUAWEI

- As shown in the figure, virtual interfaces of virtual systems and the public system are connected to form virtual links. You can consider virtual systems and the public system as independent devices, and virtual interfaces as communication interfaces between them. Virtual systems can communicate with each other and with the public system after their virtual interfaces are added to security zones and routes and policies are configured for device communications.

Virtual System Communication with the Public System

- Users on network segment 10.3.0.0/24 in virtual system A access the Internet server 3.3.3.3 through GE1/0/1 of the public system.

Destination IP Address	Outbound Interface	Next Hop
0.0.0.0/0	Virtual-if0	0.0.0.0
10.3.0.0/24	GE1/0/2	10.1.1.2
...	...	...

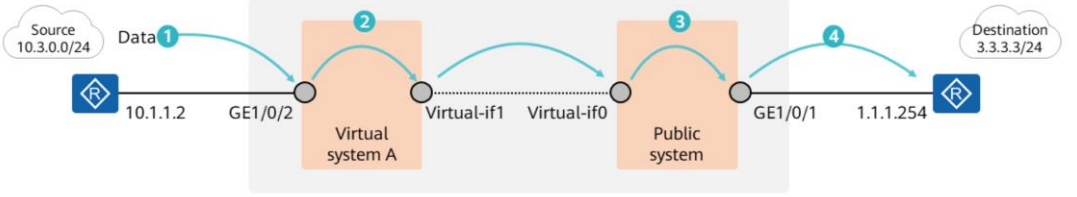
Routing table of virtual system A

Forwards data based on the routing table and establishes a session.

Destination IP Address	Outbound Interface	Next Hop
0.0.0.0/0	GE1/0/1	1.1.1.254
...	...	...
...	...	...

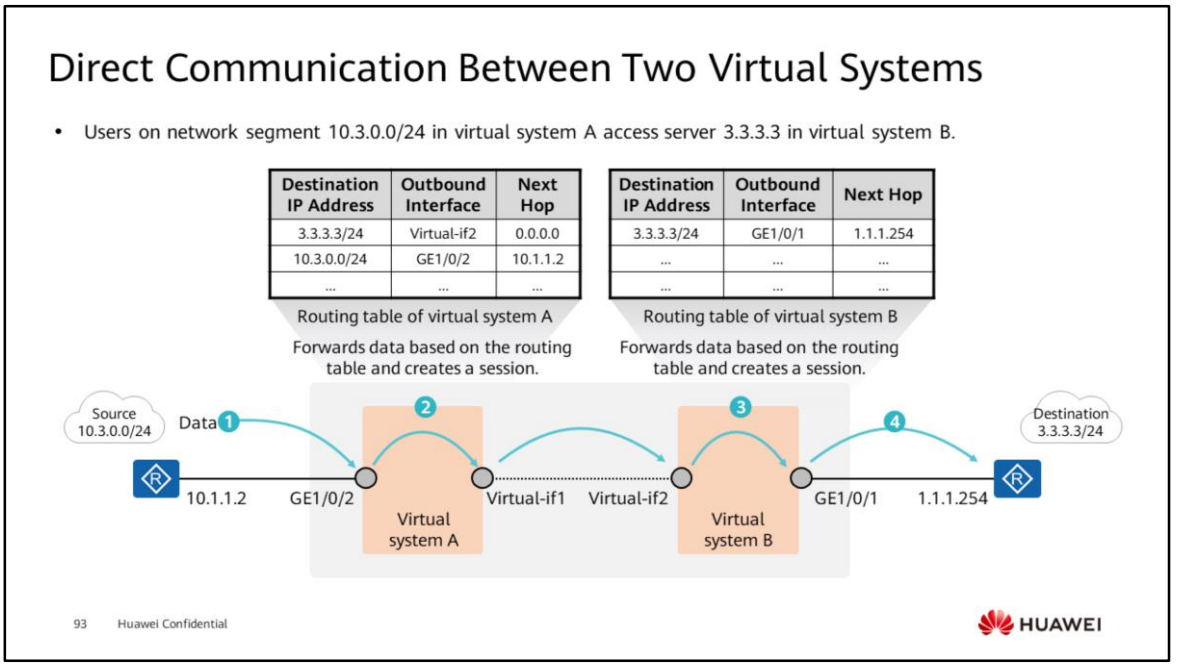
Routing table of the public system

Forwards data based on the routing table and establishes a session.



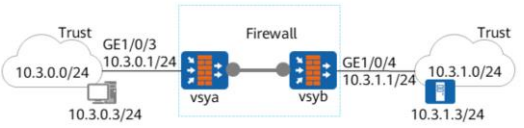
HUAWEI

- Communication between a virtual system and the public system involves two scenarios: from a virtual system to the public system, and from the public system to a virtual system. The packet forwarding processes in the two scenarios are slightly different.
- This slide uses the access from a virtual system to the public system as an example. Packets are processed in both the virtual system and public system according to the firewall's packet forwarding process. As such, you must perform key configurations such as security policies and routes in both the virtual system and public system.



- On a firewall, virtual systems are isolated by default. As such, hosts attached to different virtual systems cannot communicate with each other. To enable communication between two hosts attached to different virtual systems, configure security policies and routes. In this example, virtual system A initiates an access request to virtual system B. The request packet enters virtual system A, which then processes the packet according to the firewall's packet forwarding process. Then, the request packet enters virtual system B, which also processes the packet according to the firewall's forwarding process.
- As both virtual systems need to process the packet according to the firewall's packet forwarding process, you must perform key configurations such as security policies and routes in both virtual systems.

Example for Configuring Communication Between Virtual Systems (1)



- As shown in the figure, users in vsysa need to access the server in vsysb.
- The configuration procedure is as follows:
  - **Configure the route for communication between virtual systems vsysa and vsysb in the public system.**
  - Configure a security policy in vsysa.
  - Configure a security policy in vsysb.

Configure the route for communication between virtual systems vsysa and vsysb in the public system.

```
<FW> system-view
[FW] ip route-static vpn-instance vsysa 10.3.1.0 24 vpn-instance vsysb
```

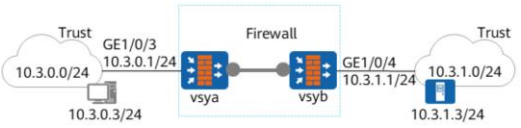
Set an IP address for the public system's virtual interface Virtual-if0 and add the interface to the Trust zone. The IP address of Virtual-if0 can be any value as long as it does not conflict with the IP address of any other interface.

```
[FW] interface Virtual-if 0
[FW-Virtual-if0] ip address 172.16.0.1 24
[FW-Virtual-if0] quit
[FW] firewall zone trust
[FW-zone-trust] add interface Virtual-if0
[FW-zone-trust] quit
```


HUAWEI

- The preceding configuration allows only the unidirectional communication from vsysa to vsysb. If hosts in vsysb need to access hosts in vsysa, you must configure the routes and security policies for access from vsysb to vsysa.

Example for Configuring Communication Between Virtual Systems (2)



- As shown in the figure, users in vsysa need to access the server in vsysb.
- The configuration procedure is as follows:
  - Configure the route for communication between virtual systems vsysa and vsysb in the public system.
  - **Configure a security policy in vsysa.**
  - Configure a security policy in vsysb.

Set an IP address for vsysa's virtual interface Virtual-if1 and add the interface to the Untrust zone. The IP address can be any value as long as it does not conflict with the IP address on any other interface.

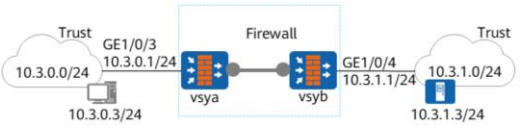
```
[FW-vsysa] interface Virtual-if 1
[FW-vsysa-Virtual-if1] ip address 172.16.1.1 24
[FW-vsysa-Virtual-if1] quit
[FW-vsysa] firewall zone untrust
[FW-vsysa-zone-untrust] add interface Virtual-if1
[FW-vsysa-zone-untrust] quit
```

Configure a policy for users in vsysa to access the server in vsysb.

```
[FW-vsysa] security-policy
[FW-vsysa-policy-security] rule name to_server
[FW-vsysa-policy-security-rule-to_internet] source-zone trust
[FW-vsysa-policy-security-rule-to_internet] destination-zone untrust
[FW-vsysa-policy-security-rule-to_internet] source-address 10.3.0.0 24
[FW-vsysa-policy-security-rule-to_internet] destination-address 10.3.1.3 32
[FW-vsysa-policy-security-rule-to_internet] action permit
[FW-vsysa-policy-security-rule-to_internet] quit
[FW-vsysa-policy-security] quit
```


HUAWEI

Example for Configuring Communication Between Virtual Systems (3)



- As shown in the figure, users in vsysa need to access the server in vsysb.
- The configuration procedure is as follows:
  - Configure the route for communication between virtual systems vsysa and vsysb in the public system.
  - Configure a security policy in vsysa.
  - **Configure a security policy in vsysb.**

Set an IP address for vsysb's virtual interface Virtual-if2 and add the interface to the Untrust zone. The IP address can be any value as long as it does not conflict with the IP address on any other interface.

```
[FW-vsysb] interface Virtual-if 2
[FW-vsysb-Virtual-if1] ip address 172.16.2.1 24
[FW-vsysb-Virtual-if1] quit
[FW-vsysb] firewall zone untrust
[FW-vsysb-zone-untrust] add interface Virtual-if2
[FW-vsysb-zone-untrust] quit
```

Configure a policy for users in vsysa to access the server in vsysb.

```
[FW-vsysb] security-policy
[FW-vsysb-policy-security] rule name vsysa_to_server
[FW-vsysb-policy-security-rule-vsysa_to_vsysb] source-zone untrust
[FW-vsysb-policy-security-rule-vsysa_to_vsysb] destination-zone trust
[FW-vsysb-policy-security-rule-vsysa_to_vsysb] source-address 10.3.0.0 24
[FW-vsysb-policy-security-rule-vsysa_to_vsysb] destination-address 10.3.1.3
32
[FW-vsysb-policy-security-rule-vsysa_to_vsysb] action permit
```

Verify the configuration.

```
PC>ping 10.3.1.3
Ping 10.3.1.3: 32 data bytes, Press Ctrl_C to break
From 10.3.1.3: bytes=32 seq=1 ttl=127 time=79 ms
```


HUAWEI

Quiz

1. (Multiple-answer question) DHCP snooping is a DHCP security feature. Which of the following attacks can DHCP snooping defend against? ( )
   A. Starvation attacks by changing the CHADDR field
   B. Bogus DHCP server attacks
   C. TCP flag attacks
   D. Man-in-the-middle attacks and IP/MAC spoofing attacks
2. (True or false) On a firewall, virtual systems are isolated by default. Hosts in different virtual systems cannot communicate with each other. ( )
   A. True
   B. False

HUAWEI

1. ABD
2. A

Summary

- Port isolation can isolate interfaces in a VLAN. Two port isolation modes are available: Layer 2 isolation and Layer 3 interworking, and Layer 2 and Layer 3 isolation.
- MAC address entries of a switch are classified into static, blackhole, and dynamic MAC address entries.
- Port security enables a switch to convert dynamic MAC addresses learned by an interface into secure MAC addresses. Secure MAC addresses are usually used together with security protection actions.
- Enabling MAC address flapping detection on a switch helps engineers quickly troubleshoot loops on the switch.
- MACsec defines a method for secure data communication based on Ethernet and ensures data transmission security through hop-by-hop data encryption between devices.
- The difference between traffic suppression and storm control is that traffic suppression only limits the rate of various packets and discards excess packets, whereas storm control takes different actions, such as shutting down an interface or blocking packets based on the packet rate.
- DHCP snooping plays an important role in preventing network attacks on terminals that automatically obtain IP addresses on the Ethernet. You can configure the DHCP snooping trusted interface and DHCP snooping binding table to prevent DHCP-based network attacks.
- An IPSG-enabled switch checks packets against the binding table to prevent IP spoofing attacks.
- Deploying firewalls in hot standby mode can improve network reliability, and deploying firewall virtual systems can logically divide physical devices to improve resource utilization.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

MPLS Fundamentals and Configuration

HUAWEI

Foreword

- In the mid-1990s, Internet traffic grew rapidly. Traditional IP packets were forwarded by routers based on routing tables. However, due to hardware technology limitations, the forwarding performance of routers is low. As a result, this mode becomes a bottleneck for network data forwarding.
- As such, the Multiprotocol Label Switching (MPLS) protocol is proposed. It is designed to speed up packet forwarding on routers. On an MPLS network, IP packet headers are analyzed on network edges only, and MPLS uses labels to guide packet forwarding inside the MPLS network. Label-based forwarding features high efficient, shortening the packet processing time.
- With the continuous improvement of devices' hardware performance, the advantage of MPLS in improving the data forwarding speed is gradually weakened. However, as MPLS supports multi-layer label nesting and forwarding-and-control separation inside a device, MPLS is applicable to many scenarios, such as virtual private network (VPN) and quality of service (QoS).
- This course describes the principles and label-based forwarding process of MPLS, as well as the method for configuring static label switched paths (LSPs).

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Understand basic MPLS concepts and terms.
  - Understand the principles of MPLS.
  - Configure static LSPs.
  - Describe the MPLS forwarding process.

HUAWEI

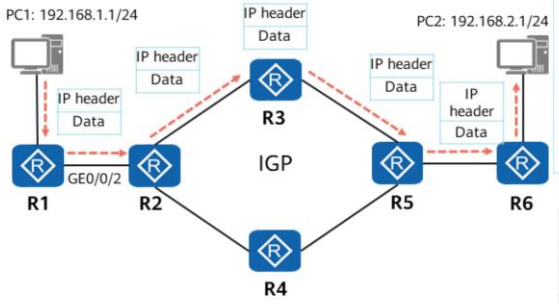
Contents

1. **MPLS Basics**
   - MPLS Overview
   - MPLS Terms
   - MPLS Labels
2. MPLS Forwarding
3. Static LSP Configuration

HUAWEI

Traditional IP Routing and Forwarding

- Traditional IP forwarding uses the hop-by-hop forwarding mode, in which a packet is decapsulated by all routers that receive the packet. Each router needs to obtain the network layer information and search its routing table for a route based on the longest match rule. The repeated processes of decapsulating packets, searching routing tables, and re-encapsulating the packets on routers result in low forwarding performance.



Characteristics of traditional IP routing and forwarding:

- All routers need to know the network-wide routes.
- The length of an IP header is not fixed, and the processing efficiency is low.
- Traditional IP forwarding is connectionless and cannot provide good end-to-end QoS guarantee.

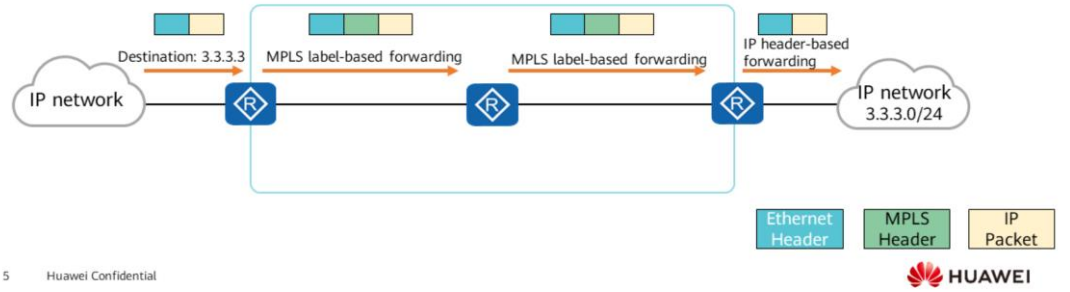
R1 routing table

Destination/Mask	Protocol	Preference	Cost	Next Hop	Interface
192.168.1.0/24	Direct	0	0	192.168.1.254	GE0/0/0
192.168.0.0/16	Direct	0	0	192.168.12.1	GE0/0/2
0.0.0.0/0	Static	60	0	192.168.12.2	GE0/0/2

HUAWEI

Basic MPLS Concepts

- MPLS is located between the data link layer and the network layer in the TCP/IP protocol stack and can provide services for all network layers.
- An MPLS header is added between a data-link-layer header and a network-layer header, and data can be forwarded quickly based on the Label field in the MPLS header.
- This course describes the application of MPLS on IP networks only.



HUAWEI

- MPLS is derived from the Internet Protocol version 4 (IPv4). Core MPLS technologies can be extended to support multiple network protocols, such as the Internet Protocol version 6 (IPv6), Internet Packet Exchange (IPX), Appletalk, DECnet, and Connectionless Network Protocol (CLNP). MPLS uses label-based forwarding to replace IP forwarding. A label is a short connection identifier of fixed length that is meaningful only to a local end.

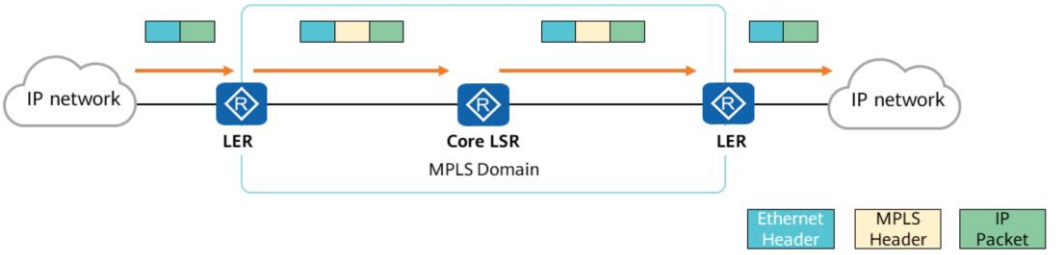
Contents

1. **MPLS Basics**
   - MPLS Overview
   - MPLS Terms
   - MPLS Labels
2. MPLS Forwarding
3. Static LSP Configuration

HUAWEI

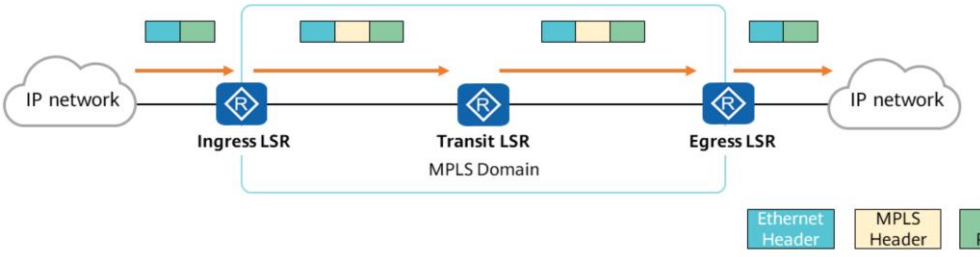
MPLS Terms - LSR and MPLS Domain

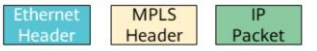
- MPLS domain: consists of a series of consecutive network devices that run MPLS.
- Label switching router (LSR): a routing device, such as a router or switch, that runs MPLS. An LSR that resides at the edge of an MPLS domain and connects to a non-MPLS network is called a label edge router (LER). An LSR that resides inside an MPLS domain is called a core LSR.



MPLS Terms - LSR Classification

- In addition to location-based LSR classification, LSRs can also be classified according to data processing modes:
  - Ingress LSR: generally pushes an MPLS header into an IP packet and generates an MPLS packet.
  - Transit LSR: generally pushes a label, swaps a label, or pops out a label in a received MPLS packet, and then forwards the packet inside the MPLS domain.
  - Egress LSR: generally pops out an MPLS header to restore the IP packet.



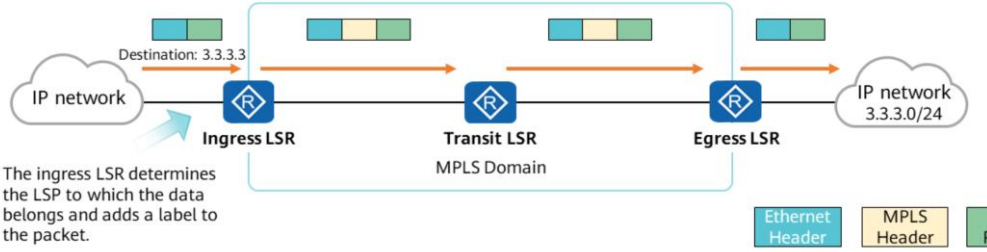


HUAWEI

- MPLS label operations will be introduced in following courses.

MPLS Terms - FEC

- Forwarding equivalence class (FEC): a set of packets with similar or identical characteristics and forwarded in the same way by LSRs.
  - On an MPLS network, FECs can be classified based on destination IP addresses, network masks, DSCP values, and so on.
  - When a packet enters an MPLS domain, the ingress LSR determines the LSP to which the packet belongs.
  - MPLS labels match FECs. A mechanism must be available for LSRs to obtain label information about FECs.

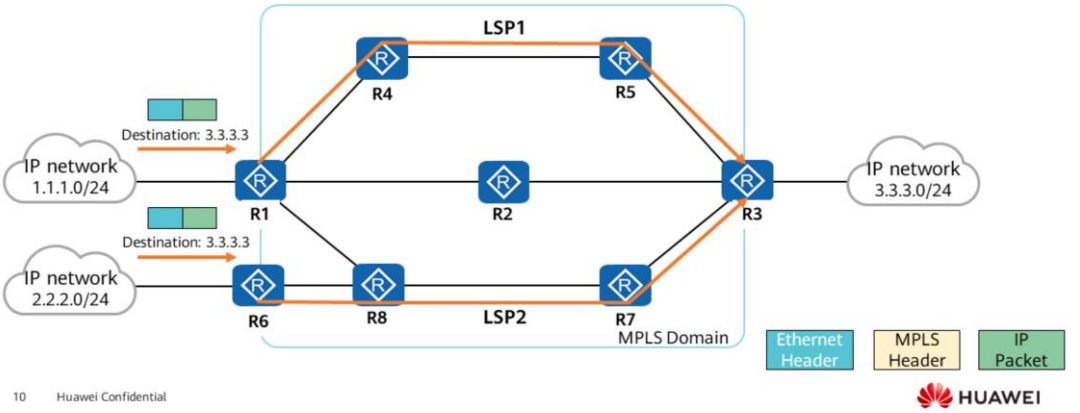


HUAWEI

- In traditional IP forwarding that uses the longest match algorithm, all packets that match the same route belong to the same FEC.
- In MPLS, the most common example of FEC is: Packets whose destination IP addresses match the same IP route are considered to belong to the same FEC.

MPLS Terms - LSP

- Label switched path (LSP): a path through which a labeled packet traverses an MPLS domain to reach a destination.
- The packets of the same FEC usually traverse an MPLS domain through the same LSP. Therefore, an LSR always uses the same label to forward the packets of the same FEC.



HUAWEI

- An LSP is composed of an ingress LSR, an egress LSR, and a variable number of transit LSRs. Therefore, an LSP can be considered as an ordered set of these LSRs.
- An LSP must be established before a packet is forwarded; otherwise, the packet fails to traverse an MPLS domain.
- LSPs can be established in static or dynamic mode.
- An LSP is a unidirectional path from the start point to the end point. If bidirectional data communication is required, an LSP for return traffic needs to be established between the two ends.

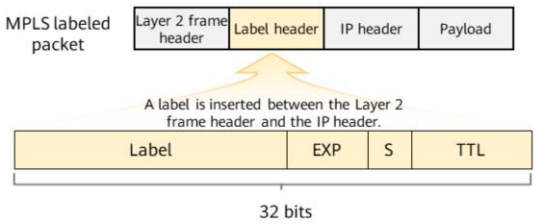
Contents

1. **MPLS Basics**
   - MPLS Overview
   - MPLS Terms
   - MPLS Labels
2. MPLS basic principles
3. Static LSP Configuration

HUAWEI

MPLS Labels

- Before an IP packet enters an MPLS domain, the ingress LSR pushes an MPLS header (also called an MPLS label) into the packet to form an MPLS labeled packet. An MPLS labeled packet can contain one or more labels.



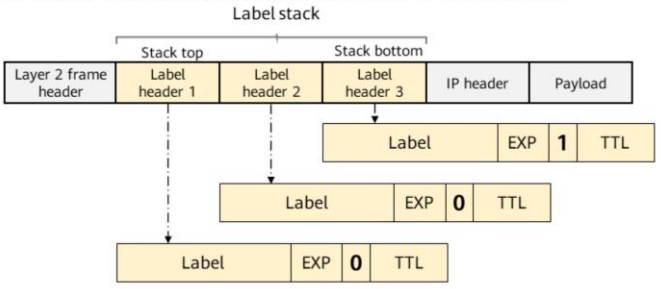
- Label: carries a label value and is 20 bits long.
- Experimental use (EXP): is mainly used for class of service (CoS) and is 3 bits long.
- Bottom of stack (S): indicates whether a label header is the last label and is 1 bit long. If its value is 1, the current label header is at the bottom of the label stack. If its value is 0, the current label header is not at the bottom and is followed by other label headers.
- Time to live (TTL): prevents a labeled packet from being infinitely forwarded when a loop occurs. It is 8 bits long and has the same meaning as the TTL field in an IP packet header.

HUAWEI

- The EXP field is defined in early MPLS standards and is an experimental field. Actually, this field is mainly used for CoS. To avoid ambiguity, this field is renamed Traffic Class in RFC 5462.

MPLS Label Stack

- MPLS allows one or more layers of label headers to be added in a packet. The ordered set of these label headers is called label stack.
- When there are multiple labels in the label stack:
  - The label closest to the Layer 2 header is the top label, and the S field in the label is 0.
  - The label closest to the IP header is the bottom label, and the S field in the label is 1.



HUAWEI

- When the upper layer is the MPLS label stack, the Type field in the Ethernet header is 0x8847, and the Protocol field in the PPP header is 0x8281.

Label Space

- A label is a short and fixed-length identifier that has only local significance. A label space means a label value range. The label value range and planning are as follows:

Label Value	Description
0 to 15	Special label values. For example, 0 indicates an IPv4 explicit null label, and 3 indicates an implicit null label.
16 to 1023	Label space shared by static LSPs and static CR-LSPs.
1024 to 1048575	Label space for dynamic signaling protocols, such as Label Distribution Protocol (LDP), Resource Reservation Protocol-Traffic Engineering (RSVP-TE), and Multiprotocol Extensions for Border Gateway Protocol (MP-BGP). Each dynamic signaling protocol uses an independent and contiguous label space, which is not shared with other dynamic signaling protocols.

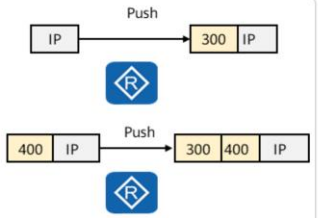
HUAWEI

- The label spaces of different LSRs are independent of each other, indicating that each router can use the entire label space.

MPLS Label Processing

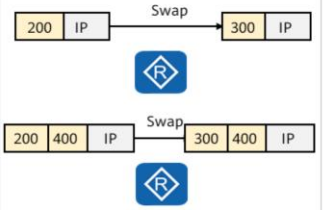
- LSRs can perform the following operations on labels: push, swap, and pop.

Push



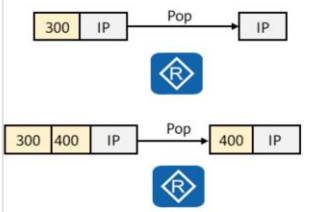
When an IP packet enters an MPLS domain, the ingress adds a label between the Layer 2 frame header and IP header of the packet. When the packet reaches a transit node, the transit node can also add a new label at the top of the label stack.

Swap



When the packet is forwarded inside the MPLS domain, a transit node searches the label forwarding table and replaces the top label with the label that is assigned by the next hop.

Pop



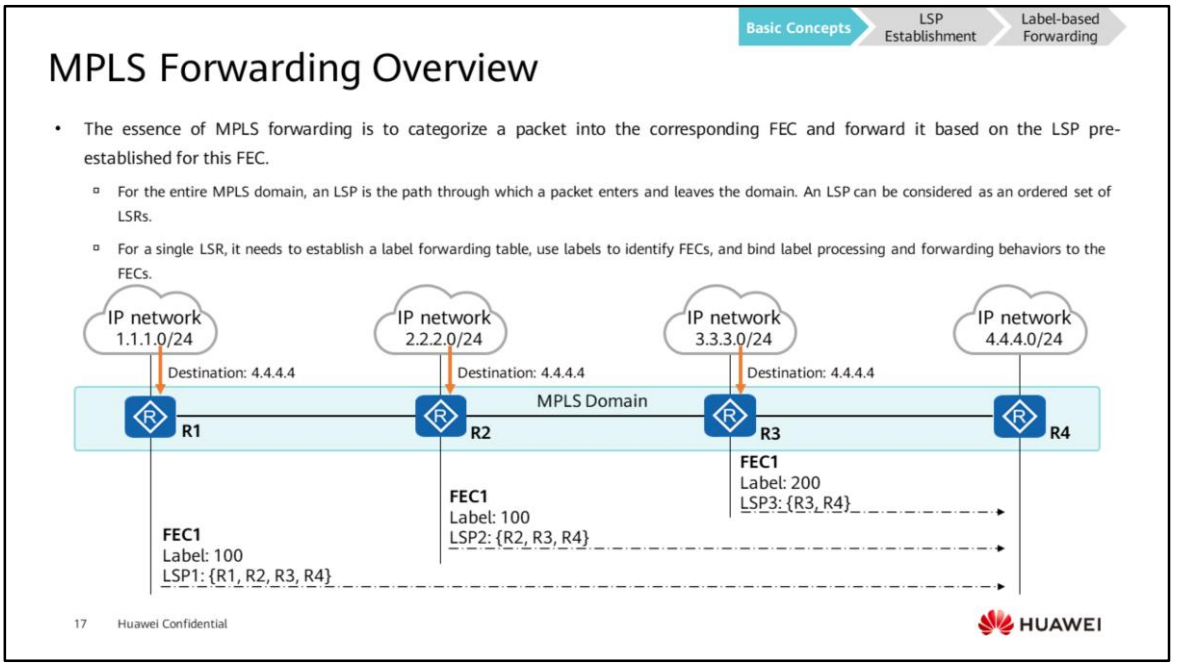
Before the packet leaves the MPLS domain, the egress removes the label of the MPLS packet.

HUAWEI

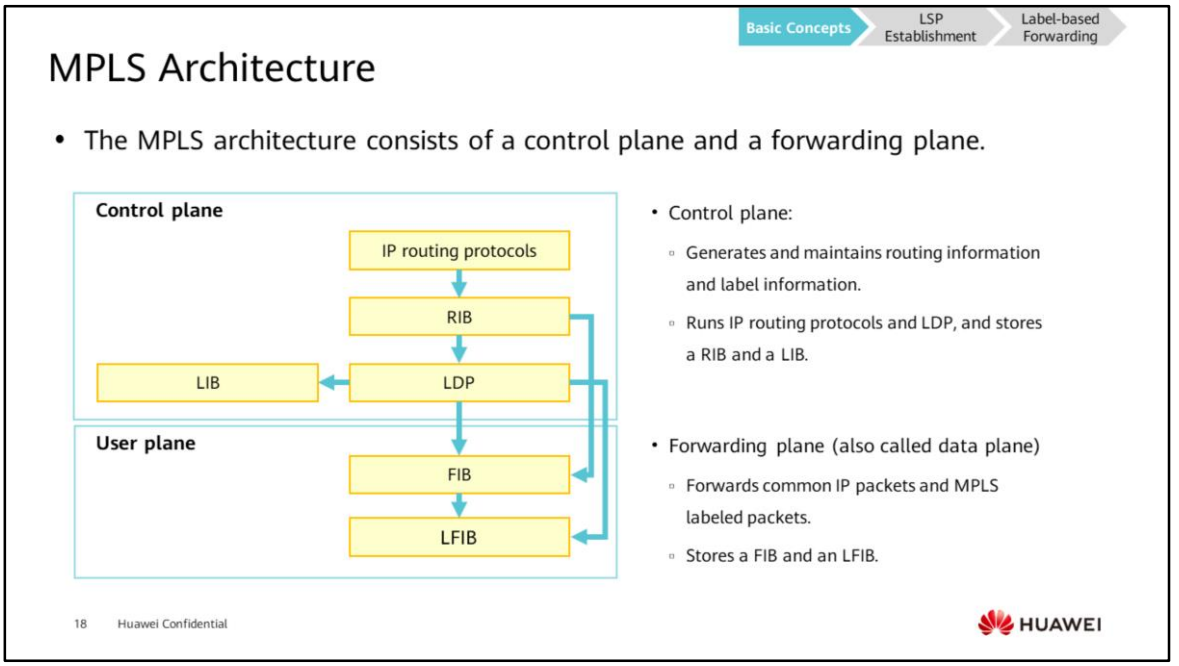
Contents

1. MPLS Basics
2. **MPLS Forwarding**
3. Static LSP Configuration

HUAWEI

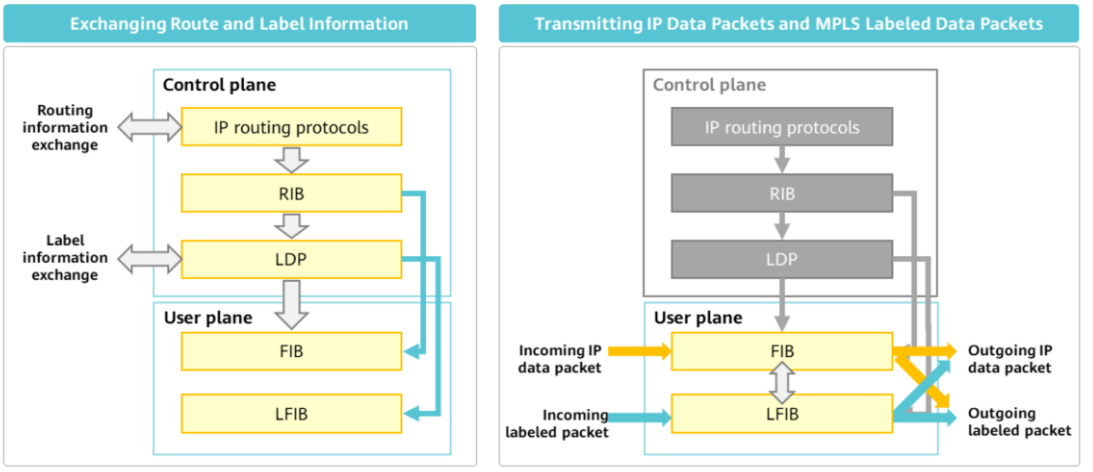


- If the ingress LSRs of packets belonging to the same FEC are different, the LSPs for forwarding the packets are different.
- An LSR uses the same way to process packets in the same FEC, regardless of where the packets' inbound interfaces are the same.
- An LSP is composed of the forwarding actions of LSRs, and the label forwarding table determines the forwarding action. Therefore, establishing a label forwarding table can also be considered as establishing an LSP.
- As shown in the figure, the three packets belong to the same FEC, FEC1, because they have the same destination. However, as their ingress LSRs are different, the packets are forwarded along different LSPs (LSP1, LSP2, and LSP3, respectively). The labels assigned by different LSRs to the same FEC can be the same or different, because labels are valid only on their local LSRs.



- Control plane:
  - The control plane is connectionless. It generates and maintains routing and label information.
  - The control plane includes:
    - Routing information base (RIB): stores static routes, direct routes, and routes generated by IP routing protocols. Routes can be selected from the RIB to guide packet forwarding.
    - Label information base (LIB): stores and manages labels statically configured and dynamically generated by label switching protocols (such as LDP and RSVP).
- Forwarding plane
  - The forwarding plane, also called the data plane, is connection-oriented. It forwards common IP packets and MPLS labeled packets.
  - The forwarding plane includes:
    - Forwarding information base (FIB): stores forwarding information that is generated based on the routing information extracted from the RIB. The forwarding information is used to guide common IP packet forwarding.
    - Label forwarding information base (LFIB): stores label-based forwarding information to guide MPLS labeled packet forwarding.

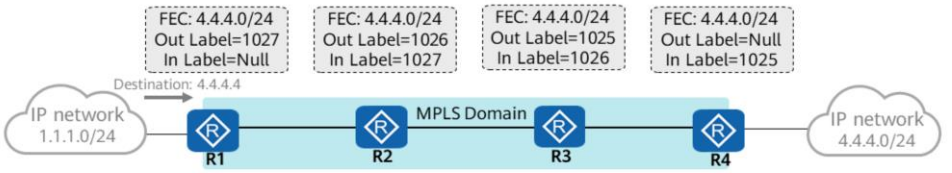
Control Plane and Forwarding Plane



HUAWEI

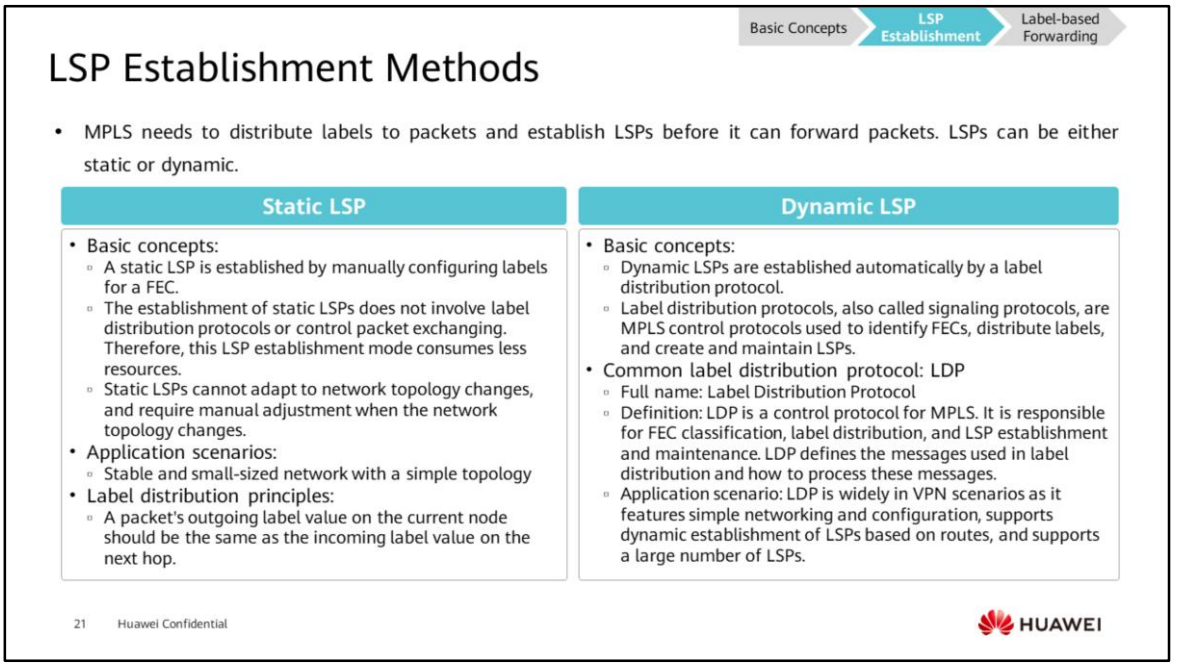
LSP Establishment Principles

- When the network layer protocol is IP, the route for a FEC must exist in the IP routing table of an LSR. Otherwise, the label forwarding entry of the FEC does not take effect.
- An LSR uses labels to identify a FEC, and it can correctly process an incoming packet of a FEC only if the packet carries the correct label.

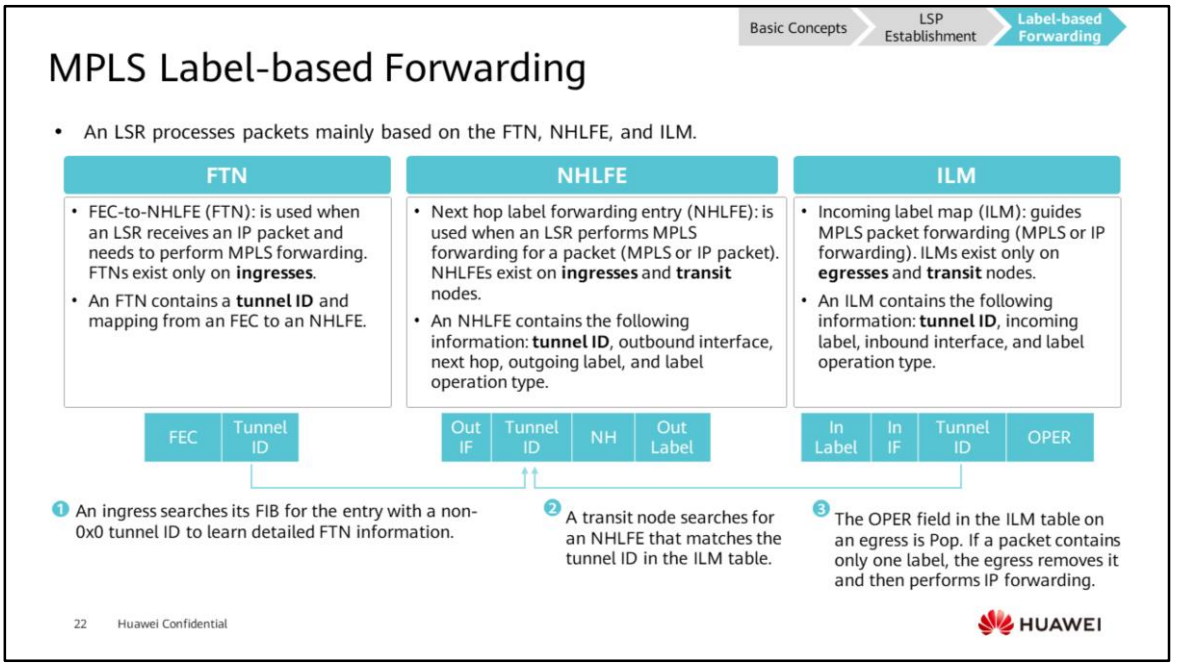


- A device provides an incoming label and an outgoing label for each FEC. The incoming label indicates the one that needs to be carried in a received packet. The outgoing label indicates the one that needs to be carried in a sent packet.
- Assume that a packet destined for 4.4.4.0/24 is sent from R1 to R2. R1 is the upstream LSR of R2, and R3 is the downstream LSR of R2. To ensure that a labeled packet is correctly processed and forwarded, the outgoing label carried in the packet sent from R1 to R2 must be the same as the incoming label on R2 for the FEC to 4.4.4.0/24. Similarly, the outgoing label on R2 for the FEC to 4.4.4.0/24 must be the same as the incoming label of R3.

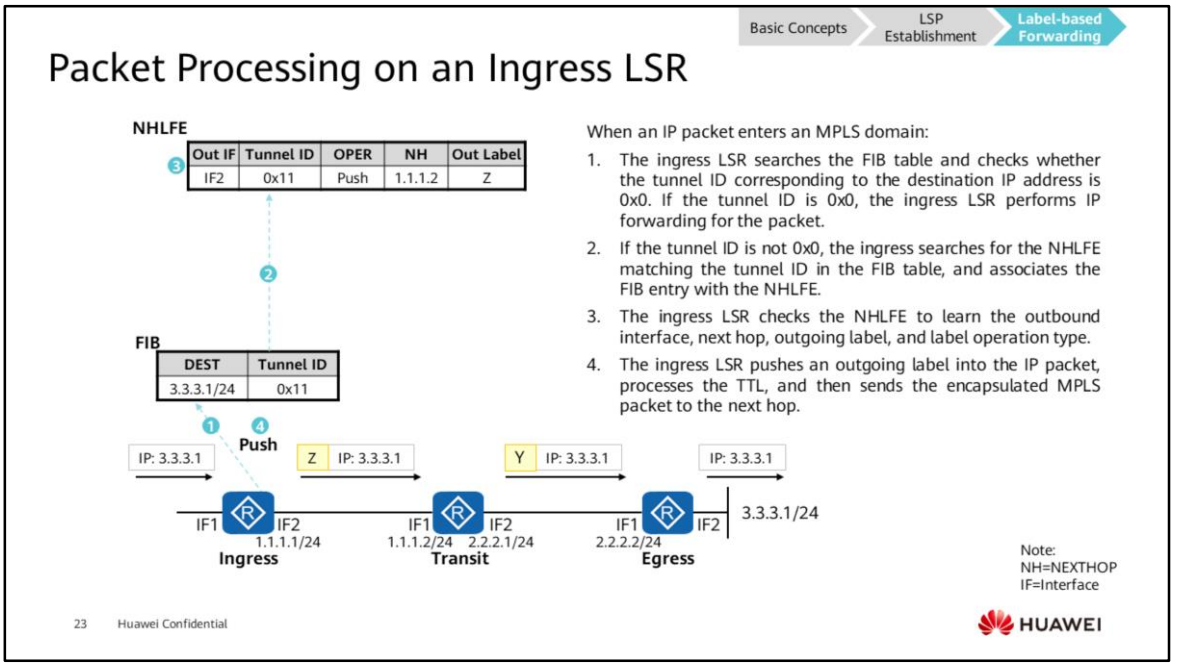
HUAWEI



- Static LSP:
  - A static LSP is meaningful only to the local node, and the local node cannot be aware of the entire LSP.
- Dynamic LSP:
  - Other label distribution protocols:
    - Resource Reservation Protocol-Traffic Engineering (RSVP-TE): an extension based on RSVP. RSVP-TE is used to establish constraint-based routed LSPs (CR-LSPs). Unlike LDP LSPs, CR-LSPs support parameters, such as bandwidth reservation requests, bandwidth constraints, link colors, and explicit paths.
    - Multiprotocol Border Gateway Protocol (MP-BGP): an extension based on BGP. MP-BGP distributes labels to MPLS VPN routes and inter-AS VPN labeled routes.



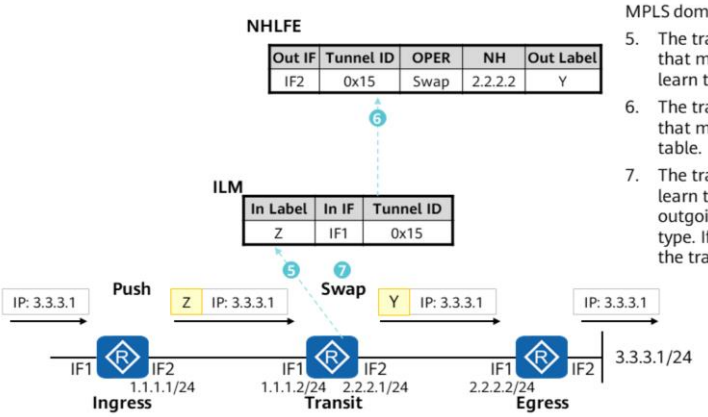
- Tunnel ID: an ID automatically allocated to a tunnel, providing a unified interface for upper-layer applications (such as VPN and route management) that use the tunnel. A tunnel ID is 32 bits long and is valid only on the local device. During MPLS forwarding, LSRs find matching FIB entries, ILMs, and NHLFEs based on tunnel IDs.



- An ingress LSR searches the FIB table (to learn FTN information) and NHLFE table to guide packet forwarding.
- When an IP packet enters an MPLS domain, the ingress searches the FIB to check whether the tunnel ID corresponding to the destination IP address is 0x0.
    - If the tunnel ID is 0x0, the ingress LSR performs IP forwarding for the packet.
    - If the tunnel ID is not 0x0, the ingress LSR performs MPLS forwarding for the packet.

Basic Concepts > LSP Establishment > Label-based Forwarding

Packet Processing on a Transit LSR



When an MPLS packet is forwarded in an MPLS domain:

5. The transit LSR searches for the ILM that matches the MPLS label value to learn the tunnel ID.
6. The transit LSR searches for the NHLFE that matches the tunnel ID in the ILM table.
7. The transit LSR checks the NHLFE to learn the outbound interface, next hop, outgoing label, and label operation type. If the label operation type is Swap, the transit LSR performs label swapping.

Note:
NH=NEXTHOP
IF=Interface

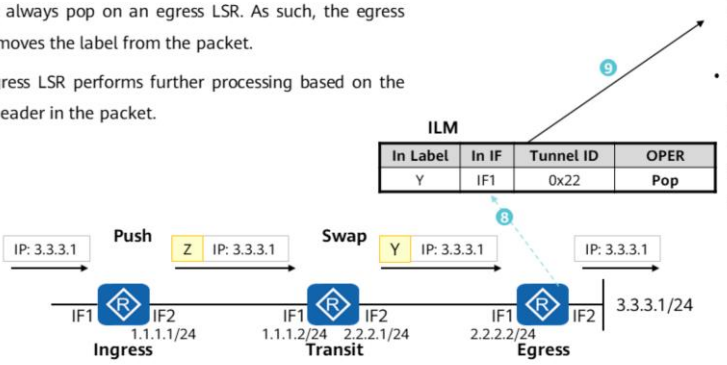
HUAWEI

- A transit LSR searches for ILMs and NHLFEs to guide MPLS packet forwarding.

Packet Processing on an Egress LSR

When an MPLS packet leaves an MPLS domain:

8. An egress LSR searches the ILM table for the operation type that matches the FEC of the packet. The operation type is always pop on an egress LSR. As such, the egress LSR removes the label from the packet.
9. The egress LSR performs further processing based on the inner header in the packet.



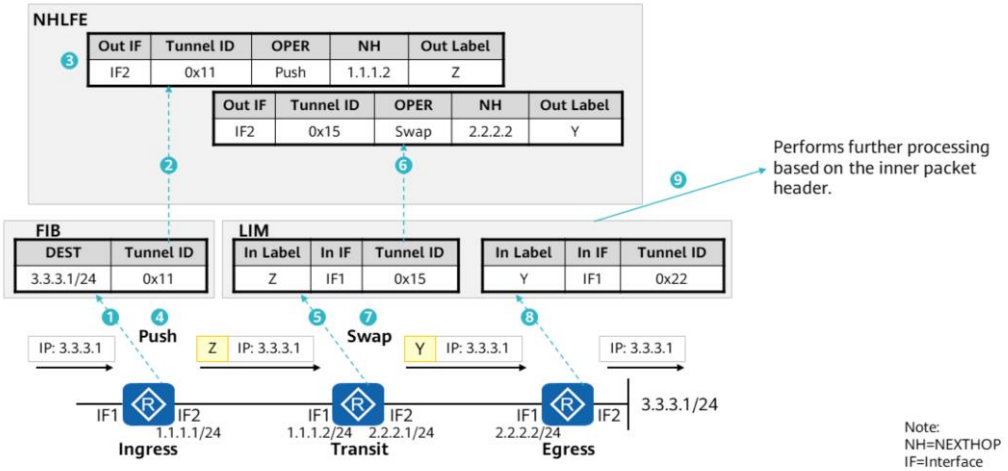
- S = 1: The current label is the bottom label, and IP forwarding is performed for the packet.
- S = 0: The current label is not the bottom label, and a lower-layer label exists. In this case, the packet is performed based on the lower-layer label.

Note:
NH=NEXTHOP
IF=Interface

HUAWEI

- The egress LSR searches the ILM table to guide MPLS packet forwarding.

Detailed MPLS Packet Forwarding Process



Note:
NH=NEXTHOP
IF=Interface

HUAWEI

Contents

1. MPLS Basics
2. MPLS Forwarding
3. **Static LSP Configuration**

HUAWEI

Basic MPLS Configuration Commands

1. Set an LSR ID.

   [Huawei] **mpls lsr-id** *lsr-id*

   The **mpls lsr-id** command sets an LSR ID. An LSR ID uniquely identifies an LSR on a network. By default, no LSR ID is set for an LSR, and an LSR ID can only be manually set. To improve network reliability, you are advised to use the IP address of a loopback interface on an LSR as the LSR ID and plan the LSR IDs of all LSRs on the network in a unified manner before configuration.

2. Enable MPLS.

   [Huawei] **mpls**

   The **mpls** command enables MPLS on the current node and displays the MPLS view.

   [Huawei-GigabitEthernet0/0/0] **mpls**

   Enable MPLS on an interface. Before running this command, enable MPLS globally.

HUAWEI

Static LSP Configuration Commands (1)

1. Configure a static LSP on an ingress LSR.

   [Huawei] **static-lsp ingress** *lsp-name* **destination** *ip-address* { *mask-length* | *mask* } { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* }* **out-label** *out-label*

   The **static-lsp ingress** command configures a static LSP on an ingress.
   - You are advised to specify the **next-hop** parameter to configure a static LSP. This ensures that the local routing table contains the routing entry that exactly matches the specified destination IP address and next-hop IP address. If the outbound interface of the LSP is an Ethernet interface, you must specify **nexthop** *next-hop-address* to ensure normal forwarding along the LSP.
   - The value of *out-label* ranges from 16 to 1048575.

2. Configure a static LSP on a transit LSR.

   [Huawei] **static-lsp transit** *lsp-name* [**incoming-interface** *interface-type interface-number*] **in-label** *in-label* { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* }* **out-label** *out-label*

   The **static-lsp transit** command configures a static LSP on a transit node.
   - The configuration rules of the next hop and outbound interface are the same as those for an ingress LSR.
   - The value of *in-label* ranges from 16 to 1023.
   - The value of *out-label* ranges from 16 to 1048575.

HUAWEI

- An outgoing label occupies the label space of the downstream LSR, but the label distribution mode used by the downstream space is uncertain. As such, the value of an outgoing label ranges from 16 to 1048575.
- An incoming label occupies the label space of the current LSR. When a static LSP is used, the value of an incoming label ranges from 16 to 1023.

Static LSP Configuration Commands (2)

3. Configure a static LSP on an egress LSR.

   [Huawei] **static-lsp egress** *lsp-name* [**incoming-interface** *interface-type interface-number*] **in-label** *in-label*

   The **static-lsp egress** command configures a static LSP on an egress LSR.

   - The value of *in-label* ranges from 16 to 1023.

4. Check information about static LSPs.

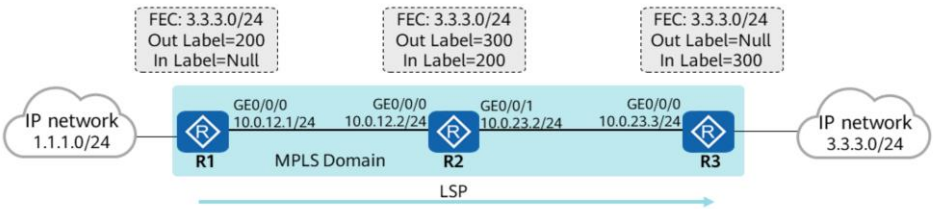
   [Huawei] **display mpls static-lsp** [*lsp-name*] [{ **include** | **exclude** } *ip-address mask-length*] [**verbose**]

   The **display mpls static-lsp** command displays information about static LSPs.

HUAWEI

Static LSP Configuration Example (1)

- Use case: An IGP has been deployed on R1, R2, and R3. The networks 1.1.1.0/24 and 3.3.3.0/24 can communicate with each other. In this scenario, configure static LSPs to allow the two networks to communicate through MPLS. The following figure shows the distributed labels.
- Configuration roadmap:
  - Enable MPLS on the devices and interfaces.
  - Configure static LSPs as planned.

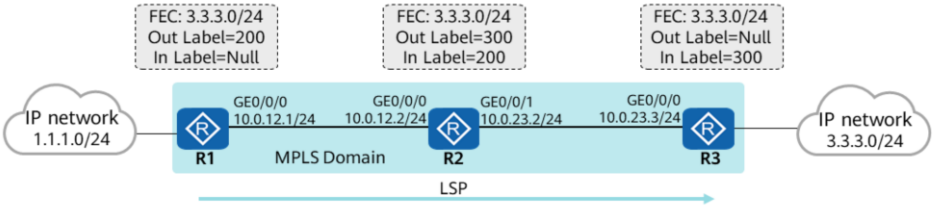


HUAWEI

Static LSP Configuration Example (2)

1. The LSR IDs of the three routers are 10.1.1.1, 10.1.1.2, and 10.1.1.3, respectively. The following example uses R1 to describe the configuration. Enable MPLS globally and on interfaces on R1.

```
[R1]mpls lsr-id 10.1.1.1
[R1]mpls
Info: Mpls starting, please wait... OK!
[R1-mpls]quit
[R1]interface GigabitEthernet 0/0/0
[R1-GigabitEthernet0/0/0]mpls
[R1-GigabitEthernet0/0/0]quit]
```



HUAWEI

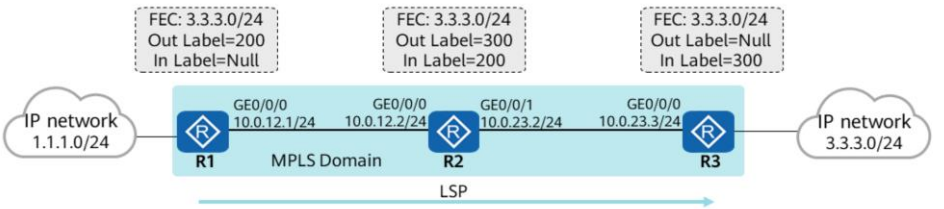
Static LSP Configuration Example (3)

2. Configure a static LSP from R1 to R3.

```
[R1] static-lsp ingress 1to3 destination 3.3.3.0 24 nexthop 10.0.12.2 out-label 200
```

```
[R2] static-lsp transit 1to3 incoming-interface GigabitEthernet 0/0/0 in-label 200 nexthop 10.0.23.3 out-label 300
```

```
[R3] static-lsp egress 1to3 incoming-interface GigabitEthernet 0/0/0 in-label 300
```



HUAWEI

Checking the Configuration

```
[R1]display mpls lsp
------------------------------------------------------------------------------
                LSP Information: STATIC LSP
------------------------------------------------------------------------------
FEC                 In/Out Label   In/Out IF         Vrf Name
3.3.3.0/24          NULL/200       -/GE0/0/0
```

```
[R2]display mpls lsp
------------------------------------------------------------------------------
                LSP Information: STATIC LSP
------------------------------------------------------------------------------
FEC                 In/Out Label   In/Out IF         Vrf Name
3.3.3.0/24          200/300        GE0/0/0/GE0/0/1
```

```
[R3]display mpls lsp
------------------------------------------------------------------------------
                LSP Information: STATIC LSP
------------------------------------------------------------------------------
FEC                 In/Out Label   In/Out IF         Vrf Name
3.3.3.0/24          300/NULL       GE0/0/0/-
```

The ping test from the host on the 1.1.1.0/24 network segment to the host on the 3.3.3.0/24 network segment is successful.

HUAWEI

Packet Header Obtaining Analysis

No.	Time	Source	Destination	Protocol	Length	Info
5	12.109000	1.1.1.1	3.3.3.3	ICMP	102	Echo (ping) request

Frame 5: 102 bytes on wire (816 bits), 102 bytes captured (816 bits) on interface 0
Ethernet II, Src: HuaweiTe_b1:15:3e (00:e0:fc:b1:15:3e), Dst: HuaweiTe_49:20:bb (00:e0:fc:49:20:bb)
MultiProtocol Label Switching Header, Label: 300, Exp: 0, S: 1, TTL: 254
Internet Protocol Version 4, Src: 1.1.1.1, Dst: 3.3.3.3
Internet Control Message Protocol

No.	Time	Source	Destination	Protocol	Length	Info
6	12.141000	3.3.3.3	1.1.1.1	ICMP	98	Echo (ping) reply

Frame 6: 98 bytes on wire (784 bits), 98 bytes captured (784 bits) on interface 0
Ethernet II, Src: HuaweiTe_49:20:bb (00:e0:fc:49:20:bb), Dst: HuaweiTe_b1:15:3e (00:e0:fc:b1:15:3e)
Internet Protocol Version 4, Src: 3.3.3.3, Dst: 1.1.1.1
Internet Control Message Protocol

According to the obtained packet information, the packets from the host on the 1.1.1.0/24 network segment to the host on the 3.3.3.0/24 network segment are forwarded based on MPLS labels.
The packets from the host on the 3.3.3.0/24 network segment to the host on the 1.1.1.0/24 network segment are forwarded based on IP packet headers.

HUAWEI

Quiz

1. (Multiple-answer question) Which of the following statements about an MPLS header are correct? ( )
   A. An MPLS header is 32 bits long.
   B. The value of the Label field in an MPLS header ranges from 0 to 65535.
   C. MPLS can implement multi-layer MPLS header nesting.
   D. The max_hop field in an MPLS header can be used to prevent labeled packets from being infinitely forwarded.
2. (True or false) The values of the incoming and the outgoing label both range from 16 to 1023 for a static LSP to be configured on a transit LSR. ( )
   A. True
   B. False

HUAWEI

1. AC
2. B

Summary

- MPLS was originally proposed to solve the problem of low forwarding speed of traditional IP routers. MPLS implements fast forwarding through label headers.
- An MPLS label is a short and fixed-length identifier that has only local significance. It is used to uniquely identify the FEC to which a packet belongs. Label operations on an LSR include push, swap, and pop.
- MPLS consists of the control plane and data plane. The control plane transmits routing information and distributes labels, and the data plane forwards data.
- With the development of technologies, the advantage of MPLS in the data forwarding speed is weakened gradually, but MPLS is widely used in the VPN area due to its pros.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

MPLS LDP Fundamentals and Configuration

HUAWEI

Foreword

- Multiprotocol Label Switching (MPLS) implements data forwarding based on short and fixed-length labels carried in packets.
- A fundamental concept in MPLS is that two LSRs must agree on the meaning of the labels used to forward traffic between them. The Label Distribution Protocol (LDP) can be used by an LSR to send its label binding information to other LDP-capable LSRs, helping implement correct forwarding of labeled packets.
- This course describes the principles, features, and basic configurations of LDP.

HUAWEI

- LDP mentioned in this course refers to that defined in RFC 3036 for the first time. This protocol has been replaced by RFC 5036.
- Other label distribution protocols include MP-BGP and RSVP.

Objectives

- On completion of this course, you will be able to:
  - Understand LDP's basic concepts and working mechanisms.
  - Describe the MPLS label distribution control mode, advertisement mode, and retention mode.
  - Perform basic LDP configurations.

HUAWEI

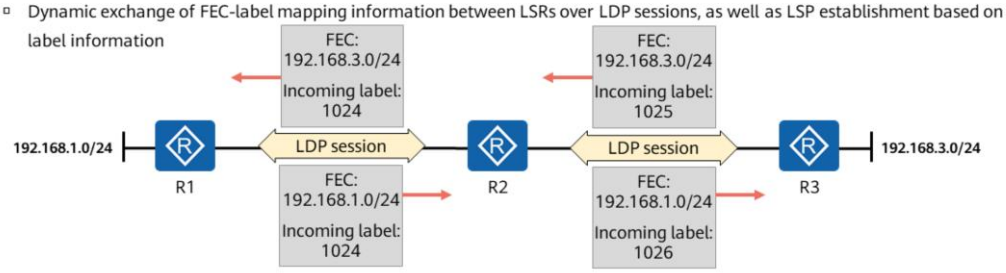
Contents

1. **Basic LDP Concepts**
2. LDP Principles
3. Basic LDP Configurations

HUAWEI

LDP Overview

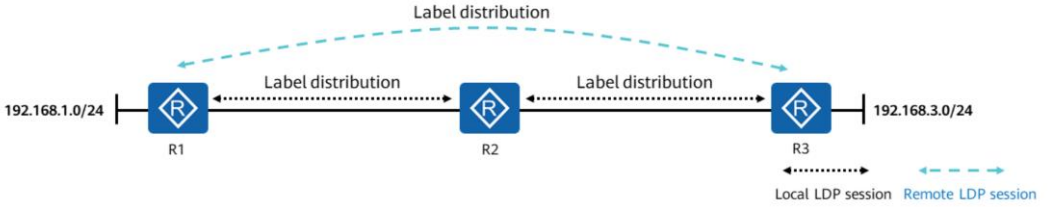
- LDP is an MPLS control protocol, which is similar to a signaling protocol on a traditional network. LDP is responsible for FEC classification, label distribution, and LSP establishment and maintenance. LDP defines the messages used in label distribution as well as the message processing procedures.
- The working process of LDP involves:
  - LDP session establishment between LSRs
  - Dynamic exchange of FEC-label mapping information between LSRs over LDP sessions, as well as LSP establishment based on label information



HUAWEI

LDP Session, Adjacency, and Peer

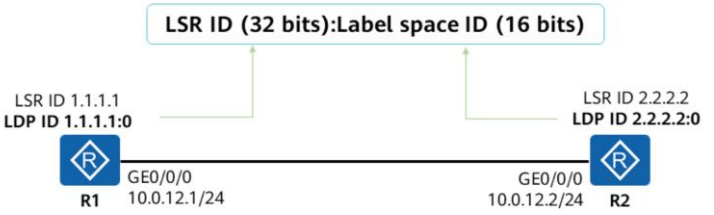
- An LDP session must be established before LSRs can exchange label binding information. LDP sessions are classified into the following types:
  - Local LDP session: can be established between two LSRs that are directly connected.
  - Remote LDP session: can be established between two LSRs that are directly or indirectly connected.
- An adjacency is established between two LSRs after they exchange Hello messages.
- After an adjacency is established between two LSRs, they exchange LDP session messages to establish an LDP session. An LDP peer relationship is then established between them.



HUAWEI

LSR ID and LDP ID

- Each LDP-capable LSR must have an LDP ID, in addition to an LSR ID.
  - An LDP ID is 48 bits long and consists of a 32-bit LSR ID and a 16-bit label space ID.
  - An LDP ID is presented in the format of "LSR ID:Label space ID", for example, 2.2.2.2:0.
- The meaning of a label space ID varies according to its value:
  - 0: indicates a device-based label space.
  - Non-zero value: indicates an interface-based label space.



HUAWEI

- This course takes the device-based label space as an example.

LDP Messages

- LDP-capable LSRs exchange LDP messages to discover peers, establish and maintain sessions, and manage labels.

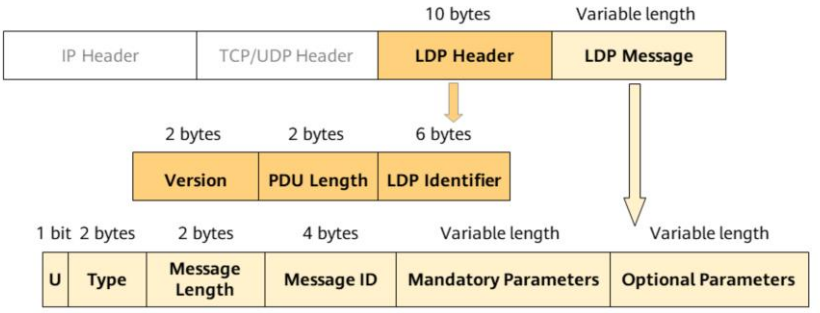
Message Type	Message Name	Transport Layer Protocol	Function
Discovery message	Hello	UDP	Advertises local LSRs and discovers peers in the LDP discovery process.
Session message	Initialization	TCP	Negotiates parameters in an LDP session establishment process.
	KeepAlive		Monitors the TCP connection integrity of LDP sessions.
Advertisement message	Address		Advertises interface addresses.
	Address Withdraw		Withdraws interface addresses.
	Label Mapping		Advertises FEC-label mapping information.
	Label Request		Requests label mappings for FECs.
	Label Abort Request		Aborts undone Label Request messages.
	Label Withdraw		Withdraws FEC-label mappings.
	Label Release		Releases labels.
Notification message	Notification		Informs LDP peers of errors.

HUAWEI

- LDP messages are classified into four types by function: discovery, session, advertisement, and notification.
  - Discovery messages: announce and maintain the presence of LSRs on a network. Hello messages belong to this category.
  - Session messages: establish, maintain, and terminate sessions between LDP peers. Initialization and KeepAlive messages belong to this category.
  - Advertisement messages: generate, change, and delete label mappings for FECs.
  - Notification messages: provide advisory information and signal error information.
- LDP messages are carried over UDP or TCP, with the port number being 646. Discovery messages, which are used to discover peers, are carried over UDP. Other LDP messages must be transmitted in a reliable and ordered manner. Therefore, LDP uses TCP to establish sessions. Session, advertisement, and notification messages are transmitted over TCP.

LDP Packet Encapsulation

- An LDP packet consists of an LDP header and an LDP message.
  - An LDP header carries information such as the LDP version and packet length.
  - An LDP message carries information such as the message type and message length.



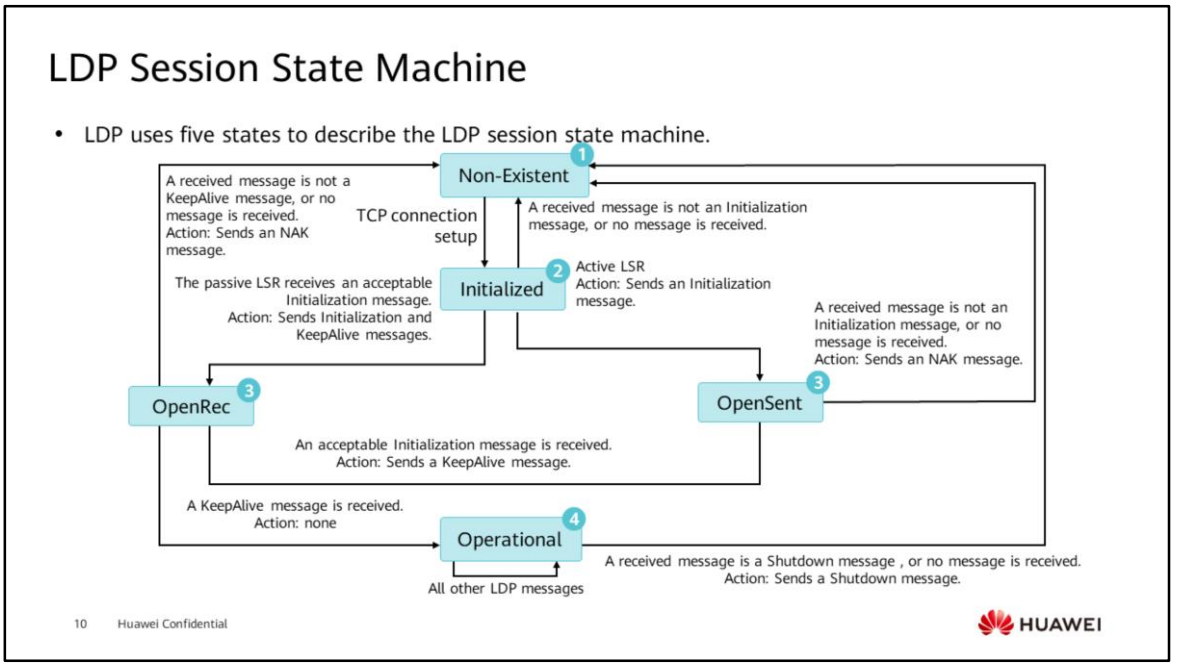
HUAWEI

- An LDP header is 10 bytes long. It consists of three parts: Version, PDU Length, and LDP Identifier.
  - The Version field occupies 2 bytes. It indicates the LDP version number. The current version number is 1.
  - The PDU Length field occupies 2 bytes. It indicates the packet length in bytes, excluding the Version and PDU Length fields.
  - The LDP Identifier field (that is, LDP ID) occupies 6 bytes. The first 4 bytes uniquely identify an LSR, and the last 2 bytes identify the label space of the LSR.
- An LDP message consists of five parts.
  - The U field occupies 1 bit, which is an unknown message. When an LSR receives an unknown message, the LSR returns a notification message to the message originator if the U field is 0, but ignores the message and does not respond with a notification message if the U field is 1.
  - Message Length occupies 2 bytes. It indicates the total length of Message ID, Mandatory Parameters, and Optional Parameters, in bytes.
  - Message ID occupies 32 bits. It identifies a message.
  - Each of the Mandatory Parameters and Optional Parameters fields has a variable length.
  - Message Type indicates a specific message type. Currently, common messages defined by LDP include Notification, Hello, Initialization, KeepAlive, Address, Address Withdraw, Label Mapping, Label Request, Label Abort Request, Label Withdraw, and Label Release.

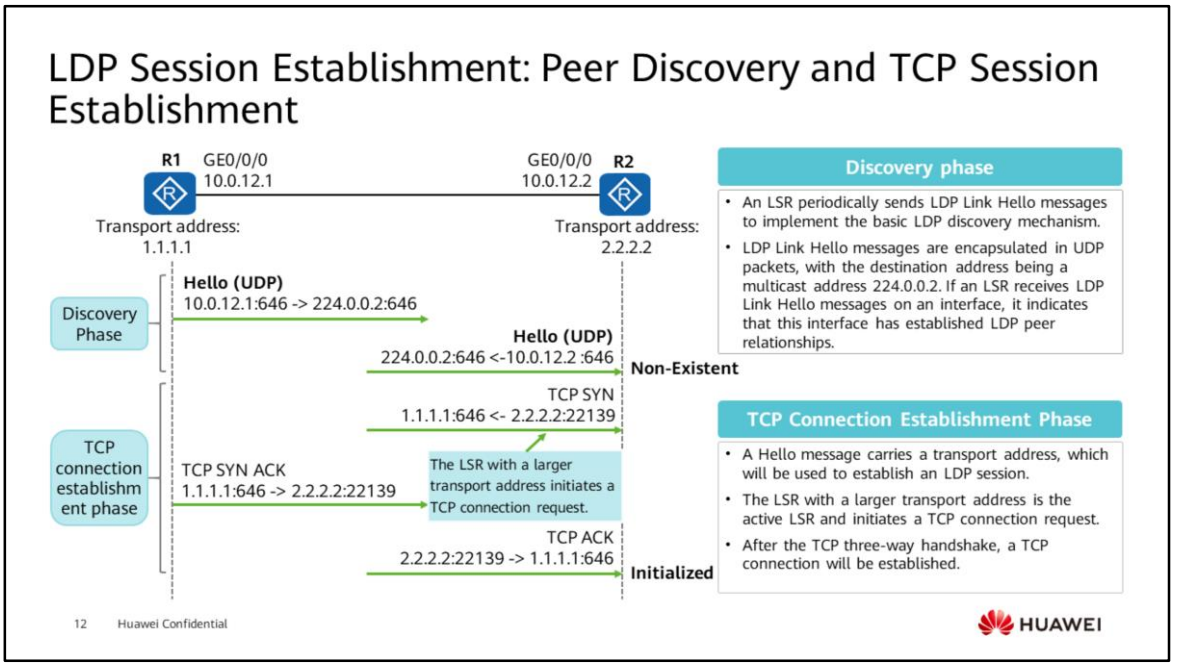
Contents

1. Basic LDP Concepts
2. **LDP Principles**
   - LDP Session Establishment
   - LDP-based Label Distribution
   - LDP Working Process
3. Basic LDP Configurations

HUAWEI

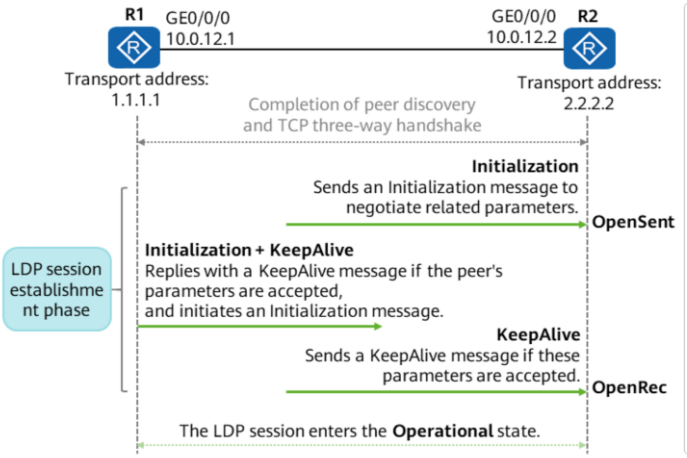


- The LDP session negotiation process can be described through the state machine. As shown in the figure, there are five states. They are Non-Existent, Initialized, OpenRec, OpenSent, and Operational.
    - Non-Existent: It is the initial state of an LDP session. In this state, both LSRs send Hello messages to elect the active LSR. After a TCP connection establishment success event is received, the state changes to Initialized.
    - Initialized: In this state, the active LSR sends an Initialization message to the passive LSR, sets the session state to OpenSent, and waits for an Initialization message. The passive LSR waits for the Initialization message sent by the active LSR. If the parameters in the received Initialization message are accepted, the passive LSR sends Initialization and KeepAlive messages, and sets the session state to OpenRec. When the active and passive LSRs receive any non-initialization message or the waiting period times out, both of them set the session state to Non-Existent.
    - OpenSent: It is a state after the active LSR sends an Initialization message. In the OpenSent state, the active LSR waits for the passive LSR to respond to the Initialization and KeepAlive messages. If the parameters in the received Initialization message are accepted, the active LSR sets the session state to OpenRec. However, if the parameters are not accepted or the Initialization message times out, the active LSR tears down the TCP connection and sets the session state to Non-Existent.



- In addition to the basic discovery mechanism, the extended discovery mechanism is supported, which can be used to discover indirectly connected remote adjacencies. For details, see RFC5036.
- LDP transport addresses are used to establish TCP connections with peers.
  - Before establishing an LDP session, two LSRs need to establish a TCP connection to exchange LDP packets.
  - A transport address of an LSR is contained in LDP Hello messages, through which an LSR can learn the transport addresses of its peers.
  - After two LSRs discover each other and learn each other's transport address through Hello messages, the LSRs attempt to perform the TCP three-way handshake (based on the transport addresses), and exchange LDP Initialization messages, Label Mapping messages, and so on. All these messages use the transport addresses of the two ends as source and destination IP addresses.
  - An LSR must have a route to the transport address of its peer.
  - By default, the transport address for a device on a public network is the LSR ID of the device, and the transport address for a device on a private network is the primary IP address of an interface on the device.
  - The **mpls ldp transport-address** command can be run in the interface view to change a transport address.

LDP Session Establishment - Session Establishment and Maintenance



- After a TCP connection is established, R2 (active LSR with a larger transport address) sends an LDP Initialization message to negotiate parameters related to LDP session establishment.
- These parameters include the LDP version, label distribution mode, KeepAlive timer value, maximum PDU length, and label space.
- After R1 receives the Initialization message, it replies with a KeepAlive message if it accepts R2's parameters. To improve transmission efficiency, R1 also sends an Initialization message.
- After R2 receives the Initialization message, it replies with a KeepAlive message if it accepts R1's parameters.
- After both ends receive each other's KeepAlive message, the session is established successfully. They periodically send KeepAlive messages to maintain the session.

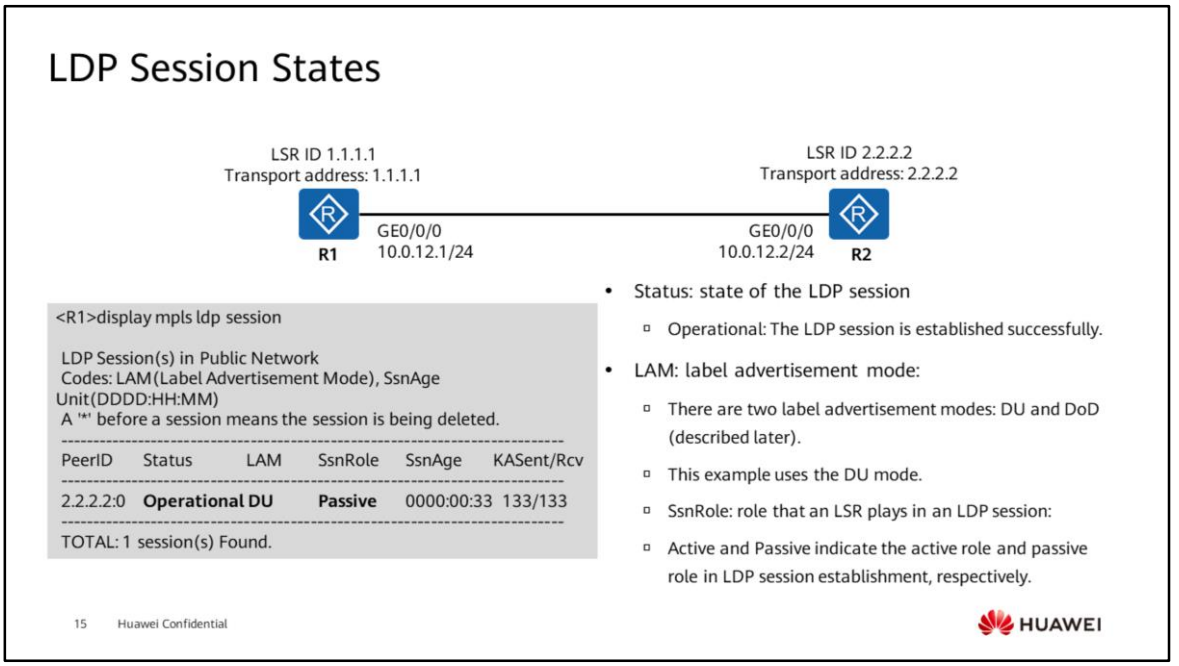
HUAWEI

LDP Peer State

```
<R1>display mpls ldp peer

LDP Peer Information in Public network
A '*' before a peer means the peer is being deleted.
------------------------------------------------------------------------------
PeerID            TransportAddress   DiscoverySource
------------------------------------------------------------------------------
2.2.2.2:0         2.2.2.2            GigabitEthernet0/0/0
------------------------------------------------------------------------------
TOTAL: 1 Peer(s) Found.
```

- PeerID: LDP ID of the peer
  - 2.2.2.2: LSR ID of the peer
  - 0: device-based label space
- TransportAddress: transport address of the peer
  - 2.2.2.2: IP address used to establish a TCP connection

HUAWEI



- LDP session states:
  - NonExistent: initial state of an LDP session. In this state, the two ends send Hello messages to each other. After the TCP connection establishment success event is triggered, the session enters the Initialized state.
  - Initialized: The LDP session is being initialized.
  - OpenSent: The active LSR sends an Initialization message to the passive LSR and waits for a reply.
  - Open Recv: LDP peers at both ends of the LDP session wait for a KeepAlive message from each other after the session enters the initialization state. If they receive each other's KeepAlive message, the LDP session enters the Operational state.
  - Operational: The LDP session is established successfully.

Contents

1. Basic LDP Concepts
2. **LDP Principles**
   - LDP Session Establishment
   - LDP-based Label Distribution
   - LDP Working Process
3. Basic LDP Configurations

HUAWEI

Label Advertisement and Management

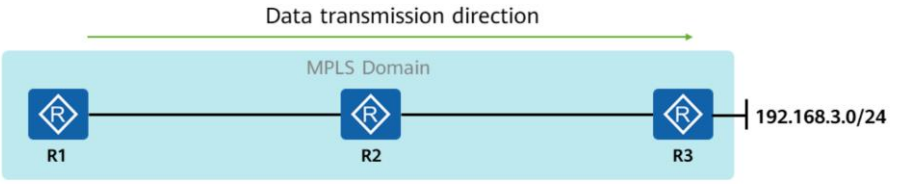
- On an MPLS network, a downstream LSR determines the bindings between labels and FECs and advertises the bindings to its upstream LSR.
- To establish LSPs, LDP sends Label Request and Label Mapping messages to advertise the bindings between labels and FECs.
- Label advertisement and management are determined by the label advertisement mode, label distribution control mode, and label retention mode.

Item	Mode	Default Mode	Description
Label advertisement mode	Downstream unsolicited (DU)	Yes	An LSR assigns and distributes labels to a FEC without receiving Label Request messages from its upstream LSR.
	Downstream on demand (DoD)	No	An LSR assigns and distributes labels to a FEC only after receiving Label Request messages from its upstream LSR.
Label distribution control mode	Independent	No	A local LSR assigns and binds a label to a FEC and then advertises the binding to the upstream LSR, without waiting for the label distributed by the downstream LSR.
	Ordered	Yes	An LSR sends the label mapping of a FEC to its upstream device only if the LSR has received Label Mapping messages from the next hop of the FEC or if the LSR is the egress of the FEC.
Label retention mode	Liberal	Yes	An LSR retains all label mappings received from a peer, regardless of whether the peer is its next hop.
	Conservative	No	An LSR retains the label mappings received from a peer only if the peer is its next hop.

HUAWEI

Upstream and Downstream

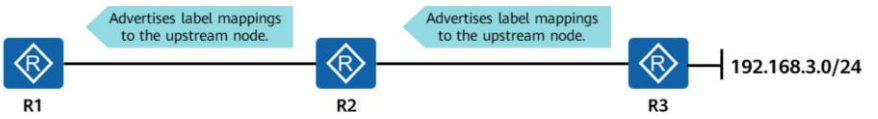
- MPLS determines the upstream and downstream relationships based on the data forwarding direction. Labeled packets are sent from an upstream LSR, and received and processed by a downstream LSR.
- As shown in the figure, for the LSP to 192.168.3.0/24, R3 is the downstream LSR of R2, and R1 is the upstream LSR of R2.



HUAWEI

Label Advertisement Mode - DU

- DU mode
  - An LSR assigns and distributes labels to a FEC without having to receive Label Request messages from its upstream LSR.
  - An LSR actively advertises the labels of a FEC to its upstream peer without having to receive Label Request messages from the peer.

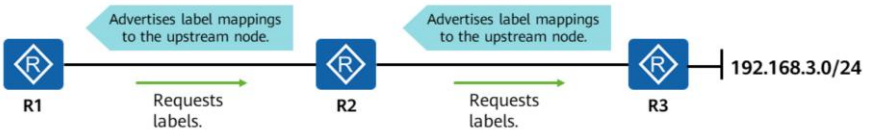


HUAWEI

- Label assignment: An LSR assigns a label from the local label space and binds it with a FEC.
- Label distribution: An LSR notifies the upstream LSR of the binding between labels and FECs.
- When the DU label advertisement mode is used, an LSR can assign labels to all its peers by default. Specifically, each LSR can distribute label mappings to all its peers, regardless of whether the LSR is an upstream or a downstream one. If an LSR distributes labels only to upstream peers, it must identify its upstream and downstream nodes based on routing information before sending Label Mapping messages. An upstream node cannot send Label Mapping messages to its downstream node.

Label Advertisement Mode - DoD

- DoD mode
  - An LSR assigns and distributes labels to a FEC only after receiving Label Request messages from its upstream LSR.
  - Generally, a Label Request message is triggered when an access request to a particular FEC arises.



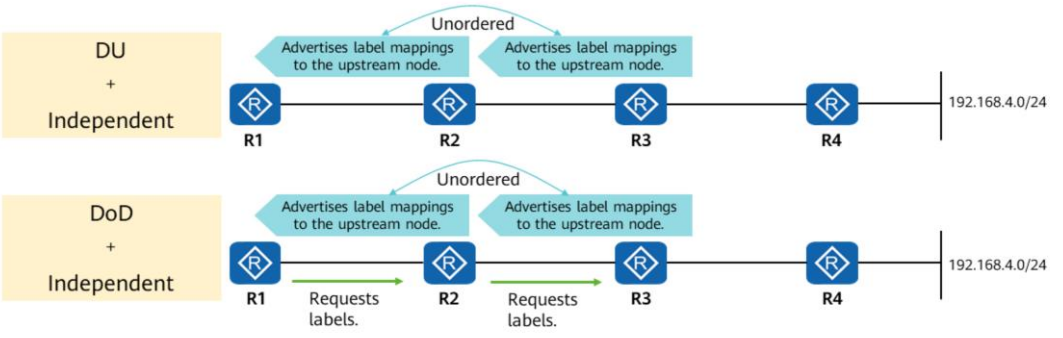
R1 requires to access the 192.168.3.0/24 network segment.

HUAWEI

- An LSR advertises label mappings to an upstream peer only after receiving Label Request messages from the upstream peer.

Label Distribution Control Mode - Independent

- Independent mode
  - A local LSR assigns and binds a label to a FEC and then advertises the binding to the upstream LSR, without waiting for the label distributed by the downstream LSR.

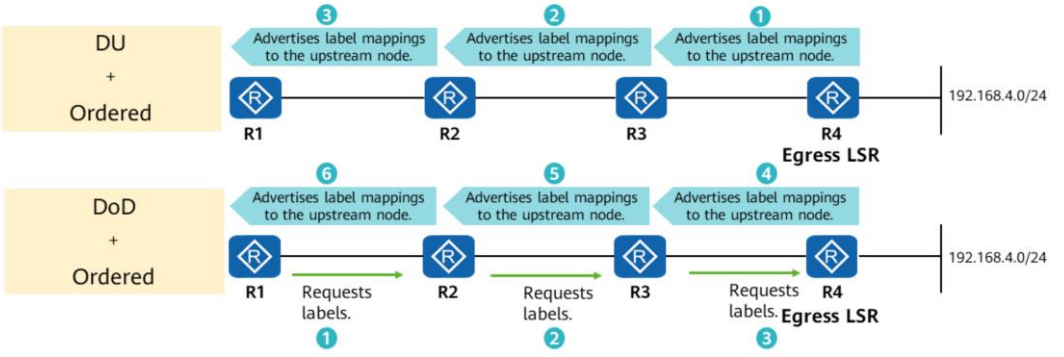


HUAWEI

- The label distribution control mode works with the label advertisement mode:
  - If the network shown in the figure uses the DU label advertisement mode, R2 and R3 can actively notify the upstream LSR of the label binding for the FEC 192.168.4.0/24 even if the upstream LSR does not send Label Request messages and R2 and R3 do not receive label binding information from the downstream LSR.
  - If the network uses the DoD label advertisement mode, R2 and R3 can notify the upstream LSR of the label binding for the FEC 192.168.4.0/24 given that R2 and R3 have received Label Request messages from the upstream LSR, regardless of whether R2 and R3 have received label binding information from the downstream LSR.

Label Distribution Control Mode - Ordered

- Ordered mode
  - An LSR sends the label mapping of a FEC to its upstream device only if the LSR has received Label Mapping messages from the downstream of the FEC or if the LSR is the egress of the FEC.

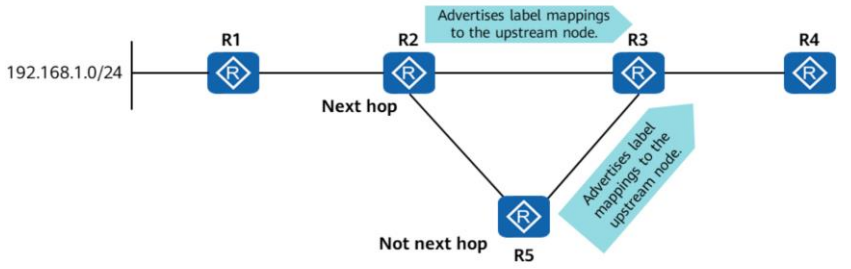


HUAWEI

- In ordered label distribution control mode, an LSR can send a Label Mapping message to its upstream node only when the LSR receives Label Mapping messages of a FEC from the downstream of the FEC or when the LSR is the egress of an LSP.
  - If the network shown in the figure uses the DU label advertisement mode, an LSR sends the label binding information of the FEC 192.168.4.0/24 to its upstream node only after the LSR receives the label binding information of the FEC from its downstream node, even if the upstream node has sent Label Request messages. Therefore, the initiator for LSP establishment must be an egress LSR (R4 in this example).
  - If the network uses the DoD label advertisement mode, an LSR advertises the label binding information of the FEC 192.168.4.0/24 to the upstream node only after the LSR receives Label Request messages from the upstream node as well as the label binding information of the FEC from the downstream node. Therefore, a Label Request message can be initiated by the ingress LSR (R1) only. After a Label Request is sent hop by hop to the egress LSR (R4), R4 advertises a Label Mapping message to the upstream LSR to establish an LSP.

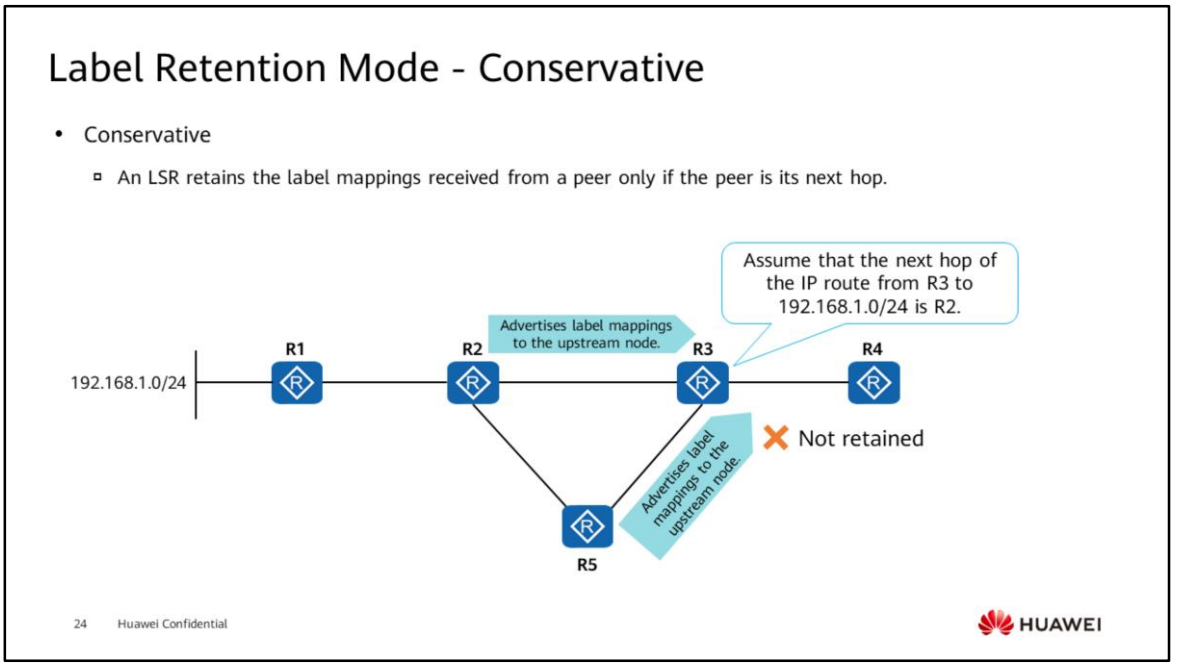
Label Retention Mode - Liberal

- Liberal mode
  - An LSR can receive label mappings from its next hop or non-next hop nodes.
  - An LSR retains all label mappings received from a peer, regardless of whether the peer is its next hop.



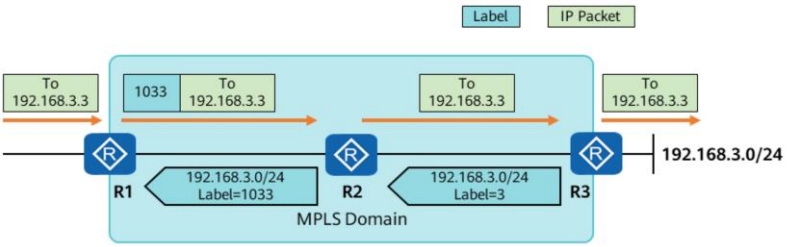
HUAWEI

- If MPLS is deployed on an IP network, an LSR uses the IP routing table to determine whether a label mapping is received from the next hop.
- In liberal mode, a new LSP can be quickly established when routes change, because all received labels are retained, which is the biggest advantage of this mode. The disadvantage is that unnecessary label mappings are distributed and maintained.
  - In DU label advertisement mode, if the liberal label retention mode is used, R3 retains the labels of the FEC 192.168.1.0/24 sent by all LDP peers (R2 and R5 in this example), regardless of whether R2 and R5 are the next hops of the routes to 192.168.1.0/24 in the IP routing table.
  - In DoD label advertisement mode, if the liberal label retention mode is used, an LSR requests labels from all LDP peers. However, the DoD label advertisement mode is generally used together with the conservative label retention mode.



- The advantage of the conservative mode is that only the labels that will be used to forward data are retained and maintained, thereby saving the label space.
  - In DU label advertisement mode, an LSR may receive Label Mapping messages for the same network segment (FEC) from multiple LDP peers. As shown in the figure, R3 receives Label Mapping messages for the network segment 192.168.1.0/24 from both R2 and R5. If the conservative label retention mode is used, R3 retains only the label sent by the next hop R2 and discards the label sent by the non-next hop R5.
  - In DoD label advertisement mode, an LSR uses routing information to determine its next hop and requests labels only from the next hop.
- If the next hop of a FEC changes, either of the following situations occurs:
  - In liberal label retention mode, the LSR can use an existing label advertised by a non-next hop LSR to quickly establish an LSP. The liberal mode requires more memory and label space.
  - In conservative label retention mode, the LSR retains the labels advertised by the next hop only. This mode saves memory and label space but consumes more time to reestablish the LSP.
  - An LSR that has a limited label space usually uses the conservative mode and DoD mode.

PHP



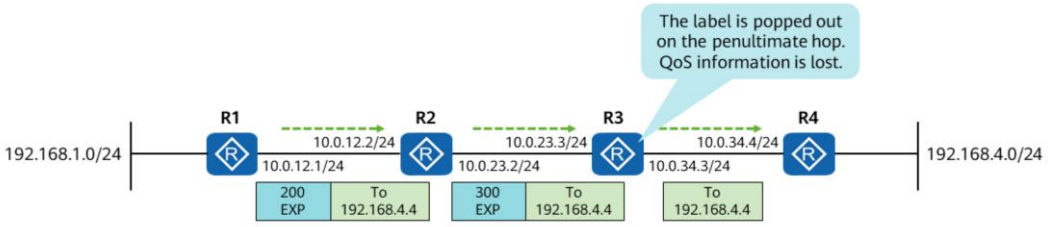
- Penultimate hop popping (PHP): If PHP is enabled, an egress assigns a special label (3), to a local route. This label is called an implicit null label. When an LSR forwards a labeled packet and finds that the outgoing label value is 3, the LSR removes the top label from the packet and forwards the inner data to the downstream LSR.

HUAWEI

- During label advertisement, R3 is the egress of the FEC 192.168.3.0/24. During label distribution, R3 assigns label 3 to the FEC and advertises the label binding information to R2.
- During data forwarding, R2, as the penultimate hop to 192.168.3.0, finds that the outgoing label value is 3. Then, R2 removes the label header and forwards the IP packet to R3. R3 only needs to query the FIB once to obtain the corresponding forwarding information, improving the forwarding efficiency.

Implicit Null Label and Explicit Null Label (1)

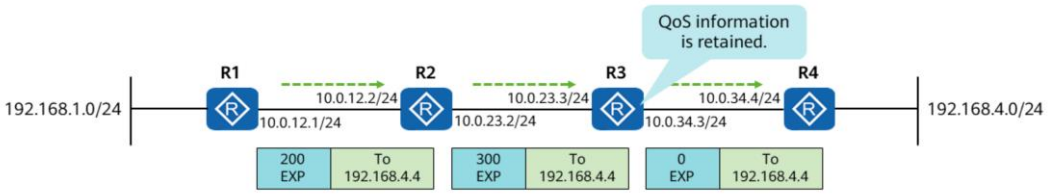
- By default, an egress assigns implicit null labels, that is, label 3, to the penultimate hop.
- However, if QoS is deployed, after the label is popped out, the priority in the label is lost.



HUAWEI

Implicit Null Label and Explicit Null Label (2)

- In the explicit null label mechanism, an egress assigns label 0 to the penultimate hop.
- When R3 forwards a labeled packet of which the outgoing label is 0, R3 does not pop out the label header, and therefore QoS information is retained. When R4 receives a packet with label 0, it directly pops out the label without searching for an ILM entry.
- By default, an egress assigns implicit null labels. You can run the label advertise explicit-null command to enable the egress to assign explicit null labels to the penultimate hop.



HUAWEI

- Run the **label advertise** { *explicit-null* | *implicit-null* | *non-null* } command in the MPLS view to configure the label to be assigned to the penultimate hop.
- You can specify one of the following parameters:
  - **implicit-null**: is the default value. If this parameter is set, an egress assigns an implicit null label with the value of 3 to the penultimate hop.
  - **explicit-null**: If this parameter is set, an egress assigns an explicit null label with the value of 0 to the penultimate hop. The explicit-null parameter can be set if MPLS QoS attributes need to be used.
  - **non-null**: If this parameter is set, an egress assigns a common label with a value greater than or equal to 16 to the penultimate hop.

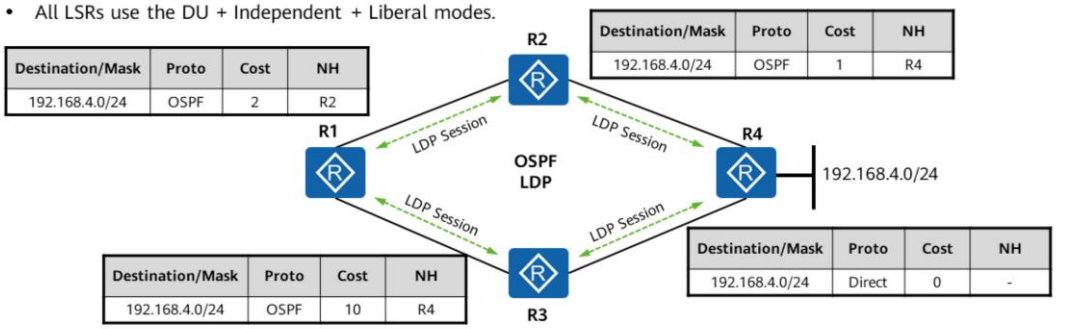
Contents

1. Basic LDP Concepts
2. **LDP Principles**
   - LDP Session Establishment
   - LDP-based Label Distribution
   - LDP Working Process
3. Basic LDP Configurations

HUAWEI

Networking Overview

- OSPF has been deployed on the network, and devices can learn routes from each other.
- MPLS and LDP have been enabled on devices and interfaces, and local LDP sessions have been established between neighboring devices.
- All LSRs use the DU + Independent + Liberal modes.

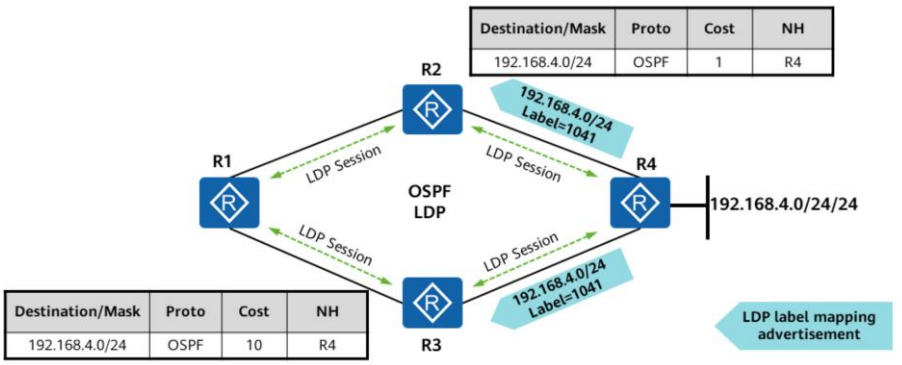


HUAWEI

- Currently, Huawei devices use the DU + Ordered + Liberal modes by default.
- For a packet that enters the MPLS domain from R1 and is destined for 192.168.4.0/24, R1 is the ingress LSR, and R4 is the egress LSR.

Label Distribution - Egress LSR

- R4 is directly connected to the network segment 192.168.4.0/24. R4 actively assigns labels, such as 1041, to routes destined for this network segment and advertises label mapping information to its LDP peers (R2 and R3) through LDP packets.



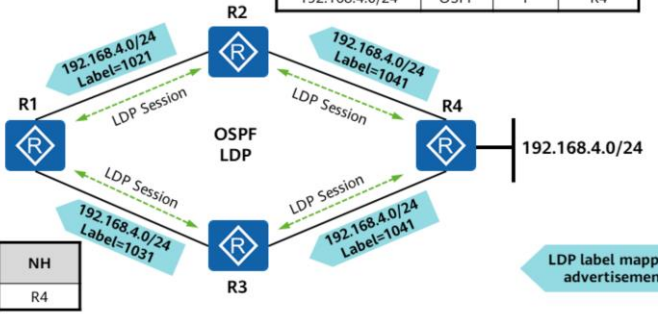
HUAWEI

- Note: By default, 32-bit host IP routes are used to trigger LSP establishment. You can manually trigger the establishment of an LSP with non-32-bit host IP routes.

Label Distribution - Transit LSR

- Take R2 as an example. In its routing table, the next hop of the route 192.168.4.0/24 is R4. When R2 receives a Label Mapping message for the route 192.168.4.0/24 from R4, R2 assigns label 1021 to the route because the message is sent by a downstream LDP peer, and advertises the label mapping to the LDP peer, for example, R1. This process also applies to R3.

Destination/Mask	Proto	Cost	NH
192.168.4.0/24	OSPF	1	R4



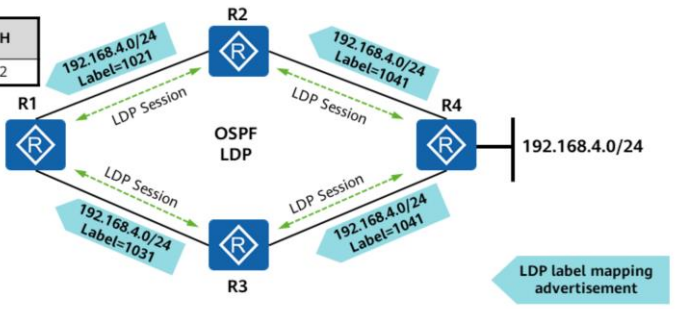
Destination/Mask	Proto	Cost	NH
192.168.4.0/24	OSPF	10	R4

HUAWEI

Label Distribution - Ingress LSR

- After R1 receives the label mappings for the route 192.168.4.0/24 advertised by R2 and R3, R1 stores both of the label mappings. However, R1 only uses label 1021 advertised by R2, because the next hop of the route to 192.168.4.0/24 is R2, as shown in R1's routing table.

Destination/Mask	Proto	Cost	NH
192.168.4.0/24	OSPF	2	R2



HUAWEI

- Note: If R2 fails, OSPF routes re-converge. The next hop of the route 192.168.4.0/24 in the routing table of R1 is switched to R3. In this case, R1 uses the label advertised by R3 for 192.168.4.0/24.

Label-based Forwarding - Ingress LSR

- As an ingress LSR, R1 pushes a label into each received IP packet, and forwards packets based on labels.

Routing table of R1

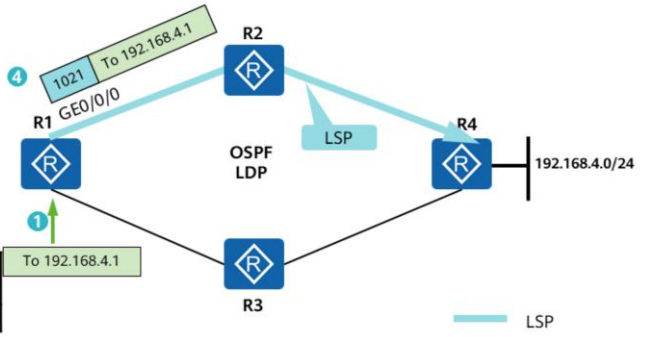
Destination/Mask	Proto	Cost	NH
192.168.4.0/24	OSPF	2	R2

FIB of R1 ②

Destination/Mask	Proto	Cost	NH	Tunnel ID
192.168.4.0/24	OSPF	2	R2	0x12

NHLFE table of R1 ③

Tunnel ID	Out intf	OPER	NH	Out Label
0x12	GE0/0/0	push	R2	1021



HUAWEI

- When R1 receives an IP packet destined for 192.168.4.1, it searches the FIB for a forwarding entry matching the destination IP address of the packet, and finds that the tunnel ID in the matching entry is not 0. As such, R1 continues to search for an NHLFE matching the tunnel ID, pushes a label to the IP packet, and forwards the packet. The outbound interface is GE 0/0/0, the next hop is R2, and the outgoing label is 1021. Therefore, R1 adds a label header to the packet and forwards the packet.

Label-based Forwarding - Transit LSR

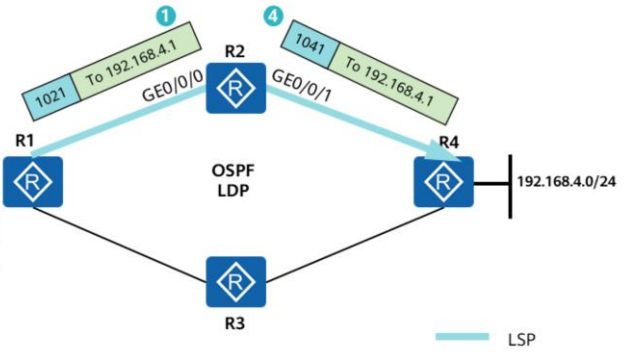
- As a transit LSR, R2 needs to swap labels in received IP packets and forward the packets.

ILM table of R2 ②

In Label	Tunnel ID
1021	0x12

NHLFE table of R2 ③

Tunnel ID	Out intf	OPER	NH	Out Label
0x12	GE0/0/1	Swap	R4	1041



HUAWEI

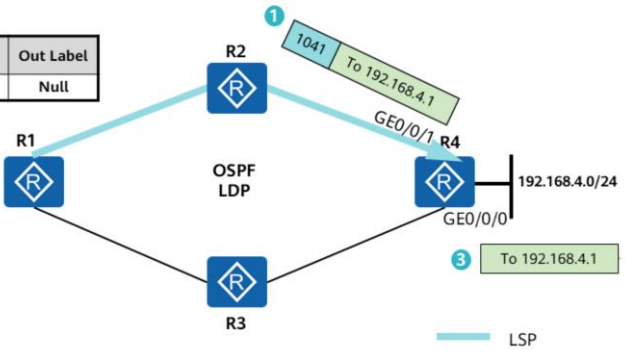
- When R2 receives a packet with label 1021, it searches for a matching ILM entry and an NHLFE matching the ILM entry. Then, R2 changes the label of the packet to 1041 and forwards the packet through the matching outbound interface.

Label-based Forwarding - Egress LSR

- As an egress LSR, R4 needs to perform the pop operation on received IP packets to remove labels and forward the packets through IP.

ILM table of R4 ②

In Label	Tunnel ID	Out intf	OPER	NH	Out Label
1041	0x12	GE0/0/0	Pop	-	Null



HUAWEI

- When R4 receives a packet with label 1041, it searches for a matching ILM entry and finds that the operation type is pop. R4 then performs a pop operation to remove the outer label from the packet. The packet then becomes a standard IP packet, and therefore R4 performs the standard IP forwarding on the packet.
- When R4 forwards the packet, it searches the LFIB and FIB. How can the forwarding efficiency be improved on the egress LSR (R4)?

Summary of an LDP-Capable LSR's Operations on an MPLS Network

- An LSR runs an IGP (such as OSPF or IS-IS) to construct a routing table and FIB.
- LDP assigns labels to route prefixes (FECs) in the routing table based on the label assignment mode it uses.
- LDP advertises the labels assigned to the route prefixes to LDP peers through LDP Label Mapping messages based on the label advertisement mode it uses.
- An LSR stores the labels that it assigns to route prefixes and the labels that LDP peers advertise to the route prefixes, and associates the labels with information such as outbound interfaces and next-hop addresses (label forwarding entries).
- When an LSR forwards a labeled packet destined, the LSR always uses the outgoing label advertised by the downstream LDP peer. The downstream peer is the next-hop device to the destination network in the routing table.

HUAWEI

Contents

1. Basic LDP Concepts
2. LDP Principles
3. **Basic LDP Configurations**

HUAWEI

Basic LDP Configuration Commands (1)

1. Enable LDP.

   [Huawei] **mpls ldp**

   The **mpls ldp** command enables LDP and displays the LDP view.

   [Huawei-GigabitEthernet0/0/0] **mpls ldp**

   Enable LDP on an interface. Before running this command, enable LDP globally.

2. Configure a remote LDP peer.

   [Huawei] **mpls ldp remote-peer** *remote-peer-name*

   The **mpls ldp remote-peer** command creates a remote peer and displays the remote peer view.

   [Huawei-mpls-ldp-remote-PeerName] **remote-ip** *ip-address*

   Set the **remote-ip** *ip-address* parameter to the IP address of a remote LDP peer.

HUAWEI

Basic LDP Configuration Commands (2)

3. Configure a policy for triggering LSP establishment.

[Huawei-mpls] **lsp-trigger** { **all** | **host** | **ip-prefix** *ip-prefix-name* | **none** }

The **lsp-trigger** command configures the routes (**static and IGP routes**) that are used to trigger LSP establishment, which are IP routes with a 32-bit mask by default.

- **all**: All static and IGP routes are used to trigger LSP establishment. If this parameter is set, a significant of LSPs will be established, consuming label resources excessively and slowing down network-wide LSP convergence. Therefore, setting this parameter is not recommended.
- **host**: IP routes with a 32-bit mask are used to trigger LSP establishment.
- **ip-prefix** *ip-prefix-name*: Routes matching a specified IP prefix list are used to trigger LSP establishment.
- **none**: LSP establishment cannot be triggered.

4. Configure a label advertisement mode.

[Huawei-GigabitEthernet0/0/0] **mpls ldp advertisement** { **dod** | **du** }

The default label advertisement mode is downstream unsolicited (DU).

- If the label advertisement mode is DU, the label retention mode is liberal.
- If the label advertisement mode is DoD, the label retention mode is conservative.

HUAWEI

- BGP routes can also be used to trigger LDP LSP establishment. This trigger policy is not covered in this course.

Basic LDP Configuration Commands (3)

5. Configure an LDP label distribution control mode.

   [Huawei-mpls-ldp] **label distribution control-mode** { **independent** | **ordered** }

   The default LDP label distribution control mode is ordered.

6. Configure PHP.

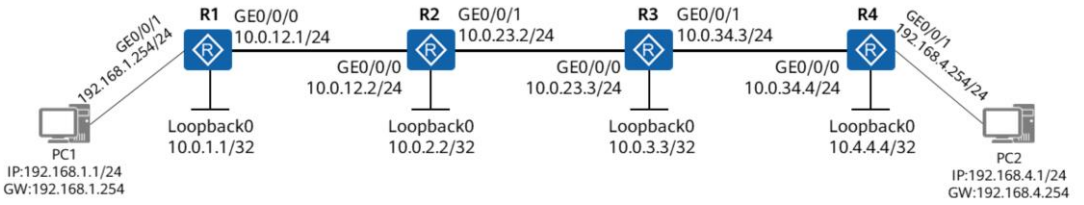
   [Huawei-mpls] **label advertise** { **explicit-null** | **implicit-null** | **non-null** }

   By default, an egress distributes implicit null labels to a penultimate hop.

   - **explicit-null**: An egress assigns explicit null label to the penultimate hop.
   - **implicit-null**: An egress assigns implicit null labels to the penultimate hop.
   - **non-null**: An egress assigns common labels to the penultimate hop.

HUAWEI

Configuration Examples

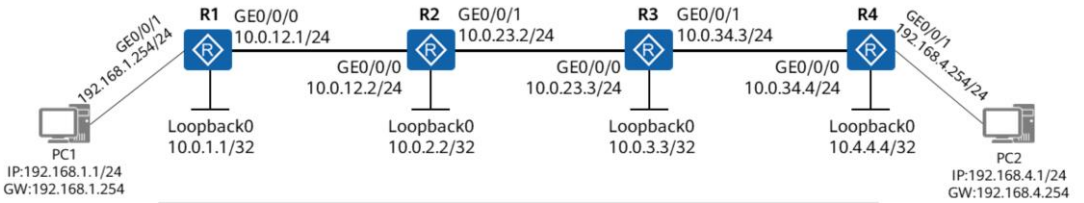


Scenario: R1, R2, R3, and R4 run an IGP to implement IP interworking.

Requirement: Configure MPLS and LDP to implement mutual access between the network segments 192.168.1.0/24 and 192.168.4.0/24 through MPLS forwarding.

HUAWEI

Configuration Procedure (1)

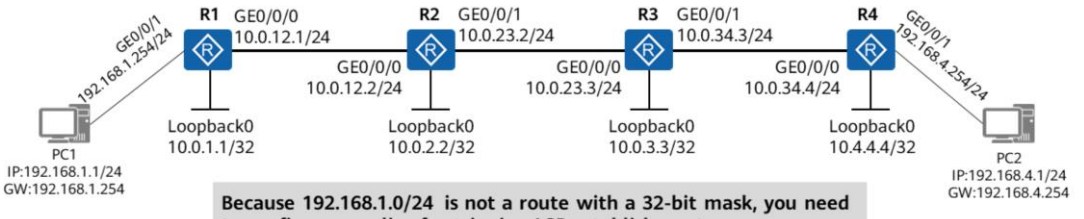


Enable basic MPLS and LDP functions on the devices. The following example uses R1.

```
[R1]mpls
[R1-mpls]quit
[R1]mpls ldp
[R1-mpls-ldp]quit
[R1]interface GigabitEthernet 0/0/0
[R1-GigabitEthernet0/0/0]mpls
[R1-GigabitEthernet0/0/0]mpls ldp
[R1-GigabitEthernet0/0/0]quit
```


HUAWEI

Configuration Procedure (2)



Because 192.168.1.0/24 is not a route with a 32-bit mask, you need to configure a policy for trigging LSP establishment.

```
[R1]ip ip-prefix ldp permit 192.168.1.0 24
[R1-mpls]lsp-trigger ip-prefix ldp
```

Perform similar configuration on R4.

```
[R4]ip ip-prefix ldp permit 192.168.4.0 24
[R4-mpls]lsp-trigger ip-prefix ldp
```


HUAWEI

Checking the Configuration - Checking LSP Information

Check information about the LDP LSPs created on R1.
[R1]display mpls ldp lsp
LDP LSP Information

DestAddress/Mask	In/OutLabel	UpstreamPeer	NextHop	OutInterface
10.0.2.2/32	1024/3	10.0.2.2	10.0.12.2	GE0/0/0
10.0.3.3/32	1025/1025	10.0.2.2	10.0.12.2	GE0/0/0
192.168.1.0/24	3/NULL	10.0.2.2	192.168.1.254	GE0/0/1
*192.168.1.0/24	Liberal/1027		DS/10.0.2.2	
192.168.4.0/24	1027/1028	10.0.2.2	10.0.12.2	GE0/0/0

HUAWEI

Quiz

1. (Single-answer question) Which of the following commands can be used to display a label distributed for a specific FEC? ( )
   A. display mpls ldp
   B. display mpls ldp interface
   C. display mpls lsp
   D. display mpls ldp session
2. (Single-answer question) What is the default combination of label advertisement mode, label distribution control mode, and label retention mode on Huawei devices? ( )
   A. DU + Independent + Conservative
   B. DU + Ordered+ Liberal
   C. DoD + Independent+ Liberal
   D. DoD + Ordered + Conservative

HUAWEI

1. C
2. B

Summary

- MPLS supports multiple label distribution protocols, among which LDP is widely used.
- LDP is a process in which LSRs negotiate the meaning of labels. LDP uses discovery, session, advertisement, and notification packets to establish sessions and distribute labels.
- LDP determines label advertisement and management based on the label advertisement mode, label distribution control mode, and label retention mode. By default, Huawei datacom devices use the DU label advertisement mode + ordered label distribution control mode + liberal label retention mode.
- LDP can directly map network-layer routing information to label information in order to establish LSPs. LSRs are connected according to the incoming label, next hop, and outgoing label corresponding to a specified FEC in the local forwarding table. In this manner, the LSP crossing the entire MPLS domain can be formed.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

MPLS VPN Fundamentals and Configuration

HUAWEI

Foreword

- Virtual private network (VPN) is a technology that enables VPN users to communicate with each other over a public network.
- MPLS VPN is a Layer 3 VPN (L3VPN) technology widely used in the industry. Note that in this course, MPLS VPN refers to BGP/MPLS IP VPN.
- This course describes the basic concepts, working process, and typical configuration methods of MPLS VPN.

HUAWEI

- Unless otherwise specified, MPLS VPN refers to BGP/MPLS IP VPN.

Objectives

- On completion of this course, you will be able to:
  - Describe the MPLS VPN model.
  - Understand basic concepts of MPLS VPN.
  - Describe the MPLS VPN route transmission and label distribution process.
  - Describe the MPLS VPN data forwarding process.
  - Perform basic MPLS VPN configurations.

HUAWEI

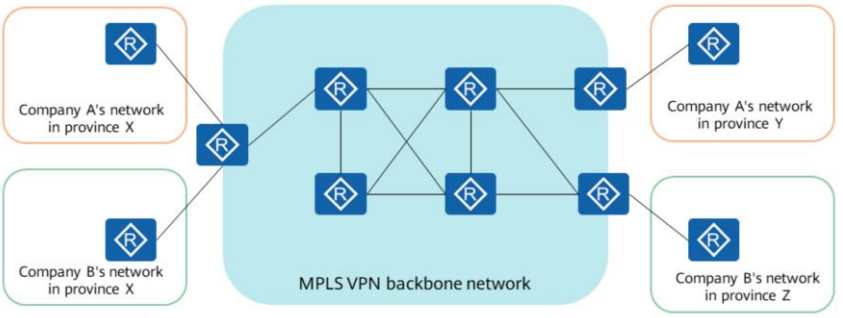
Contents

1. **MPLS VPN Overview**
2. MPLS VPN Route Exchange
3. MPLS VPN Packet Forwarding
4. MPLS VPN Configuration and Implementation

HUAWEI

MPLS VPN Definition

- BGP/MPLS IP VPNs are usually constructed by a carrier to provide VPN users with VPN services, and thus implementing route transmission and data forwarding between user networks.
- MPLS VPN uses BGP to advertise VPN routes and uses MPLS to forward VPN packets on the carrier's backbone network (IP network). BGP/MPLS IP VPN, short to MPLS VPN, is a common L3VPN technology.

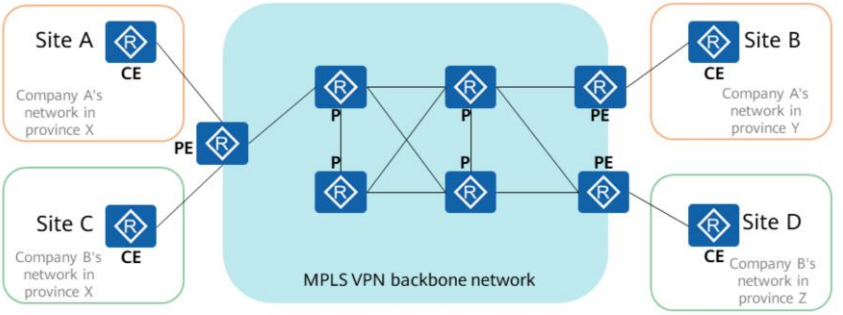


HUAWEI

- MPLS VPN backbone networks can also be constructed by enterprises themselves, with technical implementation similarly to that of carriers. This course focuses on the scenario where enterprises purchase MPLS VPN services from carriers.

Network Architecture of MPLS VPN

- A MPLS VPN consists of three parts: customer edge (CE), provider edge (PE), and provider (P). The PE and P are carrier devices, whereas the CE is a MPLS VPN user device.
- A site represents a MPLS VPN user, consisting of CEs and other user devices.

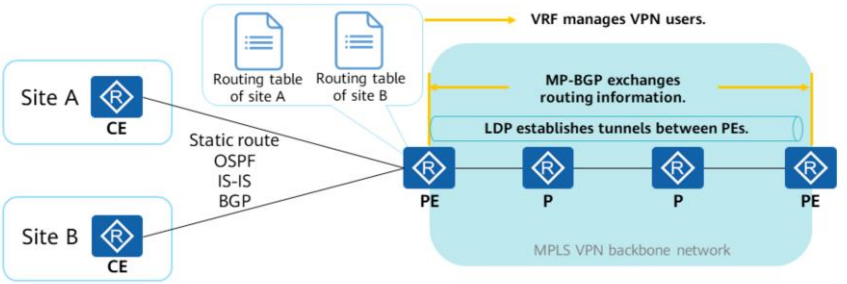


HUAWEI

- CE: an edge device on a user network. A CE provides interfaces that are directly connected to a carrier network. A CE can be a router, switch, or host. CEs are usually unaware VPNs and do not need to support MPLS.
- PE: an edge device on a carrier network and directly connected to a CE. On an MPLS network, PEs process all VPN services, and therefore PEs must have high performance.
- P: a backbone router on a carrier network and not directly connected to a CE. P devices need only to provide basic MPLS forwarding capabilities and do not maintain VPN information.
- The meaning of a site can be understood from the following aspects:
  - A site is a group of IP systems that can communicate without using carrier networks.
  - Sites are classified based on topological relationships between devices rather than the geographical locations of devices. In the preceding figure, the networks in provinces X and Y of company A need to communicate through the backbone network of the carrier. Therefore, the two networks are considered as two sites. If a physical private line is added between the CEs on the networks of provinces X and Y, the two networks can communicate without the need of the carrier network. In this case, the two networks are considered as one site.
- Relationship between sites and VPNs:
  - Sites connected to the same service provider network can be classified into different collections based on configured policies. Only sites that belong to the same collection can access each other, and this collection is defined as a VPN.
  - Devices at a site can belong to multiple VPNs. In other words, a site can belong to more than one VPN.

Technical Architecture of MPLS VPN

- MPLS VPN is not a single VPN technology. Rather, it is a comprehensive solution that combines multiple technologies, including:
  - MP-BGP: transmits site routing information between PEs.
  - LDP: establishes tunnels between PEs.
  - VRF: manages VPN users on PEs.
  - Static route, IGP, or BGP: exchanges routing information between PEs and CEs.



HUAWEI

- Multiprotocol Extensions for BGP (MP-BGP): an extended BGP protocol that supports multiple address families. For details, see related courses.

Why MPLS VPN?

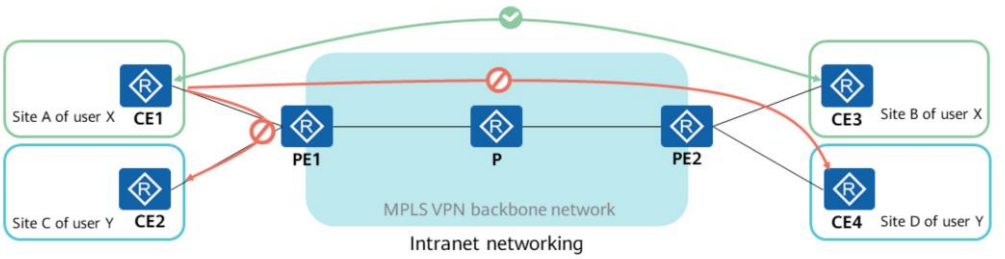
- For VPN users:
  - Unaware of VPN existence and therefore do not need to deploy and maintain VPNs, simplifying enterprise O&M and reducing O&M costs.
  - Usually deployed on a carrier's dedicated network for MPLS VPN and therefore enjoy a certain degree of security.
- For carriers:
  - MPLS adds a connection-oriented control plane to a connectionless IP network, which facilitates IP network management and operations.
  - Supports overlapping address spaces and overlapping VPNs, enabling flexible networking and good scalability.
  - Easily supports MPLS TE to properly adjust and control existing network resources, minimizing carriers' CAPEX.

HUAWEI

- MPLS traffic engineering (MPLS TE): steers traffic to constrained LSPs for forwarding so that the traffic is transmitted along specified paths. MPLS TE fully uses network resources and provides bandwidth and QoS guarantee without the need for hardware upgrades. It minimizes network costs.

Common MPLS VPN Networking Schemes

- The following common networking schemes are available to meet different requirements of VPN users:
  - Intranet: All users on a VPN are in a group isolated from users of other VPNs. Users at the same VPN site can communicate with each other, but users in different VPN sites cannot.
  - Extranet: applies to the scenario where a VPN user wants to provide some resources of the local site for users of other VPNs.
  - Hub&Spoke: can be used if a carrier wants to use a central access control device on a VPN to control the mutual access of users.



Intranet networking

HUAWEI

- Intranet networking is the simplest and most typical MPLS VPN networking scheme. The following technical implementation of MPLS VPN will be described based on this networking scheme.

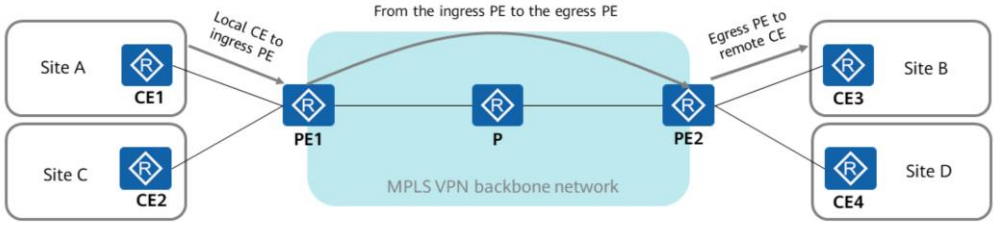
Contents

1. MPLS VPN Overview
2. **MPLS VPN Route Exchange**
3. MPLS VPN Packet Forwarding
4. MPLS VPN Configuration and Implementation

HUAWEI

MPLS VPN Route Advertisement

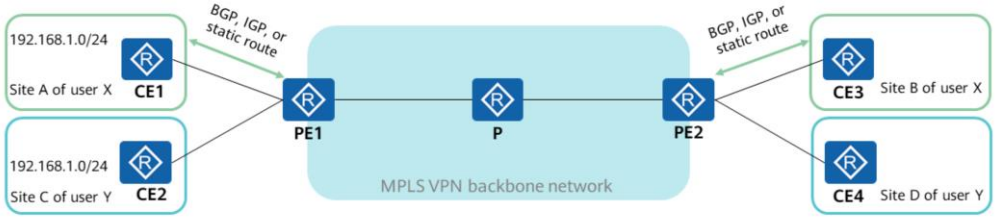
- Route exchange must be completed between different sites to allow different sites on the same VPN to communicate with each other. On a basic MPLS VPN, only CEs and PEs are involved in VPN route advertisement. P routers only need to maintain the routes of the backbone network, without the need to know any VPN routes. VPN route advertisement involves the following three phases:
  - From the local CE to the ingress PE
  - From the ingress PE to the egress PE
  - From the egress PE to the remote CE



HUAWEI

Route Exchange Between CEs and PEs

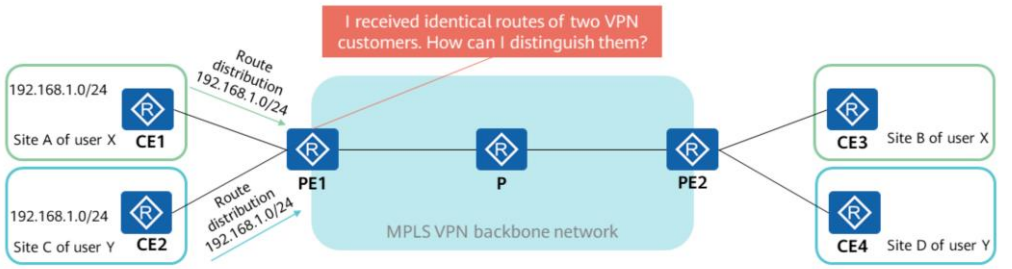
- As shown in the figure, customers X and Y belong to different VPNs, and each of them has two sites. Routes need to be exchanged between the two sites (belong to the same VPN).
- The CEs and PEs can use static routes, OSPF, IS-IS, or BGP to exchange routes. No matter which routing protocol is used, CEs and PEs exchange standard IPv4 routes.
- The local CEs exchange routes with the ingress PE in the same way as the egress PE exchange routes with the remote CEs.



HUAWEI

Route Distribution from the Ingress PE to the Egress PE (1)

- After receiving routes from CEs, PEs need to independently store routes of different VPNs and solve the problem where different customers use overlapping IP address spaces.

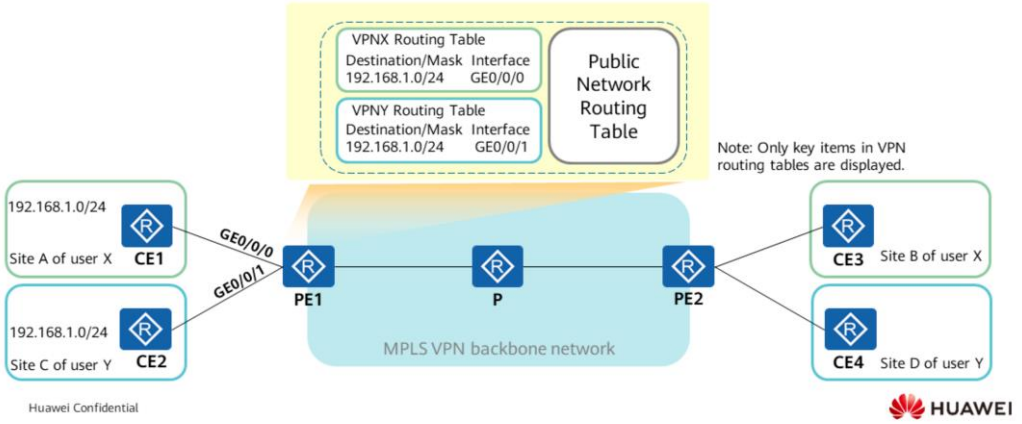


HUAWEI

- A VPN is a private network. Different VPNs independently manage their own address ranges, which are also called address spaces. Address spaces of different VPNs may overlap. For example, in the preceding figure, both user X and user Y use 192.168.1.0/24, indicating that the address spaces overlap. VPNs can use overlapping address spaces in the following situations:
  - They do not share the same site.
  - They share a same site, but devices at the site do not communicate with devices using overlapping address spaces at the other sites of the VPNs.

VRF

- Virtual routing and forwarding (VRF), also called VPN instance, is a key technology in the MPLS VPN architecture. Each VRF uses independent routing and forwarding entries to logically isolate VPNs.



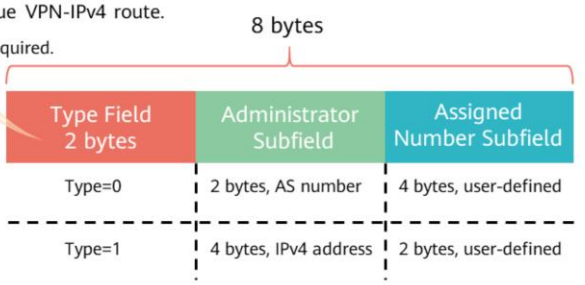
HUAWEI

- For details about VRFs, see the related HCIP-Datacom-Core course.

RD

- After a PE receives IPv4 address prefixes from CEs in different VPNs, the PE differentiates these address prefixes based on VPN instance configurations. However, VPN instances are only locally significant. A PE cannot transmit VPN instance information to the peer PE. The route distinguisher (RD) is introduced to resolve this issue.
  - An RD is 8 bytes long and used to distinguish IPv4 prefixes with the same address space.
  - After a PE receives an IPv4 route from a CE, the PE adds an RD to the IPv4 prefix of the route and converts the route to a globally unique VPN-IPv4 route.

No configuration is required.

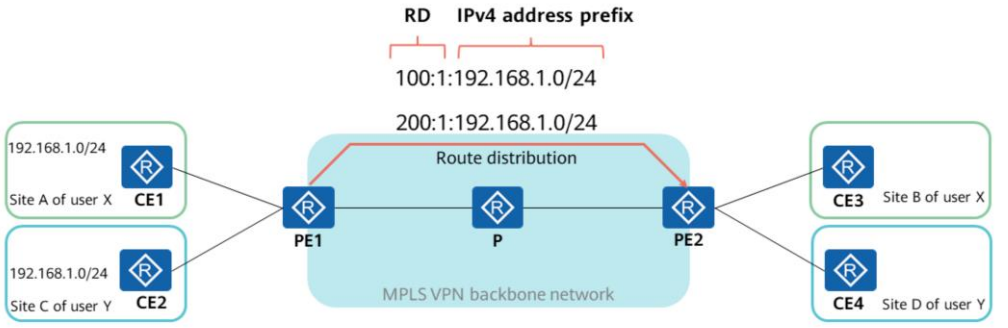


HUAWEI

- When configuring an RD, you need to specify only the Administrator and Assigned Number subfields in the RD.
- Four types of RD configuration formats are available. The following two types are commonly used:
  - 16-bit AS number:32-bit user-defined number (for example 100:1)
  - 32-bit IPv4 address:16-bit user-defined number (for example, 172.1.1.1:1)
- The RD structure enables each carrier to allocate RDs independently. In some application scenarios, however, RDs must be globally unique to ensure normal routing.

VPN-IPv4 Address

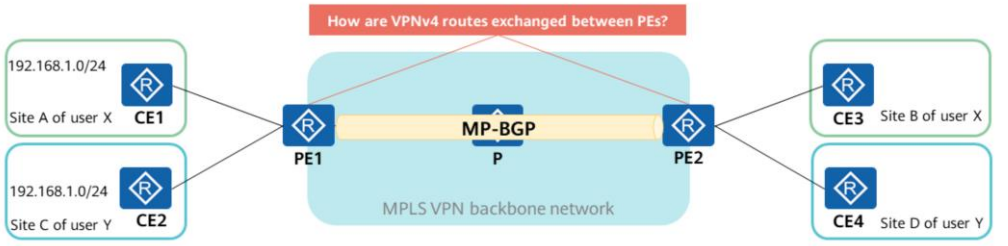
- A VPN-IPv4 address — also called a VPNv4 address — consists of 12 bytes, including an 8-byte RD and a 4-byte IPv4 address prefix.



HUAWEI

Route Distribution from the Ingress PE to the Egress PE (2)

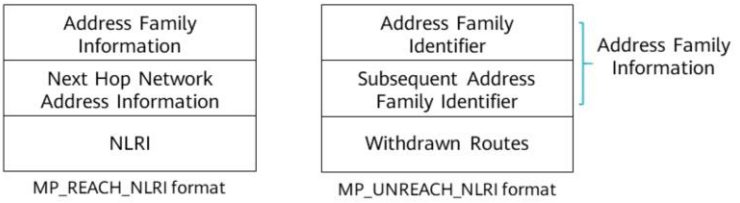
- PEs establish BGP peer relationships and use BGP to exchange routes with each other. Why is BGP used?
  - BGP uses TCP as its transport layer protocol to improve reliability. It enables PEs that are not directly connected to directly exchange routes.
  - BGP is easy to expand, facilitating VPN route distribution between PEs.
  - The number of routes to be exchanged between PEs may be large. BGP sends only updated routes, enabling more routes to be exchanged while conserving link bandwidth resources.
- Traditional BGP-4 cannot process VPNv4 routes.



HUAWEI

MP-BGP

- MPLS VPN uses MP-BGP defined in RFC 2858 (Multiprotocol Extensions for BGP-4) to correctly process VPN routes.
- MP-BGP uses address families to differentiate network layer protocols. An address family can be a traditional IPv4 address family or any other address family, such as a VPN-IPv4 address family or an IPv6 address family.
- The following two path attributes are added for MP-BGP:
  - MP_REACH_NLRI: Multiprotocol Reachable NLRI, used to advertise reachable routes and next hop information.
  - MP_UNREACH_NLRI: Multiprotocol Unreachable NLRI used to withdraw unreachable routes.

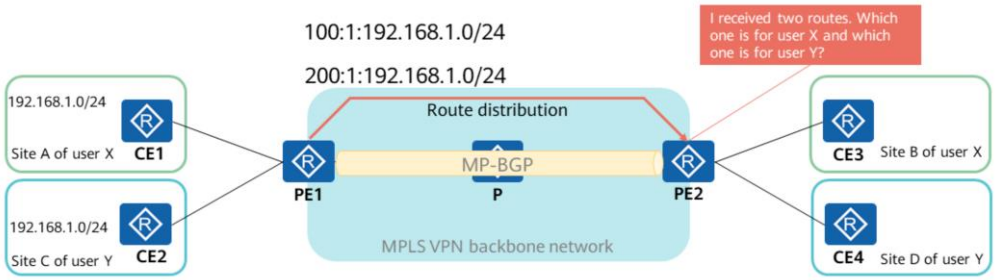


HUAWEI

- NLRI: Network Layer Reachability Information
- For the values of address families, see RFC 3232 (Assigned Numbers).
- MP_REACH_NLRI is used to advertise reachable routes and next hop information. It consists of one or more 3-tuples <Address Family Information, Next Hop Network Address Information, NLRI>.
  - Address Family Information: consists of a 2-byte Address Family Identifier (AFI) and a 1-byte Subsequent Address Family Identifier (SAFI).
    - The AFI identifies the network layer protocol, corresponding to the address family value defined by "Address Family Number" in RFC 3232. For example, 1 indicates IPv4 and 2 indicates IPv6.
    - The SAFI indicates the NLRI type. If the AFI value is 1 and the SAFI value is 128, the address in the NLRI is an MPLS-labeled VPN-IPv4 address.
  - Next Hop Network Address Information: consists of the 1-byte length of the next hop network address and the variable-length next hop network address.
  - NLRI: consists of one or more 3-tuples <length, label, prefix>. This part will be described in detail in the following slides.
- MP_UNREACH_NLRI is used to instruct a peer to delete unreachable routes. The format of this attribute is as follows:
  - AFI: same as that in the MP_REACH_NLRI attribute
  - SAFI: NLRI type, same as that in the MP_REACH_NLRI attribute
  - Withdrawn Routes: an unreachable route list, consisting of one or more NLRI fields A BGP speaker can withdraw a route by adding the NLRI same as that in a previously advertised reachable route to the Withdrawn Routes field.

Route Distribution from the Ingress PE to the Egress PE (3)

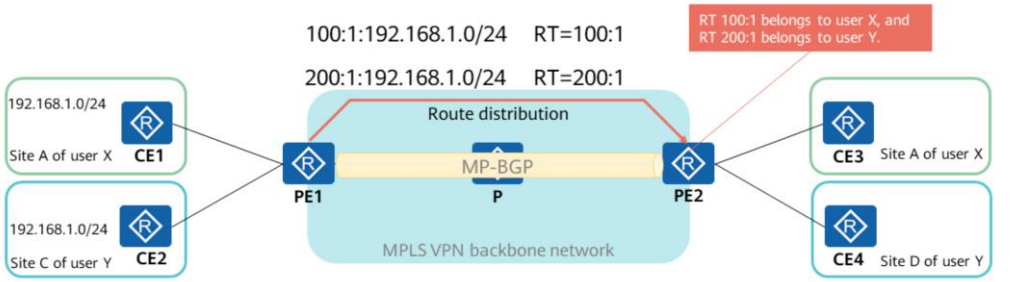
- After MP-BGP distributes VPNv4 routes to the remote PE, the remote PE needs to import the VPNv4 routes to the correct VPN instances.
- MPLS VPN uses a BGP extended community attribute — VPN target, or route target — to control the export and import of VPN routes.
- The local PE adds RTs to its VPNv4 routes before advertising the routes. After receiving the VPNv4 routes, the remote PE imports the routes to the corresponding VPN instances according to the RTs.



HUAWEI

RT

- On a PE, each VPN instance is bound to one or more VPN targets. There are two types of VPN targets:
  - Export target (ERT): After learning IPv4 routes from directly connected sites, the local PE converts the routes to VPN-IPv4 routes and adds the ERT attribute to these routes. The ERT attribute, as an extended community attribute of BGP, is distributed with these routes.
  - Import target (IRT): After receiving a VPN-IPv4 route distributed by another PE, the PE checks the ERT attribute of the route. If the ERT is identical with the IRT of a VPN instance on the PE, the PE adds the route to the routing table of the VPN instance.

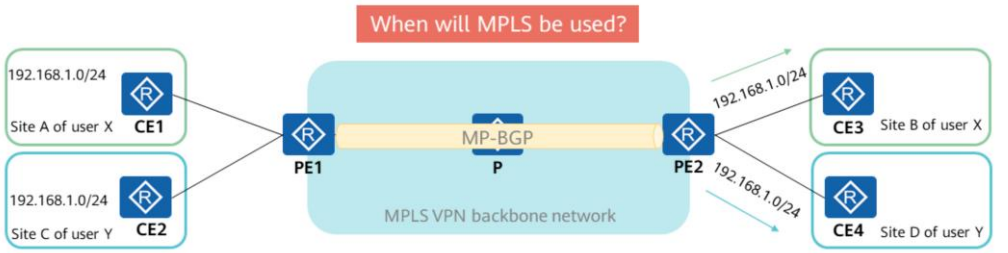


HUAWEI

- Similar to an RD, an RT consists of three fields: Type, Administrator, and Assigned Number. The length of an RT is also 8 bytes.
- When configuring a VPN target, you need to specify only the Administrator and Assigned Number subfields in the VPN target. VPN targets have the same configuration formats as RDs.

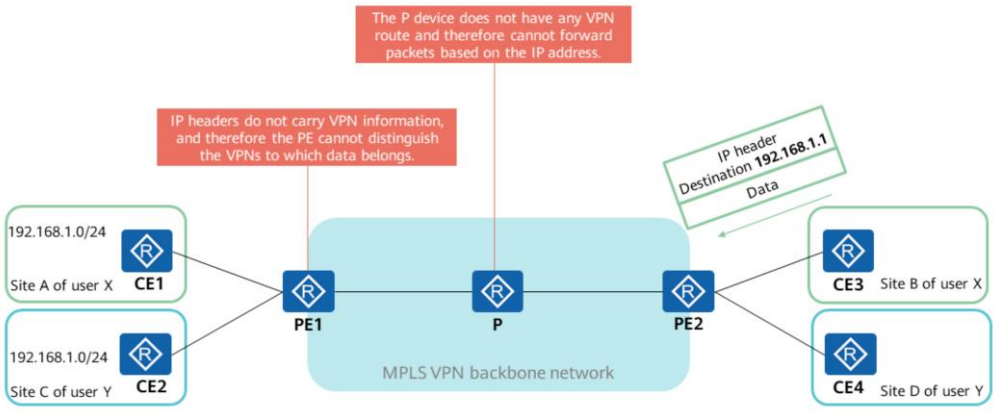
Route Distribution from the Ingress PE to the Egress PE (4)

- After the egress PE imports a VPNv4 route into a correct VPN instance based on the RT carried in the VPNv4 route, the PE removes the RD from the VPNv4 route and distributes the IPv4 route to the corresponding CE.
- The CEs at site B and site D can learn the routes destined for their respective remote sites. Different sites of the same user (in the same VPN) can communicate with each other through similar operations.



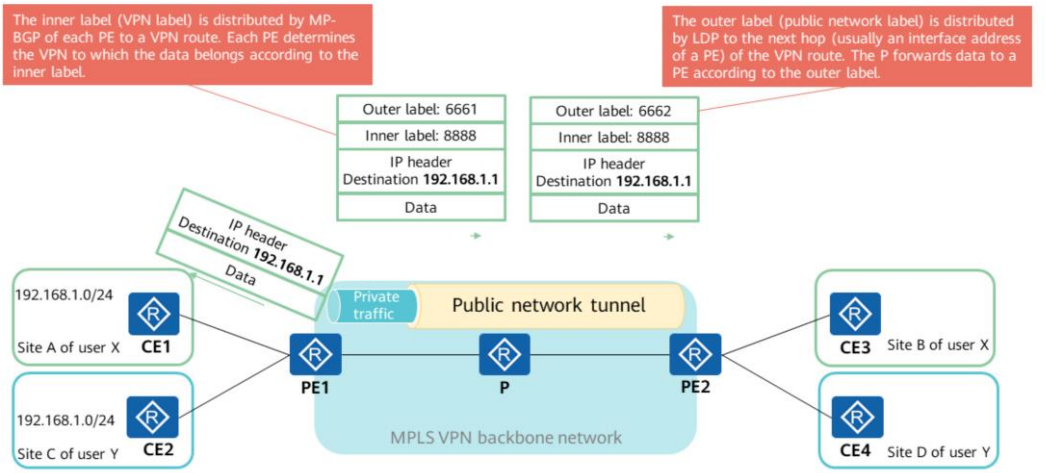
HUAWEI

Problems Encountered During Data Forwarding



HUAWEI

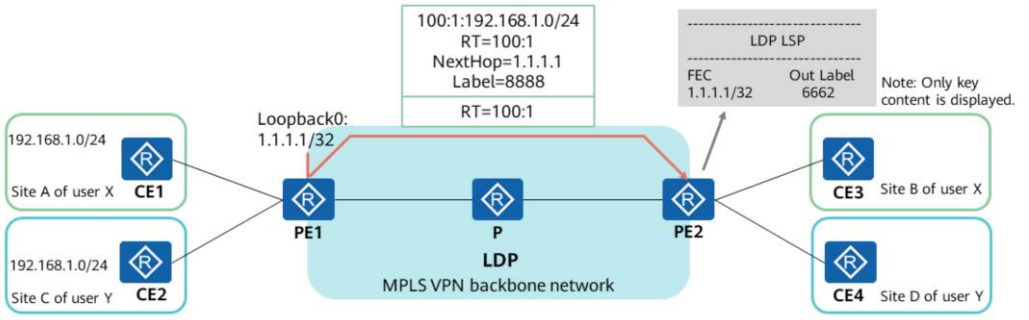
Solving Problems Through Labels



HUAWEI

Route Distribution from the Ingress PE to the Egress PE (5)

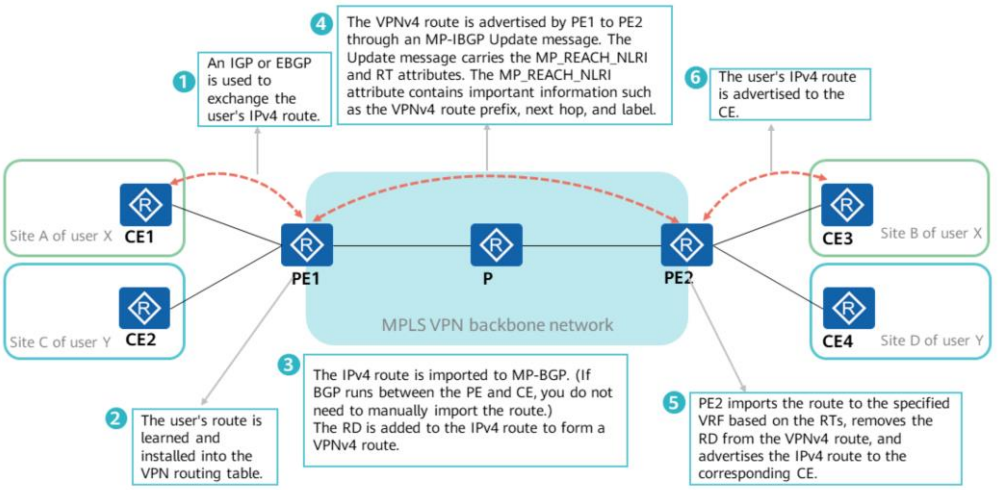
- The PEs and P distribute public network labels through LDP to establish LSPs (public network tunnels) between PEs.
- When transmitting VPNv4 routes through MP-BGP, the ingress PE carries VPN labels to differentiate data of different VPNs.
- After receiving VPNv4 routes, the egress PE performs VPN route leaking and tunnel recursion to select routes.



HUAWEI

- A PE device distributes MPLS labels in either of the following ways:
  - One label per route: Each route in a VRF is assigned one label. When many routes exist on the network, the Incoming Label Map (ILM) maintains these entries, requiring high router capacity.
  - One label per instance: Each VPN instance is assigned one label. All the routes of a VPN instance share the same label, reducing the number of labels required.
- VPN route leaking: a process of matching VPNv4 routes against the VPN targets of local VPN instances. After a PE receives a VPNv4 route, the PE directly matches the route against the VPN targets of local VPN instances, without selecting the optimal route or checking whether a desired tunnel exists.
- Tunnel recursion: A public network tunnel is required to transmit VPN traffic from one PE to the other PE over the public network. After VPN route leaking, the route must be successfully recursed to an LSP based on the destination IPv4 prefix before the route is added to the routing table of the corresponding VPN instance. This means that the next hop of the IPv4 route must match an LSP.

Route Exchange Process on a MPLS VPN



HUAWEI

Section Summary

- On a MPLS VPN, VPN routes need to be transmitted between PEs and CEs and between PEs.
  - PEs and CEs can use BGP, IGP, or static routes to exchange IPv4 routes.
  - PEs use MP-BGP to exchange VPNv4 routing information, including:
    - RD: forms a VPNv4 prefix together with an IPv4 prefix.
    - RT: controls route import and export between PEs.
    - Label: inner label (VPN label), used to differentiate data of different VPNs on a PE during data forwarding.

HUAWEI

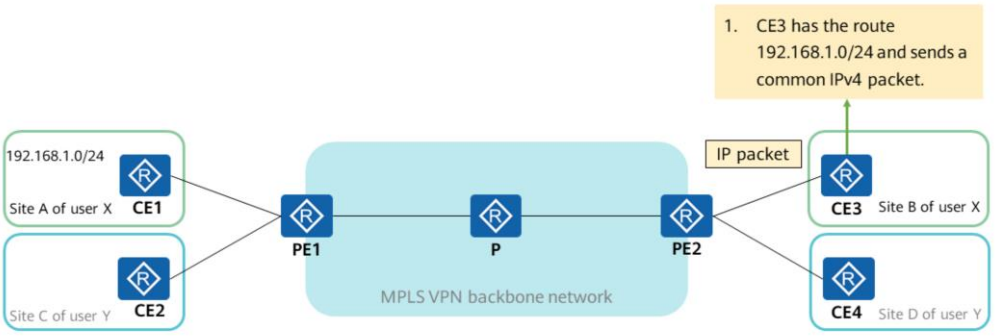
Contents

1. MPLS VPN Overview
2. MPLS VPN Route Exchange
3. **MPLS VPN Packet Forwarding**
4. MPLS VPN Configuration and Implementation

HUAWEI

Packet Forwarding Process (1)

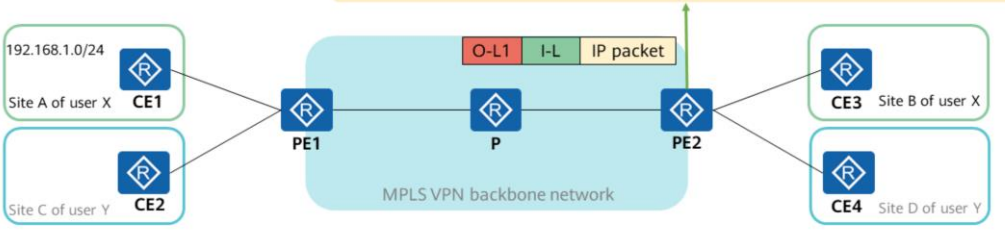
- The following example assumes that site B of user X wants to access 192.168.1.0/24 of site A. The packet forwarding process is as follows:



HUAWEI

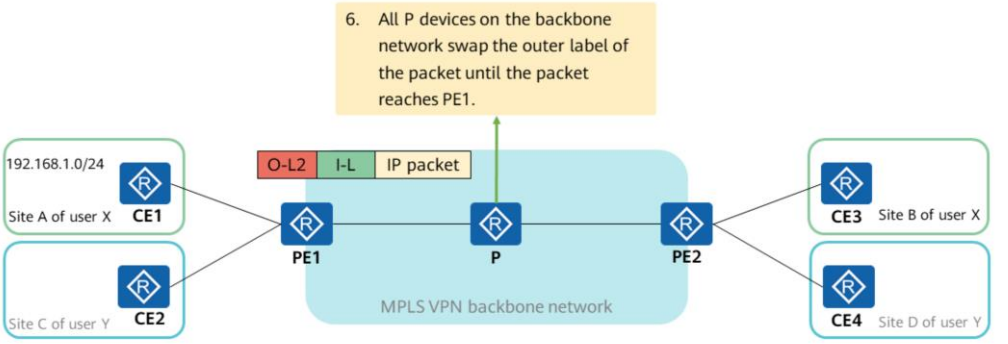
Packet Forwarding Process (2)

2. PE2 finds the VPN instance based on the inbound interface of the packet and searches the forwarding table of the VPN instance.
3. PE2 searches for the corresponding tunnel ID against the packet's destination IPv4 prefix.
4. PE2 finds the tunnel based on the tunnel ID and adds an inner label (I-L) to the packet.
5. PE2 adds an outer MPLS label (O-L1) to the packet and sends the packet out through the MPLS tunnel.



HUAWEI

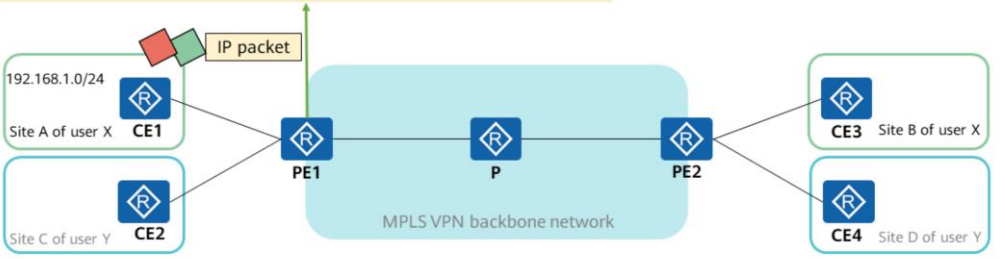
Packet Forwarding Process (3)



HUAWEI

Packet Forwarding Process (4)

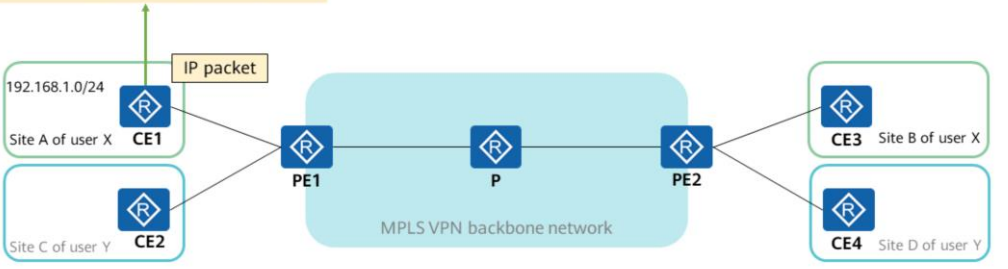
7. After receiving the packet, PE1 sends the packet to the MPLS module for processing. The MPLS module removes the outer label.
8. PE1 continues to process the inner label. Specifically, it determines the next hop based on the inner label, removes the inner label, and sends the native IPv4 packet to CE1.



HUAWEI

Packet Forwarding Process (5)

9. After receiving the IPv4 packet, CE1 performs regular IPv4 processing.



HUAWEI

Contents

1. MPLS VPN Overview
2. MPLS VPN Route Exchange
3. MPLS VPN Packet Forwarding
4. **MPLS VPN Configuration and Implementation**

HUAWEI

Configuration Commands - VPN Instance Configuration (1)

1. Create a VPN instance or enter the VPN instance view.

   [PE] **ip vpn-instance** *vpn-instance-name*

   By default, no VPN instance is configured.
2. Enable the VPN instance IPv4 address family or enter the VPN instance IPv4 address family view.

   [PE-vpn-instance-InstanceName] **ipv4-family**

   By default, the VPN instance IPv4 address family is disabled.
3. Configure an RD for the VPN instance address family.

   [PE-vpn-instance-InstanceName] **route-distinguisher** *route-distinguisher*

   The common formats of RDs are as follows:
   - 2-byte AS number:4-byte user-defined number, for example, 100:1.
   - IPv4 address:2-byte user-defined number, for example, 192.168.122.15:1.

   The RD value must be globally unique regardless of the format.

   An RD configured for a VPN instance address family cannot be modified or deleted. To modify the RD, you need to first disable the corresponding VPN instance address family, or delete the VPN instance and then reconfigure one.

HUAWEI

Configuration Commands - VPN Instance Configuration (2)

4. Configure VPN targets for the VPN instance.

   [PE-vpn-instance-InstanceName] **vpn-target** *vpn-target*&<1-8> [**both** | **export-extcommunity** | **import-extcommunity**]

   The **vpn-target** command configures export or import VPN targets for the VPN instance address family.

   - The format of a VPN target is the same as that of an RD.
   - A maximum of eight VPN targets can be configured at a time using the **vpn-target** command. To configure more VPN targets, run the **vpn-target** command multiple times.

5. Bind an interface to the VPN instance.

   [PE-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 belongs to the root instance and is not bound to any VPN instance. Binding an interface to a VPN instance or deleting the binding will result in the deletion of the IP address of the interface, Layer 3 features, and IP-related routing protocols. These features must be re-configured if needed.

HUAWEI

Configuration Commands - MP-BGP Configuration

1. Configure basic BGP functions.

   [PE] **bgp** { *as-number-plain* | *as-number-dot* }

   [PE-bgp] **peer** *ipv4-address* **as-number** *as-number*

   [PE-bgp] **peer** *ipv4-address* **connect-interface loopback** *interface-number*

   A PE must use a loopback interface address with a 32-bit mask to set up an MP-IBGP peer relationship with the peer PE so that VPN routes can be recursed to tunnels.

2. Enable the PE to exchange VPNv4 routes with a specific MP-BGP peer.

   [PE-bgp] **ipv4-family vpnv4** [**unicast**]

   [PE-bgp-af-vpnv4] **peer** *ipv4-address* **enable**

   By default, only the peer relationships in the BGP IPv4 unicast address family view are automatically enabled. In other words, after the **peer as-number** command is run in the BGP view, the system automatically configures the **peer enable** command. In other address family views, however, peering must be enabled manually.

3. Configure the PE to filter VPNv4 routes

   [PE-bgp-af-vpnv4] **policy vpn-target**

   The **policy vpn-target** command enables the PE to filter received VPN routes based on VPN targets. By default, this function is enabled. In some specific networking scenarios, you need to manually disable the filtering function.

HUAWEI

Configuration Commands - Route Configuration Between a PE and CE

1. Use EBGP to transmit routes between a PE and a CE.

   [PE-bgp] **ipv4-family vpn-instance** *vpn-instance-name*

   [PE-bgp-InstanceName] **peer** *ipv4-address* **as-number** *as-number*

   On the PE, you need to enter the VPN instance IPv4 address family view and configure the CE as a BGP peer.

   On the CE, common EBGP is configured, and VPN routes are imported to BGP using the import or network mode.

2. Use an IGP to transmit routes between the PE and CE. (OSPF is used as an example).

   [PE] **ospf** *process-id* [**router-id** *router-id*] **vpn-instance** *vpn-instance-name*

   [PE-ospf-processid] **import-route bgp** [**permit-ibgp**] [**cost** *cost* | **route-policy** *route-policy-name* | **tag** *tag* | **type** *type*] *

   [PE-bgp] **ipv4-family vpn-instance** *vpn-instance-name*

   [PE-bgp] **import-route ospf** *process-id* [**med** *med* | **route-policy** *route-policy-name*] *

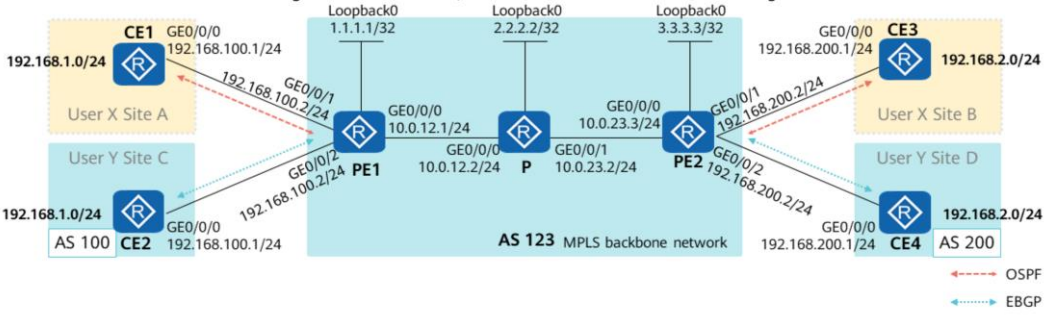
   An OSPF process to be bound to the VPN instance needs to be created on the PE, and an OSPF neighbor relationship needs to be established between the PE and its connected CE. In addition, OSPF and BGP need to import routes from each other.

   On the CE, common OSPF is configured.

HUAWEI

MPLS VPN Configuration Example - Networking Requirements

- The two sites of each user (user X and user Y) need to be interconnected through MPLS VPN. User X and user Y correspond to VPNX and VPNY, respectively.
- The interconnection interfaces, AS numbers, and IP addresses are shown in the following figure.
- Sites of user X use OSPF to exchange route with the PEs, and sites of user Y uses BGP to exchange routes with the PEs.



HUAWEI

MPLS VPN Configuration Example - Configuration Roadmap

1. Configure the MPLS VPN backbone network.
   1.1 Configure an IGP to implement IP connectivity on the backbone network.
   1.2 Configure MPLS and MPLS LDP to establish MPLS LSPs for transmitting VPN data through the public network.
   1.3 Configure MP-BGP to establish MP-BGP peer relationships for transmitting VPNv4 routes.
2. Configure VPN user access.
   2.1 Create VPN instances and set parameters (RTs and RDs).
   2.2 Bind the interfaces to the VPN instances.
   2.3 Configure route exchange between the PEs and their connected CEs.

HUAWEI

MPLS VPN Configuration Example - Data Planning

- The MPLS backbone network uses single-area OSPF to implement route interworking. MPLS LDP is enabled on all interconnection interfaces of PEs and Ps.
- The following table lists the VPN-related configurations on the PEs.

Item	Description			
	PE1		PE2	
VPN name	VPNX	VPNY	VPNX	VPNY
RD	100:1	200:1	100:1	200:1
IRT	100:321	200:234	100:123	200:432
ERT	100:123	200:432	100:321	200:234
Interface	GE0/0/1	GE0/0/2	GE0/0/1	GE0/0/2
MP-BGP	Source interface: Loopback0		Source interface: Loopback0	

HUAWEI

MPLS VPN Backbone Network Configuration (1)

1.1 Configure OSPF on the MPLS VPN backbone network to enable route connectivity within the backbone network. The OSPF configuration on PE1 is used as an example.

```
[PE1]ospf 100 router-id 1.1.1.1
[PE1-ospf-100]area 0
[PE1-ospf-100-area-0.0.0.0]network 10.0.12.1 0.0.0.0
[PE1-ospf-100-area-0.0.0.0]network 1.1.1.1 0.0.0.0
```

1.2 Configure MPLS and LDP on PE1, P, and PE2. The configuration on PE1 is used as an example.

```
[PE1]mpls lsr-id 1.1.1.1
[PE1]mpls
Info: Mpls starting, please wait... OK!
[PE1-mpls]mpls ldp
[PE1-mpls-ldp]Interface GigabitEthernet 0/0/0
[PE1-GigabitEthernet0/0/0]mpls
[PE1-GigabitEthernet0/0/0]mpls ldp
```


HUAWEI

MPLS VPN Backbone Network Configuration (2)

1.3 Establish an MP-BGP peer relationship between PE1 and PE2. The configuration on PE1 is used as an example.

```
[PE1]bgp 123
[PE1-bgp]router-id 1.1.1.1
[PE1-bgp]peer 3.3.3.3 as-number 123
[PE1-bgp]peer 3.3.3.3 connect-interface Loopback 0
# Enter the BGP-VPNv4 address family view and enable the VPNv4 address family for peer 3.3.3.3.
[PE1-bgp]ipv4-family vpnv4 unicast
[PE1-bgp-af-vpnv4]peer 3.3.3.3 enable
```


HUAWEI

- By default, only the peer relationships in the BGP IPv4 unicast address family view are automatically enabled. In other words, after the **peer as-number** command is run in the BGP view, the system automatically configures the peer enable command. In other address family views, however, peering must be enabled manually.

MPLS VPN Backbone Network Configuration - Configuration Verification

Check whether LDP LSPs are established.

```
[PE1]display mpls lsp
------------------------------------------------------------------------------
LSP Information: LDP LSP
------------------------------------------------------------------------------
FEC             In/Out Label        In/Out IF      Vrf Name
3.3.3.3/32      NULL/1025           -/GE0/0/0
1.1.1.1/32      3/NULL              -/-
```

```
[PE2]display mpls lsp
------------------------------------------------------------------------------
LSP Information: LDP LSP
------------------------------------------------------------------------------
FEC             In/Out Label        In/Out IF      Vrf Name
3.3.3.3/32      3/NULL              -/-
1.1.1.1/32      NULL/1024           -/GE0/0/0
```

Check the MP-BGP peer status. The following example uses the command output on PE1.

```
[PE1]display bgp vpnv4 all peer
 BGP local router ID : 1.1.1.1
 Local AS number : 123
 Total number of peers : 1                 Peers in established state : 1

  Peer        V          AS  MsgRcvd  MsgSent  OutQ     Up/Down       State Pre                fRcv
  3.3.3.3     4         123  16        18       0        00:14:20      Established  0
```


HUAWEI

VPN User Access Configuration (1)

2.1 Create VPN instances and set RDs and RTs as planned. The following uses PE1 as an example.

```
[PE1]ip vpn-instance VPNX
[PE1-vpn-instance-VPNX]route-distinguisher 100:1
[PE1-vpn-instance-VPNX-af-ipv4] vpn-target 100:321 import-extcommunity
 IVT Assignment result:
Info: VPN-Target assignment is successful.
[PE1-vpn-instance-VPNX-af-ipv4] 100:123 export-extcommunity
 EVT Assignment result:
Info: VPN-Target assignment is successful.
[PE1-vpn-instance-VPNX-af-ipv4] quit
[PE1-vpn-instance-VPNX]quit
[PE1]ip vpn-instance VPNY
[PE1-vpn-instance-VPNY]route-distinguisher 200:1
[PE1-vpn-instance-VPNY-af-ipv4]vpn-target 200:234 import-extcommunity
[PE1-vpn-instance-VPNY-af-ipv4]vpn-target 200:432 export-extcommunity
[PE1-vpn-instance-VPNY-af-ipv4]quit
[PE1-vpn-instance-VPNY]quit
```


HUAWEI

VPN User Access Configuration (2)

2.2 Bind the interfaces to the VPN instances.

```
[PE1]interface GigabitEthernet 0/0/1
[PE1-GigabitEthernet0/0/1]ip binding vpn-instance VPNX
Info: All IPv4 related configurations on this interface are removed!
Info: All IPv6 related configurations on this interface are removed!
[PE1-GigabitEthernet0/0/1]ip address 192.168.100.2 24
[PE1-GigabitEthernet0/0/1]interface GigabitEthernet 0/0/2
[PE1-GigabitEthernet0/0/2]ip binding vpn-instance VPNY
Info: All IPv4 related configurations on this interface are removed!
Info: All IPv6 related configurations on this interface are removed!
[PE1-GigabitEthernet0/0/2]ip address 192.168.100.2 24
```


HUAWEI

VPN User Access Configuration (3)

2.3.1 Configure route exchange between CE1 and PE1 and between CE3 and PE2. The configuration on PE1 is used as an example.

```
# Create an OSPF process and bind it to the instance.
[PE1]ospf 2 vpn-instance VPNX
[PE1-ospf-2]area 0
[PE1-ospf-2-area-0.0.0.0]network 192.168.100.0 0.0.0.255
[PE1-ospf-2-area-0.0.0.0]quit
```

```
# Configure route import between OSPF and MP-BGP.
[PE1]ospf 2 vpn-instance VPNX
[PE1-ospf-2]import-route bgp
[PE1-ospf-2]quit
[PE1]bgp 123
[PE1-bgp]ipv4-family vpn-instance VPNX
[PE1-bgp-VPNX]import-route ospf 2
```

2.3.2 Configure route exchange between CE2 and PE1 and between CE4 and PE2. The configurations on CE2 and PE1 are used as an example.

```
# Configure EBGP on CE2 and import the direct route
192.168.1.0/24.
[CE2]BGP 200
[CE2-bgp]peer 192.168.100.2 as-number 123
[CE2-bgp]network 192.168.1.0 24
[CE2-bgp]quit
```

```
# Configure an EBGP peer for the VPN instance VPNY on
PE1.
[PE1]bgp 123
[PE1-bgp]ipv4-family vpn-instance VPNY
[PE1-bgp-VPNY]peer 192.168.100.1 as-number 200
```


HUAWEI

Configuration Verification (1)

Check the routes learned on the CEs for user X in VPNX.

```
[CE1]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
192.168.1.0/24      Direct  0    0           D   192.168.1.254   GigabitEthernet0/0/1
192.168.2.0/24      OSPF    10   4           D   192.168.100.2   GigabitEthernet0/0/0
192.168.100.0/24    Direct  0    0           D   192.168.100.1   GigabitEthernet0/0/0
192.168.200.0/24    O_ASE   150  1           D   192.168.100.2   GigabitEthernet0/0/0
```

```
[CE3]dis ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
192.168.1.0/24      OSPF    10   4           D   192.168.200.2   GigabitEthernet0/0/0
192.168.2.0/24      Direct  0    0           D   192.168.2.254   GigabitEthernet0/0/1
192.168.100.0/24    O_ASE   150  1           D   192.168.200.2   GigabitEthernet0/0/0
192.168.200.0/24    Direct  0    0           D   192.168.200.1   GigabitEthernet0/0/0
```


HUAWEI

Configuration Verification (2)

Check the routes learned on the CEs for user Y in VPNY.

```
[CE2]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
192.168.1.0/24      Direct  0    0           D   192.168.1.254   GigabitEthernet0/0/1
192.168.2.0/24      EBGP    255  0           D   192.168.100.2   GigabitEthernet0/0/0
192.168.100.0/24    Direct  0    0           D   192.168.100.1   GigabitEthernet0/0/0
```

```
[CE4]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
192.168.2.0/24      Direct  0    0           D   192.168.1.254   GigabitEthernet0/0/1
192.168.1.0/24      EBGP    255  0           D   192.168.100.2   GigabitEthernet0/0/0
192.168.200.0/24    Direct  0    0           D   192.168.100.1   GigabitEthernet0/0/0
```


HUAWEI

Configuration Verification (3)

```
[PE2] display bgp vpnv4 vpn-instance VPNX routing-table 192.168.1.0 24

 BGP local router ID : 3.3.3.3
 Local AS number : 123
 VPN-Instance VPNX, Router ID 3.3.3.3:
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 192.168.1.0/24:
 Label information (Received/Applied): 1026/NULL
 From: 1.1.1.1 (1.1.1.1)
Relay token: 0x1
 Original nexthop: 1.1.1.1
```

```
[PE2]display mpls lsp
------------------------------------------------------------------------------
LSP Information: LDP LSP
------------------------------------------------------------------------------
FEC                 In/Out Label   In/Out IF    Vrf Name
1.1.1.1/32          NULL/1024      -/GE0/0/0
1.1.1.1/32          1024/1024      -/GE0/0/0
```

Use the data from 192.168.2.0/24 to 192.168.1.0/24 as an example. The outer label is 1024, which is allocated by MPLS LDP. The inner label is 1026, which is allocated by MP-BGP.

No.	Time	Source	Destination	Protocol	Length	Info
5	12.109000	**192.168.2.254**	**192.168.1.254**	ICMP	102	Echo (ping) request

Frame 5: 102 bytes on wire (816 bits), 102 bytes captured (816 bits) on interface 0
Ethernet II, Src: HuaweiTe_b1:15:3e (00:e0:fc:b1:15:3e), Dst: HuaweiTe_49:20:bb (00:e0:fc:49:20:bb)
MultiProtocol Label Switching Header, Label: **1024**, Exp: 0, S: 0, TTL: 254
MultiProtocol Label Switching Header, Label: **1026**, Exp: 0, S: 1, TTL: 254
Internet Protocol Version 4, Src: 192.168.2.254, Dst: 192.168.1.254
Internet Control Message Protocol

HUAWEI

Quiz

1. (Single-answer question) Which of the following route targets is carried when MP-BGP transmits VPNv4 routes? ( )
   A. Export RT
   B. Implied RT
   C. Import RT
   D. Extended RT
2. (Multiple-answer question) In basic MPLS VPN networking, which of the following configurations are required for a VPN instance? ( )
   A. Configure an import RT.
   B. Configure an export RT.
   C. Configure an RD.
   D. Bind an interface to the VPN instance.

HUAWEI

1. A
2. ABCD

Summary

- MPLS VPNs are usually constructed by a carrier to provide VPN services for different users. In this manner, routes and data of different users can be transmitted over the MPLS VPN and are isolated from each other.
- MPLS VPN control plane:
  - VRFs and RDs are used to isolate VPN routes of different users and construct unique VPNv4 routes.
  - RTs are to control the import and export of VPNv4 routes.
  - MP-BGP is used to transmit VPNv4 route prefixes, labels, and RTs.
  - MPLS LDP is used to establish LSPs over the public network.
- MPLS forwarding plane: Data is forwarded based on inner and outer labels. The inner labels differentiate VPN data, whereas the outer labels are used for public network transport.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

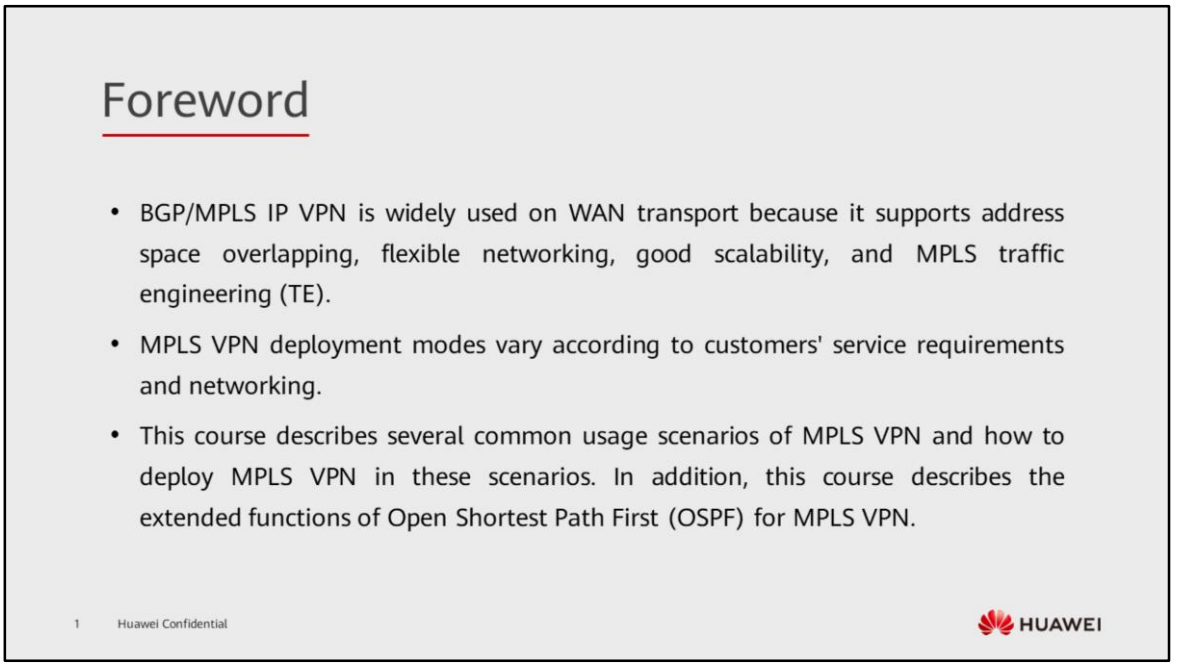
Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

MPLS VPN Deployment and Application

HUAWEI



- Note: Unless otherwise specified, MPLS VPN in this document indicates BGP/MPLS IP VPN.

Objectives

- Upon completion of this course, you will be able to:
  - Understand the usage scenarios and networking types of MPLS VPN.
  - Deploy the intranet solution with MPLS VPN.
  - Deploy the MPLS VPN Hub&Spoke solution.
  - Understand extended functions of OSPF for MPLS VPN.

HUAWEI

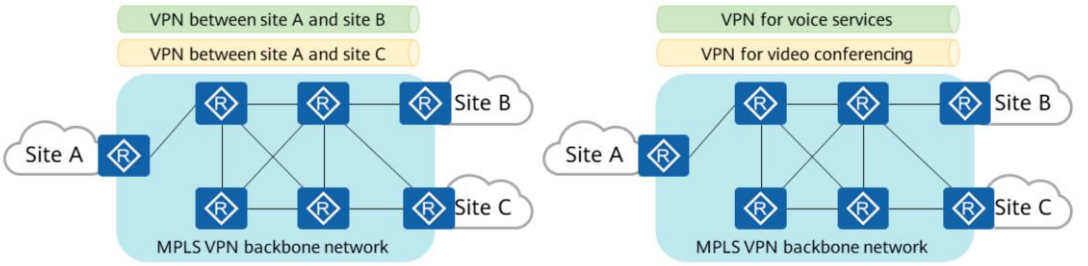
Contents

1. **MPLS VPN Applications and Networking Overview**
2. Typical Usage Scenarios and Deployment of MPLS VPN
3. OSPF VPN Extension

HUAWEI

Typical MPLS VPN Applications

- Currently, MPLS VPN is mainly used for enterprise interconnection and virtual service networks.
  - Enterprise interconnection: MPLS VPN connects the geographically different IP networks of branches, employees on business trips, and partners.
  - Virtual service network: Multiple services, such as VoIP and IPTV, can run on the same physical network. A VPN is established for each type of service to isolate services.

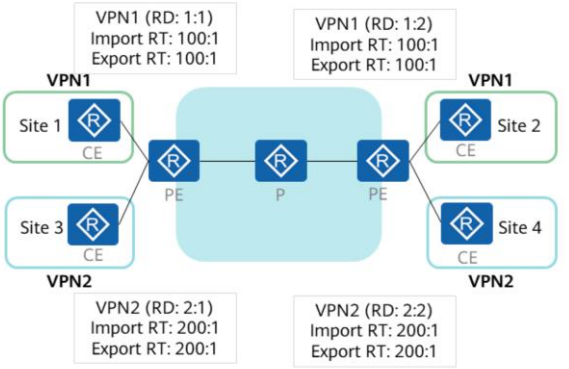


HUAWEI

- The advantages of MPLS VPN include but are not limited to the following:
  - One-hop access and network-wide connectivity can be implemented, and heterogeneous media interconnection is supported. Unlike the traditional leased line, which uses the same medium for a connection between each pair of user devices, the MPLS VPN can provide universal services conveniently.
  - Elastic bandwidth can be implemented. Traffic policing technology is used to ensure the required minimum bandwidth of users and implements best effort scheduling for burst traffic. In addition, the basic bandwidth can be soft expanded. That is, the basic bandwidth can be expanded within a range based on user requirements.
  - The MPLS VPN technology ensures the dedicated bandwidth of each VPN to meet the requirements of different users, traffic models, and QoS of various services.

Basic MPLS VPN Networking: Intranet

- When the intranet networking solution is used, all users in a VPN are in a user group isolated from the users of other VPNs and can forward traffic to each other. Users in a VPN cannot communicate with users outside the VPN. The sites of a VPN usually belong to the same organization.

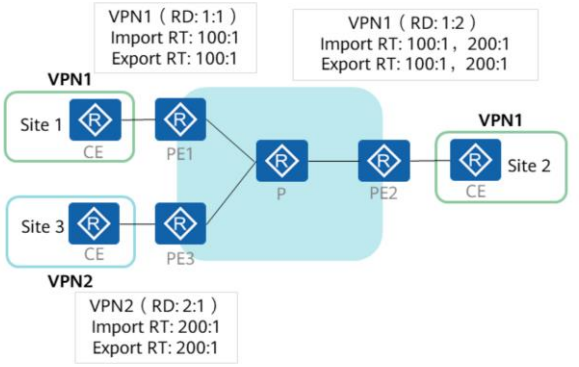


- A PE needs to create a VPN instance for each site and set a unique route distinguisher (RD) for each site.
- Import and export route targets (RTs) are set on PEs to prevent mutual communication between sites in different VPNs.

HUAWEI

Basic MPLS VPN Networking: Extranet

- When the extranet networking solution is used, VPN users can share network resources in some sites with other VPN users.



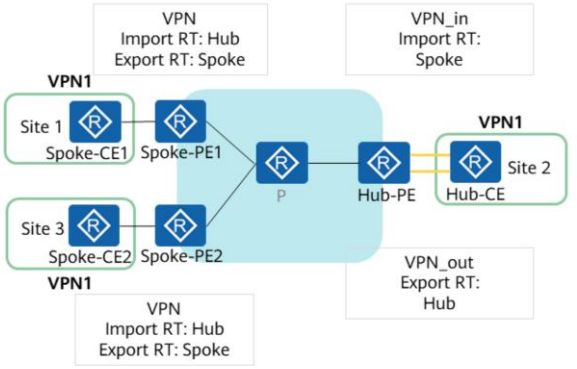
As shown in the figure, site 2 is a shared site that can be accessed by VPN1 and VPN2 users. The following requirements must be met:

- PE2 can receive VPNv4 routes advertised by PE1 and PE3.
- PE1 and PE3 can accept VPNv4 routes advertised by PE2.
- PE2 advertises neither VPNv4 routes received from PE1 to PE3 nor VPNv4 routes received from PE3 to PE1.

HUAWEI

Basic MPLS VPN Networking: Hub&Spoke (1)

- In the Hub&Spoke solution, one site can be configured as the hub site, and the other sites can be configured as spoke sites. Mutual access between sites must pass through the hub site. Data transmission between sites is centrally managed and controlled by the hub site.

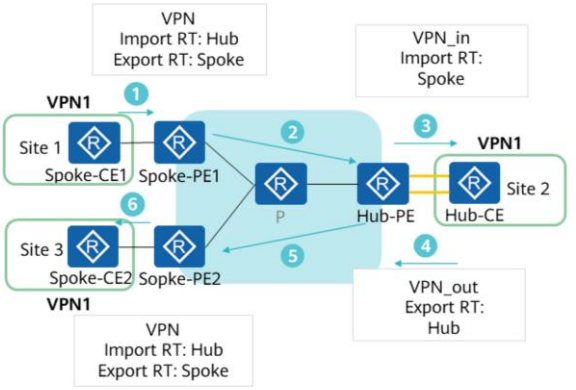


- A spoke site needs to advertise routes to a hub site, and then the hub site advertises the routes to other spoke sites. Spoke sites do not directly exchange routing information.
- For Spoke-PEs, set the export RT to "Spoke" and the import RT to "Hub."
- The Hub-PE needs to use two interfaces or sub-interfaces (bound to two VPN instances that are created). One interface or sub-interface is used to receive the routes from the Spoke-PE, and the import RT of the VPN instance is "Spoke." The other is used to advertise routes to Spoke-PEs, and the export RT of the VPN instance is "Hub."

HUAWEI

Basic MPLS VPN Networking: Hub&Spoke (2)

- The process of advertising routes from site 1 to site 2 is as follows:

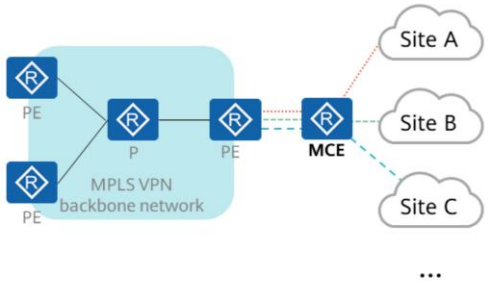


1. Spoke-CE1 advertises the route to Spoke-PE1.
2. Spoke-PE1 advertises the route to the Hub-PE through IBGP.
3. The Hub-PE imports the route into the VPN_in routing table through the import RT attribute of the VPN instance (VPN_in) and advertises the route to the Hub-CE.
4. The Hub-CE learns the route and advertises the route to the VPN instance (VPN_out) of the Hub-PE.
5. The Hub-PE advertises the route to Spoke-PE2 with the export RT attribute of VPN_out.
6. Spoke-PE2 advertises the route to Spoke-CE2.

HUAWEI

MCE Networking

- When a private network is divided into VPNs based on services or networks, services of different VPN users must be completely isolated. In this case, configuring a CE for each VPN increases device and maintenance costs.
- A multi-VPN-instance CE (MCE) device can function as a CE for multiple VPN instances on an MPLS VPN network, reducing the investment on network devices.

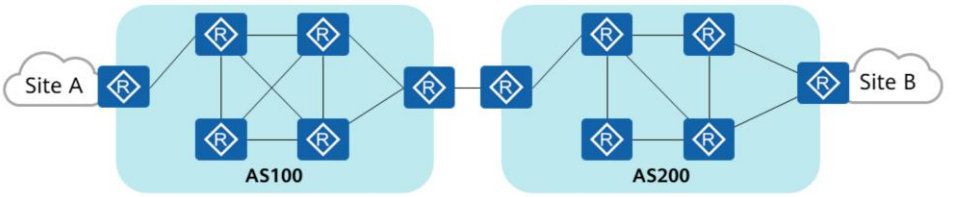


- The MCE device extends some functions of the PE to the CE. By binding different interfaces to VPNs, the MCE creates and maintains an independent routing and forwarding table (multi-VRF) for each VPN.
- The MCE can be connected to the corresponding PE through physical interfaces, sub-interfaces, or logical interfaces. On the PE, these interfaces must be bound to the corresponding VPN instances.

HUAWEI

Inter-AS MPLS VPN Networking

- With the wide application of the MPLS VPN solution, the number and geographical scope of terminal users are increasing. The number of sites in an enterprise is increasing. It is common to connect a geographical location to another service provider, for example, between different MANs of a carrier, the backbone networks of the carriers that cooperate with each other may span different autonomous systems (ASs).
- Generally, the MPLS VPN architecture runs in an AS. The routing information of any VPN can only be flooded in one AS as needed. The inter-AS MPLS VPN solution is used to deploy the MPLS VPN between ASs.

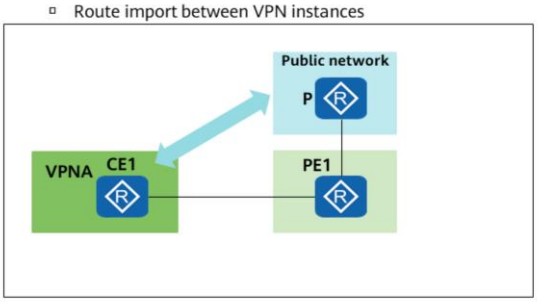


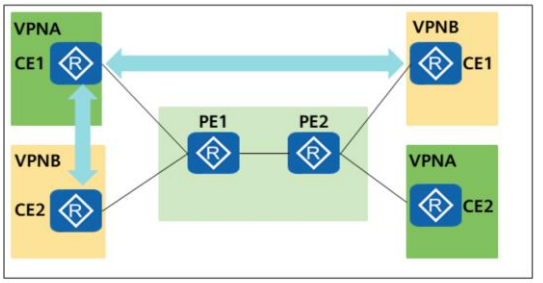
HUAWEI

- RFC2547 defines three inter-AS VPN solutions:
  - Inter-AS VPN Option A (inter-provider backbones Option A) mode: The inter-AS VPN manages its own VPN routes through dedicated interfaces between AS boundary routers (ASBRs). This mode is also called VRF-to-VRF.
  - Inter-AS VPN Option B (inter-provider backbones Option B) mode: ASBRs use MP-EBGP to advertise labeled VPNv4 routes. This mode is also called EBGP redistribution of labeled VPN-IPv4 routes.
  - Inter-AS VPN Option C (inter-provider backbones Option C): PEs use multi-hop MP-EBGP to advertise VPNv4 routes, which are also called multi-hop EBGP redistribution of labeled VPN-IPv4 routes.

Networking for Route Import Between Instances

- In BGP/MPLS IP VPN networking, VPN users cannot communicate with public network users. In addition, users in one VPN can communicate with those in another VPN only if the two VPNs have matching VPN targets. To enable communication in both cases, configure route import between instances. Route import between instances is classified into the following types:
  - Route import between VPN and public network instances
  - Route import between VPN instances





HUAWEI

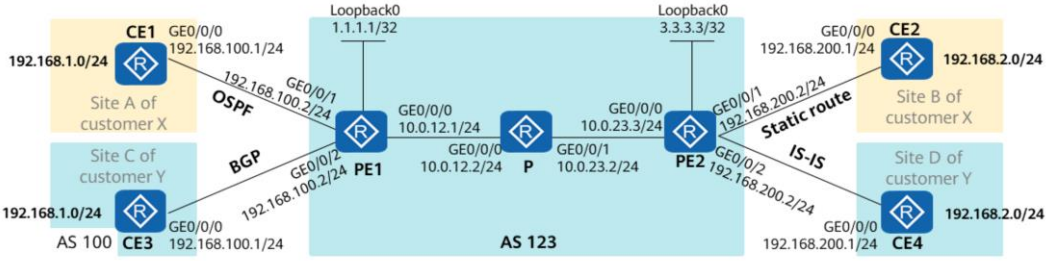
Contents

1. MPLS VPN Applications and Networking Overview
2. **MPLS VPN Deployment in Typical Scenarios**
   - Intranet Scenario
   - Hub&Spoke Scenario
   - Route Import Between VPN Instances Scenario
3. OSPF VPN Extension

HUAWEI

Deploying MPLS VPN in the Intranet Scenario

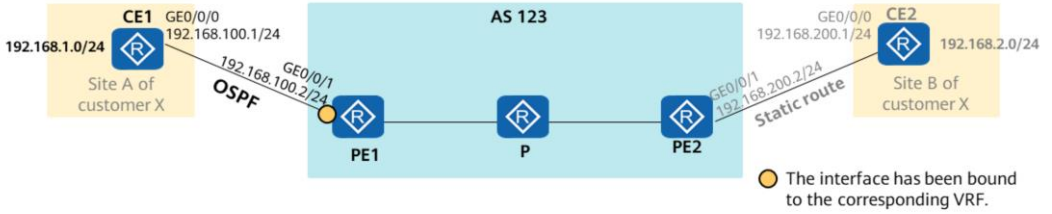
- As shown in the figure, customer X and customer Y have two sites each. The two sites need to be interconnected through MPLS VPN, which corresponds to VPNX and VPNY, respectively.
- Interconnection interfaces, AS numbers, and IP addresses are shown in the figure. CEs and PEs exchange routing information using the protocol or method shown in the figure.



HUAWEI

- Note: This course describes only the non-inter-AS MPLS VPN deployment.

Deploying OSPF Between PEs and CEs (1)

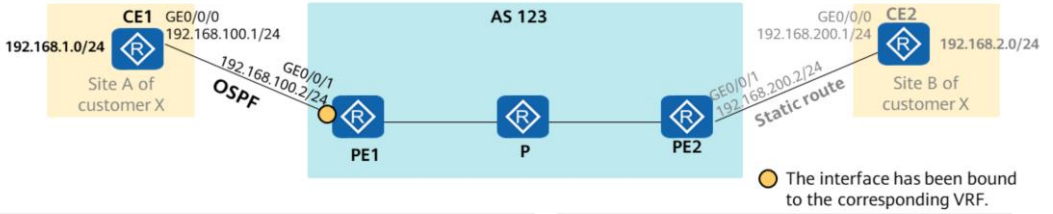



```
[CE1] ospf 1
[CE1-ospf-1] area 0
[CE1-ospf-1-area-0.0.0.0] network 192.168.100.0 0.0.0.255
[CE1-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
```

The OSPF configuration on CE1 is similar to the traditional OSPF configuration. CE1 does not need to support VRF.

HUAWEI

Deploying OSPF Between PEs and CEs (2)




```
[PE1] ospf 1 vpn-instance VPNX
[PE1-ospf-1] area 0
[PE1-ospf-1-area-0.0.0.0] network 192.168.100.0 0.0.0.255
[PE1-ospf-1-area-0.0.0.0] quit
[PE1-ospf-1] import bgp
```

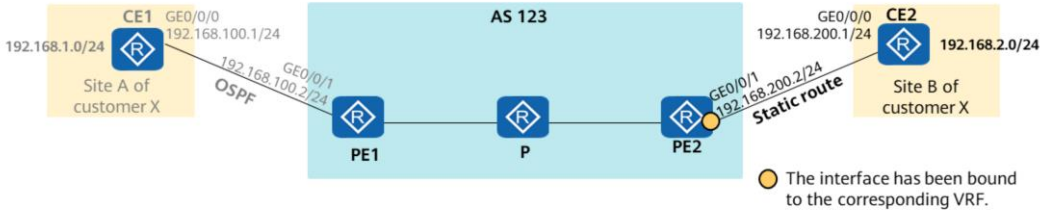
The OSPF process used by PE1 to interconnect with CE1 must be bound to the corresponding VPN instance.
Import the BGP routes in the routing table of VPNX on PE1 (mainly the customer routes learned by PE1 through BGP and destined for site B) into OSPF so that these routes can be advertised to CE1 through OSPF.

```
[PE1] bgp 123
[PE1-bgp] ipv4-family vpn-instance VPNX
[PE1-bgp] import-route ospf 1
```

Import the OSPF routes learned by OSPF process 1 in the routing table of VPNX on PE1 to BGP. Then, convert the customer routes destined for site A to BGP VPNv4 routes and advertise them to PE2.

HUAWEI

Deploying Static Routes Between PEs and CEs



[CE2] **ip route-static** 192.168.1.0 24 192.168.200.2
[CE2] **ip route-static** 192.168.100.0 24 192.168.200.2

A static route to each network segment at site A needs to be configured on CE2.

[PE2] **ip route-static vpn-instance** VPNX 192.168.2.0 24 192.168.200.1

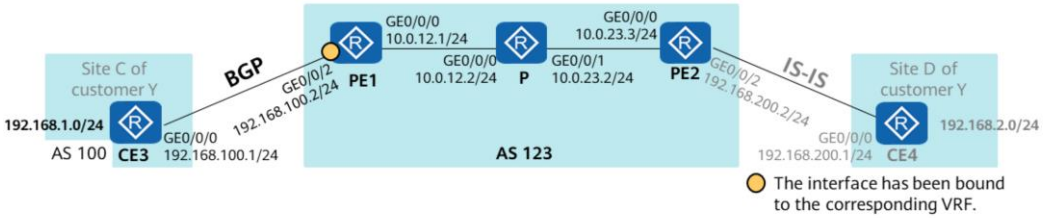
A static route to each network segment at site B needs to be configured on PE2.

[PE2] **bgp** 123
[PE2-bgp] **ipv4-family vpn-instance** VPNX
[PE2-bgp] **import-route static**

Import the static route in the routing table of VPNX on PE2 to BGP so that the static route can be converted into a BGP VPNv4 route and advertised to PE1.

HUAWEI

Deploying EBGP Between PEs and CEs



[CE3] **bgp** 100
[CE3-bgp] **peer** 192.168.100.2 **as-number** 123
[CE3-bgp] **network** 192.168.1.0 24

CE3 only needs to perform common BGP configurations and does not need to support VRF.

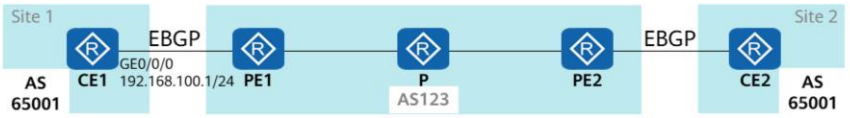
[PE1] **bgp** 123
[PE1-bgp] **ipv4-family vpn-instance** VPNY
[PE1-bgp-VPNY] **peer** 192.168.100.1 **as-number** 100

When a PE and a CE use BGP to exchange customer routes, you do not need to manually import routes on the PE. In this example, after PE1 learns a customer route from CE3 using BGP, PE1 automatically converts the route to a VPNv4 route and advertises the route to PE2. After PE1 learns the route to site D from PE2 using BGP, PE1 automatically converts the route to an IPv4 route and advertises the IPv4 route to CE3.

HUAWEI

BGP Configuration in Special Scenarios: AS Number Replacement

- In an MPLS VPN scenario, if EBGP runs between a PE and a CE to exchange routing information, the AS numbers of the two sites may be the same.



  - If CE1 sends a VPN route to PE1 through EBGP and PE2 forwards the route to CE2, CE2 will discard the route due to repetitive AS numbers. As a result, site 1 and site 2 that belong to the same VPN cannot communicate with each other.
  - You can run the **peer substitute-as** command on each PE to enable the AS number replacement function. That is, the PE replaces the AS number of the VPN site where the CE resides in the received private network route with the local AS number. The peer CE then does not discard the route with the repetitive AS number.

```
[PE1] bgp 123
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] peer 192.168.100.1 substitute-as
```

When sending a BGP route to CE1, PE1 replaces 65001 with the local AS number 123 if the AS_Path attribute contains 65001. Therefore, if a route is transmitted from CE2 to PE2 and then from PE2 to PE1, the AS_Path attribute of the BGP route is {123,123} when PE1 transmits the route to CE1.

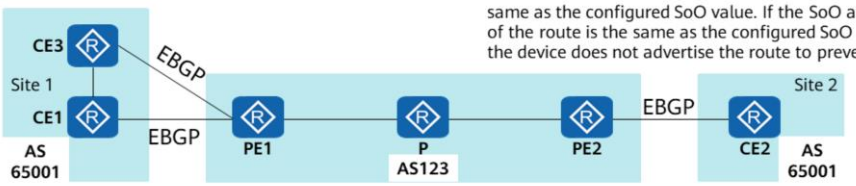
HUAWEI

BGP Configuration in Special Scenarios: SoO

- In a CE multi-homing scenario, if BGP AS number replacement is enabled, routing loops may occur. Therefore, the site of origin (SoO) feature is required to prevent the routing loops.
  - Both CE1 and CE3 belong to site 1. CE2 belongs to site 2. The AS numbers of sites 1 and 2 are both 65001. EBGP runs between PEs and CEs. To ensure that the PEs and CEs learn routes from each other, configure AS number replacement on PE1 and PE2.
  - CE1 transmits an intra-site route to PE1, and PE1 transmits the route to CE3. Because AS number replacement is configured, CE3 receives the route, which may cause a routing loop.

```
[PE1] bgp 123
[PE1-bgp] ipv4-family vpn-instance vpn1
[PE1-bgp-vpn1] peer 192.168.100.1 soo 200:1
[PE1-bgp-vpn1] peer 192.168.200.1 soo 200:1
```

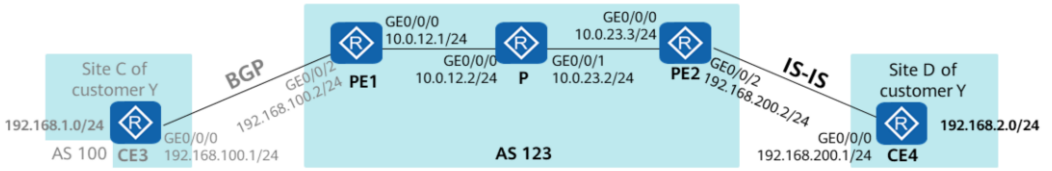
After the SoO attribute is configured for the BGP peer:
- When a BGP route is received from a peer, the SoO attribute is carried in the path attribute and advertised to other BGP peers.
- Before advertising a BGP route to a peer, the device checks whether the SoO attribute of the route is the same as the configured SoO value. If the SoO attribute of the route is the same as the configured SoO value, the device does not advertise the route to prevent loops.



HUAWEI

- Note: 192.168.100.1 and 192.168.200.1 are the IP addresses of the interfaces on CE1 and CE3, respectively, that are used to set up the BGP peer relationship with PE1.

Deploying IS-IS Between PEs and CEs



[CE4] **isis** 1
[CE4-isis-1] **network-entity 49.0001.0000.0000.1111.00**
[CE4-isis-1] **is-level level-2**
[CE4-isis-1] **quit**
[CE4] **interface GigabitEthernet** 0/0/0
[CE4-GigabitEthernet0/0/0] **isis enable** 1
[CE4-GigabitEthernet0/0/0]quit
[CE4] **interface GigabitEthernet** 0/0/1
[CE4-GigabitEthernet0/0/1] **isis enable** 1
GE0/0/1 is the interface connected to network segment 192.168.2.0/24.

[PE2] **isis** 1 **vpn-instance** VPNY
[PE2-isis-1] **network-entity 49.0002.0000.0000.2222.00**
[PE2-isis-1] **is-level level-2**
[PE2-isis-1] **import-route bgp level-2**
[PE2-isis-1] **quit**
[PE2] interface GigabitEthernet 0/0/2
[PE2-GigabitEthernet0/0/2] **isis enable** 1
[PE2] **bgp** 123
[PE2-bgp] **ipv4-family vpn-instance** VPNY
[PE2-bgp] **import-route isis 1**

HUAWEI

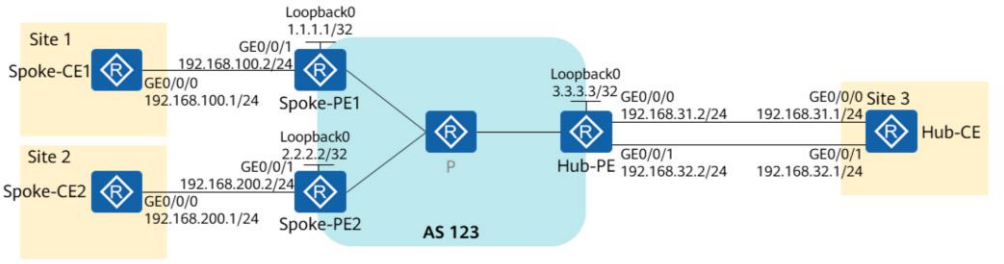
Contents

1. MPLS VPN Applications and Networking Overview
2. **MPLS VPN Deployment in Typical Scenarios**
   - Intranet Scenario
   - Hub&Spoke Scenario
   - Route Import Between VPN Instances Scenario
3. OSPF VPN Extension

HUAWEI

Deploying MPLS VPN in the Hub&Spoke Scenario

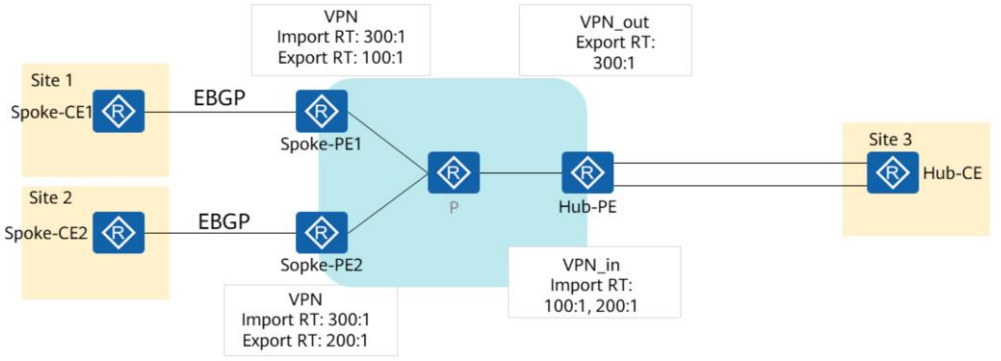
- Hub&Spoke networking solutions are as follows:
  - Method 1: EBGP runs between the Hub-CE and Hub-PE, and between the Spoke-PE and Spoke-CE.
  - Method 2: An IGP runs between the Hub-CE and Hub-PE, and between the Spoke-PE and Spoke-CE.
  - Method 3: EBGP runs between the Hub-CE and Hub-PE, and an IGP runs between the Spoke-PE and Spoke-CE.
- The Hub-CE and Hub-PE cannot use an IGP when the Spoke-PE and Spoke-CE using EBGP to deploy the MPLS VPN in the Hub&Spoke networking.



HUAWEI

VRF Configuration

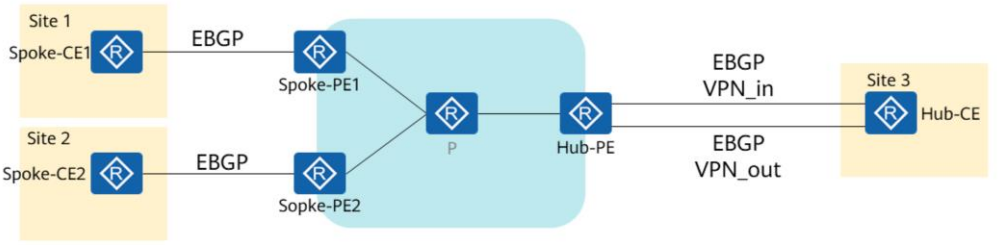
- Create a VPN instance on the Spoke-PE. The RT configuration is shown in the figure.
- Create VPN_in and VPN_out on the Hub-PE to import private network routes and export private network routes to the Spoke-PE, respectively. The RT configuration is shown in the figure.



HUAWEI

Deployment Method 1: Route Advertisement Process

- Spoke-CEs and Spoke-PEs exchange routing information through EBGP. After an EBGP connection is set up, Spoke-CEs and Spoke-PEs advertise routes to BGP.
- Two EBGP connections are set up between the Hub-PE and Hub-CE to separately advertise and accept private network routes.

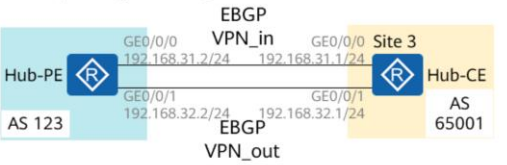


HUAWEI

- The process of advertising routes from Spoke-CE1 to Spoke-CE2 is as follows:
  - Spoke-CE1 advertises a route to Spoke-PE1 through EBGP.
  - Spoke-PE1 advertises the route to the Hub-PE through IBGP.
  - The Hub-PE imports the route into the VPN_in routing table through the import RT attribute of the VPN instance (VPN_in), and then advertises the route to the Hub-CE through EBGP.
  - The Hub-CE learns the route through the EBGP connection and advertises the route to the VPN instance (VPN_out) of the Hub-PE through another EBGP connection.
  - The Hub-PE advertises the route with the export RT attribute of VPN_out to all Spoke-PEs.
  - Spoke-PE2 advertises the route to Spoke-CE2 through EBGP.

Deployment Method 1: Configuration Between Hub-PE and Hub-CE

- The Hub-PE advertises the routes learned from spoke sites to the hub site through the EBGP connection corresponding to VPN_in.
- Hub-CE advertises the routes to spoke sites through EBGP corresponding to VPN_out.




```
# Set up two EBGP connections between the Hub-CE and Hub-PE.
[Hub-CE] bgp 65001
[Hub-CE-bgp] peer 192.168.31.2 as-number 123
[Hub-CE-bgp] peer 192.168.32.2 as-number 123
```

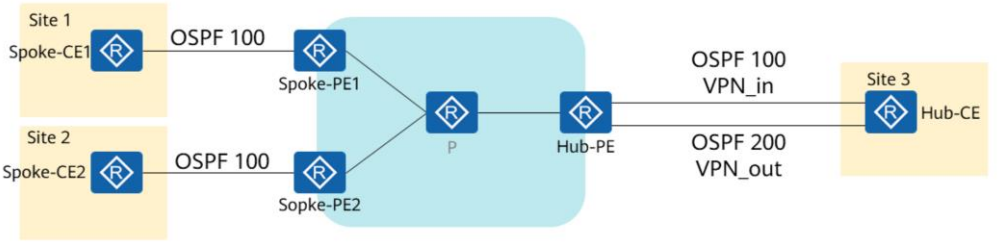
Because the routes that the Hub-CE sends to the Hub-PE through the EBGP connection corresponding to VPN_out may contain AS 123, these routes will be discarded by the Hub-PE. To prevent such a problem, the Hub-PE must be manually configured to allow repetitive local AS numbers.

```
# Establish two EBGP connections between the Hub-PE and Hub-CE.
[Hub-PE] bgp 123
[Hub-PE-bgp] ipv4-family vpn-instance VPN_in
[Hub-PE-bgp-VPN_in] peer 192.168.31.1 as-number 65001
[Hub-PE-bgp-VPN_in]quit
[Hub-PE-bgp] ipv4-family vpn-instance VPN_out
[Hub-PE-bgp-VPN_out] peer 192.168.32.1 as-number 65001
[Hub-PE-bgp-VPN_out] peer 192.168.32.1 allow-as-loop
```


HUAWEI

Deployment Method 2: Route Advertisement Process

- The following example uses OSPF as an IGP.
  - Spoke-CEs and Spoke-PEs exchange routing information through OSPF process 100.
  - The Hub-PE uses two OSPF processes to establish OSPF neighbor relationships with the Hub-CE, which separately advertise and accept private network routes.

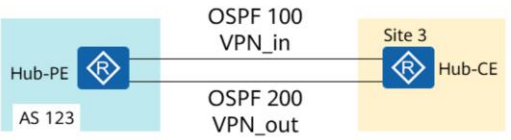


HUAWEI

- The process of advertising routes from Spoke-CE1 to Spoke-CE2 is as follows:
  - Spoke-CE1 advertises a route to Spoke-PE1 through OSPF 100.
  - Spoke-PE1 advertises the route to the Hub-PE through IBGP.
  - The Hub-PE imports the route to the VPN_in routing table through the import RT attribute of the VPN instance (VPN_in). After the BGP route is imported into OSPF 100, the route transmitted from Spoke-PE1 is advertised to the Hub-CE.
  - The Hub-CE receives the route through OSPF 100. After route import is configured, the route is advertised to OSPF 200, and then OSPF 200 advertises the route to the Hub-PE.
  - The VPN instance (VPN_out) of the Hub-PE imports the route of OSPF 200 multi-instance and advertises the route with the export RT attribute to all Spoke-PEs.
  - Spoke-PE2 advertises the route to Spoke-CE2 through OSPF 100.

Deployment Method 2: Configuration Between the Hub-PE and Hub-CE

- The Hub-PE advertises the routes learned from the spoke site to the hub site through OSPF process 100 corresponding to VPN_in.
- The Hub-CE advertises the route to the Hub-PE through OSPF (process 200) corresponding to VPN_out, and then advertises the route to all spoke sites.



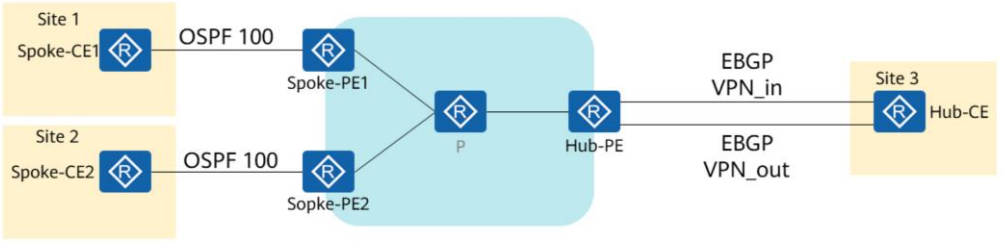

```
# Configure OSPF and BGP to import routes from each other on the Hub-PE.
[Hub-PE] OSPF 100 vpn-instance VPN_in
[Hub-PE-ospf-100]import-route bgp
[Hub-PE-ospf-100]quit
[Hub-PE]bgp 100
[Hub-PE-bgp]ipv4-family vpn-instance VPN_out
[Hub-PE-bgp-VPN_out]import-route ospf 200
```

```
# Import routes from OSPF 200 to OSPF 100 on the Hub-CE.
[Hub-CE]OSPF 200
[Hub-CE-ospf-200]import-route OSPF 100
```


HUAWEI

Deployment Method 3: Route Advertisement Process

- Use OSPF as an example. The Spoke-CEs and Spoke-PEs exchange routing information through an OSPF neighbor relationship (process 100).
- Establish two EBGP connections between the Hub-PE and Hub-CE to accept and advertise private network routes, respectively. The configurations of the Hub-PE and Hub-CE are similar to those in method 1.

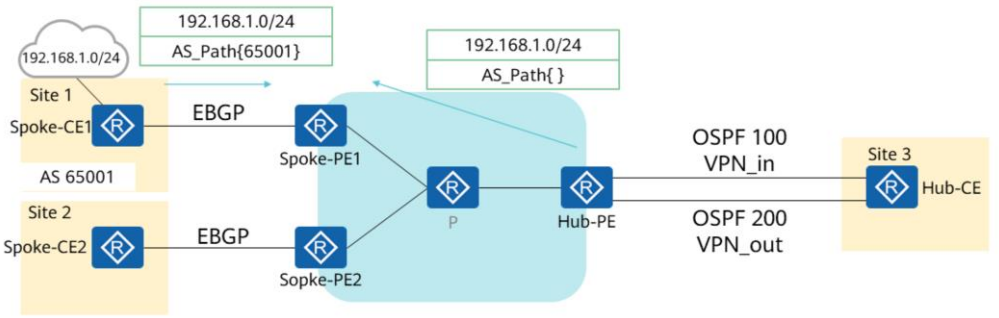


HUAWEI

- The process of advertising routes from Spoke-CE1 to Spoke-CE2 is as follows:
  - Spoke-CE1 advertises a route to Spoke-PE1 through OSPF 100.
  - Spoke-PE1 advertises the route to the Hub-PE through IBGP.
  - The Hub-PE imports the route into the VPN_in routing table through the import RT attribute of the VPN instance (VPN_in), and then advertises the route to the Hub-CE through EBGP.
  - The Hub-CE learns the route through the EBGP connection and advertises the route to the VPN instance (VPN_out) of the Hub-PE through another EBGP connection.
  - The Hub-PE advertises the route with the export RT attribute of VPN_out to Spoke-PE2.
  - Spoke-PE2 advertises the route to Spoke-CE2 through OSPF 100.

Why Is There No Method 4?

- The Hub-CE and Hub-PE **cannot** use an IGP, when the Spoke-PE and Spoke-CE **cannot** use EBGP to deploy the MPLS VPN in the Hub&Spoke networking.



HUAWEI

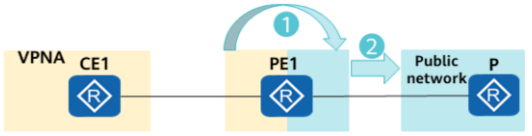
- The following takes the advertisement of the route destined for 192.168.1.0/24 from Spoke-CE1 to Spoke-CE2 as an example. The process is as follows:
  - Spoke-CE1 advertises a route to Spoke-PE1 through EBGP.
  - Spoke-PE1 advertises the route to the Hub-PE through IBGP.
  - The Hub-PE imports the route to the VPN_in routing table through the import RT attribute of the VPN instance (VPN_in) and advertises the route to the Hub-CE through OSPF 100.
  - Hub-CE learns the route through OSPF 100 and advertises the route to the Hub-PE through OSPF 200.
  - The VPN instance (VPN_out) of the Hub-PE imports the route of OSPF 200 and advertises the route with the export RT attribute of VPN_out to all Spoke-PEs.
  - The VPN instance on Spoke-PE2 imports the route based on the import RT. Spoke-PE2 advertises the route to Spoke-CE2 through EBGP.
- The VPN instance (VPN_out) on the Hub-PE advertises the route to Spoke-PE2 and Spoke-PE1 at the same time with the export RT. The route is imported by the Hub-PE through an IGP (OSPF 200). Because the IGP route does not carry the AS_Path attribute, the AS_Path attribute is null. The AS_Path of the route destined for 192.168.1.0/24 from Spoke-CE1 is not null. Therefore, the route returned by the Hub-PE takes precedence over the route from Spoke-CE1. As a result, route flapping occurs.

Contents

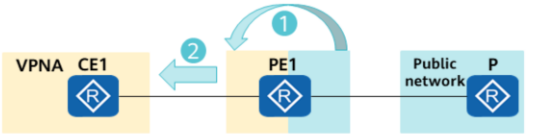
1. MPLS VPN Applications and Networking Overview
2. **MPLS VPN Deployment in Typical Scenarios**
   - Intranet Scenario
   - Hub&Spoke Scenario
   - Route Import Between VPN Instances Scenario
3. OSPF VPN Extension

HUAWEI

Configuring Interworking Between VPN and Public Network Instances



1. Configure the PE to import VPN routes to the local public routing table.
2. Configure the PE to import the local public routing table into other routing protocols so that the P device can obtain routing information.



1. Configure the PE to import public routes to the local VPN routing table.
2. Configure the PE to import the local private routing table into other routing protocols so that the CE device can obtain routing information.

- Import different types of VPN routes to the public network instance's corresponding routing tables.

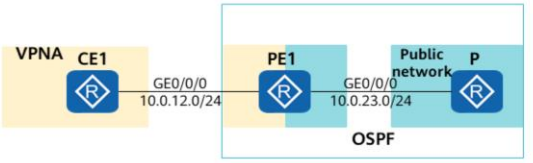
[Huawei] **ip import-rib vpn-instance** *vpn-instance-name* **protocol** { **static** | **isis** *process-id* | **ospf** *process-id* } [**valid-route**] [**route-policy** *route-policy-name* | **route-filter** *route-filter-name*]

- Import different types of routes from the public network instance to a VPN instance's corresponding routing tables.

[Huawei-vpn-instance-VPNA-af-ipv4] **import-rib public protocol** { **direct** | **vlink-direct-route** | { **static** | **isis** *process-id* | **ospf** *process-id* } [**valid-route**] } [**route-policy** *route-policy-name*]

HUAWEI

Example for Configuring Interworking Between VPN and Public Network Instances



1. Import routes of a specified VPN instance to the public network instance.

```
[PE1]ip import-rib vpn-instance VPNA protocol direct
[PE2]ospf 1
[PE1-ospf-1]import-route direct
```

2. Import routes of the public network instance to a specified VPN instance.

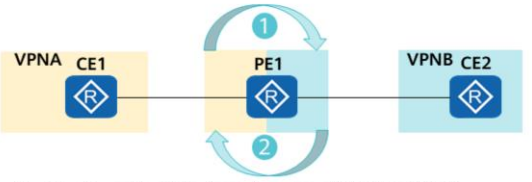
```
[PE1]ip vpn-instance VPNA
[PE1-vpn-instance-VPNA]ipv4-family
[PE1-vpn-instance-VPNA-af-ipv4] import-rib public protocol ospf 1
```

3. Verify the configuration: The P device can successfully ping the interface address of CE1 in a specified VPN instance.

```
<P>ping 10.0.12.1
  PING 10.0.12.1: 56  data bytes, press CTRL_C to break
    Reply from 10.0.12.1: bytes=56 Sequence=1 ttl=255 time=5 ms
    Reply from 10.0.12.1: bytes=56 Sequence=2 ttl=255 time=4 ms
    Reply from 10.0.12.1: bytes=56 Sequence=3 ttl=255 time=3 ms
    Reply from 10.0.12.1: bytes=56 Sequence=4 ttl=255 time=5 ms
    Reply from 10.0.12.1: bytes=56 Sequence=5 ttl=255 time=4 ms
```


HUAWEI

Configuring Interworking Between VPN Instances



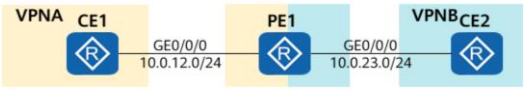
1. Configure the PE to import routes of VPNA to VPNB.
2. Configure the PE to import routes of VPNB to VPNA.

- Configure route import between VPN instances.

[Huawei-vpn-instance-VPNA-af-ipv4] **import-rib vpn-instance** *vpn-instance-name* **protocol** { **direct**| { **static** | **isis** *process-id* | **ospf** *process-id* } [**valid-route**] } [**route-policy** *route-policy-name* | **route-filter** *route-filter-name*]

HUAWEI

Example for Configuring Interworking Between VPN Instances



1. Import private routes of VPNA to VPNB.

```
[PE1]ip vpn-instance VPNB
[PE1-vpn-instance-VPNB]ipv4-family
[PE1-vpn-instance-VPNB-af-ipv4]  import-rib vpn-instance VPNA protocol direct
```

2. Import private routes of VPNB to VPNA.

```
[PE1]ip vpn-instance VPNA
[PE1-vpn-instance-VPNA]ipv4-family
[PE1-vpn-instance-VPNA-af-ipv4]  import-rib vpn-instance VPNB protocol direct
```

3. Verify the configuration: CE2 can successfully ping the interface address of CE1 in a specified VPN instance.

```
<CE2>ping 10.0.12.1
  PING 10.0.12.1: 56  data bytes, press CTRL_C to break
    Reply from 10.0.12.1: bytes=56 Sequence=1 ttl=255 time=5 ms
    Reply from 10.0.12.1: bytes=56 Sequence=2 ttl=255 time=4 ms
    Reply from 10.0.12.1: bytes=56 Sequence=3 ttl=255 time=3 ms
    Reply from 10.0.12.1: bytes=56 Sequence=4 ttl=255 time=5 ms
    Reply from 10.0.12.1: bytes=56 Sequence=5 ttl=255 time=4 ms
```


HUAWEI

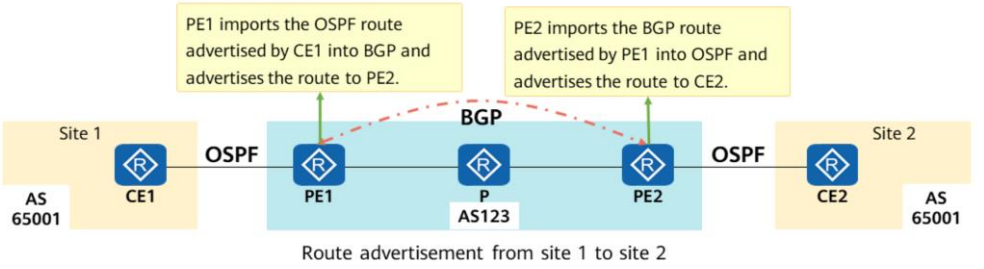
Contents

1. MPLS VPN Applications and Networking Overview
2. MPLS VPN Deployment in Typical Scenarios
3. **OSPF VPN Extension**
   - Interoperability Between OSPF and BGP
   - OSPF Loop Prevention
   - OSPF Sham Link

HUAWEI

OSPF/BGP in MPLS VPN

- When OSPF is deployed between the PE and CE to exchange routing information, if the standard BGP/OSPF exchange process (BGP/OSPF interoperation for short) is used on the PE to exchange routing information, the remote PE directly generates Type 5 LSAs when importing BGP into the OSPF process of the VPN instance. Different sites consider the routes of other sites as AS external routes (AS_external).
- To solve the problem of OSPF routing information loss caused by standard BGP/OSPF interoperability, BGP and OSPF are extended accordingly.



Route advertisement from site 1 to site 2

HUAWEI

- In actual applications, if two sites that need to communicate are in the same AS, each site should consider the route of the other site as an inter-area route rather than an AS external route.

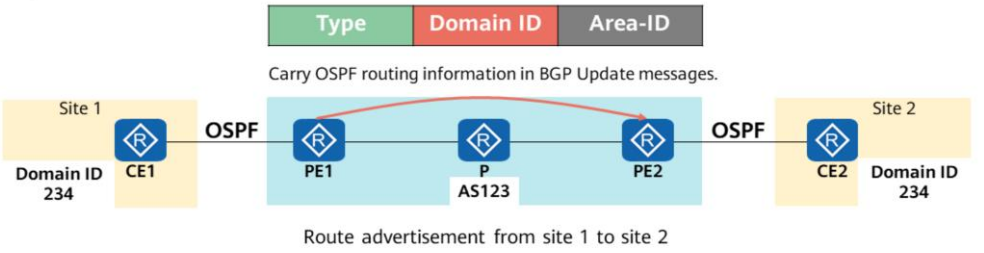
BGP Extended Community Attributes

- To retain OSPF routing information, BGP adds some community attributes that can carry OSPF routing information.
  - Domain ID: identifies a domain.
  - Route Type: contains the area ID and route type of the OSPF route imported to BGP.
    - Area-ID: ID of the VRF OSPF process of the PE that establishes an adjacency with a CE.
    - Route Type: type of the imported OSPF route:
      - 1 or 2: intra-area route, that is, the route calculated by the PE based on Type 1 and Type 2 LSAs
      - 3: inter-area route
      - 5: OSPF external route, that is, the route calculated by the PE through Type 5 LSAs. When the value of the Route Type field is 5, the value of the Area-ID field must be 0.0.0.0.
      - 7: NSSA route, that is, the route calculated by the PE through Type 7 LSAs

HUAWEI

Domain ID

- When OSPF routes are imported into BGP on a PE, the PE adds the domain ID attribute to the BGP routes according to the local configuration. The domain ID is transmitted as the extended community attribute of BGP.
- When a PE imports a BGP route to OSPF, if the domain ID carried in the BGP route is the same as the local domain ID, the two sites belong to the same OSPF routing domain. If they are different, they are considered not in the same routing domain.



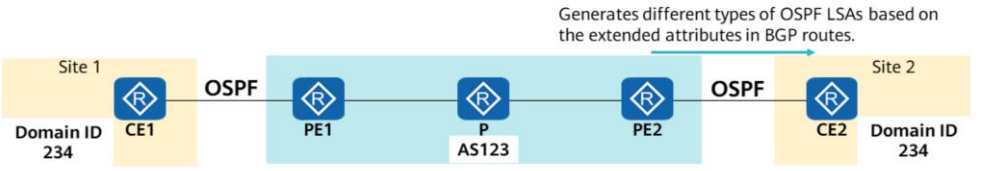
HUAWEI

- The domain ID can be configured using the domain-id command in the view of the OSPF process bound to the VRF instance.
  - By default, the domain ID is 0 (NULL). If different OSPF domains use NULL as the domain ID, these OSPF domains cannot be distinguished. Consequently, the routes between different OSPF domains are considered as intra-area routes.
  - If an OSPF domain is configured with a non-zero domain ID, NULL is no longer the domain ID of the OSPF domain.
- It is recommended that all OSPF instances related to the same VPN use the same domain ID or the default domain ID.

Domain ID and Route Type

- Based on the domain ID and route type in the BGP route, a PE generates different types of OSPF LSAs and advertises them to the OSPF process of the VRF.

Whether the domain ID is the same as the local domain ID	Route Type	Types of OSPF LSAs generated by PEs
Yes	1, 2, 3	3
	5, 7	5, 7
No	1, 2, 3, 5, 7	5, 7



Route advertisement from site 1 to site 2

HUAWEI

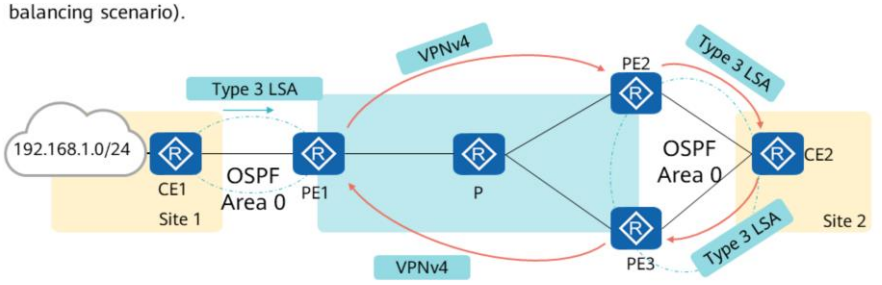
Contents

1. MPLS VPN Applications and Networking Overview
2. MPLS VPN Deployment in Typical Scenarios
3. **OSPF VPN Extension**
   - Interoperability Between OSPF and BGP
   - OSPF Loop Prevention
   - OSPF Sham Link

HUAWEI

Type 3 Routing Loop Prevention: Case

- The following figure shows an example of Type 3 LSA routing loops.
  - Site 1 and site 2 belong to VPN1.
  - Site 1 is connected to PE1 on the backbone network through OSPF area 0.
  - Site 2 is connected to PE2 and PE3 on the backbone network through OSPF area 0 (in the dual-homing load balancing scenario).

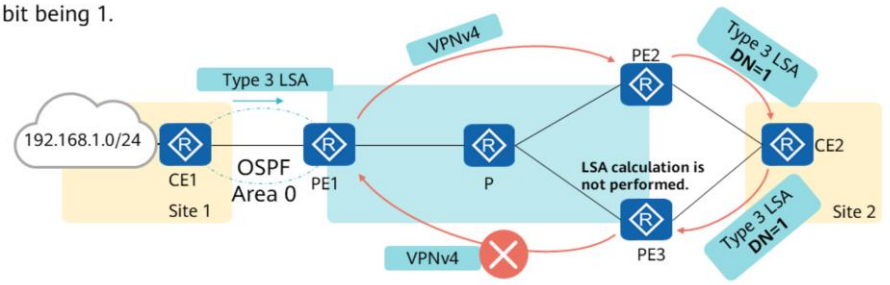


HUAWEI

- The loop generation process is as follows:
  - CE1 at site 1 advertises a route destined for 192.168.1.0/24 to PE1 using a Type 3 LSA.
  - PE1 imports the route of the OSPF VPN1 process to BGP and advertises the route to PE2 and PE3 through MP-IBGP.
  - PE2 is configured to import routes from BGP to OSPF. Therefore, PE2 generates Type 3 LSAs and sends them to CE2. CE2 then advertises the Type 3 LSAs received from PE2 to PE3.
  - PE3 receives two routes destined for 192.168.1.0/24. One is advertised by PE1, and the other is imported by PE2. By default, IGP (OSPF) routes have a higher priority than IBGP routes. Therefore, PE3 selects OSPF routes.
  - PE3 advertises the optimal route learned from OSPF to PE1 through MP-IBGP.
  - In this case, PE1 has two routes to the destination network segment 192.168.1.0/24. One is learned from CE1 through OSPF, and the other is learned from PE3 through MP-IBGP. The following problems may occur:

Type 3 Routing Loop Prevention: DN Bit

- To prevent Type 3 LSA loops, the OSPF multi-instance process uses an unused bit in the LSA Options field as a flag bit, which is called the DN bit. The DN bit is used to prevent Type 3 LSA loops.
- When performing SPF calculation, the OSPF instance process on a PE ignores Type 3 LSAs with the DN bit being 1.

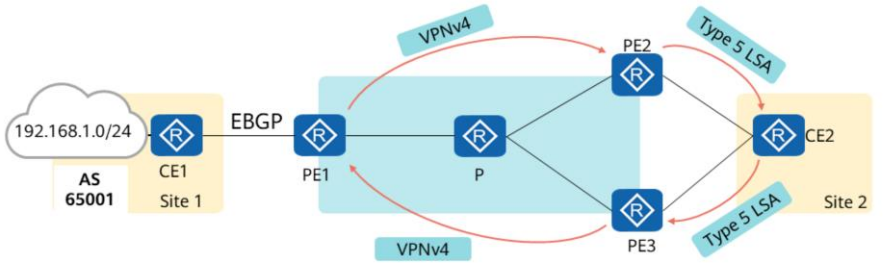


HUAWEI

- By default, the DN bit in LSAs generated by OSPF is set to 1. You can run **the dn-bit-set disable** command to disable OSPF from setting the DN bit in LSAs.

Type 5/7 Routing Loop Prevention: Case

- The following figure shows an example of a Type 5 LSA routing loop.
  - Site 1 and site 2 belong to VPN1.
  - Site 1 is connected to PE1 on the backbone network through EBGP.
  - Site 2 is connected to PE2 and PE3 on the backbone network through OSPF.

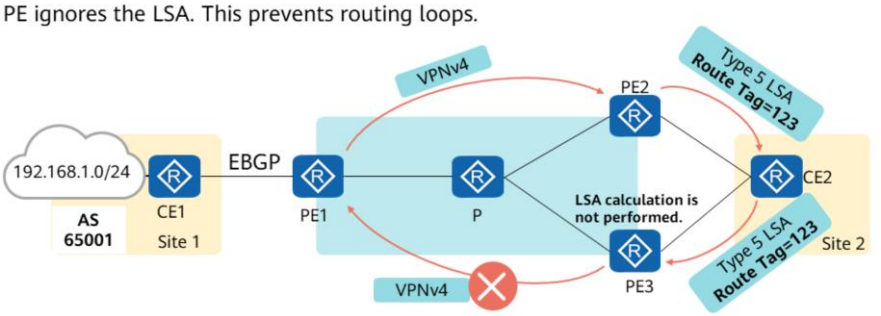


HUAWEI

- The loop generation process is as follows:
  - CE1 advertises a route destined for 192.168.1.0/24 to PE1 through EBGP. The AS_Path of the route is 65001.
  - PE1 advertises the route to PE2 and PE3 through MP-IBGP.
  - PE2 imports BGP routes to the OSPF VPN1 process and advertises a Type 5 LSA destined for 192.168.1.0/24 to CE2.
  - CE2 advertises the Type5 LSA to PE3.
  - PE3 preferentially selects an OSPF route (the OSPF route has a higher priority than the IBGP route) and advertises an Update message to PE1 through MP-IBGP.
  - PE1 receives the MP-IBGP Update message from PE3. The MP-IBGP route advertised by PE3 has a higher priority than the EBGP route advertised by CE1 because the MP-IBGP route is an IGP (OSPF) route imported by BGP on PE3 and its AS_Path is null. PE1 prefers the route advertised by PE3.
  - In this case, a routing loop is formed: PE3 -> CE2 -> PE2 -> PE1 -> PE3.
- Because PE1 does not preferentially select the route learned from CE1, PE1 withdraws the route advertised to PE2. The imported BGP route is also withdrawn in the OSPF VPN instance process on PE2. Then, both CE2 and PE3 withdraw the OSPF routes. The BGP route advertised by PE3 to PE1 is also withdrawn. On PE1, the route learned from CE1 becomes the optimal route. As a result, route flapping occurs.
- The generation and elimination of Type 7 LSA-related loops are similar to those of Type 5 LSA-related loops, and are not described here.

Type 5/7 Route Loop Prevention: VPN Route Tag

- A VPN route tag can be used to prevent Type 5 and Type 7 routing loops.
- When a PE generates Type 5 or Type 7 LSAs based on received BGP VPN routes, the LSAs carry VPN route flags. If a PE finds that the VPN route tag in an LSA is the same as the locally configured one, the PE ignores the LSA. This prevents routing loops.



HUAWEI

- The VPN route tag is not transmitted in the BGP extended community attribute. The VPN route tag is valid only on the PEs that receive BGP routes and generate OSPF LSAs.
- By default, the VPN route tag is calculated based on the AS number of BGP. If BGP is not configured, the default value is 0.
- You can run the route-tag command to set a VPN route tag.

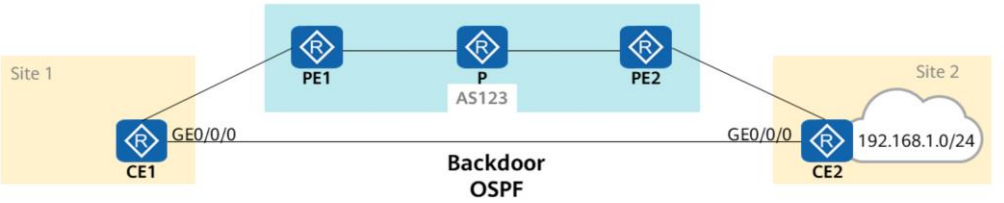
Contents

1. MPLS VPN Applications and Networking Overview
2. MPLS VPN Deployment in Typical Scenarios
3. **OSPF VPN Extension**
   - Interoperability Between OSPF and BGP
   - OSPF Loop Prevention
   - OSPF Sham Link

HUAWEI

Sham Link Usage Scenarios

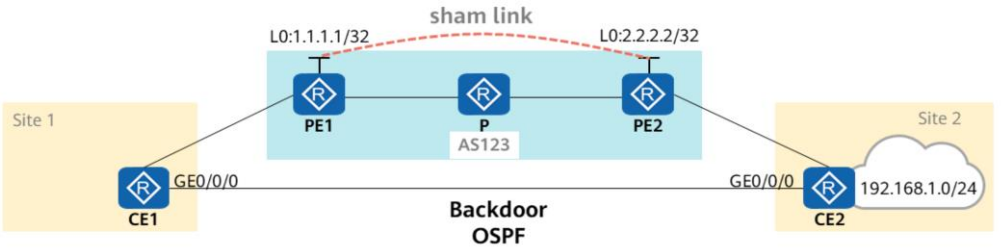
- Generally, BGP peers use BGP extended community attributes to carry routing information on the MPLS VPN backbone network. OSPF running on the peer PE can use the information to generate Type 3 LSAs from the PE to CE. These Type 3 LSAs are inter-area routes.
- If a backdoor link is added between CE1 and CE2 and OSPF is run to exchange routes, the route learned through the backdoor link is an intra-area route.
- Because intra-area routes take precedence over inter-area routes, the backdoor link is preferentially selected. To allow the backdoor link as a backup link, use the sham link.



HUAWEI

Working Mechanism of Sham Link

- The sham link creates an intra-area link between two PEs. When LSAs are flooded on a sham link, all OSPF route types remain unchanged and are not changed to Type 3 or 5 LSAs.
- A sham link is considered as a link between two VPN instances. The addresses of the two ends of the link are the addresses of the PEs, which are used as the source and destination addresses of the connection. The source and destination addresses of a sham link are loopback interface addresses with 32-bit masks. The loopback interface must be bound to a VPN instance and advertised through BGP.



HUAWEI

- Multiple sham links of the same OSPF process can share the same endpoint address, but different OSPF processes cannot have two sham links with the same endpoint address.

Sham Link Configuration Example

1. Create an interface on the PE to set up a sham link. The configuration of PE2 is similar to that of PE1.

```
[PE1]interface LoopBack0
[PE1-LoopBack0] ip binding vpn-instance VPNA
[PE1-LoopBack0]ip address 1.1.1.1 32
# Advertise the routes in the BGP VPN address family.
[PE1-bgp-VPNA]network  1.1.1.1 32
```

2. Configure the sham link on PE1. The configuration of PE2 is similar to that of PE1.

```
[PE1-ospf-1]area 0
[PE1-ospf-1-area-0.0.0.0]sham-link 1.1.1.1 2.2.2.2
```

3. Adjust the cost value to ensure that the cost value of the backdoor link is greater than that of the sham link.

```
[CE1-GigabitEthernet0/0/0]ospf cost  1000
```


HUAWEI

- When configuring a sham link, you can specify the route cost of the sham link. The default value is 1.

Verifying the Configuration of the Sham Link

1. For details about common OSPF VPN configurations, see the preceding configuration examples.

```
<PE1>display ospf sham-link  area  0

        OSPF Process 1 with Router ID 1.1.1.1

 Sham-Link: 1.1.1.1 --> 2.2.2.2
 Neighbor ID: 2.2.2.2 ,    State: Full ,    GR status: Normal
 Area: 0.0.0.0
 Cost: 1 , State: P-2-P , Type: Sham
 Timers: Hello 10 , Dead 40 , Retransmit 5 , Transmit Delay 1
```

2. Check OSPF routes on CE1. The command output shows that the peer route has been learned as an intra-area route.

```
<CE1>display  ospf routing

 Routing for Network
 Destination        Cost      Type       NextHop         AdvRouter       Area
 1.1.1.1/32         0         Stub       1.1.1.1         10.1.12.1       0.0.0.0
 10.1.12.0/24       1         Transit    10.1.12.1       10.1.12.1       0.0.0.0
 192.168.1.0/24     3         Stub       10.0.12.1       2.2.2.2         0.0.0.0
 10.1.23.0/24       3         Transit    10.1.12.2       10.1.23.2       0.0.0.0
```


HUAWEI

Quiz

1. (Multiple-answer question) On an MPLS VPN network, when a PE imports VPN routes learned from other PEs to OSPF, which of the following LSAs may be generated? ( )
   A. Type 1 LSA
   B. Type 3 LSA
   C. Type 5 LSA
   D. Type 7 LSA
2. (True or false) When a CE transmits routes to a PE through BGP, the routes may carry the SoO attribute. ( )
   A. True
   B. False

HUAWEI

1. BCD
2. B

Summary

- MPLS VPN has different networking solutions in different scenarios. The common networking solutions are intranet, extranet, and Hub&Spoke. In addition, MPLS VPN networking can be classified into inter-AS and intra-AS networking based on whether the MPLS VPN backbone network is an inter-AS network.
- PEs and CEs can use static, OSPF, IS-IS, or BGP routes to exchange routing information. OSPF provides the following extended features for MPLS VPN:
  - Domain ID: identifies whether the routes imported to a VPN instance belong to the same OSPF domain.
  - DN bit: used to prevent routing loops because of Type 3 LSAs.
  - VPN route tag: is used to prevent routing loops caused by Type 5 or Type 7 LSAs.
  - Sham link: controls OSPF route selection in special scenarios.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Inter-AS MPLS L3VPN

HUAWEI

Foreword

- As the MPLS VPN solution has been widely adopted, the user quantity and network scale are increasing rapidly, and the number of sites in an enterprise keeps increasing. Sites in different geographical locations or autonomous systems (ASs) need to be interconnected across ASs.
- To support the exchange of VPN routing information between ASs, the existing protocol needs to be extended and the MPLS VPN architecture needs to be modified to provide an interconnection model different from the basic MPLS VPN architecture, that is, the inter-AS MPLS VPN.
- RFC4364 proposed three inter-AS VPN solutions: inter-AS VPN Option A, inter-AS VPN Option B, and inter-AS VPN Option C.
- This document describes the fundamentals and configurations of the three inter-AS MPLS VPN solutions.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the fundamentals of the three inter-AS VPN solutions.
  - Perform basic configurations of the three inter-AS VPN solutions.
  - Illustrate the application scenarios of the three inter-AS VPN solutions.

HUAWEI

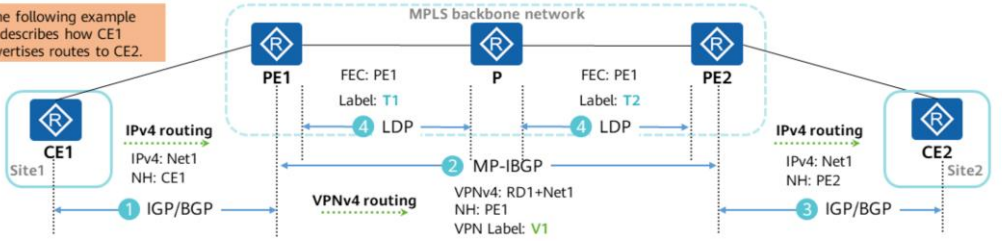
Contents

1. **Background of Inter-AS VPN Solutions**
2. Fundamentals and Configurations of Inter-AS VPN Option A
3. Fundamentals and Configurations of Inter-AS VPN Option B
4. Fundamentals and Configurations of Inter-AS VPN Option C

HUAWEI

Fundamentals of Intra-AS MPLS VPN: Route Advertisement and Label Distribution

The following example describes how CE1 advertises routes to CE2.

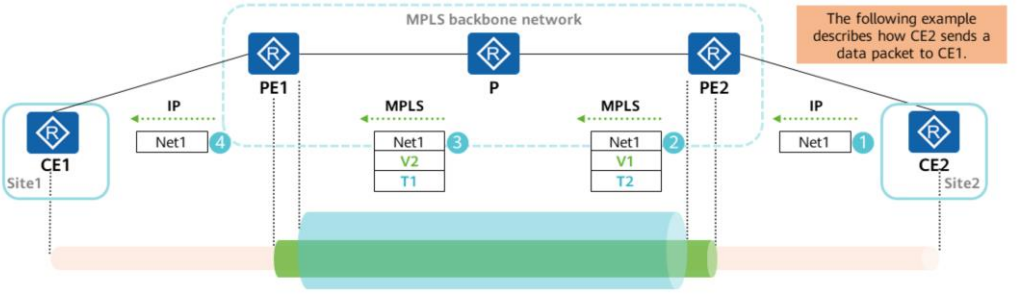


1. CE1 and PE1 run IGP/BGP to exchange routing information (IPv4 routes).
2. After PE1 receives a route from CE1, PE1 converts the route into a VPNv4 route, advertises the VPNv4 route to PE2 using MP-IBGP, and allocates a VPN label V1 to the route.
3. After PE2 receives the VPNv4 route, PE2 converts the route into an IPv4 route and advertises the route to CE2 using IGP/BGP.
4. PEs and the P on the MPLS backbone network run LDP to allocate tunnel labels. Assume that the labels allocated by PE1 and allocated by P to PE1 are T1 and T2 respectively.

Only PEs are aware of VPN routing information, while the P device is not.
The backbone network uses MPLS tunnel forwarding to avoid routing blackholes.

HUAWEI

Fundamentals of Intra-AS MPLS VPN: Data Forwarding



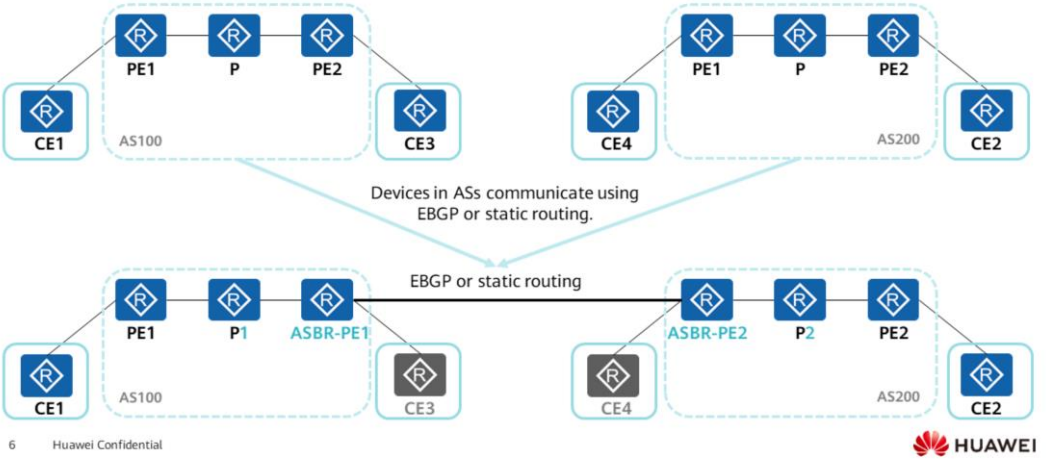
1. CE2 sends an IPv4 packet with the destination address Net1.
2. After PE2 receives the packet, it encapsulates the packet with a VPN label V1 and then an outer label T2 and forwards the packet to the P device.
3. The P device swaps the outer label T2 with T1 and forwards the packet to PE1.
4. PE1 removes all labels from the packet and forwards the packet to CE1.

IPv4
VPN
LDP LSP

HUAWEI

Background of Inter-AS MPLS VPN Solutions

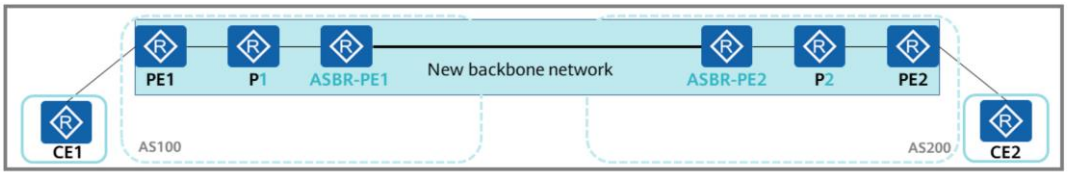
- As the enterprise scale keeps increasing, enterprise sites in different locations may belong to different ASs. As shown in the figure, when CEs in two different ASs need to communicate with each other, the inter-AS VPN technology is required.



HUAWEI

- To facilitate understanding, the devices connecting ASs in the original AS are called ASBR-PEs, and P devices are numbered differently. In addition, CE3 and CE4 will not be discussed.

Problems Caused by Inter-AS MPLS VPN



- Compared with intra-AS MPLS VPN, inter-AS MPLS VPN has similar fundamentals but faces the following problems:
  - LDP does not run between ASs, as a result of which, outer tunnels cannot be established between ASs.
  - PEs do not run IGP, cannot establish BGP peer relationships between each other by default, and so cannot directly advertise VPNv4 routes to each other.
- The solutions to the preceding problems are as follows:
  - ASBRs exchange IPv4 routes and forward IPv4 data packets. This solution is easy to understand.
  - ASBRs exchange VPNv4 routes and forward packets carrying one MPLS label.
  - PEs exchange VPNv4 routes and forward packets carrying multiple MPLS labels.

Option A
Option B
Option C

HUAWEI

Inter-AS MPLS VPN Solutions

- The following describes three inter-AS MPLS VPN solutions:
  - Inter-Provider Backbones Option A (inter-AS VPN Option A)
    - VPN instances spanning multiple ASs are bound to dedicated interfaces of ASBRs to manage their own VPN routes. This solution is also called VRF-to-VRF.
    - The configuration is simple, without requiring MPLS to run between ASBRs.
  - Inter-Provider Backbones Option B (inter-AS VPN Option B)
    - ASBRs advertise labeled VPNv4 routes to each other using MP-EBGP. This solution is also called EBGP redistribution of labeled VPNv4 routes.
    - With this solution, you do not need to create a different interface for each VPN.
  - Inter-Provider Backbones Option C (inter-AS VPN Option C)
    - PEs or route reflectors (RRs) advertise labeled VPNv4 routes to each other using multi-hop MP-EBGP. This solution is also called multi-hop EBGP redistribution of labeled VPNv4 routes.
    - The configuration is complex because ASBRs do not maintain or advertise VPNv4 routes.

HUAWEI

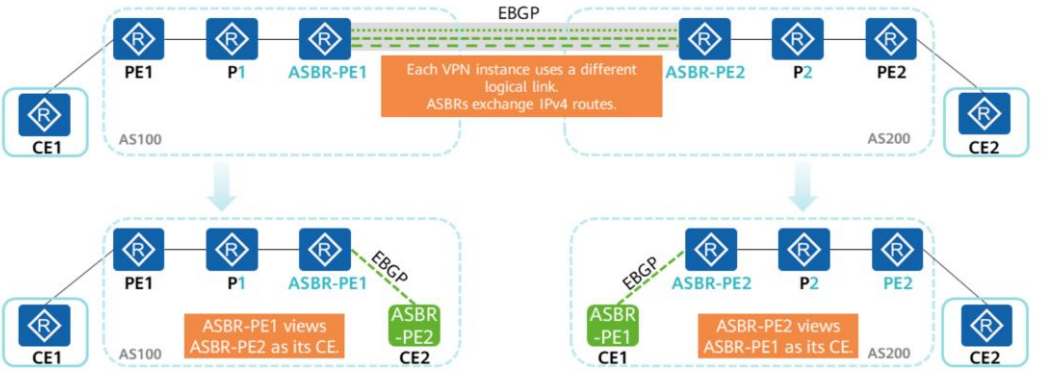
Contents

1. Background of Inter-AS VPN Solutions
2. **Fundamentals and Configurations of Inter-AS VPN Option A**
   - Fundamentals
   - Basic Configurations
3. Fundamentals and Configurations of Inter-AS VPN Option B
4. Fundamentals and Configurations of Inter-AS VPN Option C

HUAWEI

Inter-AS VPN Option A Overview

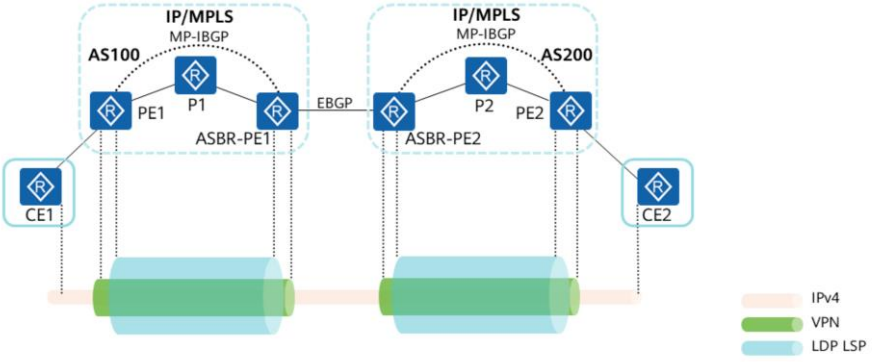
- As a basic BGP/MPLS IP VPN application in the inter-AS scenario, Option A does not need special configurations and MPLS does not need to run between ASBRs. In this mode, ASBRs of two ASs directly connect to each other and function as PEs in the ASs. Each ASBR views the peer ASBR as its CE and advertises IPv4 routes to the peer ASBR using EBGP.



HUAWEI

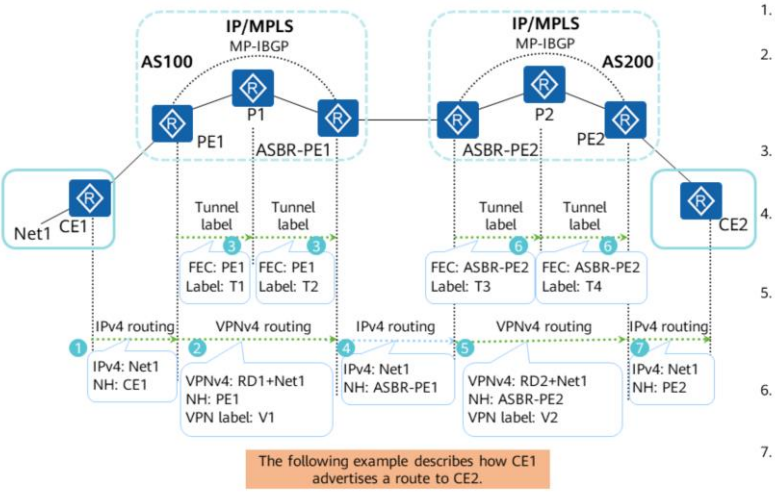
Inter-AS VPN Option A Topology

- Two ASBR-PEs are connected through multiple physical interfaces (or sub-interfaces), each interface is bound to a VPN, and each ASBR-PE views the peer ASBR-PE as a CE. Therefore, the interfaces connecting the two ASBR-PEs need to be bound to VRFs, and VPNv4 routes need to be converted into common IPv4 routes using EBGP and transmitted from one AS to another AS. Therefore, the two ASBRs need to be interconnected, but MPLS does not need to be enabled.



HUAWEI

Inter-AS VPN Option A: Control Plane

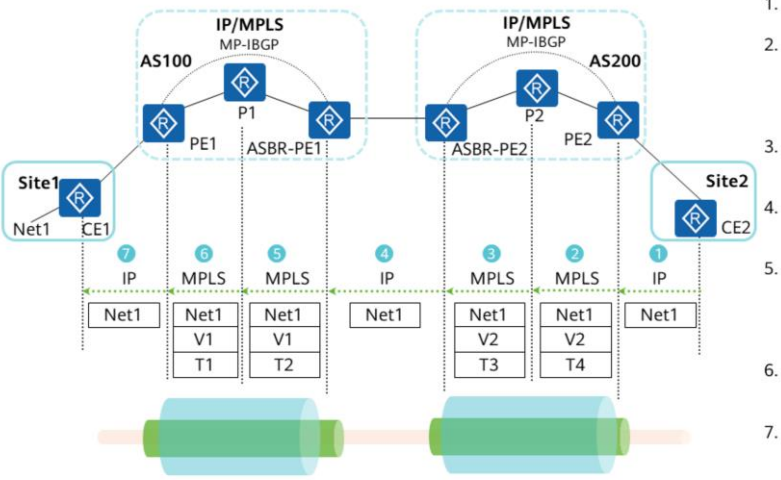


1. CE1 sets the next hop of an IPv4 route to itself and advertises the route to PE1.
2. PE1 converts the IPv4 route into a VPNv4 route, sets the next hop of the route to PE1, allocates a VPN label V1 to the route, and advertises the route to ASBR-PE1.
3. PE1 and P1 allocate tunnel labels T1 and T2 respectively to the routes destined for PE1.
4. ASBR-PE1 converts the VPNv4 route into an IPv4 route, sets the next hop to ASBR-PE1, and advertises the route to ASBR-PE2.
5. ASBR-PE2 converts the IPv4 route into a VPNv4 route, sets the next hop of the route to ASBR-PE2, allocates a VPN label V2 to the route, and advertises the route to PE2.
6. ASBR-PE2 and P2 allocate tunnel labels T3 and T4 respectively to the routes destined for ASBR-PE2.
7. PE2 converts the VPNv4 route into an IPv4 route, sets the next hop to PE2, and advertises the route to CE2.

HUAWEI

- The numbers in this example are only for ease of understanding, and do not represent the actual processing sequence on the devices.

Inter-AS VPN Option A: Forwarding Plane



1. CE2 sends an IP packet destined for Net1 to PE2.
2. After PE2 receives the IP packet, it encapsulates the packet with a VPN label V2 and then an outer label T4 and forwards the packet to P2.
3. P2 swaps the outer label T4 with T3 and forwards the packet to ASBR-PE2.
4. ASBR-PE2 removes all labels from the packet and forwards the packet to ASBR-PE1.
5. After ASBR-PE1 receives the IP packet, it encapsulates the packet with a VPN label V1 and then an outer label T2 and forwards the packet to P1.
6. P1 swaps the outer label T2 with T1 and forwards the packet to PE1.
7. PE1 removes all labels from the packet and forwards the packet to CE1.

HUAWEI

- The LDP PHP behavior is not considered in data forwarding.

Inter-AS VPN Option A Characteristics

- Inter-AS VPN Option A is easy to configure and understand.
- When there are a small number of VPNs on a network, inter-AS VPN Option A is applicable.
- An ASBR views the peer ASBR as a CE and uses a VRF interface to connect to the peer ASBR. ASBRs in two ASs exchange VPNv4 routes.
- At the data layer, VPN user data is exchanged between ASBRs in two ASs in the form of IP packets, that is, packets do not carry any label header.
- The scalability of inter-AS VPN Option A is poor. ASBRs need to manage all VPN routes, and a VPN instance needs to be configured for each VPN. This results in numerous VPN-IPv4 routes on the ASBRs. In addition, because common IP forwarding is implemented between ASBRs, each inter-AS VPN requires a different interface, which can be a sub-interface, physical interface, or bundled logical interface. This poses high requirements for ASBRs. If a VPN spans multiple ASs, the intermediate ASs must support VPN services. This requires complex configurations and greatly affects the intermediate ASs. If only a few inter-AS VPN instances are used, inter-AS VPN Option A is recommended.

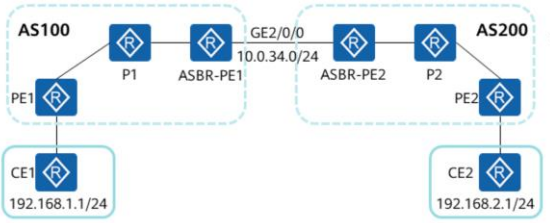
HUAWEI

Contents

1. Background of Inter-AS VPN Solutions
2. **Fundamentals and Configurations of Inter-AS VPN Option A**
   - Fundamentals
   - Basic Configurations
3. Fundamentals and Configurations of Inter-AS VPN Option B
4. Fundamentals and Configurations of Inter-AS VPN Option C

HUAWEI

Inter-AS VPN Option A Configuration Example (1)



- CE1 and CE2 belong to the same VPN named **vpna**. CE1 and CE2 need to communicate with each other using Option A.

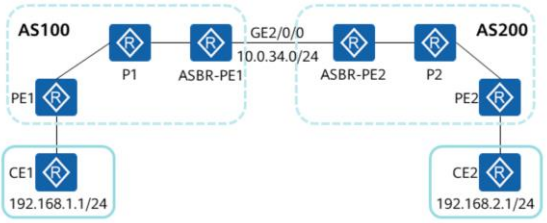
Device	Loopback0	RD	RT
PE1	10.0.1.1/32	100:1	1:1
ASBR-PE1	10.0.3.3/32	100:3	1:1
ASBR-PE2	10.0.4.4/32	200:4	2:2
PE2	10.0.6.6/32	200:6	2:2

Data plan

- Configuration roadmap:
  - Configure basic MPLS capabilities and MPLS LDP on the MPLS backbone network to establish LDP LSPs in each AS. (The configuration details are not provided here.)
  - Establish an MP-IBGP peer relationship between the PE and ASBR-PE in each AS to exchange VPN routing information. (The configuration details are not provided here.)
  - Configure a VPN instance on the PE connected to the CE in each AS and establish EBGP peer relationships between the PE and CE to exchange VPN routing information. (The configuration details are not provided here.)
  - Create a VPN instance on each ASBR-PE, bind the instance to the interface connected to the other ASBR-PE (viewed as a CE), and establish an EBGP peer relationship between the ASBR-PEs to exchange VPN routing information.

HUAWEI

Inter-AS VPN Option A Configuration Example (2)



1. Take ASBR-PE1 as an example to create a VPN instance and bind it to the interface connected to ASBR-PE2.

```
[ASBR-PE1] ip vpn-instance vpna
[ASBR-PE1-vpn-instance-vpna] ipv4-family
[ASBR-PE1-vpn-instance-vpna-af-ipv4] route-distinguisher 100:3
[ASBR-PE1-vpn-instance-vpna-af-ipv4] vpn-target 1:1 both
[ASBR-PE1-vpn-instance-vpna-af-ipv4] quit
[ASBR-PE1-vpn-instance-vpna] quit
[ASBR-PE1] interface gigabitethernet 2/0/0
[ASBR-PE1-GigabitEthernet2/0/0] ip binding vpn-instance vpna
[ASBR-PE1-GigabitEthernet2/0/0] ip address 10.0.34.3 24
[ASBR-PE1-GigabitEthernet2/0/0] quit
```

2. Take ASBR-PE1 as an example to establish an EBGP peer relationship between ASBR-PE1 and ASBR-PE2.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] ipv4-family vpn-instance vpna
[ASBR-PE1-bgp-vpna] peer 10.0.34.4 as-number 200
[ASBR-PE1-bgp-vpna] import-route direct
[ASBR-PE1-bgp-vpna] quit
[ASBR-PE1-bgp] quit
```

ASBR-PE1 views ASBR-PE2 as a CE and establishes an EBGP peer relationship with ASBR-PE2 in the VPN instance address family view to transmit IPv4 routes.

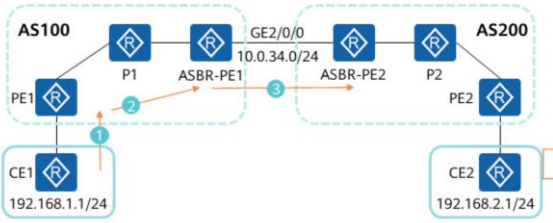
3. Take ASBR-PE1 as an example to establish an MP-IBGP peer relationship between ASBR-PE1 and PE1.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] peer 10.0.1.1 as-number 100
[ASBR-PE1-bgp] peer 10.0.1.1 connect-interface loopback 0
[ASBR-PE1-bgp] ipv4-family vpnv4
[ASBR-PE1-bgp-af-vpnv4] peer 10.0.1.1 enable
[ASBR-PE1-bgp-af-vpnv4] quit
[ASBR-PE1-bgp] quit
```

ASBR-PE1 establishes a VPNv4 peer relationship with PE1 to transmit VPNv4 routes.

HUAWEI

Verifying the Configuration (1)



1. Check CE1's routing information on PE1.

```
<PE1>display bgp vpnv4 all routing-table
 Total number of routes from all PE: 1
 Route Distinguisher: 100:1
     Network         NextHop        MED        LocPrf    PrefVal Path/Ogn
 *>  192.168.1.0     10.0.11.1      0                    0         65000i
 VPN-Instance vpna, Router ID 10.0.12.1:
 Total Number of Routes: 1
     Network         NextHop        MED        LocPrf    PrefVal Path/Ogn
 *>  192.168.1.0     10.0.11.1      0                    0         65000i
```

The next hop is CE1.

2. Check CE1's routing information on ASBR-PE1.

```
<ASBR-PE1>display bgp vpnv4 all routing-table 192.168.1.0
Total routes of Route Distinguisher(100:1): 1
 BGP routing table entry information of 192.168.1.0/24:
 Label information (Received/Applied): 1026/NULL
 From: 10.0.1.1 (10.0.1.1)
 Route Duration: 00h05m07s
 Relay IP Nexthop: 10.0.23.2
Relay token: 0x1
 Original nexthop: 10.0.1.1
 Ext-Community:RT <1 : 1>
 AS-path 65000, origin igp, MED 0, localpref 100, pref-val 0, valid,
internal, best, select, pre 255, IGP cost 2
```

The next hop is PE1.

3. Check CE1's routing information on ASBR-PE2.

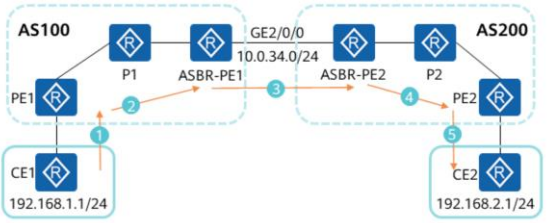
```
<ASBR-PE2>display bgp vpnv4 all routing-table 192.168.1.0
 Total routes of Route Distinguisher(100:4): 1
 BGP routing table entry information of 192.168.1.0/24:
 Label information (Received/Applied): NULL/1027
 From: 10.0.34.3 (10.0.3.3)
 Route Duration: 00h13m13s
 Direct Out-interface: GigabitEthernet0/0/1
 Original nexthop: 10.0.34.3
 Ext-Community:RT <2 : 2>
```

The label of the received route is null, indicating that the route is an IPv4 route.

The next hop is ASBR-PE1.

HUAWEI

Verifying the Configuration (2)



Ping 192.168.1.1 from CE2.

```
<CE2>ping -a 192.168.2.1 192.168.1.1
  PING 192.168.1.1: 56  data bytes, press CTRL_C to break
    Reply from 192.168.1.1: bytes=56 Sequence=1  ttl=250 time=100 ms
    Reply from 192.168.1.1: bytes=56 Sequence=2  ttl=250 time=60 ms
    Reply from 192.168.1.1: bytes=56 Sequence=3  ttl=250 time=50 ms
    Reply from 192.168.1.1: bytes=56 Sequence=4  ttl=250 time=60 ms
    Reply from 192.168.1.1: bytes=56 Sequence=5  ttl=250 time=60 ms
```

4. Check CE1's routing information on PE2.

```
<PE2>display bgp vpnv4 all routing-table 192.168.1.0
 Total routes of Route Distinguisher(100:4): 1
 BGP routing table entry information of 192.168.1.0/24:
 Label information (Received/Applied): 1027/NULL
 From: 10.0.4.4 (10.0.4.4)
 Relay IP Nexthop: 10.0.56.5
 Relay IP Out-Interface: GigabitEthernet0/0/1
 Relay Tunnel Out-Interface: GigabitEthernet0/0/1
 Relay token: 0x1
 Original nexthop: 10.0.4.4
 Qos information : 0x0
 Ext-Community:RT <2 : 2>
```

ASBR-PE2 allocates label 1027 to the route.

5. Check CE1's routing information on CE2.

```
<CE2>display bgp routing-table 192.168.1.0
 BGP local router ID : 10.0.26.2
 Local AS number : 65001
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 192.168.1.0/24:
 From: 10.0.26.6 (10.0.6.6)
 Direct Out-interface: GigabitEthernet0/0/0
 Original nexthop: 10.0.26.6
 AS-path 200 100 65000, origin igp, pref-val 0, valid, external, best, select,
```

The next hop is PE2.

HUAWEI

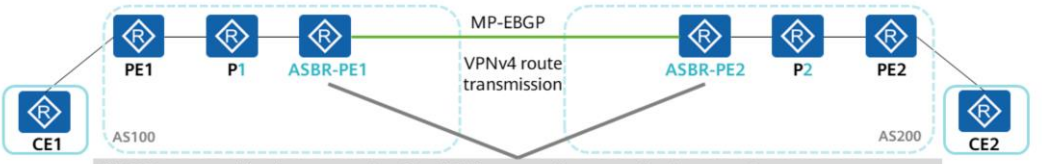
Contents

1. Background of Inter-AS VPN Solutions
2. Fundamentals and Configurations of Inter-AS VPN Option A
3. **Fundamentals and Configurations of Inter-AS VPN Option B**
   - Fundamentals
   - Basic Configurations
4. Fundamentals and Configurations of Inter-AS VPN Option C

HUAWEI

Inter-AS VPN Option B Overview

- Compared with Option A, Option B does not require VPN instances to be created on ASBR-PEs and does not require interfaces to be bound.
- In Option B, two ASBRs use MP-EBGP to exchange VPNv4 routes received from the PEs in their respective ASs.
- By default, a PE stores only the VPN routes that match the VPN targets of its local VPN instances. Therefore, ASBRs can be configured not to filter routes based on RTs.
- On a large-scale network, RRs can be deployed to transmit user-side VPN routes.

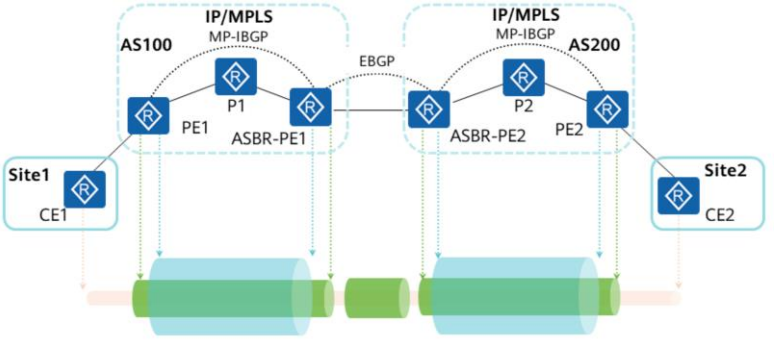


If ASBRs are not directly connected to CEs, VPN instances do not need to be created.
To disable VPN target-based filtering for received VPN routes, run the **undo policy vpn-target** command.
Interfaces connecting two ASBRs must support MPLS forwarding.

HUAWEI

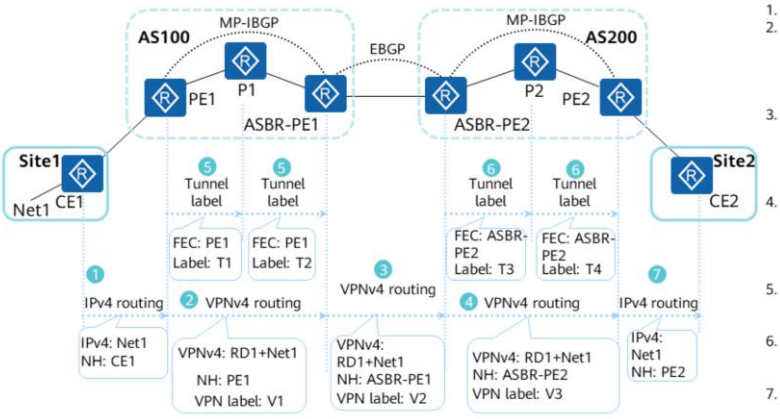
Inter-AS VPN Option B Topology

- A PE uses MP-IBGP to advertise a VPNv4 route to the ASBR-PE or VPN RR (with the ASBR-PE as the client) in the same AS. The ASBR-PE uses MP-EBGP to advertise the VPNv4 route to the peer ASBR-PE in another AS, and then the peer ASBR-PE advertises the VPNv4 route to the PE in the same AS.



HUAWEI

Inter-AS VPN Option B: Control Plane Without RRs

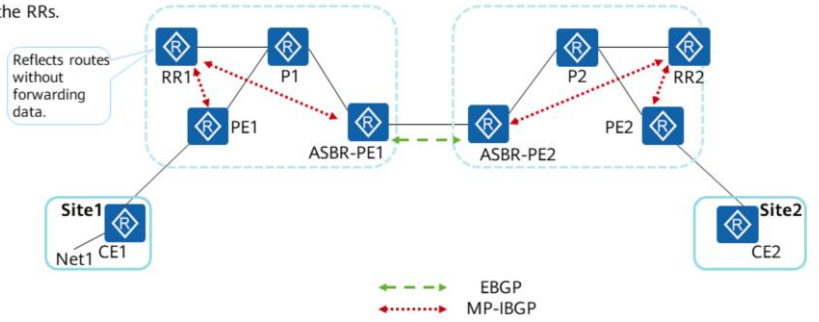


1. CE1 advertises an IPv4 route to PE1.
2. PE1 converts the IPv4 route into a VPNv4 route, sets the next hop of the route to PE1, allocates the VPN label V1 to the route, and advertises the route to ASBR-PE1.
3. ASBR-PE1 changes the next hop of the route to itself, allocates a new VPN label V2 to the route, and uses MP-EBGP to advertise the route for the prefix Net1 to ASBR-PE2.
4. ASBR-PE2 sets the next hop of the route to itself, allocates the VPN label V3 to the route, and uses MP-IBGP to advertise the route for the prefix Net1 to PE2.
5. PE1 and P1 allocate tunnel labels T1 and T2 respectively to the routes destined for PE1.
6. ASBR-PE2 and P2 allocate tunnel labels T3 and T4 respectively to the routes destined for ASBR-PE2.
7. PE2 converts the VPNv4 route into an IPv4 route, sets the next hop to PE2, and advertises the route to CE2.

HUAWEI

Inter-AS VPN Option B: Control Plane with RRs

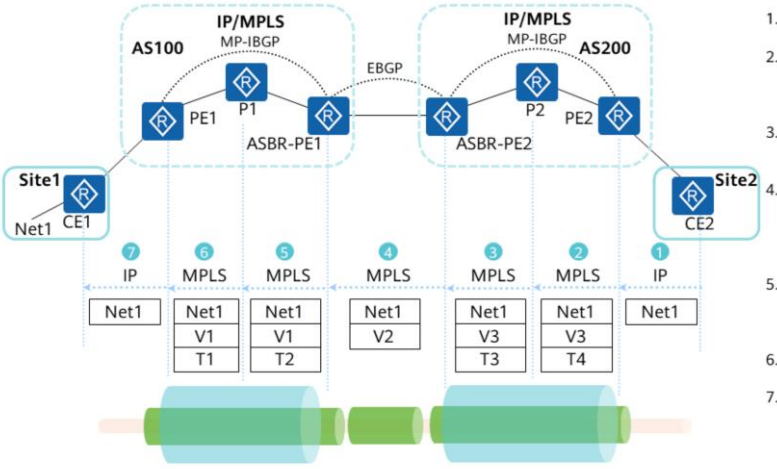
- When there are a large number of VPN instances, RRs can be deployed. As shown in the figure, the PEs and ASBRs in ASs establish MP-BGP peer relationships only with RRs, and the RRs are responsible for reflecting routes. The PEs and ASBRs do not need to establish BGP peer relationships with each other.
- The RRs are responsible only for transmitting VPNv4 routes on the control plane. During data forwarding, traffic does not pass through the RRs.



HUAWEI

- In the inter-AS VPN scenario, it is recommended that independent RRs be deployed to transmit routes only without forwarding traffic.

Inter-AS VPN Option B: Forwarding Plane



1. CE2 sends an IP packet destined for Net1 to PE2.
2. After PE2 receives the IP packet, it encapsulates the packet with a VPN label V3 and then an outer label T4 and forwards the packet to P2.
3. P2 swaps the outer label T4 with T3 and forwards the packet to ASBR-PE2.
4. ASBR-PE2 removes the outer label, swaps the VPN label V3 with V2, and forwards the packet to ASBR-PE1. In this case, the packet carries only one VPN label.
5. ASBR-PE1 swaps the VPN label V2 with V1, adds the outer label T2 to the packet, and forwards the packet to P1.
6. P1 swaps the outer label T2 with T1 and forwards the packet to PE1.
7. PE1 removes all labels from the packet and forwards the packet (a common IP packet) to CE1.

HUAWEI

Inter-AS VPN Option B Characteristics

- The advantage of Option B is that all traffic is forwarded by ASBRs. In this way, traffic is controllable, but the loads on the ASBRs are heavy because the ASBRs need to save a large number of VPNv4 routes. BGP routing policies, such as VPN target-based filtering policies, can be configured on ASBRs, so that ASBRs only save some of VPNv4 routes.
- The disadvantage is that VPN routing information is stored on and forwarded by ASBRs. If a large number of VPN routes exist, the overloaded ASBRs tend to become faulty points. Therefore, in scenarios where MP-EBGP is used, ASBRs that maintain VPN routing information generally do not perform IP forwarding on the public network.
- Option B is better than Option A when a large number of VRF instances need to communicate with each other.

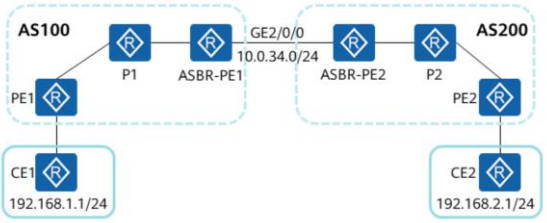
HUAWEI

Contents

1. Background of Inter-AS VPN Solutions
2. Fundamentals and Configurations of Inter-AS VPN Option A
3. **Fundamentals and Configurations of Inter-AS VPN Option B**
   - Fundamentals
   - Basic Configurations
4. Fundamentals and Configurations of Inter-AS VPN Option C

HUAWEI

Inter-AS VPN Option B Configuration Example (1)



- CE1 and CE2 belong to the same VPN named **vpna**. CE1 and CE2 need to communicate with each other using Option B.

Device	Loopback0	RD	RT
PE1	10.0.1.1/32	100:1	1:1
ASBR-PE1	10.0.3.3/32	/	/
ASBR-PE2	10.0.4.4/32	/	/
PE2	10.0.6.6/32	200:6	1:1

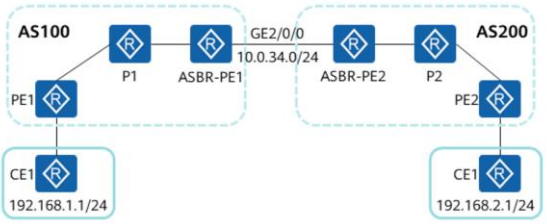
Data plan

RT values of PE1 and PE2 must match.

- Configuration roadmap
  - Configure basic MPLS capabilities and MPLS LDP on the MPLS backbone network to establish LDP LSPs in each AS. (The configuration details are not provided here.)
  - Establish an MP-IBGP peer relationship between the PE and ASBR-PE in each AS to exchange VPN routing information. (The configuration details are not provided here.)
  - Configure a VPN instance on the PE connected to the CE in each AS and establish EBGP peer relationships between the PE and CE to exchange VPN routing information. (The configuration details are not provided here.)
  - Enable MPLS on the interfaces connecting the ASBRs, establish an MP-EBGP peer relationship between the ASBRs, and configure the ASBRs not to filter received VPNv4 routes based on RTs.

HUAWEI

Inter-AS VPN Option B Configuration Example (2)



1. Take ASBR-PE1 as an example to establish an MP-EBGP peer relationship between ASBR-PE1 and ASBR-PE2, configure ASBR-PE1 not to filter received VPNv4 routes based on the VPN target, and enable ASBR-PE1 to allocate labels based on the next hop.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] peer 10.0.34.4 as-number 200
[ASBR-PE1-bgp] ipv4-family vpnv4
[ASBR-PE1-bgp-af-vpnv4] peer 10.0.34.4 enable
[ASBR-PE1-bgp-af-vpnv4] undo policy vpn-target
[ASBR-PE1-bgp-af-vpnv4] apply-label per-nexthop
[ASBR-PE1-bgp-af-vpnv4] quit
[ASBR-PE1-bgp] quit
```

By default, an ASBR allocates a label to each VPNv4 route to be advertised to its MP-BGP peers. After this command is run, ASBR-PE1 allocates a label to the routes with the same next hop and outgoing label.

2. Take ASBR-PE1 as an example to enable MPLS on the interfaces connecting ASs.

```
[ASBR-PE1] interface gigabitethernet 2/0/0
[ASBR-PE1-GigabitEthernet2/0/0] mpls
[ASBR-PE1-GigabitEthernet2/0/0] quit
```

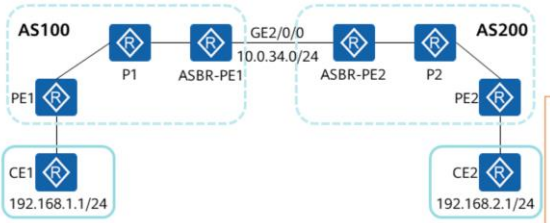
To differentiate between VPN instances, ASBRs advertise VPNv4 routes to each other. Because the VPNv4 routes carry label information, the interfaces connecting ASBRs must support label forwarding.

3. Take ASBR-PE1 as an example to establish an MP-IBGP peer relationship between ASBR-PE1 and PE1.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] peer 10.0.1.1 as-number 100
[ASBR-PE1-bgp] peer 10.0.1.1 connect-interface loopback 0
[ASBR-PE1-bgp] ipv4-family vpnv4
[ASBR-PE1-bgp-af-vpnv4] peer 10.0.1.1 enable
[ASBR-PE1-bgp-af-vpnv4] quit
[ASBR-PE1-bgp] quit
```


HUAWEI

Verifying the Configuration (1)



- The process of transmitting intra-area routing information is not described here. This example describes only the key procedure of transmitting routing information.

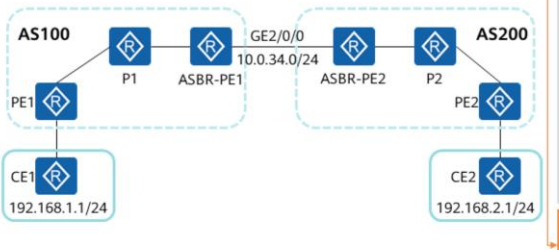
1. Check CE1's routing information on ASBR-PE1.

```
<ASBR-PE1>display bgp vpnv4 all routing-table 192.168.1.0
 BGP local router ID : 10.0.3.3
 Local AS number : 100
 Total routes of Route Distinguisher(100:1): 1
 BGP routing table entry information of 192.168.1.0/24:
 Label information (Received/Applied): 1026/1026
 From: 10.0.1.1 (10.0.1.1)
 Route Duration: 00h07m32s
 Relay IP Nexthop: 10.0.23.2
 Relay IP Out-Interface: GigabitEthernet0/0/0
 Relay Tunnel Out-Interface: GigabitEthernet0/0/0
 Relay token: 0x2
 Original nexthop: 10.0.1.1
 Qos information : 0x0
 Ext-Community:RT <1 : 1>     # The export RT value must match the
import RT value of PE2.
 AS-path 65000, origin igp, MED 0, localpref 100, pref-val 0, valid,
internal, best, select, pre 255, IGP cost 2
 Advertised to such 1 peers:
    10.0.34.4
```

PE1 allocates the label 1026 to the route. When routes are transmitted between ASs, ASBR-PE1 changes the next hop to itself. Therefore, ASBR-PE1 re-allocates the label 1026 when advertising routes to ASBR-PE2. MPLS labels are local to a device, so different devices can allocate the same label.

HUAWEI

Verifying the Configuration (2)



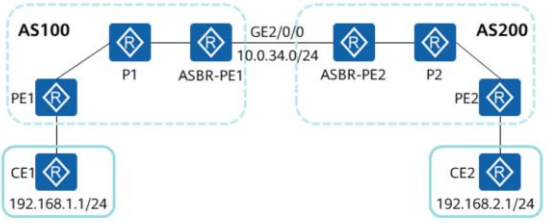
2. Check CE1's routing information on ASBR-PE2.

```
<ASBR-PE2>display bgp vpnv4 all routing-table 192.168.1.0
 BGP local router ID : 10.0.4.4
 Local AS number : 200
 Total routes of Route Distinguisher(100:1): 1
 BGP routing table entry information of 192.168.1.0/24:
 Label information (Received/Applied): 1026/1032
 From: 10.0.34.3 (10.0.3.3)
 Route Duration: 00h02m18s
 Relay Tunnel Out-Interface: GigabitEthernet0/0/1
 Relay token: 0x1
 Original nexthop: 10.0.34.3
 Qos information : 0x0
 Ext-Community:RT <1 : 1>
 AS-path 100 65000, origin igp, pref-val 0, valid, external, best, select, pre
2
55
 Advertised to such 1 peers:
    10.0.6.6
```

ASBR-PE1 allocates the label 1026 to the route. When advertising the route to PE2, ASBR-PE2 changes the next hop of the route to itself and re-allocates the label 1032 to the route.

HUAWEI

Verifying the Configuration (3)



3. Ping 192.168.1.1 from CE2.

```
<CE2>ping -a 192.168.2.1 192.168.1.1
  PING 192.168.1.1: 56  data bytes, press CTRL_C to break
    Reply from 192.168.1.1: bytes=56 Sequence=1 ttl=250 time=60 ms
    Reply from 192.168.1.1: bytes=56 Sequence=2 ttl=250 time=50 ms
    Reply from 192.168.1.1: bytes=56 Sequence=3 ttl=250 time=60 ms
    Reply from 192.168.1.1: bytes=56 Sequence=4 ttl=250 time=60 ms
```


HUAWEI

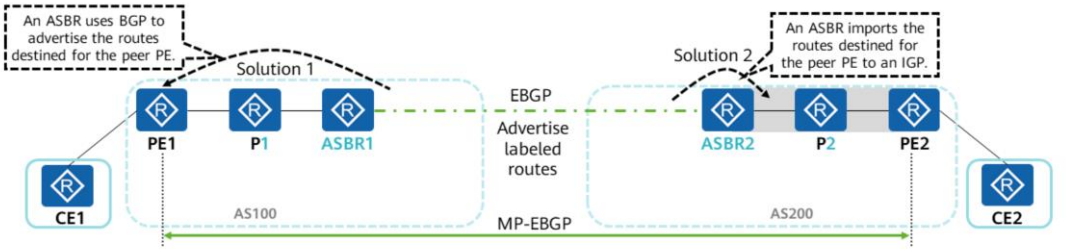
Contents

1. Background of Inter-AS VPN Solutions
2. Fundamentals and Configurations of Inter-AS VPN Option A
3. Fundamentals and Configurations of Inter-AS VPN Option B
4. **Fundamentals and Configurations of Inter-AS VPN Option C**
   - Fundamentals
   - Basic Configurations

HUAWEI

Inter-AS VPN Option C Overview

- In Option C, ASBRs do not store VPNv4 routes or advertise VPNv4 routes to each other. This is different from the implementation in Option A and Option B.
- In Option C, PEs in different ASs establish a multi-hop EBGP connection with each other to exchange VPNv4 routes. Two solutions are available to enable PEs to communicate with each other:
  - Solution 1: An ASBR uses BGP to send the routes destined for a PE in another AS to the local PE.
  - Solution 2: An ASBR imports the routes destined for a PE in another AS to an IGP.
- ASBRs do not have VPNv4 routes. To prevent routing blackholes during packet forwarding, PEs need to transmit packets over tunnels so that non-PE devices are unaware of VPN encapsulation information. Therefore, both the solutions require ASBRs to advertise labeled routes to set up outer forwarding tunnels between ASs.

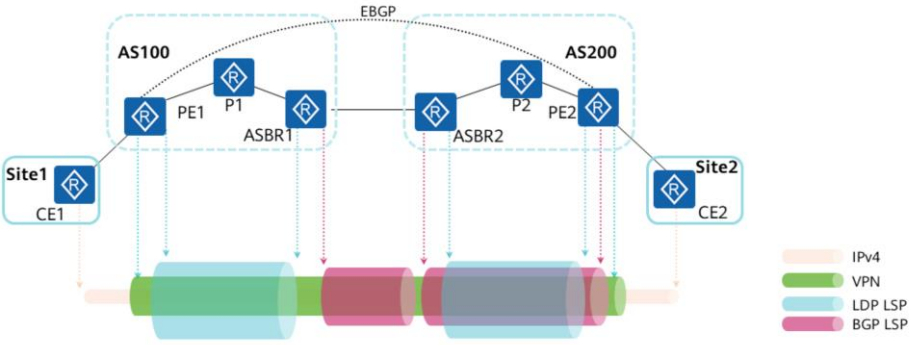


HUAWEI

Solution 1 Solution 2

Inter-AS VPN Option C Topology (Solution 1)

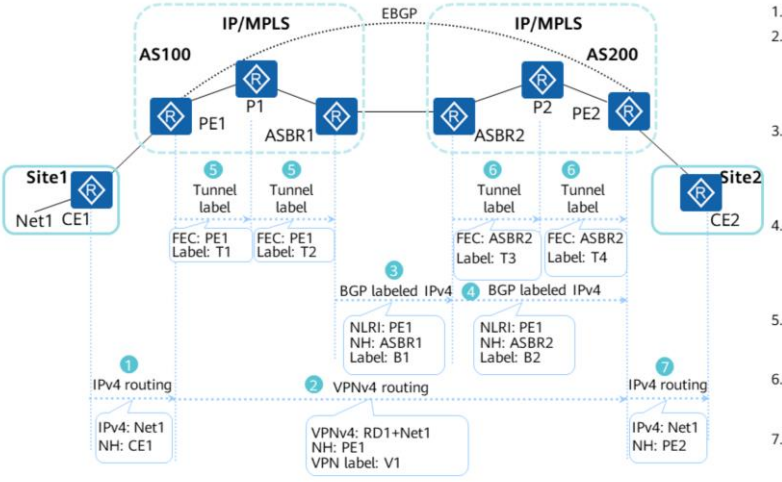
- ASBRs do not maintain or advertise VPNv4 routes. ASBRs only need to maintain labeled routes destined for PEs and use EBGP to advertise these labeled routes to the peer ASBRs in other ASs. ASBRs in the transit AS also need to use EBGP to advertise labeled IPv4 routes. In this way, a VPN LSP needs to be established between the ingress and egress PEs so that a multi-hop MP-EBGP connection can be established between the PEs to advertise VPNv4 routes.
- To further improve performance, a multi-hop MP-EBGP session can be established between VPN RRs in different ASs. When these VPN RRs advertise VPNv4 routes, the next-hop information of these routes remains unchanged. PEs establish MP-IBGP sessions only with VPN RRs.



HUAWEI

- If the P device of each AS knows the routes to the PEs of other ASs, the data forwarding process will be relatively simple. But if the P device does not know these routes, then when a PE receives VPN data from a CE, the PE will add three labels. The bottom label is the VPN label that is allocated by the peer PE and associated with VPN routes, the middle label is the label allocated by an ASBR and associated with the route to the peer PE, and the outer label is the label associated with the route to the next-hop ASBR.
- Note: For convenience, as shown in the figure above, a symmetric LSP is used for illustration, but in fact, in the working process of the control plane and the data plane, the LSP of the ASs at both ends is asymmetric.

Inter-AS VPN Option C (Solution 1): Without RRs

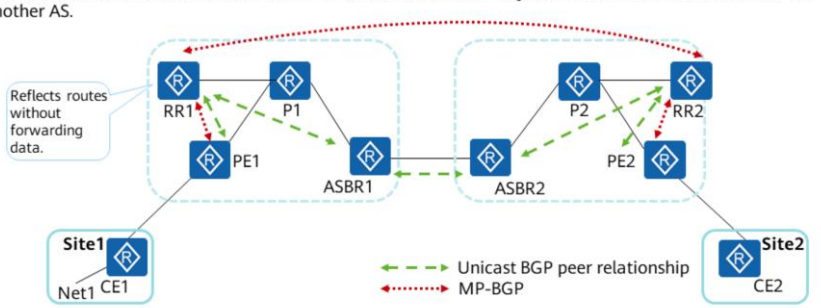


1. CE1 advertises an IPv4 route to PE1.
2. PE1 converts the IPv4 route into a VPNv4 route, sets the next hop of the route to PE1, allocates the VPN label V1 to the route, and advertises the route to PE2.
3. ASBR1 advertises a labeled IPv4 route destined for PE1 to ASBR2 through an EBGP session. The next hop of the route is ASBR1, and the label is the BGP label B1.
4. ASBR2 advertises a labeled IPv4 route destined for PE1 to PE2 through a BGP session. The next hop of the route is ASBR2, and the label is the BGP label B2.
5. PE1 and P1 allocate tunnel labels T1 and T2 respectively to the routes destined for PE1.
6. ASBR2 and P2 allocate tunnel labels T3 and T4 respectively to the routes destined for ASBR2.
7. PE2 converts the VPNv4 route into an IPv4 route, sets the next hop to PE2, and advertises the route to CE2.

HUAWEI

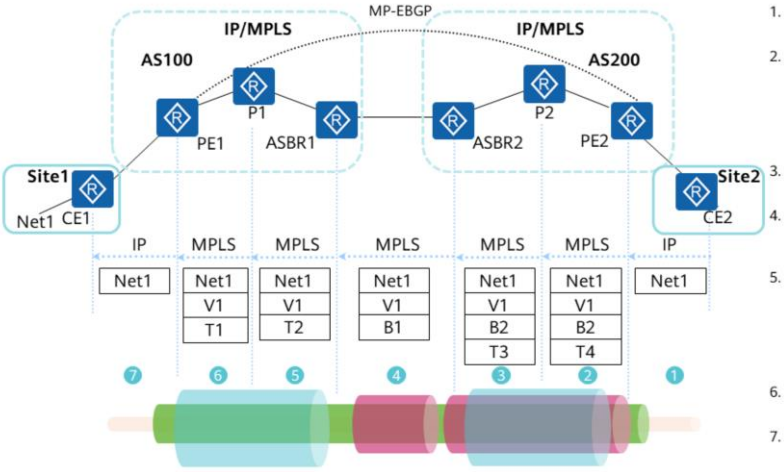
Inter-AS VPN Option C (Solution 1): With RRs

- The local PE establishes a VPNv4 peer relationship only with the local RR, and the local RR establishes a VPNv4 peer relationship with the peer RR to transmit inter-AS VPN routes.
- ASBRs and PEs establish BGP unicast IPv4 peer relationships with RRs.
  - An ASBR learns the route destined for the peer RR's loopback interface from the peer ASBR and transmits the route to the local RR so that the local RR can establish a VPNv4 peer relationship with the peer RR.
  - An ASBR learns the routes destined for the peer RR and PE's loopback interfaces from the peer ASBR and transmits the routes to the local RR, which then transmits the routes to the local PE. In this way, the local PE can establish a BGP LSP with the peer PE in another AS.



HUAWEI

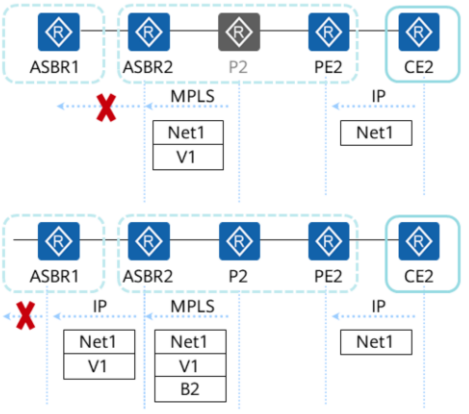
Inter-AS VPN Option C (Solution 1): Forwarding Plane



1. CE2 sends an IP packet destined for Net1 to PE2.
2. After receiving the IP packet, PE2 encapsulates the packet with the following three labels in sequence: the VPN label V1, the BGP label B2 allocated by ASBR2, and the outer label T4.
3. P2 swaps the outer label T4 with T3 and forwards the packet to ASBR2.
4. ASBR2 removes the outer label, swaps the BGP label B2 with B1, and forwards the packet to ASBR1.
5. After receiving the packet, ASBR1 removes the BGP label B1 and checks its forwarding table. It finds the label T2 associated with the route destined for PE1, adds T2 to the top of the label stack, and forwards the packet to P1.
6. P1 swaps the outer label T2 with T1 and forwards the packet to PE1.
7. PE1 removes all labels from the packet and forwards the packet to CE1.

HUAWEI

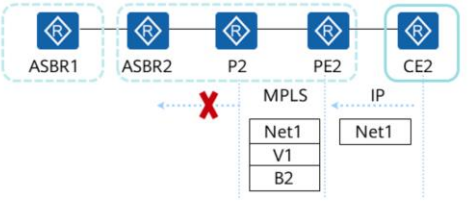
Inter-AS VPN Option C (Solution 1) FAQ (1)



- Why does ASBR2 allocate labels to the routes destined for PE1? (P2 is not considered at present.)
  - ASBR2 is unaware of VPN routes. Therefore, when a packet carrying only the VPN label (allocated by PE1) arrives at ASBR2, ASBR2 cannot identify the packet and discards it.
- Why does ASBR1 allocate labels to the routes destined for PE1?
  - ASBR1 is unaware of VPN routes. Therefore, when a packet carrying only the VPN label (allocated by PE1) arrives at ASBR1, ASBR1 cannot identify the packet and discards it.

HUAWEI

Inter-AS VPN Option C (Solution 1) FAQ (2)



- Why does P2 allocate labels to ASBR2?
  - P2 does not identify the routing information of PE1. Therefore, when a VPN packet carrying the outer BGP label (allocated by ASBR2) arrives at P2, P2 cannot identify the outer label B2 and discards the packet.

- Let's review the basic characteristics of VPN.
  - Private: It provides the same private network as a traditional private network. Resources between the VPN and the underlying bearer network are independent, that is, resources in a VPN are not used by users who do not belong to the VPN or non-VPN users. In addition, VPN security can be fully guaranteed to protect internal information of the VPN against external interference.
  - Virtual: VPN users communicate over a public network, which is also used by other non-VPN users. That is, a VPN user obtains a logical private network. A public network that carries a VPN is called a VPN backbone network.
- To achieve the two characteristics, only PEs are aware of VPN routing information. To prevent packet loss caused by routing blackholes during data forwarding, VPN data needs to be carried over various types of tunnels.
- Common tunneling technologies include MPLS, GRE, and Segment Routing over IPv6 (SRv6).

HUAWEI

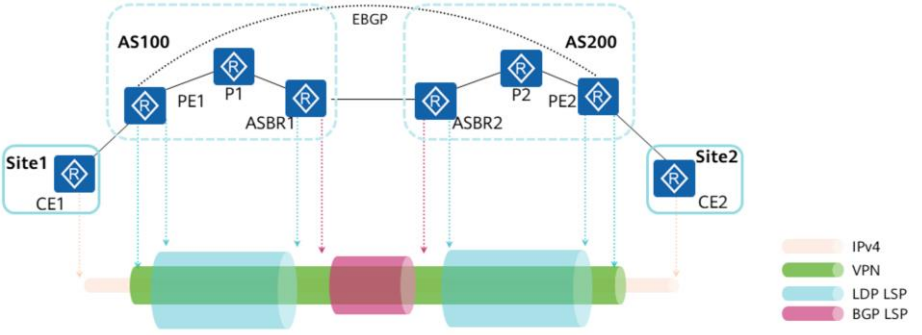
Inter-AS VPN Option C (Solution 2)

- Solution 2 is similar to solution 1.
- Three labels (VPN label, BGP LSP label, and tunnel LSP label) are used in solution 1, whereas only two labels (VPN label and tunnel label) are used in solution 2.
- In solution 1, a policy needs to be configured on the local ASBR to generate a new label. After the local ASBR receives a labeled BGP route from the peer ASBR, the local ASBR allocates a new label to the BGP route and advertises the route to the PE or RR in the same AS so that a complete BGP LSP can be established. In solution 2, MPLS needs to be configured on an ASBR to allocate labels to labeled BGP routes. On the PE in the same AS as the ASBR, you can view the LDP LSP to the peer PE instead of a BGP LSP.
- RRs can be deployed in both solution 1 and solution 2.

HUAWEI

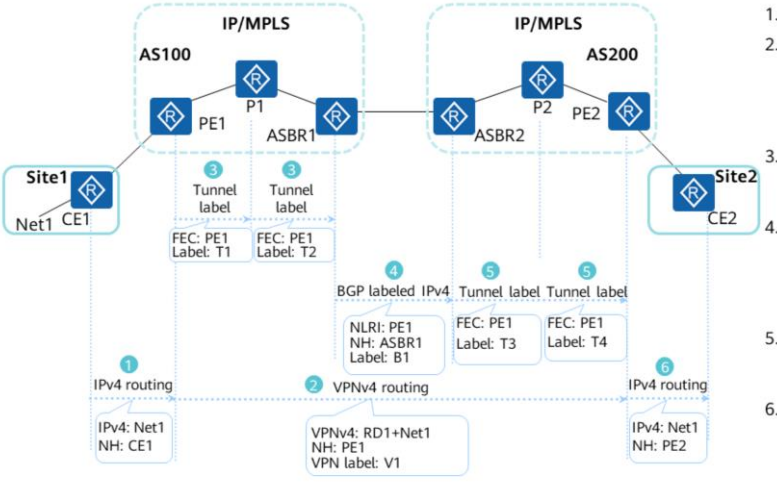
Inter-AS VPN Option C (Solution 2) Topology

- Solution 2: ASBRs do not maintain or advertise VPNv4 routes. ASBRs only need to maintain labeled routes destined for PEs and use EBGP to advertise these labeled routes to the peer ASBRs. After the peer ASBR receives a labeled BGP route, LDP triggers the generation of a label for the route and transmits the label between LDP peers in the same AS. Therefore, on a PE, you can view the LDP LSP to the peer PE.
- To further improve performance, a multi-hop MP-EBGP session can be established between VPN RRs in different ASs. A PE in an AS only needs to establish an MP-IBGP peer relationship with an RR in the AS. When advertising VPNv4 routes, the VPN RRs do not change the next-hop information. In this manner, when the peer PE forwards traffic, the traffic can be recursed to the correct tunnel.



HUAWEI

Inter-AS VPN Option C (Solution 2): Without RRs

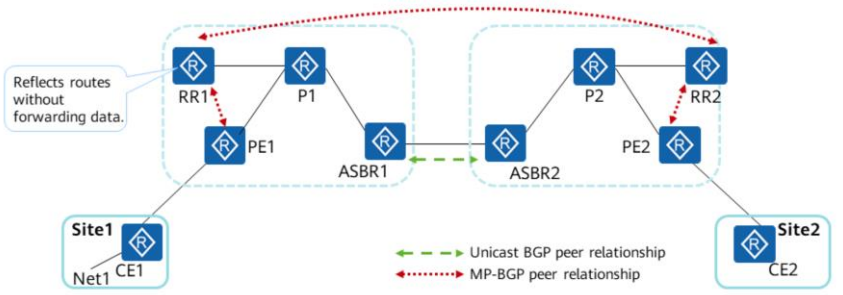


1. CE1 advertises an IPv4 route to PE1.
2. PE1 converts the IPv4 route into a VPNv4 route, sets the next hop of the route to PE1, allocates the VPN label V1 to the route, and advertises the route to PE2.
3. PE1 and P1 allocate tunnel labels T1 and T2 respectively to the routes destined for PE1.
4. ASBR1 advertises the labeled IPv4 route destined for PE1 to ASBR2 through an EBGP session. The next hop of the route is ASBR1, and the label is the BGP label B1.
5. ASBR2 and P2 allocate tunnel labels T3 and T4 respectively to the routes destined for PE1.
6. PE2 converts the VPNv4 route into an IPv4 route, sets the next hop to PE2, and advertises the route to CE2.

HUAWEI

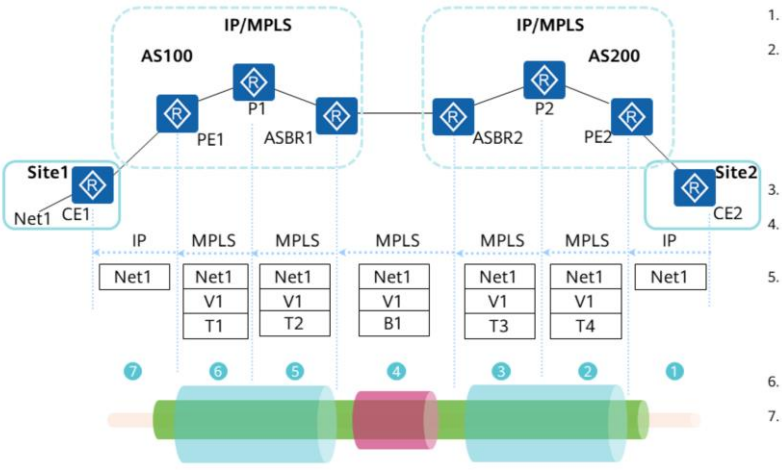
Inter-AS VPN Option C (Solution 2): Control Plane with RRs

- The local PE establishes a VPNv4 peer relationship only with the local RR, and the local RR establishes a VPNv4 peer relationship with the peer RR to transmit inter-AS VPN routes.
- The RRs are responsible only for transmitting VPNv4 routes on the control plane. Traffic on the forwarding plane does not pass through the RRs.



HUAWEI

Inter-AS VPN Option C (Solution 2) Forwarding Plane



1. CE2 sends an IP packet destined for Net1 to PE2.
2. After receiving the IP packet, PE2 encapsulates the packet with the VPN label V1. Because the next hop (PE1) of the packet destined for Net1 is not a directly connected neighbor, PE2 finds that the label of the packet destined for PE1 is T4 and encapsulates the packet with the label T4.
3. P2 swaps the outer label T4 with T3 and forwards the packet to ASBR2.
4. ASBR2 removes the outer label, swaps the label T3 with B1, and forwards the packet to ASBR1.
5. After receiving the packet, ASBR1 removes the label B1 and checks its forwarding table. It finds the label T2 associated with the route destined for PE1, adds T2 to the top of the label stack, and forwards the packet to P1.
6. P1 swaps the outer label T2 with T1 and forwards the packet to PE1.
7. PE1 removes all labels from the packet and forwards the packet to CE1.

HUAWEI

Inter-AS VPN Option C Characteristics

- VPN routes are directly exchanged between the ingress and egress PEs. The routes do not need to be stored or forwarded by intermediate devices.
- Only PEs maintain VPN routing information, and Ps and ASBRs are only responsible for packet forwarding. This means that the intermediate devices only need to support MPLS forwarding instead of MPLS VPN services. ASBRs are no longer bottlenecks. Option C, therefore, is suitable for VPNs spanning multiple ASs.
- MPLS VPN load balancing is easy to implement using Option C.
- The disadvantage of Option C is that it costs too much to manage an end-to-end connection between PEs.

HUAWEI

Comparison of Three Inter-AS VPN Solutions

Inter-AS VPN Solution	Description
Option A	• Easy configuration: MPLS is not required between ASBRs, and no special configuration is required for inter-AS connections. • Poor scalability: ASBRs need to manage all VPN routes, and a VPN instance needs to be configured for each VPN. This results in numerous VPNv4 routes on the ASBRs. In addition, because common IP forwarding is implemented between ASBRs, each inter-AS VPN requires a different interface, which can be a sub-interface, physical interface, or bundled logical interface. This poses high requirements for ASBRs. If a VPN spans multiple ASs, the intermediate ASs must support VPN services. This requires complex configurations and greatly affects the intermediate ASs. If only a few inter-AS VPN instances are used, Option A is recommended.
Option B	• Unlike Option A, Option B is not restricted by the number of links between ASBRs. • If a large number of VPN routes exist, the overloaded ASBRs tend to become faulty points. Therefore, in scenarios where MP-EBGP is used, ASBRs that maintain VPN routing information generally do not perform IP forwarding on the public network.
Option C	• VPN routes are directly exchanged between the ingress and egress PEs. The routes do not need to be stored or forwarded by intermediate devices. • Only PEs maintain VPN routing information, and Ps and ASBRs are only responsible for packet forwarding. This means that the intermediate devices only need to support MPLS forwarding instead of MPLS VPN services. ASBRs are no longer bottlenecks. Option C, therefore, is suitable for VPNs spanning multiple ASs. • MPLS VPN load balancing is easy to implement using Option C. • The disadvantage of Option C is that it costs too much to manage an end-to-end BGP LSP between PEs.

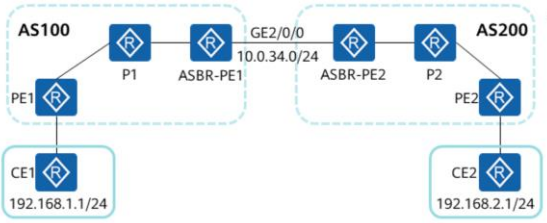
HUAWEI

Contents

1. Background of Inter-AS VPN Solutions
2. Fundamentals and Configurations of Inter-AS VPN Option A
3. Fundamentals and Configurations of Inter-AS VPN Option B
4. **Fundamentals and Configurations of Inter-AS VPN Option C**
   - Fundamentals
   - Basic Configurations

HUAWEI

Option C (Solution 1) Configuration Example (1)



- CE1 and CE2 belong to the same VPN named **vpna**. CE1 and CE2 need to communicate with each other using Option C (solution 1).

Device	Loopback0	RD	RT
PE1	10.0.1.1/32	100:1	1:1
ASBR-PE1	10.0.3.3/32	/	/
ASBR-PE2	10.0.4.4/32	/	/
PE2	10.0.6.6/32	200:6	1:1

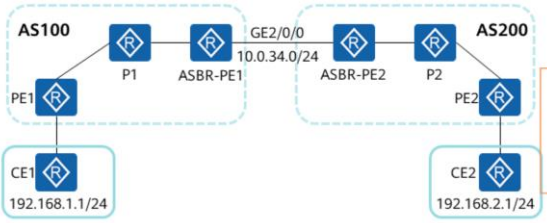
Data plan

RT values of PE1 and PE2 must match.

- Configuration roadmap
  - Configure basic MPLS capabilities and MPLS LDP on the MPLS backbone network to establish LDP LSPs in each AS. (The configuration details are not provided here.)
  - Configure a VPN instance on the PE connected to the CE in each AS and establish EBGP peer relationships between the PE and CE to exchange VPN routing information. (The configuration details are not provided here.)
  - Establish an MP-IBGP peer relationship between the PE and ASBR-PE in each AS to exchange labeled IPv4 routes.
  - Enable each ASBR-PE to exchange labeled IPv4 routes with the peer ASBR PE.
  - Configure route-policies on each ASBR-PE, so that it can allocate MPLS labels to the routes to be advertised to the peer ASBR and allocate new MPLS labels to the labeled IPv4 routes to be advertised to the PE in the same AS.
  - Establish an MP-EBGP peer relationship between PEs in different ASs and set the maximum number of hops between the PEs.

HUAWEI

Option C (Solution 1) Configuration Example (2)



1. On ASBR-PE1, create route-policies.

```
[ASBR-PE1] route-policy policy1 permit node 1
[ASBR-PE1-route-policy] apply mpls-label
[ASBR-PE1-route-policy] quit
[ASBR-PE1] route-policy policy2 permit node 1
[ASBR-PE1-route-policy] if-match mpls-label
[ASBR-PE1-route-policy] apply mpls-label
[ASBR-PE1-route-policy] quit
```

The route-policy **policy1** is used to advertise labeled routes to ASBR-PE2.

The route-policy **policy2** is used to allocate BGP labels to the labeled routes received from ASBR-PE2.

2. On ASBR-PE1, enable MPLS on the interface connected to ASBR-PE2.

```
[ASBR-PE1] interface gigabitethernet 2/0/0
[ASBR-PE1-GigabitEthernet2/0/0] mpls  # Interfaces between ASBRs support MPLS.
[ASBR-PE1-GigabitEthernet2/0/0] quit
```

3. On ASBR-PE1, apply the route-policy **policy1** to the routes advertised to ASBR-PE2 and enable the capability to exchange labeled IPv4 routes with ASBR-PE2.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] peer 10.0.34.4 route-policy policy1 export
[ASBR-PE1-bgp] peer 10.0.34.4 label-route-capability
[ASBR-PE1-bgp] quit
```

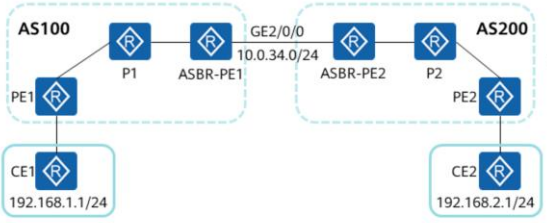
4. On ASBR-PE1, apply the route-policy **policy2** to the routes advertised to PE1 and enable the capability to exchange labeled IPv4 routes with PE1.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] peer 10.0.1.1 route-policy policy2 export
[ASBR-PE1-bgp] peer 10.0.1.1 label-route-capability
[ASBR-PE1-bgp] quit
```

In this example, the configuration of ASBR-PE1 is provided for your reference.

HUAWEI

Option C (Solution 1) Configuration Example (3)



5. Configure ASBR-PE1 to advertise the route destined for PE1's loopback interface to ASBR-PE2, which then advertises the route to PE2.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] network 10.0.1.1 32
[ASBR-PE1-bgp] quit
```

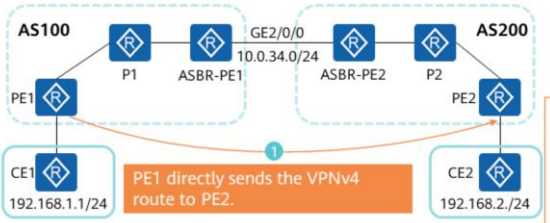
6. On PE1, enable it to establish an MP-EBGP peer relationship with PE2.

```
[PE1] bgp 100
[PE1-bgp] peer 10.0.6.6 as-number 200
[PE1-bgp] peer 10.0.6.6 connect-interface LoopBack 1
[PE1-bgp] peer 10.0.6.6 ebgp-max-hop 10
[PE1-bgp] ipv4-family vpnv4
[PE1-bgp-af-vpnv4] peer 10.0.6.6 enable
[PE1-bgp-af-vpnv4] quit
[PE1-bgp] quit
```

If there is no directly connected physical link between EBGP peers, the **peer ebgp-max-hop** command must be run to configure EBGP peers to establish TCP connections across multiple hops.

HUAWEI

Verifying the Configuration (1)



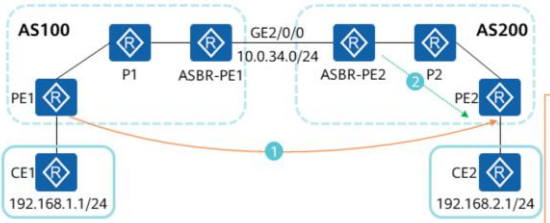
1. Check CE1's routing information on PE2.

```
<PE2>display bgp vpnv4 all routing-table 192.168.1.0
 BGP local router ID : 10.0.6.6
 Local AS number : 200
 Total routes of Route Distinguisher(100:1): 1
 BGP routing table entry information of 192.168.1.0/24:
 Label information (Received/Applied): 1029/NULL
 From: 10.0.1.1 (10.0.1.1)
 Route Duration: 00h00m14s
 Relay IP Nexthop: 10.0.56.5
 Relay IP Out-Interface: GigabitEthernet0/0/1
 Relay Tunnel Out-Interface: GigabitEthernet0/0/1
 Relay token: 0x9
 Original nexthop: 10.0.1.1
 Qos information : 0x0
 Ext-Community:RT <1 : 1>
 AS-path 100 65000, origin igp, pref-val 0, valid, external, best, select, pre
255, IGP cost 2
```

PE1 allocates the label 1029 to the route.

HUAWEI

Verifying the Configuration (2)



2. Check information about routes destined for 10.0.1.1 on PE2.

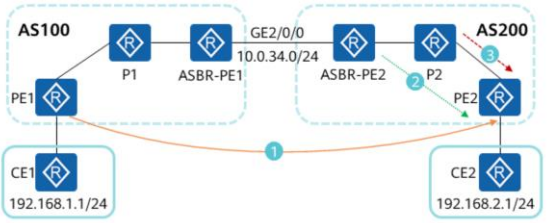
```
<PE2>display bgp routing-table 10.0.1.1
 BGP local router ID : 10.0.6.6
 Local AS number : 200
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 10.0.1.1/32:
 Label information (Received/Applied): 1032/NULL
 From: 10.0.4.4 (10.0.4.4)
 Route Duration: 00h07m52s
 Relay IP Nexthop: 10.0.56.5
 Relay IP Out-Interface: GigabitEthernet0/0/1
 Relay Tunnel Out-Interface: GigabitEthernet0/0/1
 Relay token: 0x1
 Original nexthop: 10.0.4.4
 Qos information : 0x0
 AS-path 100, origin igp, MED 2, localpref 100, pref-val 0, valid, internal,
bes
t, select, active, pre 255, IGP cost 2
 Not advertised to any peer yet
```

ASBR-PE2 allocates the BGP label 1032 to the route destined for 10.0.1.1.

The next hop of the route to 10.0.1.1 is 10.0.4.4. Therefore, PE2 performs recursive query again.

HUAWEI

Verifying the Configuration (3)



Based on steps 1, 2, and 3, the packets sent by PE2 should carry the following labels:

Label Value	Type	FEC	Label Allocation Device
1029	VPN label	192.168.1.0/24	PE1
1032	BGP label	10.0.1.1/32	ASBR-PE2
1024	LDP label	10.0.4.4/32	P2

Note: Other devices are not described here.

3. Check information about routes destined for 10.0.4.4 on PE2.

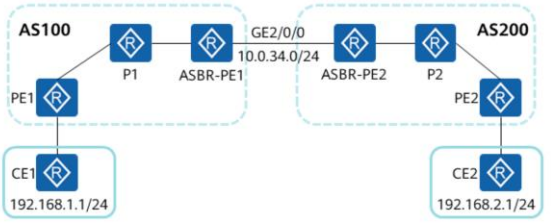
```
<PE2>display fib 10.0.4.4
  Route Entry Count: 1
Destination/Mask    Nexthop         Interface        TunnelID
10.0.4.4/32         10.0.56.5       GE0/0/1          0x1
<PE2>display tunnel-info tunnel-id 1
Tunnel ID:                  0x1
Tunnel Token:               1
Type:                       lsp
Destination:                10.0.4.4
Out Interface:              GigabitEthernet0/0/1
Out Label:                  1024
Next Hop:                   10.0.56.5
Lsp Index:                  6144
<PE2>display mpls lsp out-label 1024 verbose
  No                : 1
  Fec               : 10.0.4.4/32
  Nexthop           : 10.0.56.5
  In-Label          : NULL
  Out-Label         : 1024
  In-Interface      : ----------
  Out-Interface     : GigabitEthernet0/0/1
  LspIndex          : 6144
  Token             : 0x1
  LsrType           : Ingress
  Label Operation   : PUSH
```

Packets are transmitted to 10.0.4.4 over tunnel 1.

The label 1024 needs to be added to the packets reaching 10.0.4.4.

HUAWEI

Verifying the Configuration (4)

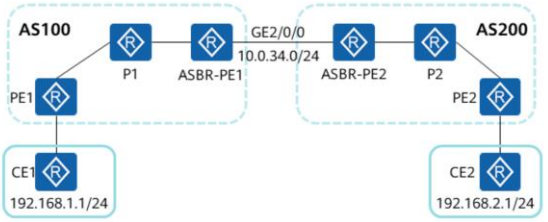


4. Ping 192.168.1.1 from CE2.

```
<CE2>ping -a 192.168.2.1 192.168.1.1
  PING 192.168.1.1: 56  data bytes, press CTRL_C to break
    Reply from 192.168.1.1: bytes=56 Sequence=1 ttl=250 time=60 ms
    Reply from 192.168.1.1: bytes=56 Sequence=2 ttl=250 time=50 ms
    Reply from 192.168.1.1: bytes=56 Sequence=3 ttl=250 time=60 ms
    Reply from 192.168.1.1: bytes=56 Sequence=4 ttl=250 time=60 ms
```


HUAWEI

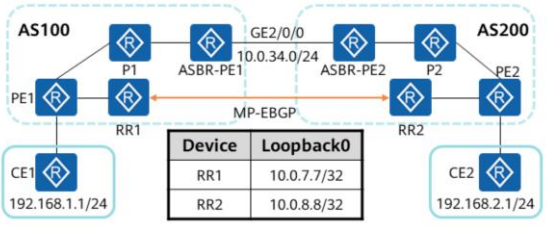
Summary of Key Configurations for Inter-AS VPN Option C (Solution 1)



Device	Key Points
PE1 PE2	Enable PEs to exchange labeled IPv4 routes with the ASBR-PEs in the same AS. Enable PEs to establish multi-hop MP-EBGP peer relationships with the PEs in other ASs.
ASBR-PE1 ASBR-PE2	Enable ASBR-PEs to exchange labeled IPv4 routes with ASBR-PEs in other ASs. Enable ASBR-PEs to exchange labeled IPv4 routes with the PEs in the same AS. Enable MPLS on the interfaces connecting ASBRs.

HUAWEI

Option C (Solution 2) Configuration Example (1)



Device	Loopback0
RR1	10.0.7.7/32
RR2	10.0.8.8/32

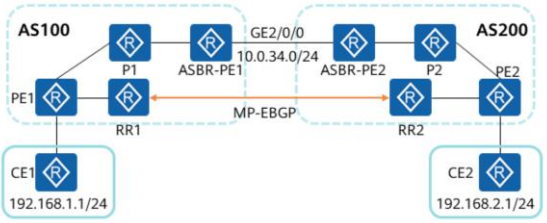
- CE1 and CE2 belong to the same VPN named **vpna**. CE1 and CE2 need to communicate with each other using Option C (solution 1). RR1 and RR2 reflect routes but do not forward traffic.

Device	Loopback0	RD	RT
PE1	10.0.1.1/32	100:1	1:1
ASBR-PE1	10.0.3.3/32	/	/
ASBR-PE2	10.0.4.4/32	/	/
PE2	10.0.6.6/32	200:6	1:1

- Configuration roadmap
  - Configure basic MPLS capabilities and MPLS LDP on the MPLS backbone network to establish LDP LSPs in each AS. (The configuration details are not provided here.)
  - Configure a VPN instance on the PE connected to the CE in each AS and establish EBGP peer relationships between the PE and CE to exchange VPN routing information.
  - Advertise the routes of a PE in an AS to the peer PE: Configure the local ASBR-PE to advertise the routes of a PE in the local AS to the peer ASBR-PE using BGP, import BGP routes into an IGP on the peer ASBR-PE so that the peer PE can use an IGP to learn the routes of the PE in the same AS.
  - Enable ASBR-PE1 and ASBR-PE2 to exchange labeled IPv4 routes.
  - Configure ASBR-PEs to establish LDP LSPs for labeled BGP routes of the public network.
  - Establish an MP-EBGP peer relationship between RRs in different ASs and set the maximum number of hops between the RRs.

HUAWEI

Option C (Solution 2) Configuration Example (2)



1. On RR1, configure RR1 to establish VPNv4 peer relationships with RR2 and PE1, and configure RR1 not to change the next hops of routes.

```
[RR1] bgp 100
[RR1-bgp] peer 10.0.8.8 as-number 200
[RR1-bgp] peer 10.0.8.8 connect-interface LoopBack 1
[RR1-bgp] peer 10.0.8.8 EBGP-max-hop 100
[RR1-bgp] peer 10.0.1.1 as-number 100
[RR1-bgp] peer 10.0.1.1 connect-interface LoopBack 1
[RR1-bgp] ipv4-family vpnv4
[RR1-bgp-af-vpnv4] undo policy vpn-target
[RR1-bgp-af-vpnv4] peer 10.0.8.8 enable
[RR1-bgp-af-vpnv4] peer 10.0.8.8 next-hop-invariable
[RR1-bgp-af-vpnv4] peer 10.0.1.1 enable
[RR1-bgp-af-vpnv4] peer 10.0.1.1 next-hop-invariable
[RR1-bgp-af-vpnv4] quit
[RR1-bgp] quit
```

Configure RR1 not to change the next hops of routes so that traffic does not pass through RR1. That is, RR1 only reflects routes without forwarding traffic.

2. On ASBR-PE1, create a route-policy.

```
[ASBR-PE1] route-policy policy1 permit node 1
[ASBR-PE1-route-policy] apply mpls-label
[ASBR-PE1-route-policy] quit
```

3. On ASBR-PE1, apply the route-policy to the routes advertised to ASBR-PE2 and enable the capability to exchange labeled IPv4 routes with ASBR-PE2.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] peer 10.0.34.4 route-policy policy1 export
[ASBR-PE1-bgp] peer 10.0.34.4 label-route-capability
[ASBR-PE1-bgp] quit
```

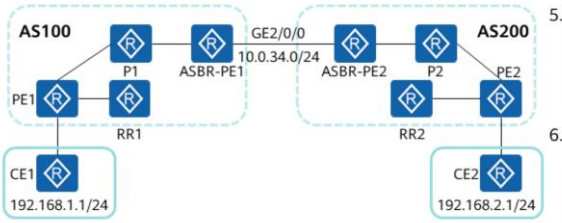
4. On ASBR-PE1, advertise the routes destined for the loopback interfaces of PE1 and RR1 to ASBR-PE2.

```
[ASBR-PE1] bgp 100
[ASBR-PE1-bgp] network 10.0.1.1 32  # Required for data forwarding
[ASBR-PE1-bgp] network 10.0.7.7 32
[ASBR-PE1-bgp] quit
```

Enable RR2 to learn the route to RR1 and establish a VPNv4 neighbor relationship.

HUAWEI

Option C (Solution 2) Configuration Example (3)



5. On ASBR-PE1, import BGP routes into an IGP.

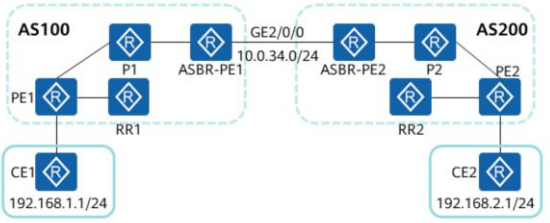
```
[ASBR-PE1] ospf 1
[ASBR-PE1-ospf-1] import-route bgp
[ASBR-PE1-ospf-1] quit
```

6. On ASBR-PE1, establish LDP LSPs for labeled BGP routes of the public network.

```
[ASBR-PE1] mpls
[ASBR-PE1-mpls] lsp-trigger bgp-label-route
[ASBR-PE1-mpls] quit
```


HUAWEI

Verifying the Configuration (1)



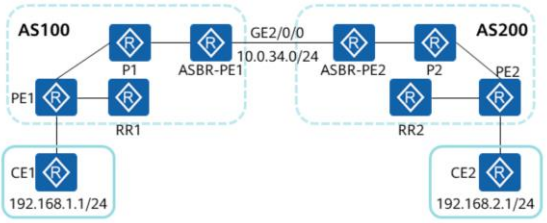
1. Check CE1's routing information on PE2.

```
<PE2>display ip routing-table vpn-instance vpna 192.168.1.0 verbose
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Table : vpna
Summary Count : 1
Destination: 192.168.1.0/24
    Protocol: IBGP              Process ID: 0
  Preference: 255                     Cost: 0
     NextHop: 10.0.1.1          Neighbour: 10.0.8.8
       State: Active Adv Relied       Age: 00h00m05s
         Tag: 0                  Priority: low
       Label: 1028               QoSInfo: 0x0
  IndirectID: 0x12
RelayNextHop: 10.0.56.5          Interface: GigabitEthernet0/0/1
    TunnelID: 0xe                Flags: RD
```

The route is learned from RR2 and its next hop is 10.0.1.1.

HUAWEI

Verifying the Configuration (2)



Based on steps 1 and 2, the packets sent by PE2 should carry the following labels:

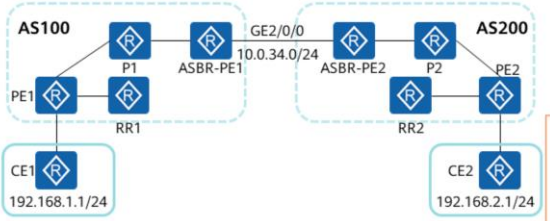
Label Value	Type	FEC	Label Allocation Device
1028	VPN label	192.168.1.0/24	PE1
1028	LDP label	10.0.1.1/32	P2

2. Check information about routes destined for 10.0.1.1 (PE1) on PE2.

```
<PE2>display fib 10.0.1.1
  Route Entry Count: 1
Destination/Mask   Nexthop        Interface        TunnelID
10.0.1.1/32        10.0.56.5       GE0/0/1          0xe
<PE2>display tunnel-info tunnel-id e
Tunnel ID:                  0xe
Tunnel Token:                14
Type:                       lsp
Destination:                10.0.1.1
Out Interface:              GigabitEthernet0/0/1
Out Label:                  1028
Next Hop:                   10.0.56.5
Lsp Index:                  6151
<PE2>display mpls lsp out-label 1028 verbose
  No                : 1
  Fec               : 10.0.1.1/32
  Nexthop           : 10.0.56.5
  In-Label          : NULL
  Out-Label         : 1028
  LspIndex          : 6151
  Token             : 0xe
  LsrType           : Ingress
  Label Operation   : PUSH
```


HUAWEI

Verifying the Configuration (3)



Based on steps 1 and 3, the packets sent by ASBR-PE2 should carry the following labels:

Label Value	Type	FEC	Label Allocation Device
1028	VPN label	192.168.1.0/24	PE1
1025	BGP label	10.0.1.1/32	ASBR-PE2

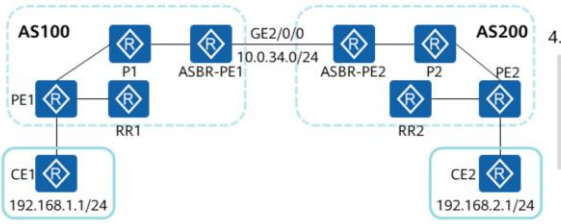
3. Check information about routes destined for 10.0.1.1 (PE1) on ASBR-PE2.

```
<ASBR-PE2>display bgp routing-table 10.0.1.1 32
BGP local router ID : 10.0.4.4
Local AS number : 200
Paths:   1 available, 1 best, 1 select
BGP routing table entry information of 10.0.1.1/32:
Label information (Received/Applied): 1025/NULL
From: 10.0.34.3 (10.0.3.3)
Route Duration: 01h30m31s
Direct Out-interface: GigabitEthernet0/0/1
Relay Tunnel Out-Interface: GigabitEthernet0/0/1
Relay token: 0x1
Original nexthop: 10.0.34.3
Qos information : 0x0
AS-path 100, origin igp, MED 2, pref-val 0, valid, external, best, select,
active, pre 255
```

Label allocated by ASBR-PE1 to the route destined for 10.0.1.1

HUAWEI

Verifying the Configuration (4)

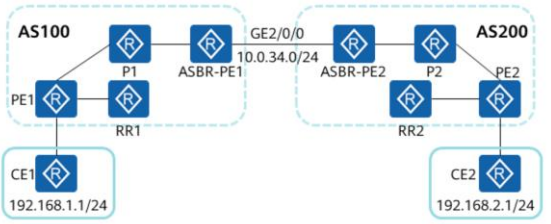


4. Ping 192.168.1.1 from CE2.

```
<CE2>ping -a 192.168.2.1 192.168.1.1
  PING 192.168.1.1: 56  data bytes, press CTRL_C to break
    Reply from 192.168.1.1: bytes=56 Sequence=1 ttl=250 time=60 ms
    Reply from 192.168.1.1: bytes=56 Sequence=2 ttl=250 time=50 ms
    Reply from 192.168.1.1: bytes=56 Sequence=3 ttl=250 time=60 ms
    Reply from 192.168.1.1: bytes=56 Sequence=4 ttl=250 time=60 ms
```


HUAWEI

Summary of Key Configurations for Inter-AS VPN Option C (Solution 2)



Device	Key Points
PE1 PE2	Enable PEs to establish MP-IBGP peer relationships with RRs in the local AS.
RR1 RR2	Enable RRs to establish MP-EBGP peer relationships with RRs in other ASs and not to change the next hops of routes. Enable RRs to establish MP-IBGP peer relationships with PEs in the same AS and not to change the next hops of routes.
ASBR-PE1 ASBR-PE2	Enable ASBR-PEs to exchange labeled IPv4 routes with ASBR-PEs in other ASs. Enable ASBR-PEs to establish LDP LSPs for labeled BGP routes of the public network. Enable MPLS on the interfaces connecting ASBRs.

HUAWEI

Quiz

1. (Single-answer question) Which of the following inter-AS VPN solutions does not require ASBRs to save CE-side routing information? ( )
   A. Option A
   B. Option B
   C. Option C
2. (Single-answer question) When only LDP or BGP is used to allocate labels, which of the following inter-AS VPN solutions may use three types of labels during packet forwarding? ( )
   A. Option A
   B. Option B
   C. Option C

HUAWEI

1. C
2. C

Summary

- When VPN routing information is exchanged between ASs, three inter-AS VPN solutions are proposed in related standards:
  - Option A: It features simple configurations as MPLS does not need to run between ASBRs. It is applicable to the scenario with a small number of VPNs.
  - Option B: It does not require interfaces to be created for each inter-AS VPN. All traffic is forwarded by ASBRs, facilitating traffic control but burdening ASBRs. When there are a large number of VPN routes, ASBRs are overloaded and may become a bottleneck.
  - Option C: It does not require ASBRs to maintain or advertise VPN-IPv4 routes because these routes are directly exchanged between PEs. Option C is applicable to scenarios where multiple ASs are spanned. The disadvantage of Option C is that it costs too much to manage an end-to-end BGP LSP between PEs.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

EVPN Fundamentals and Configuration

HUAWEI

Foreword

- Virtual Private LAN Service (VPLS) is a point-to-multipoint (P2MP) L2VPN service provided over a public network.
- VPLS cannot implement load balancing, simplify network deployment, provide flexible L2VPN deployment, or improve link efficiency in CE multi-homing scenarios. Ethernet Virtual Private Network (EVPN) solves these problems.
- Defined in RFC7432, EVPN introduces the control plane to control MAC address learning. EVPN has been continuously expanded to support both L2VPN and L3VPN.
- EVPN has been widely used in various scenarios, such as WANs, data centers (DCs), and campus networks.
- This document describes the background, fundamentals, typical route types, and application scenarios of EVPN.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the background of EVPN.
  - Know how EVPN solves VPLS problems.
  - Understand the fundamentals and typical route types of EVPN.
  - Describe the fundamentals of inter-AS EVPN.
  - Describe the typical application scenarios of EVPN.

HUAWEI

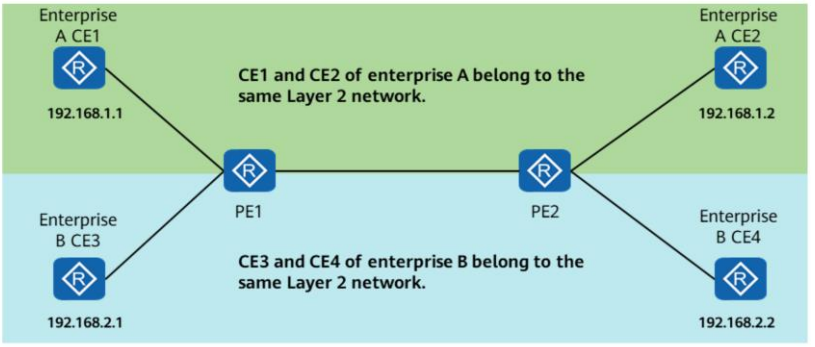
Contents

1. **EVPN Background and Terms**
2. EVPN Fundamentals
3. Inter-AS EVPN
4. Typical EVPN Application Scenarios
5. Basic EVPN Configurations

HUAWEI

VPLS Overview

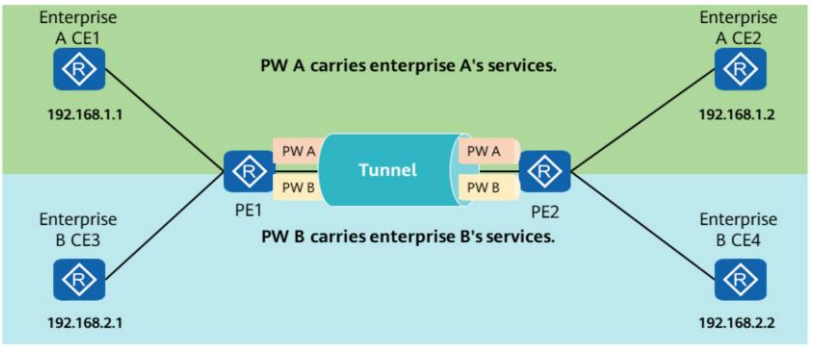
- VPLS is an Ethernet-based L2VPN technology. It provides services similar to LAN services on an MPLS network and allows users to access the network from multiple geographical locations and communicate with each other.
- Implementing VPLS involves three steps: creating tunnels and pseudo wires (PWs), creating virtual switch instances (VSIs), and binding PWs and attachment circuits (ACs) to VSIs.



HUAWEI

VPLS Fundamentals: Creating Tunnels and PWs

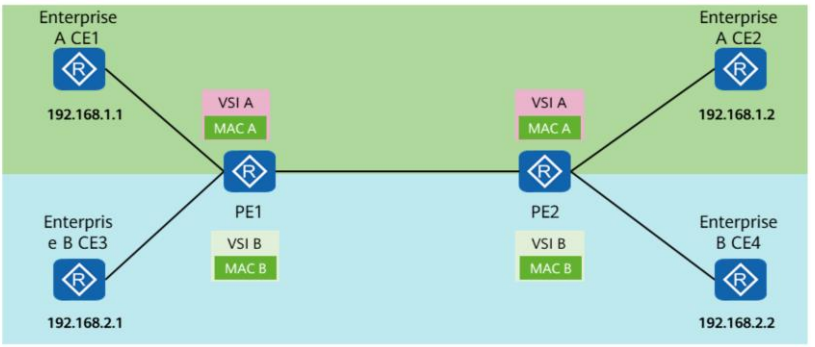
- To implement VPLS, you need to establish a tunnel between PE1 and PE2 and then create PWs to carry services of different enterprise customers.



HUAWEI

VPLS Fundamentals: Creating VSIs

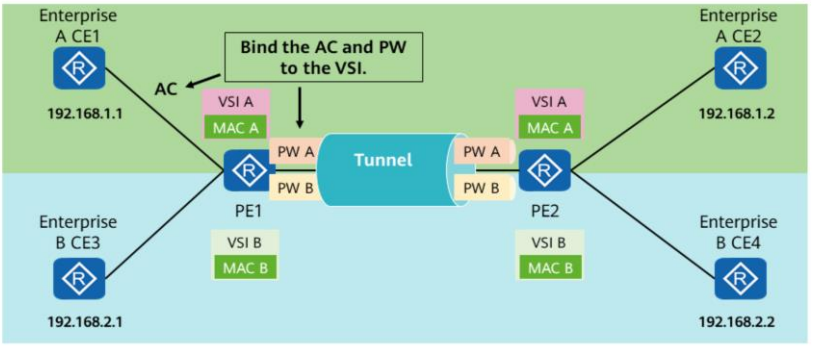
- A VSI is an independent virtual switching unit for each VPLS service. A VSI stores an independent MAC address table, forwards packets, and terminates PWs.



HUAWEI

VPLS Fundamentals: Binding PWs and ACs to VSIs

- An attachment circuit (AC) is a link between a CE and a PE. After a PW and an AC are bound to a VSI, the VSI works like a switch. The VSI records the MAC addresses learned by different interfaces and generates MAC address entries.

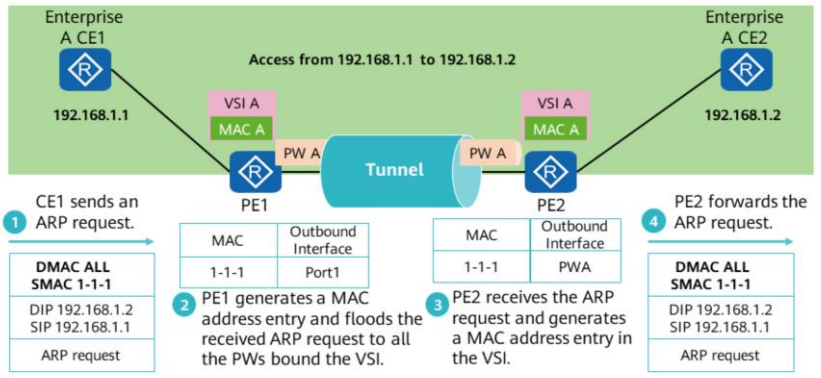


HUAWEI

- Common VPLS PW creation modes:
  - Static configuration: A large number of manual configurations are required.
  - LDP signaling (Martini)
  - BGP signaling (Kompella)

VPLS Packet Forwarding

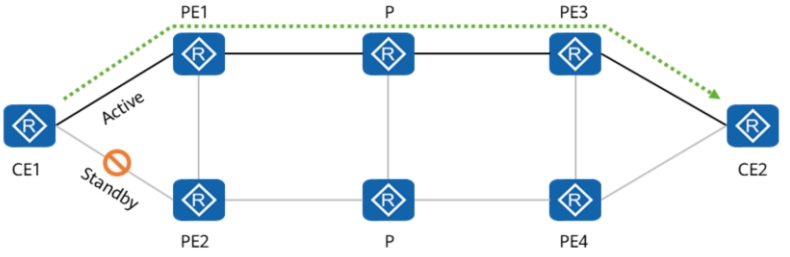
- VPLS emulates a switch. MAC address learning depends on flooding on the data plane without involving the control plane.
- The forwarding behavior is determined by examining the MAC address table. If packets match an entry in the MAC address table, packets are forwarded according to the table. Otherwise, packets are flooded for forwarding.



HUAWEI

VPLS Challenge 1: Uneven Load Balancing Due to Single-Active Access

- In VPLS, if a ring topology is achieved when a CE is dual-homed to two PEs, the loop-free forwarding path is formed by blocking a port. This is the same as the loop prevention technology (STP) of switches. Therefore, VPLS access works in active/standby mode.
- There is only one link between the CE and PE for traffic forwarding, and multiple paths cannot be formed between PEs. As a result, some links may be congested.



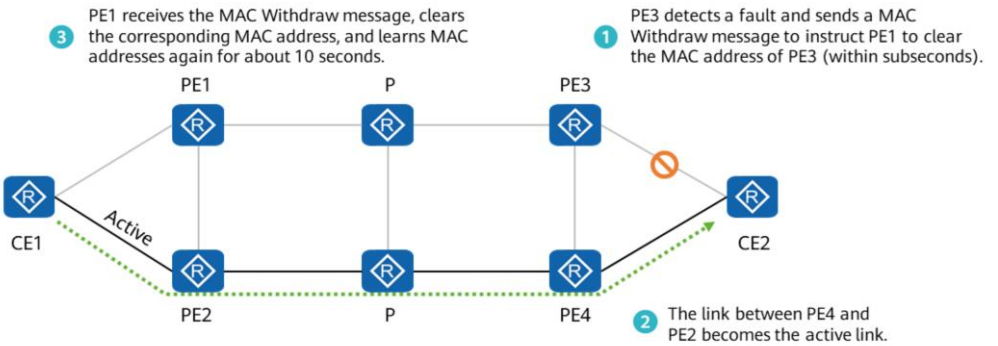
1 The standby link between the CE and PE does not forward data.

2 Multiple paths cannot be formed between PEs.

HUAWEI

VPLS Challenge 2: Slow Fault Convergence

- After detecting a link fault, PE3 sends a MAC Withdraw message to the peer PE, instructing the peer PE to delete the MAC address of PE3. PE4 changes the standby link to the active state. After receiving the Withdraw message, PE1 clears the corresponding MAC address and learn MAC addresses again. The fault convergence time is closely related to the number of MAC addresses.

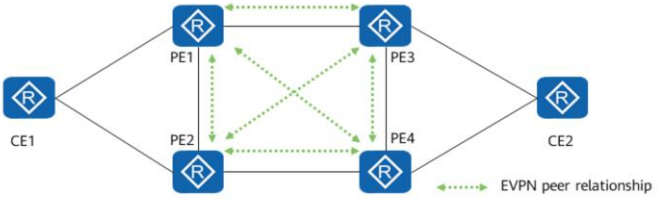


HUAWEI

EVPN Overview

- EVPN changes the traditional MAC address learning mode on the L2VPN data plane. It introduces the control plane to learn MAC addresses and IP addresses for data forwarding, implementing forwarding-control separation.
- EVPN implements active-active, fast convergence, and simplified O&M, resolving typical problems of traditional L2VPNs.
- EVPN uses Multiprotocol Border Gateway Protocol (MP-BGP) on the control plane and supports multiple types of tunnels on the data plane, such as MPLS tunnels, GRE tunnels, and Segment Routing over IPv6 (SRv6) tunnels. In this course, MPLS is used as the outer tunnel forwarding technology.

Control plane: PEs exchange BGP EVPN packets to transmit MAC addresses and IP addresses.



- Other advantages of EVPN:
  - All-active access of CEs to PEs
  - Automatic PE discovery
  - Loop prevention
  - Broadcast traffic optimization
  - Equal-cost multi-path routing (ECMP)

Data plane: Data forwarding paths are formed by IP tunnels or MPLS label forwarding paths. The data plane is only responsible for forwarding data and does not need to broadcast packets to learn MAC addresses.

HUAWEI

EVPN Terms

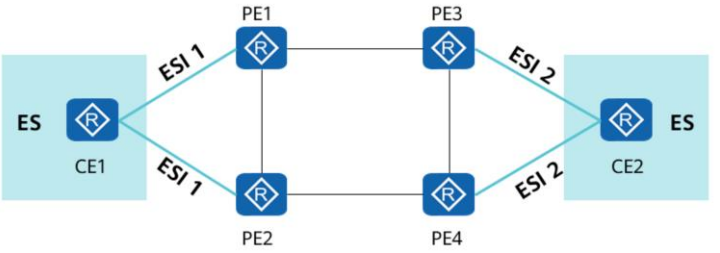
- An Ethernet segment (ES) is a group of Ethernet links that connect a user site (device or network) to a PE. An ES is identified by an Ethernet Segment Identifier (ESI).
- An EVPN instance (EVI) identifies an EVPN client.
- A MAC-VRF table is a MAC address table that belongs to an EVI on a PE.
- A route distinguisher (RD) is the unique identifier of an EVPN and identifies an EVI.
- Route targets (RTs) are used to control the import of EVPN routes.
- In a CE multi-homing scenario, only the PE that is elected as a designated forwarder (DF) forwards broadcast, unknown unicast, and multicast (BUM) traffic to CEs.
- ESI Label is an extended community attribute carried in an EVPN Type 1 route. It implements fast convergence and split horizon in multi-homing scenarios.
- BUM labels are carried in EVPN Type 3 routes to forward BUM traffic.
- Unicast labels are carried in Type 2 routes to forward unicast traffic.

HUAWEI

- For details about the terms, see RFC7432.

EVPN terms: ES and ESI

- An ES is a group of Ethernet links that connect a user site (device or network) to a PE. An ES is identified by an ESI.
- An ESI is 10 bytes long and is unique on the entire network.
- Dual-homing PEs on the same ES must have the same ESI. For example, PE1 and PE2 must be configured with the same ESI (for example, 1).



HUAWEI

ESI Format

- Type 0: The ESI is manually configured by the administrator.
- Type 1: If LACP is used between a PE and a CE, the ESI is the CE LACP System MAC address (6 bytes) + CE LACP Port Key (2 bytes) + 0x00 (1 byte).
- Type 2: The ESI is automatically generated by MSTP.
- Type 3: The ESI is system MAC address plus local discriminator.
- Type 4: The ESI is router ID plus local discriminator.
- Type 5: The ESI is AS number plus local discriminator.

Static ESI configuration example.

```
[*PE1] interface eth-trunk 10
[*PE1-Eth-Trunk10] esi 0000.1111.2222.1111.1111
```

View ESI information.

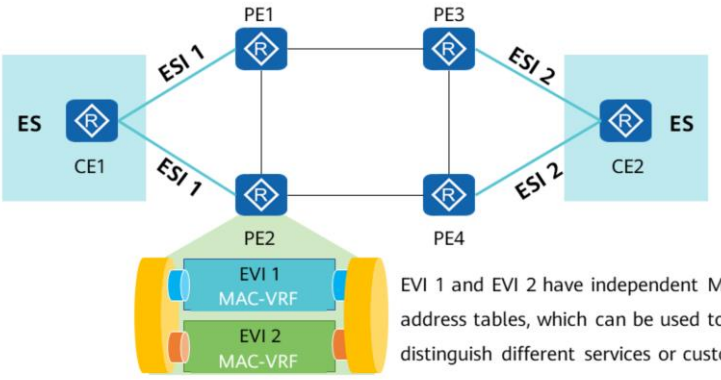
```
<PE1> display bgp evpn all esi
Number of ESI for EVPN address family: 1
        ESI                          IFName/Bridge-domain
 0000.1111.2222.1111.1111            Eth-Trunk10
```


HUAWEI

- Currently, ESIs of Type 2, Type 3, Type 4 and Type 5 are not used on Huawei devices.

Basic EVPN Terms: EVI and MAC-VRF

- An EVI identifies an EVPN client.
- A MAC-VRF table is a MAC address table that belongs to an EVI on a PE.

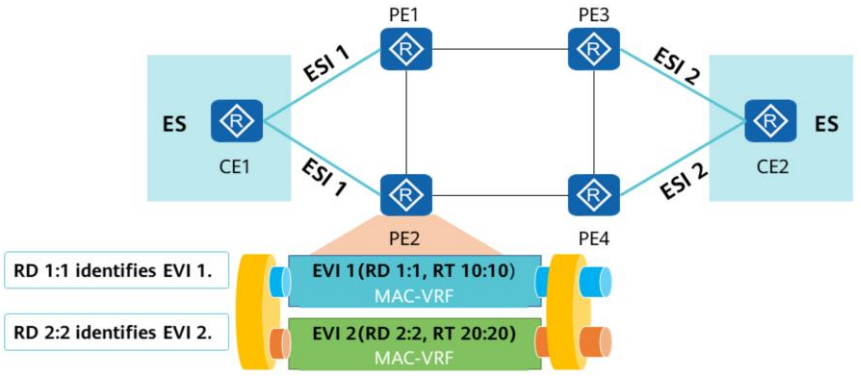


EVI 1 and EVI 2 have independent MAC address tables, which can be used to distinguish different services or customers.

HUAWEI

Basic EVPN Terms: RD and RT

- An RD is the unique identifier of an EVPN and identifies an EVI.
- RTs are used to label routes to control the import of EVPN routes.



HUAWEI

Contents

1. EVPN Background and Terms
2. **EVPN Fundamentals**
   - EVPN Route Overview and Interaction Process
   - EVPN Route Types
   - EVPN Access Principles
3. Inter-AS EVPN
4. Typical EVPN Application Scenarios
5. Basic EVPN Configurations

HUAWEI

EVPN Routes

- EVPN defines a new BGP network layer reachability information (NLRI), which is known as EVPN NLRI, to carry all EVPN routes.
- EVPN NLRI is carried by MP-BGP. MP-BGP supports multi-protocol extensions. In MP-BGP, the Address Family Identifier (AFI) defining EVPN is 25 and the Subsequent Address Family Identifier (SAFI) is 70.

EVPN NLRI format

1. Ethernet A-D route
2. MAC/IP advertisement route
3. Inclusive multicast route
4. Ethernet segment route

The Route Type field defines four types of common EVPN routes.

HUAWEI

Functions of Four Types of EVPN Routes

- RFC7432 defines four types of EVPN routes: Type 1 to Type 4. With the development of the EVPN protocol, more and more routes are redefined. The following describes four types of EVPN routes.

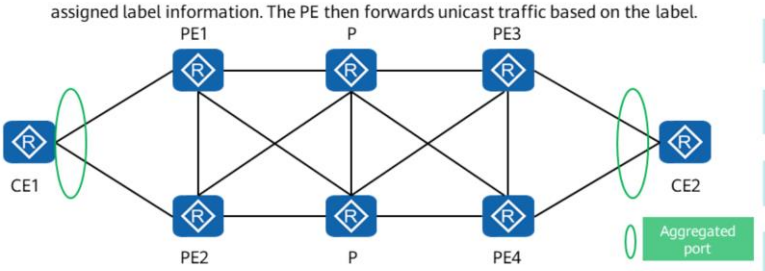
Route Type	Function	Benefits
(Type 1) Ethernet A-D route	Aliasing MAC address withdrawal in batches All-active indication ESI label advertisement	Loop prevention Fast convergence Load balancing
(Type 2) MAC/IP advertisement route	MAC address learning and advertisement MAC/IP binding MAC address mobility	ARP suppression Host migration
(Type 3) Inclusive multicast route	Auto-discovery of multicast tunnel endpoints & multicast types	BUM traffic forwarding
(Type 4) Ethernet segment route	ES member auto-discovery DF election	Support for both all-active and single-active modes

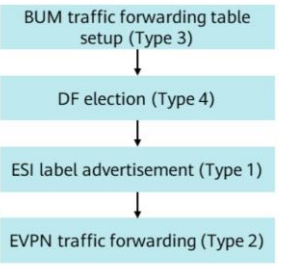
HUAWEI

- Up to now, 11 route types have been defined in the RFC and draft. Type 1 to Type 5 are relatively mature. Type 5 is in the draft phase. Type 6 to Type 11 are used for multicast traffic optimization and are not mature.

EVPN Working Process

- The EVPN process is divided into two phases:
  - Startup phase
    - EVPN peers exchange EVPN Type 3 routes to establish the BUM traffic forwarding table.
    - EVPN peers exchange Type 4 routes to complete ES discovery and DF election (in ES multi-homing scenarios).
    - EVPN peers exchange Type 1 routes and ESI labels to implement split horizon and aliasing.
  - Traffic forwarding phase: CE-side traffic triggers a PE to advertise the MAC address through a Type 2 route. The route carries assigned label information. The PE then forwards unicast traffic based on the label.





HUAWEI

- The CE all-active access scenario uses Huawei NE series routers as an example. An Eth-Trunk is configured on CE1 to connect to PE1 and PE2, and an inter-device link aggregation enhanced trunk (E-Trunk) needs to be configured on PE1 and PE2. E-Trunk will not be marked in the following figures.

EVPN Entry Overview

- EVPN maintains three tables to guide traffic forwarding: MAC-VRF table, BUM traffic forwarding table, and ES member table.
  - The MAC-VRF table is used to record the forwarding of known unicast traffic.
  - The BUM traffic forwarding table is used to guide the forwarding of broadcast, unknown unicast, and multicast traffic.
  - The ES member table is used to record information about the PEs that users access.
- The following describes the generation process and functions of the three tables.

PE1 is used as an example.

MAC-VRF

MAC	ESI	Next hop	Label
1-1-1	ESI 1	Port1	NULL
2-2-2	ESI 2	PE3	303
2-2-2	ESI 2	PE4	304

BUM traffic forwarding table

Peer	Label
PE2	102
PE3	103
PE4	104

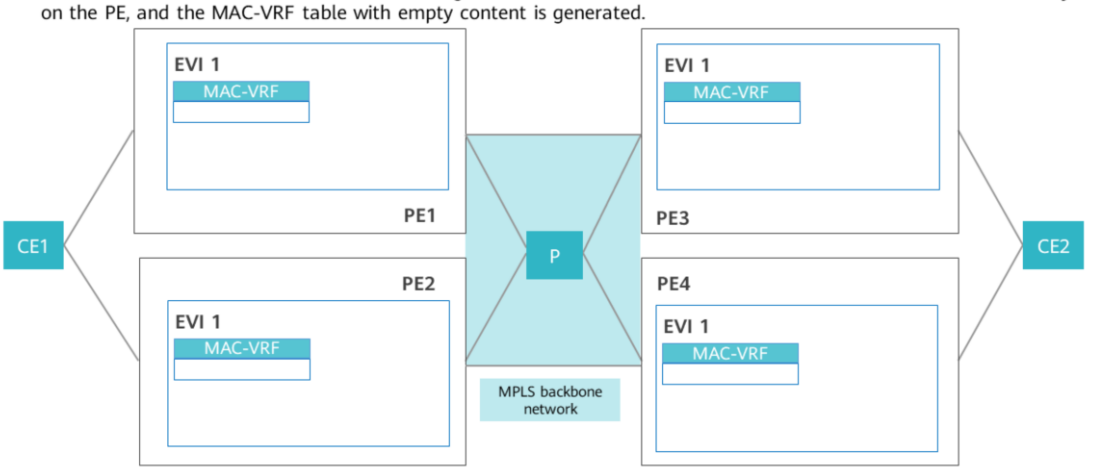
ES member table

No.	ESI	Member	Label
0	ESI 1	PE2	202
1	ESI 1	PE1	201

HUAWEI

Creating an EVPN Instance Locally

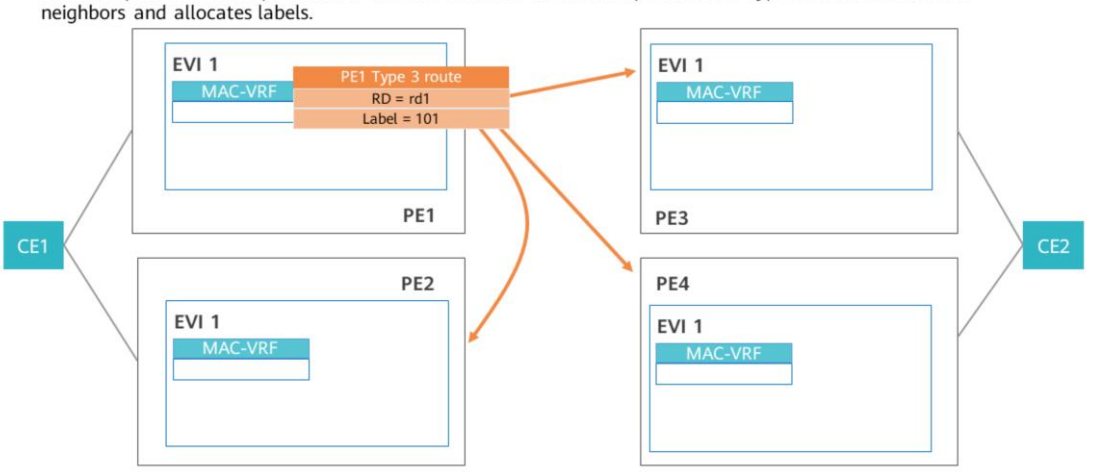
- Create an EVPN instance on each PE and configure an RD and an RT for the instance. EVPN is then activated locally on the PE, and the MAC-VRF table with empty content is generated.



- The backbone network consisting of P devices involves only outer tunnel forwarding, which will not be described in later sections.

BUM Traffic Forwarding Table (1)

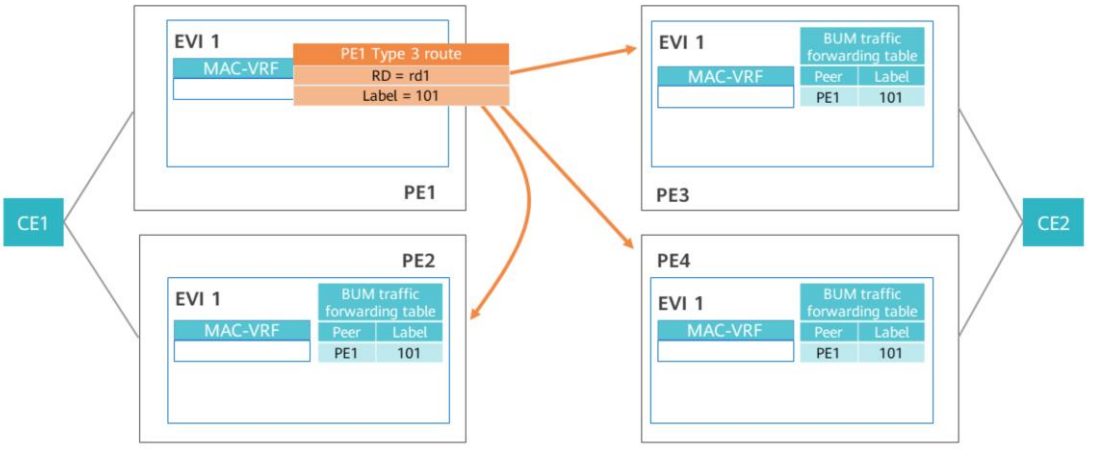
- Establish peer relationships between the PEs. Take PE1 as an example. It sends Type 3 routes to discover neighbors and allocates labels.



HUAWEI

BUM Traffic Forwarding Table (2)

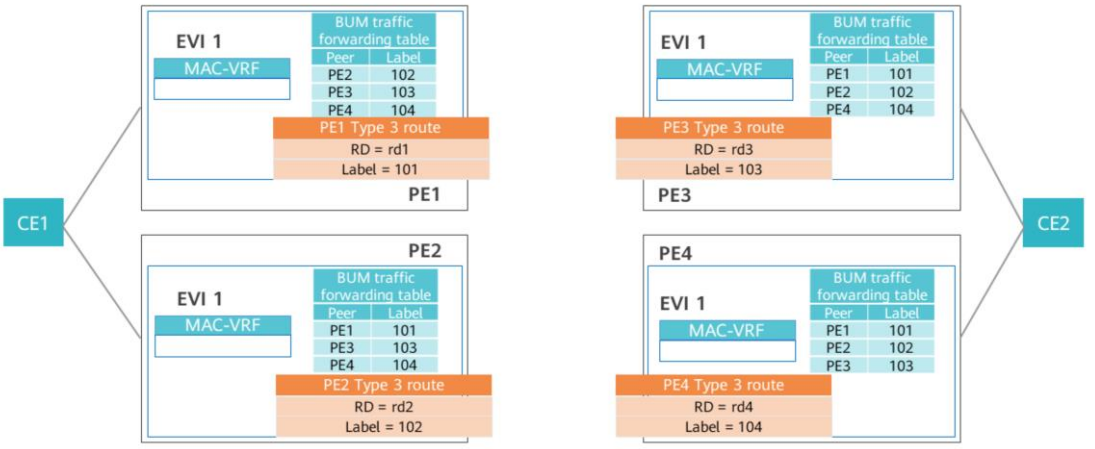
- PE2, PE3, and PE4 each generate a BUM traffic forwarding table.



HUAWEI

BUM Traffic Forwarding Table (3)

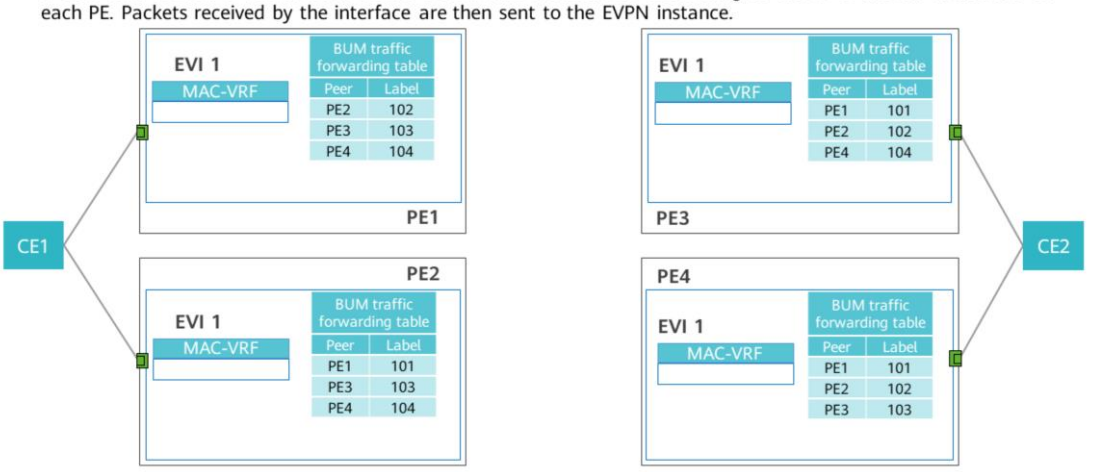
- In this process, all PEs send Type 3 routes to generate a stable BUM traffic forwarding table.



HUAWEI

Binding an Interface to an EVPN Instance

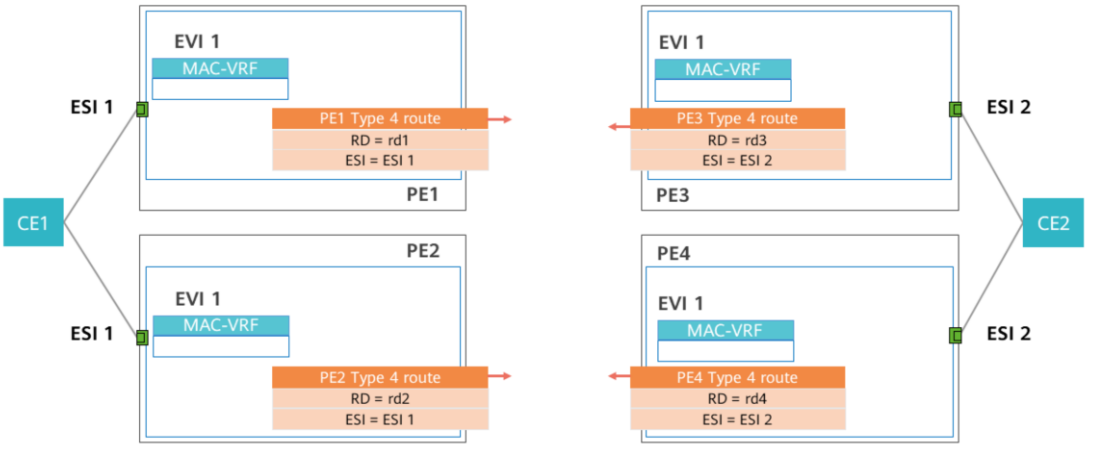
- A CE is dual-homed to PEs in active-active mode. Bind the interface connecting to the CE to the EVPN instance on each PE. Packets received by the interface are then sent to the EVPN instance.



HUAWEI

Configuring ESIs for PEs' Interfaces Connected to CEs

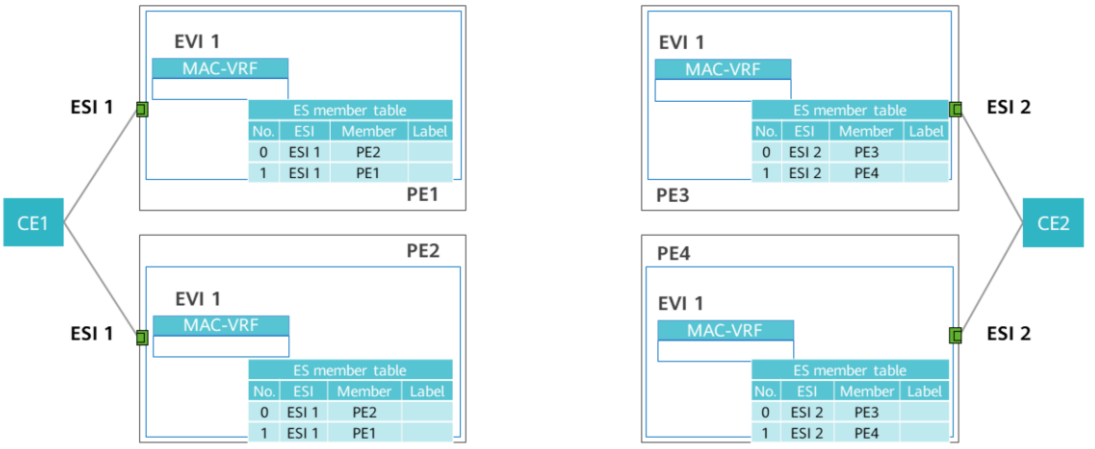
- Configure ESIs for PEs' interfaces connected to CEs. PEs exchange Type 4 routes to advertise ESIs and elect a DF.



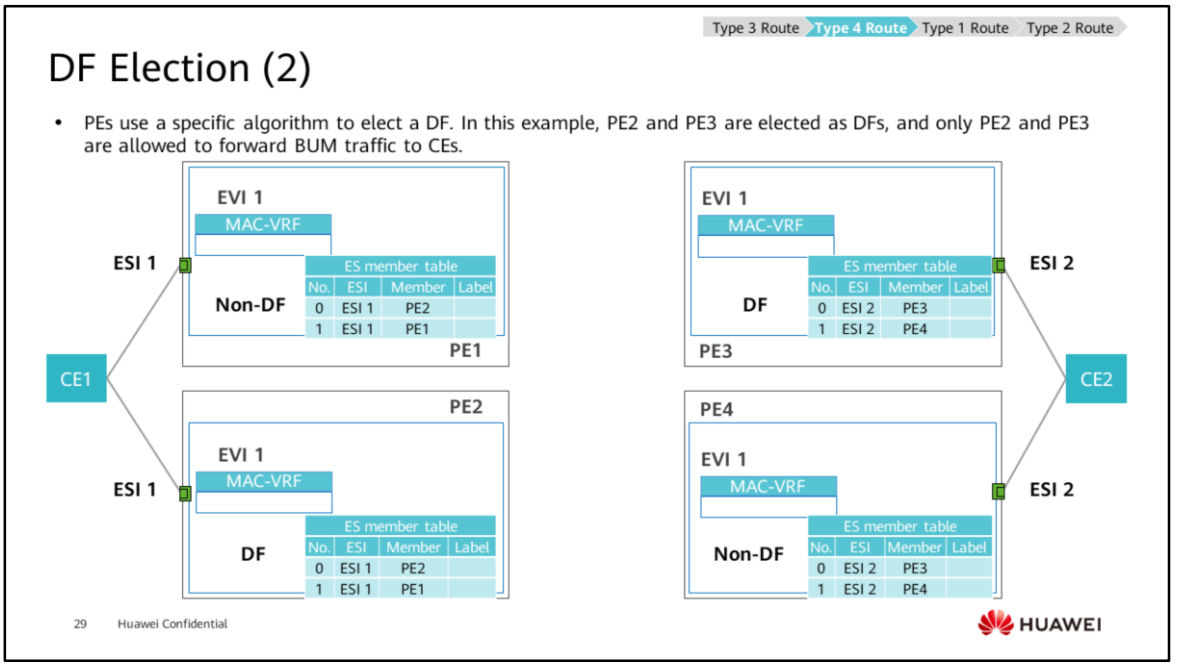
HUAWEI

DF Election (1)

- When a CE is multi-homed to multiple PEs, only one PE forwards BUM traffic to the CE. The process of selecting such a PE is DF election.



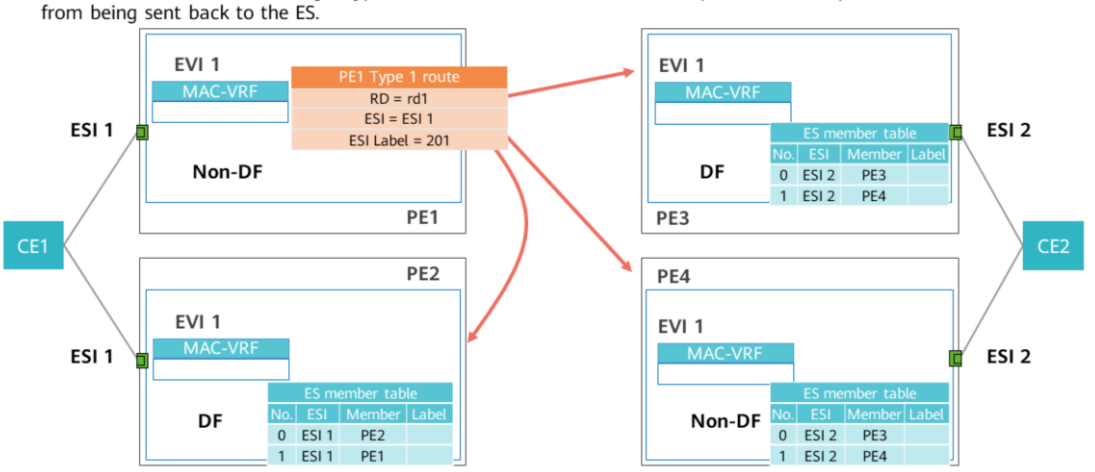
HUAWEI



- The DF election mode can be set to interface-based or VLAN-based DF election on Huawei devices. By default, interface-based DF election is enabled, which may cause traffic unbalance among multi-homing links. You can configure VLAN-based DF election so that BUM traffic from a PE to a CE can be evenly distributed to multi-homing links based on VLANs.

ESI Label Distribution: Split Horizon (1)

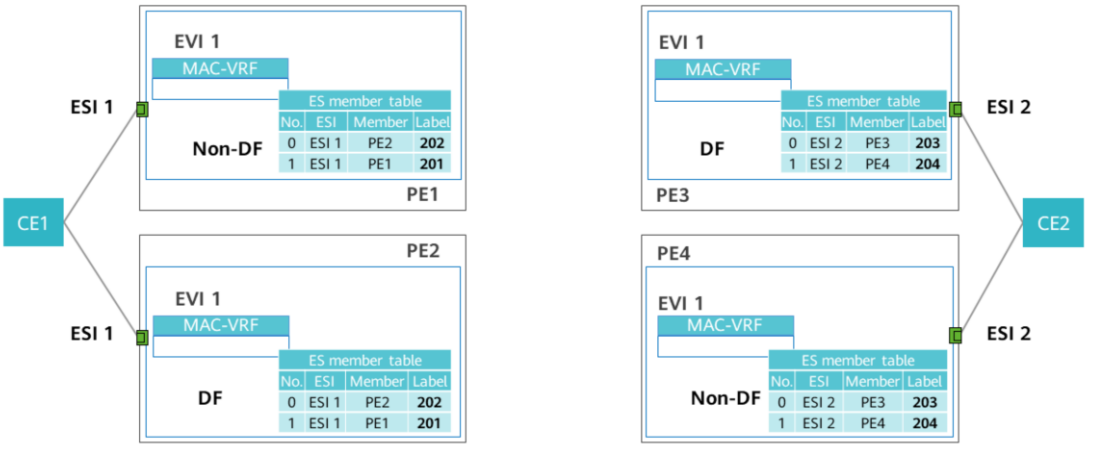
- PEs distribute ESI labels through Type 1 routes. ESI labels are used for split horizon to prevent traffic from an ES from being sent back to the ES.



HUAWEI

ESI Label Distribution: Split Horizon (2)

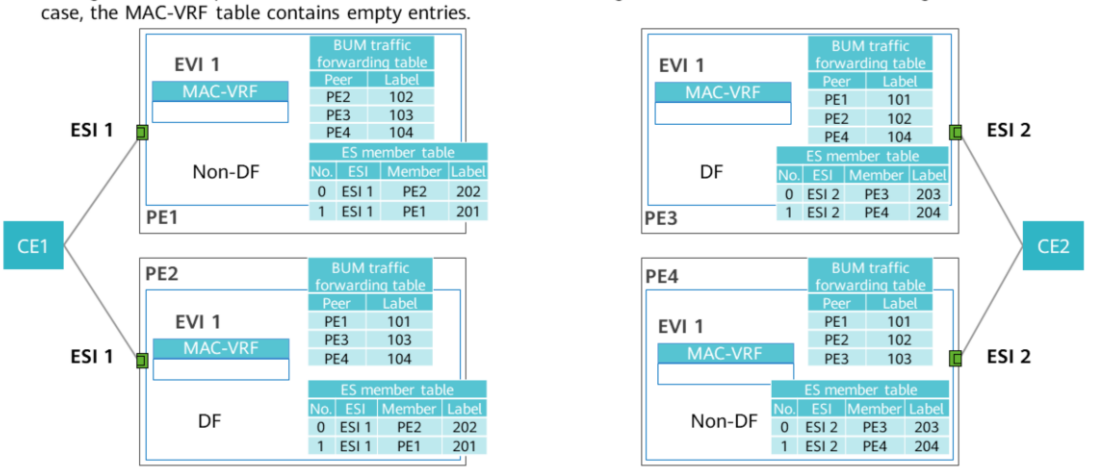
- In this process, all PEs send Type 1 routes to generate a complete ES member table.



HUAWEI

Summary of the EVPN Startup Phase

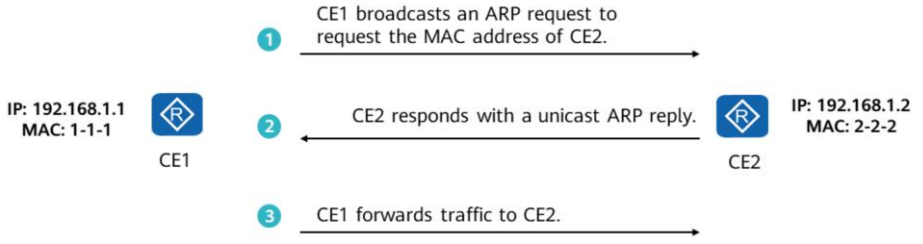
- During EVPN startup, the MAC-VRF table, BUM traffic forwarding table, and ES member table are generated. In this case, the MAC-VRF table contains empty entries.

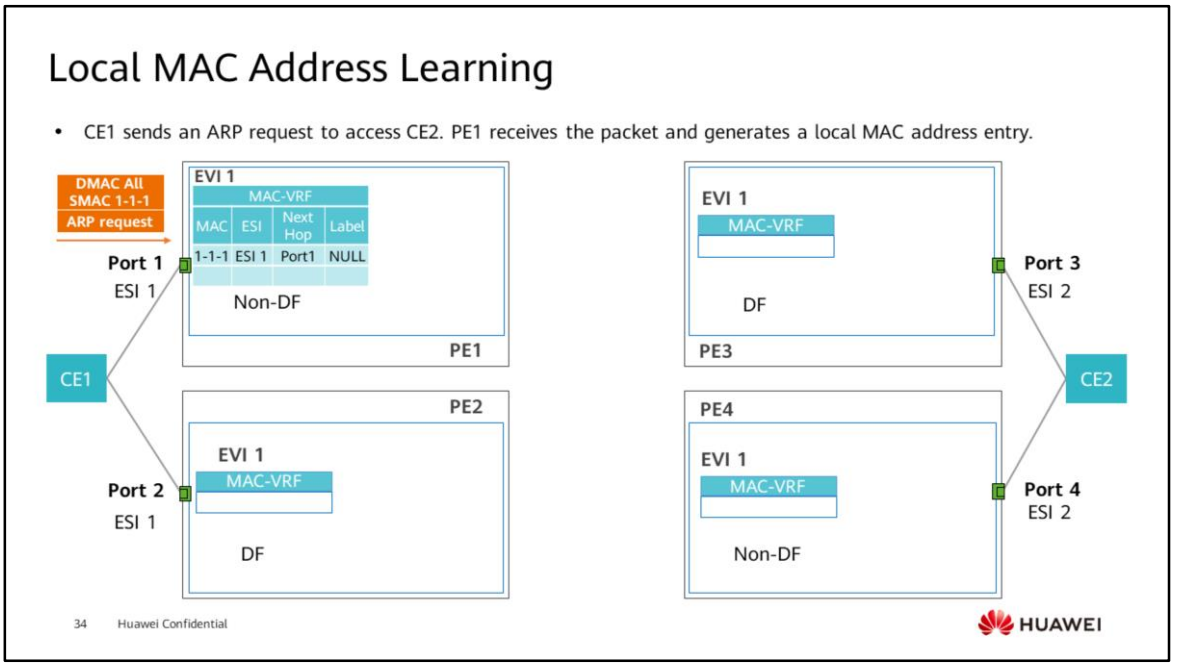


HUAWEI

EVPN Traffic Forwarding Phase

- The EVPN traffic forwarding phase starts when the CE-side user traffic is initiated.
- CE1 and CE2 are unaware of EVPN packet exchange between PEs.
- In this example, the service process of CE1 accessing CE2 is as follows:

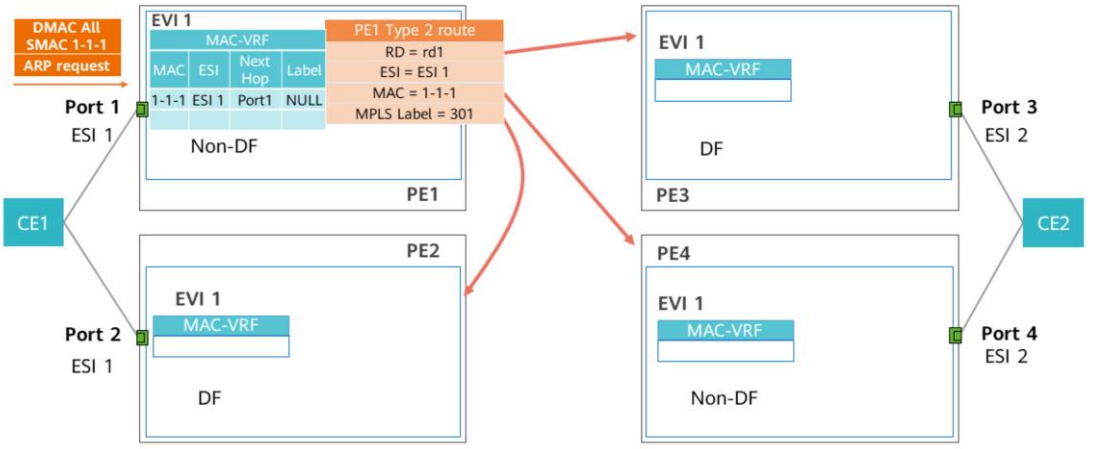




- The data packets sent by the CEs do not carry any label. Therefore, the incoming label in the MAC-VRF table of PEs is NULL.

MAC Address Advertisement (1)

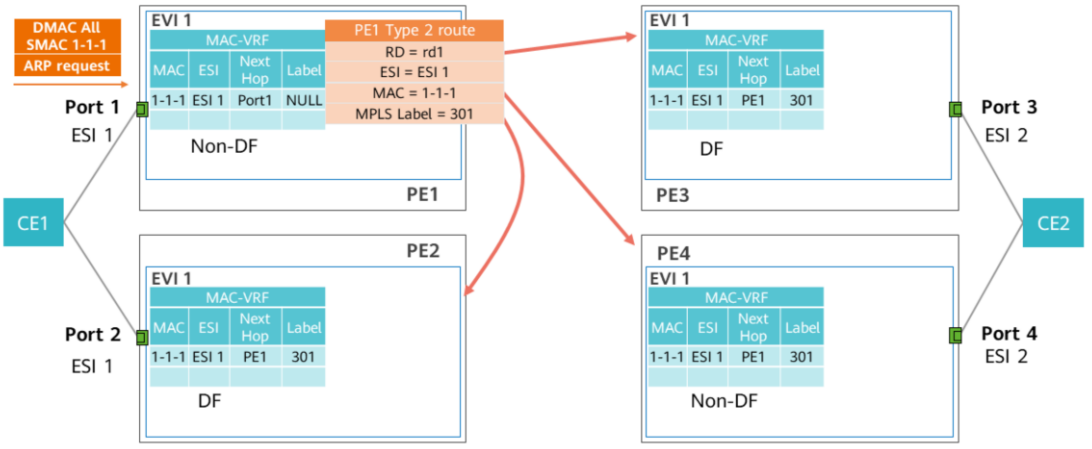
- PE1 EVPN generates a Type 2 route based on the local MAC address entry, carrying label 301 assigned by PE1.



HUAWEI

MAC Address Advertisement (2)

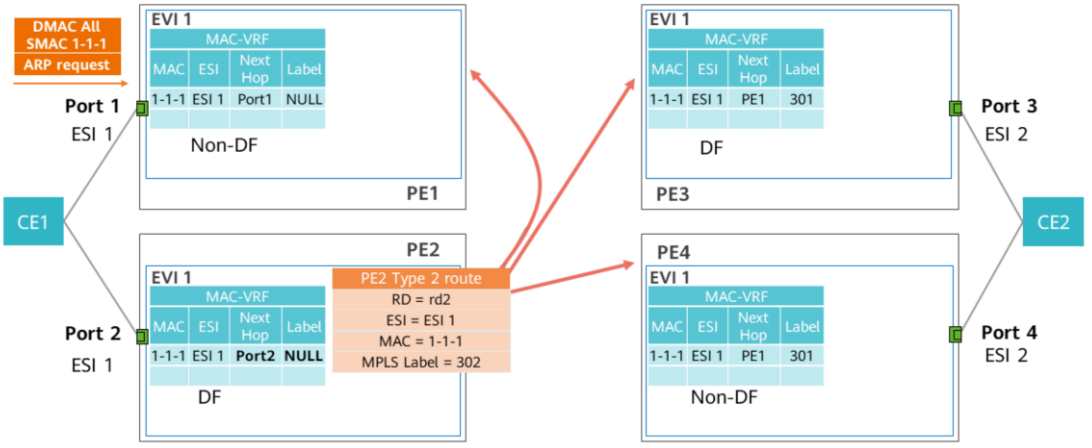
- The peer PE learns the EVPN route through MP-BGP and generates a MAC address entry.



HUAWEI

MAC Address Advertisement (3)

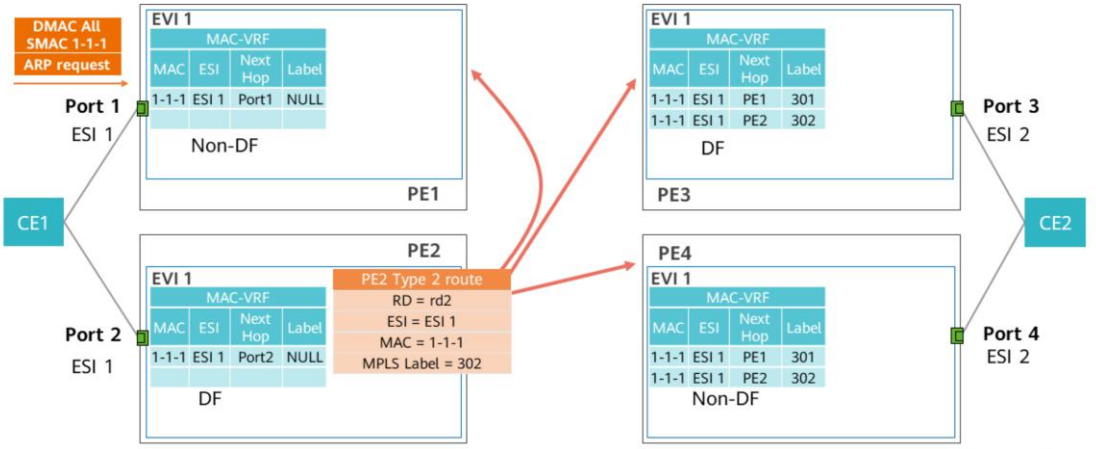
- EVPN allows CEs to access PEs in all-active mode. PE2 detects that it is directly connected to CE1, updates the optimal MAC address entry, and generates and advertises a Type 2 route.



HUAWEI

Remote MAC Address Learning

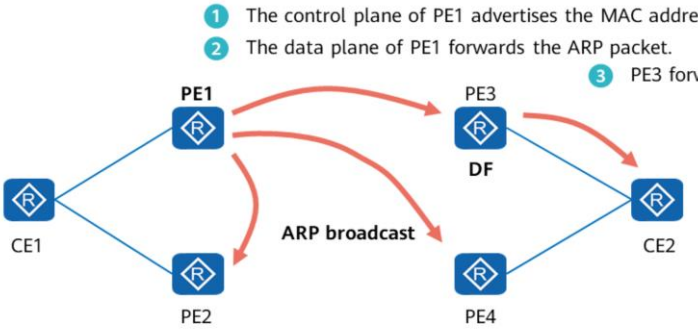
- Because PE1 and PE2 allocate different MPLS labels, PE3 and PE4 each will have two paths to CE1.



HUAWEI

ARP Broadcast Forwarding

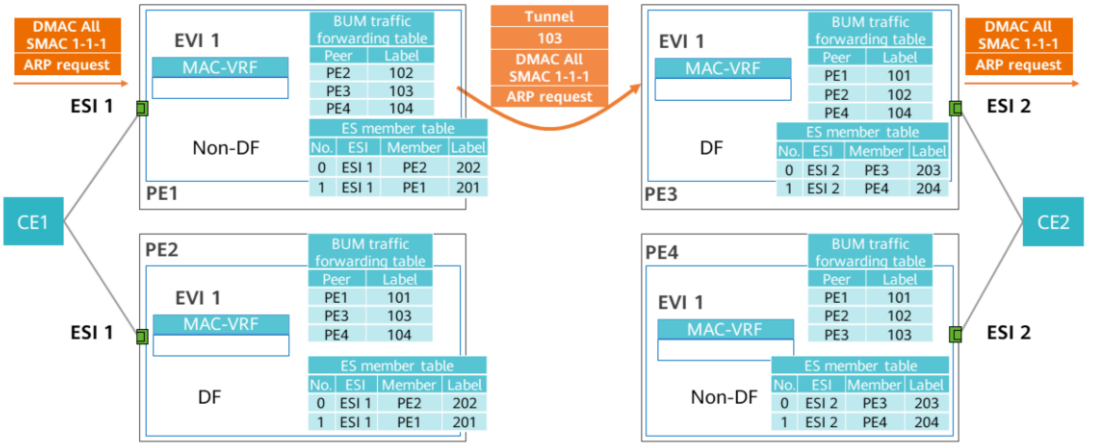
- The ARP request sent by CE1 reaches PE1. PE1 learns the MAC address of CE1 through the data plane and sends the MAC address to all neighbors through a Type 2 route.
- After the control plane behavior is completed, PE1 performs the data plane behavior, that is, forwards the ARP broadcast request. Because PE3 is the DF, it forwards the ARP broadcast packet to CE2.



HUAWEI

ARP Broadcast Forwarding: from PE1 to PE3

- The ARP packet carrying label 103 from PE1 to PE3 is forwarded based on the BUM traffic forwarding table. PE3 functions as the DF and forwards the packet through Port3.

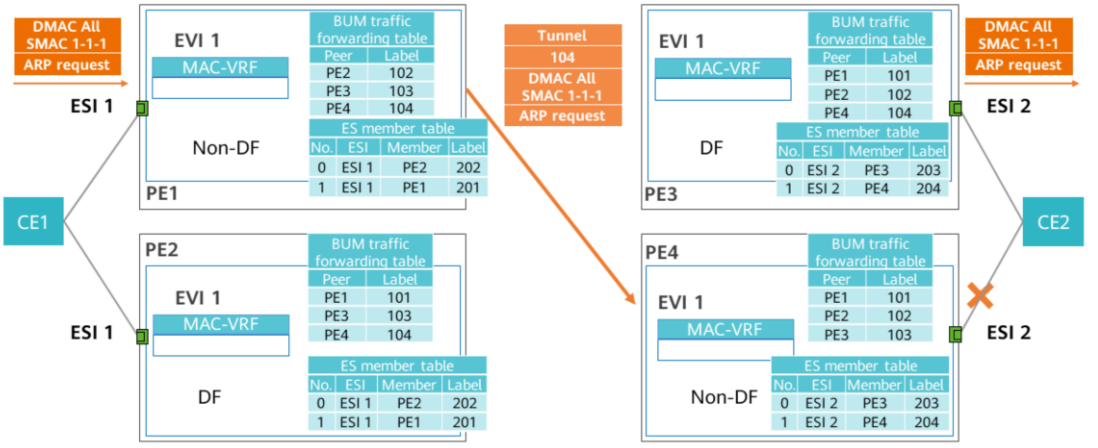


HUAWEI

- PE1 forwards the ARP packet to all members based on the BUM traffic forwarding table.

ARP Broadcast Forwarding: from PE1 to PE4

- Traffic from PE1 to PE4 carries label 104. PE4 determines that the traffic is BUM traffic based on the label and discards the traffic because PE4 is a non-DF.

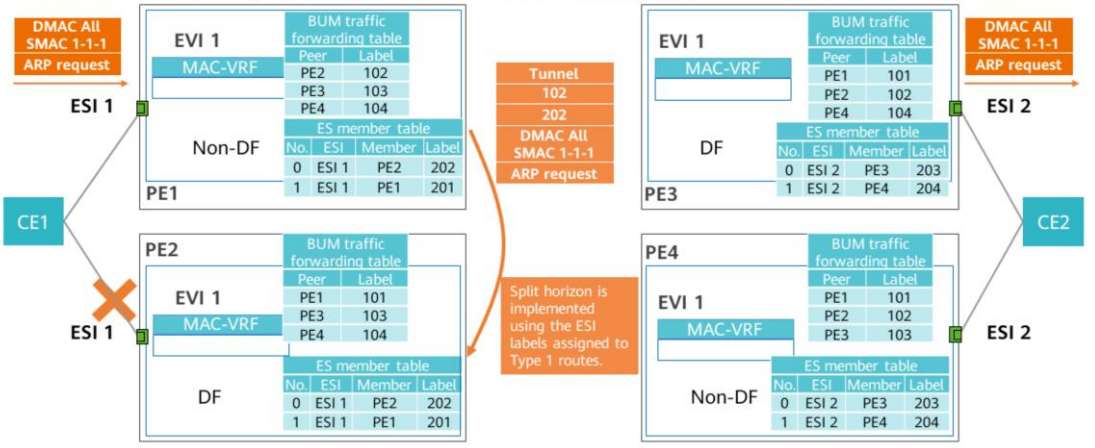


HUAWEI

- Only the router that is elected as the DF forwards BUM traffic to CEs.

ARP Broadcast Forwarding: from PE1 to PE2

- Because PE1 and PE2 belong to the same ES, traffic from PE1 to PE2 carries both the ESI label 202 and the BUM label 102. After receiving the packet, PE2 finds that the packet carrying the label 202 and discards the packet.

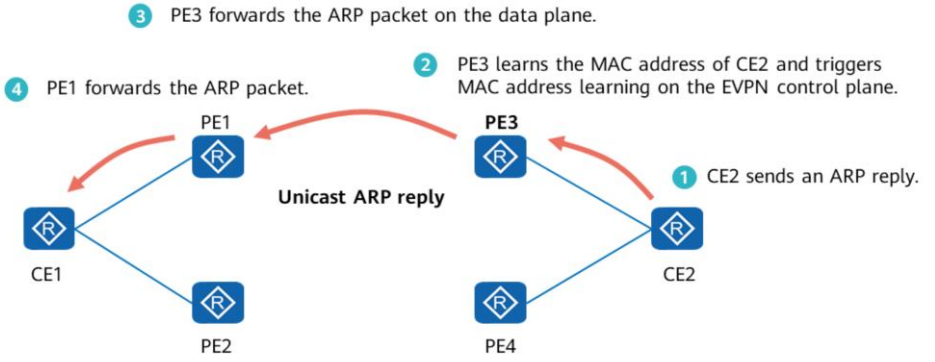


HUAWEI

- ESI labels are used to prevent CE traffic from being looped back.

Unicast ARP Reply

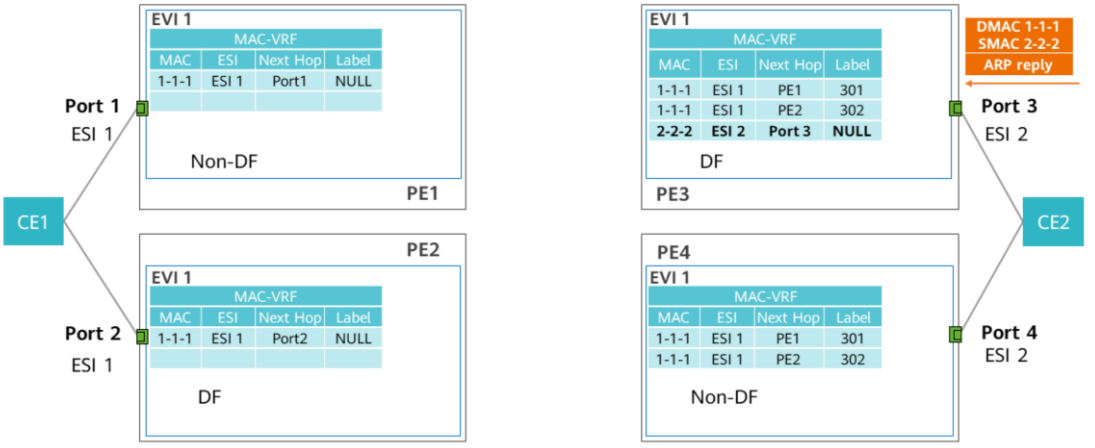
- CE2 sends a unicast ARP reply. PE3 learns the MAC address of CE2 and sends a Type 2 route to trigger the EVPN control plane behavior. PE3 searches the MAC address table to forward the unicast ARP packet to PE1. Finally, PE1 forwards the ARP packet to CE1.



HUAWEI

Unicast ARP Reply: Local MAC Address Learning

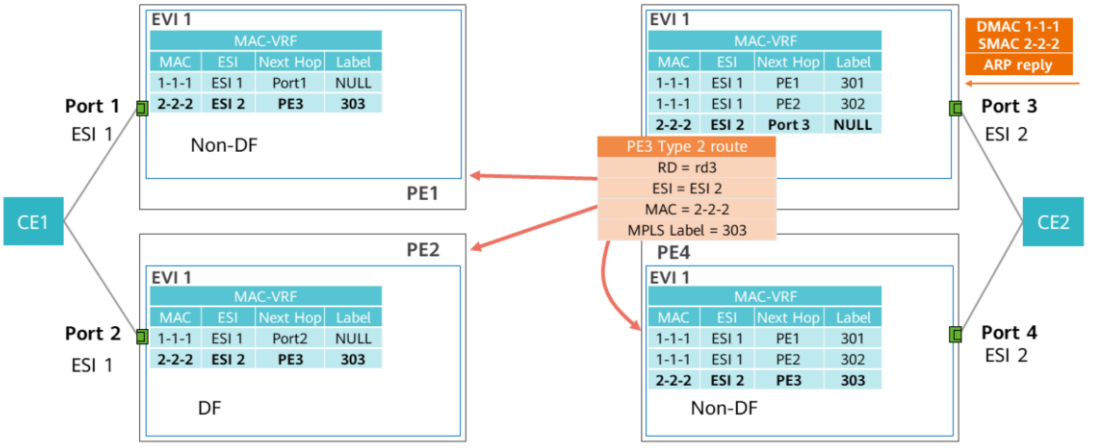
- CE2 sends a unicast ARP reply. PE3 learns the MAC address of CE2 through the data plane and generates a local MAC-VRF entry.



HUAWEI

Unicast ARP Reply: MAC Address Advertisement (1)

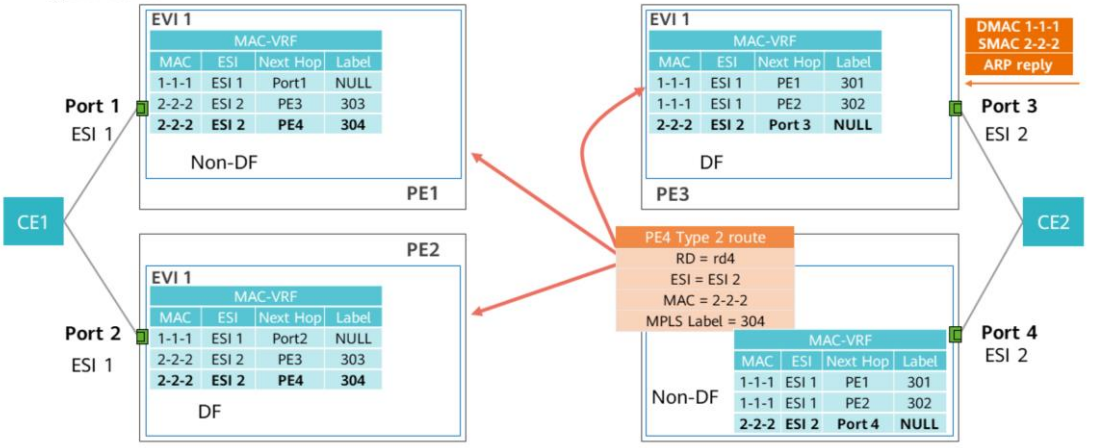
- PE3 generates and advertises a Type 2 route. After receiving the Type 2 route, the other PEs update their local MAC address entries.



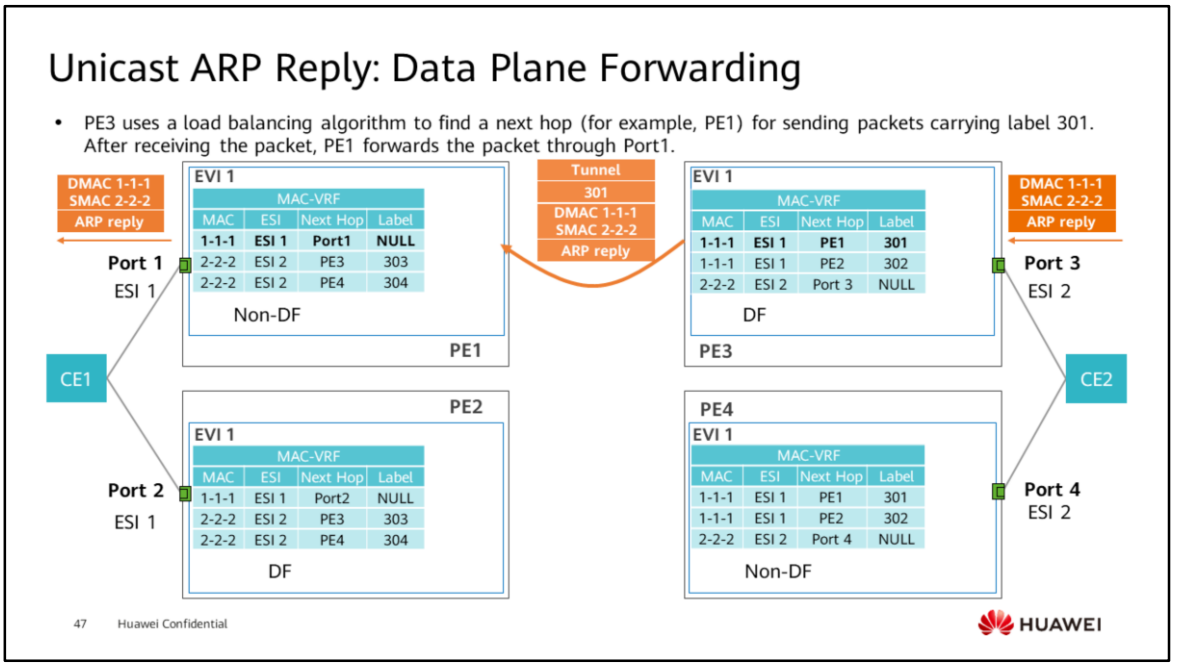
HUAWEI

Unicast ARP Reply: MAC Address Advertisement (2)

- The interface of PE4 belongs to ESI 2, so PE4 updates its MAC address table, and generates as well as advertises a Type 2 route.



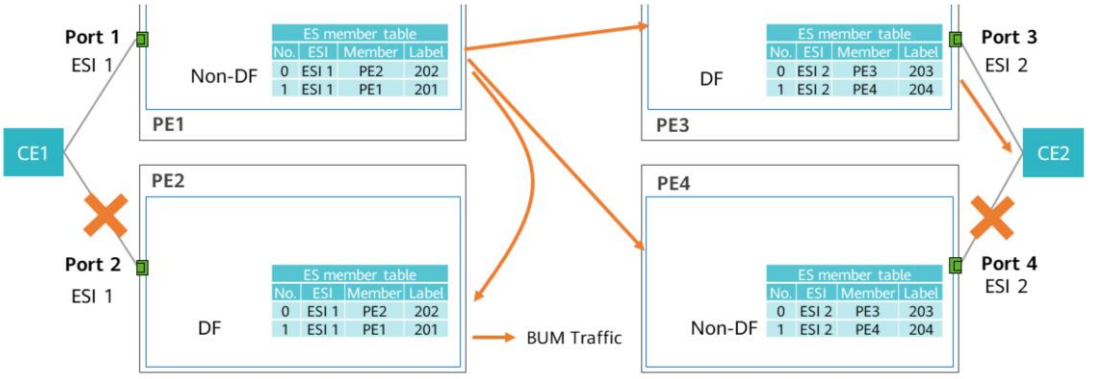
HUAWEI



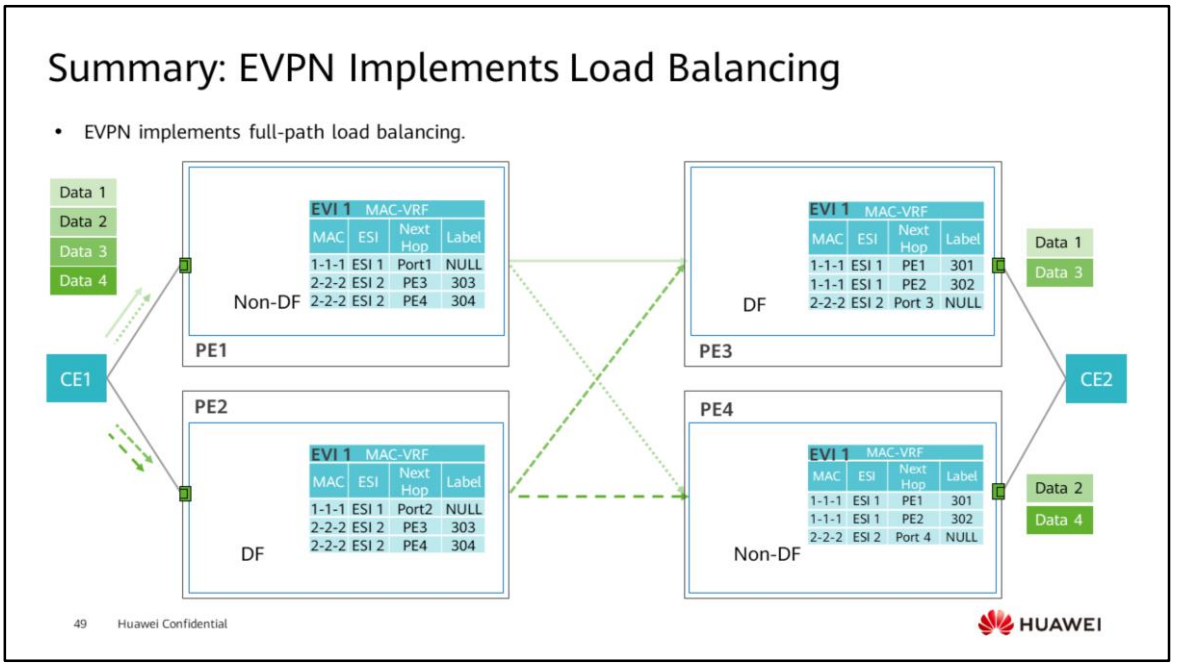
- PE3 searches the MAC-VRF table based on the destination MAC address (1-1-1) of the packet and finds that there are two next hops: PE1 and PE2. PE3 uses a load balancing algorithm to select an appropriate next hop (for example, PE1) for packet forwarding.
- After receiving the ARP reply, PE1 searches the MAC-VRF table and sends the packet through Port1.

Summary: EVPN Solves the Problems Brought by the Active-Active Mode

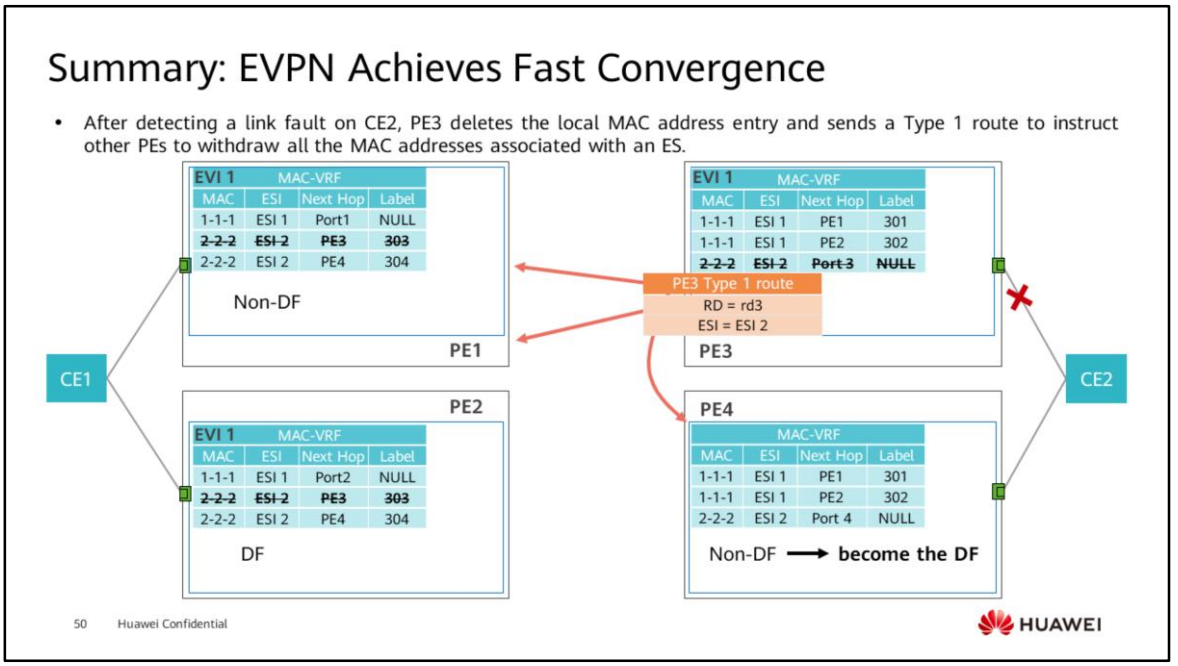
- ESIs carried in Type 1 routes prevent loops on the CE side. Type 4 routes are used for DF election, preventing multiple copies of broadcast traffic from being sent to CEs.



HUAWEI



- CE1 sends traffic to PE1 and PE2 over two active paths for load balancing. Because PE1 and PE2 each have established two paths to CE2, traffic can be load balanced between the two paths. Finally, four service data flows are sent to CE2 over different paths.



- After receiving the Type 1 route, PE4 becomes a DF.
- After receiving the Type 1 route, PE1 and PE2 update their MAC labels and withdraw the MAC routes to ES2. Traffic is automatically switched to PE4.
- If a CE is connected to multiple PEs and a connectivity fault occurs between a specific PE and the CE, the PE must send an Ethernet A-D per ES route to withdraw all the MAC addresses previously advertised.

Contents

1. EVPN Background and Terms
2. **EVPN Fundamentals**
   - EVPN Route Overview and Interaction Process
   - EVPN Route Types
   - EVPN Access Principles
3. Inter-AS EVPN
4. Typical EVPN Application Scenarios
5. Basic EVPN Configurations

HUAWEI

EVPN Route Review

- The EVPN working process describes the functions of Type 1 to Type 4 routes. This section will continue to describe the packets and application scenarios of EVPN routes.

Route Type	Function	Benefits
(Type 1) Ethernet A-D route	Aliasing MAC address withdrawal in batches All-active indication ESI label advertisement	Loop prevention Fast convergence Load balancing
(Type 2) MAC/IP advertisement route	MAC address learning and advertisement MAC/IP binding MAC address mobility	Policy based on each MAC address ARP suppression Host migration
(Type 3) Inclusive multicast route	Auto-discovery of multicast tunnel endpoints & multicast types	BUM traffic forwarding
(Type 4) Ethernet segment route	ES member auto-discovery DF election	Support for both all-active and single-active modes

HUAWEI

Key Parameters Carried in EVPN Routes

- The functions of the four types of EVPN routes are closely related to the parameters carried in the routes.
  - Type 1 routes carry ESI labels to implement split horizon.
  - Type 2 routes carry unicast labels to guide unicast traffic forwarding.
  - Type 3 routes carry BUM labels to guide BUM traffic forwarding, but Type 3 routes do not carry ES information.
  - Type 4 routes carry ESIs for DF election but do not carry any label information.

Parameter / Route Type	RD	ESI	Ethernet Tag ID	Originating Router's IP Address	MAC/IP Address	MPLS Label
Type 1	√	√	√			ESI Label
Type 2	√	√	√		√	√
Type 3	√		√	√		BUM Label
Type 4	√	√		√		

HUAWEI

Ethernet A-D Route

- Ethernet Auto-Discovery (A-D) routes are classified into two types:
  - Ethernet A-D per ES route: used for fast convergence, redundancy mode, and split horizon.
  - Ethernet A-D per EVI route: used for aliasing.

NLRI Format	Field Description
Route Distinguisher (8 bytes)	In an Ethernet A-D per ES route, this field contains the source IP address configured on a PE, for example, X.X.X.X:0. In an Ethernet A-D per EVI route, this field is the RD of an EVPN instance.
Ethernet Segment Identifier (10 bytes)	This field uniquely identifies connections between PEs and a CE. In an Ethernet A-D per ES route, this field is all Fs. In an Ethernet A-D per EVI route, this field identifies different sub-broadcast domains of an ES. If this field is all 0s, the EVI has only one broadcast domain.
Ethernet Tag ID (4 bytes)	This field identifies an Ethernet tag. According to the RFC recommendation, the last 20 bits of this field can be the VLAN ID.
MPLS Label (3 bytes)	In an Ethernet A-D per ES route, the field is all 0s. In an Ethernet A-D per EVI route, this field is the MPLS label used to forward EVPN unicast traffic in load balancing mode.

HUAWEI

ESI Label Extended Community Attribute

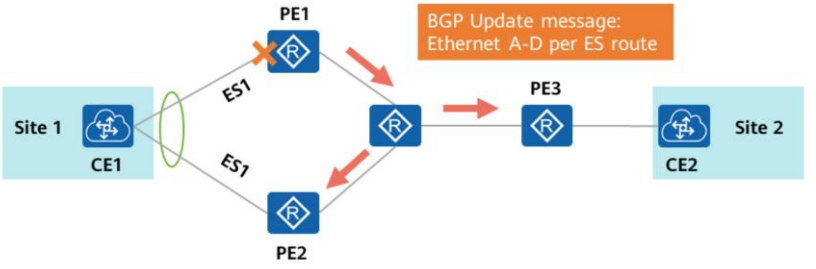
- ESI Label is an extended community attribute carried in an EVPN Type 1 route.
- In a multi-homing scenario, Type 1 routes must carry this community attribute to implement split horizon.

Extended Community Attribute	Field Description
Type (1 byte)	This field has a fixed value of 0x06.
Sub-Type (1 byte)	This field has a fixed value of 0x01.
Flag (1 byte)	The lowest bit of this field is defined as the single-active bit. The value 0 indicates the multi-active scenario.
Reserved (2 bytes)	This field is reserved and all 0s.
ESI Label (3 bytes)	This field is used by a PE to identify each ES for split horizon.

HUAWEI

Fast Convergence

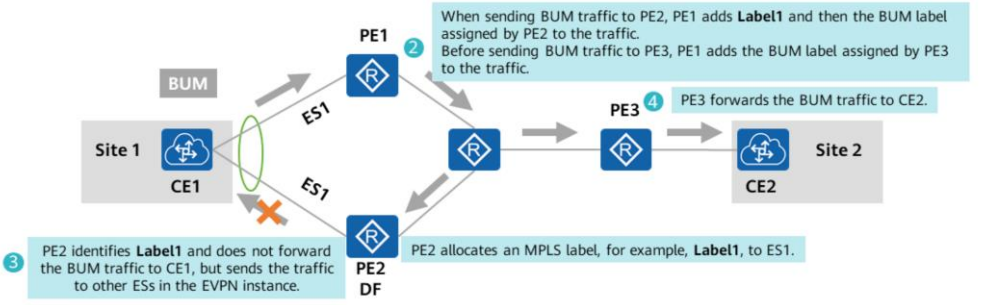
- In EVPN, MAC address learning is controlled by BGP. On a large-scale complex network, route convergence takes a long time. To address this issue, EVPN defines a mechanism to efficiently instruct peer PEs to update their forwarding tables. Specifically, PEs advertise Ethernet A-D per ES routes for all ESs.
- When a PE detects a fault on the connected CE, the PE withdraws the corresponding MAC address entry and sends an Ethernet A-D per ES route to instruct other PEs to withdraw the corresponding MAC address entry.
- An Ethernet A-D per ES route must carry the ESI Label extended community attribute.



HUAWEI

Split Horizon

- When a CE is multi-homed to PEs, if the links of the CE work in all-active mode, BUM packets sent by the CE to a PE may be looped back by another PE. This problem needs to be solved using split horizon.
- All PEs advertise Ethernet A-D per ES routes that carry the ESI Label extended community attribute to implement split horizon. As shown in the following figure, PE2 (DF) assigns an ESI label to identify ES1 and advertises the label to PE1 (non-DF) through an Ethernet A-D per ES route. PE1 adds this label to a BUM packet before sending the packet to PE2. After receiving the packet, PE2 finds that the label is assigned by itself and does not forward the packet to ES1.

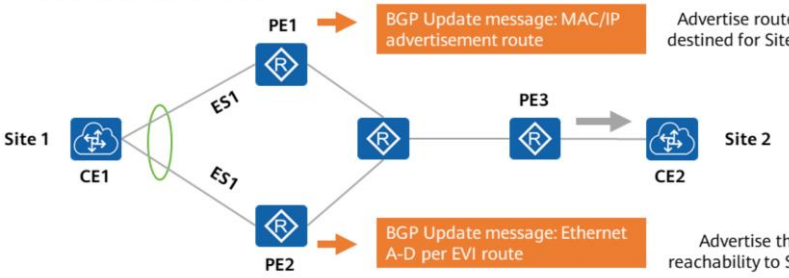


HUAWEI

Type 1 Type 2 Type 3 Type 4

Aliasing

- When a CE is multi-homed to PEs in all-active mode, some PEs may fail to learn the MAC address of the CE. As a result, the peer PE receives MAC/IP advertisement routes from only one PE, failing to implement load balancing among paths between PEs. Aliasing solves this problem using Ethernet A-D per EVI routes. Aliasing ensures that a PE is reachable even if the PE does not learn MAC addresses from the EVI/ES.
- In this example, CE1 is dual-homed to PE1 and PE2. If PE1 has learned the address of Site1 but PE2 has not, PE1 sends a MAC/IP route carrying detailed information to PE2, and PE2 advertises reachability through an Ethernet A-D per EVI route. Therefore, PE3 considers that PE1 and PE2 at Site 1 are reachable.



HUAWEI

- When a peer PE receives a MAC/IP advertisement route, it needs to check whether the advertised MAC address is reachable through other PEs. Therefore, the PE checks the address reachability based on the Ethernet A-D per EVI route.
- Note: The peer PE may first receive the Ethernet A-D per EVI route and then the Ethernet A-D per ES route. To prevent this problem, the peer PE forwards traffic only when it receives the Ethernet A-D per EVI route and Ethernet A-D per ES route at the same time.
- When a CE is multi-homed to PEs in single-active mode, the PE establishes a backup path using the Ethernet A-D per EVI route and Ethernet A-D per ES route.

MAC/IP Advertisement Route

- MAC/IP advertisement routes are used to advertise MAC and IP addresses.

NLRI Format	Field Description
Route Distinguisher (8 bytes)	This field indicates the RD of an EVPN instance.
Ethernet Segment Identifier (10 bytes)	This field uniquely identifies connections between PEs and a CE.
Ethernet Tag ID (4 bytes)	This field is the VLAN ID configured on the device.
MAC Address Length (1 byte)	This field is set to 48 bits.
MAC Address (6 bytes)	This field indicates the host MAC address carried in the route.
IP Address Length (1 byte)	This field is optional and can be set to 32 or 128 bits.
IP Address (0, 4, or 16 bytes)	This field is optional and indicates the host IP address carried in the route.
MPLS Label1 (3 bytes)	This field is assigned by the downstream router to forward Layer 2 service traffic.
MPLS Label2 (0 or 3 bytes)	This field is optional and used to forward Layer 3 service traffic.

HUAWEI

Type 1 Type 2 Type 3 Type 4

MAC Mobility Extended Community Attribute

- MAC Mobility is an extended community attribute carried in an EVPN Type 2 route.
- The least significant bit in the Flag field is defined as the Sticky/static bit. If this bit is 1, the MAC address is fixed and cannot be moved.

Extended Community Attribute	Field Description
Type (1 byte)	This field has a fixed value of 0x06.
Sub-Type (1 byte)	This field has a fixed value of 0x00.
Flag (1 byte)	The least significant bit of this field is defined as the Sticky/static bit. If this bit is 0, the MAC address can be moved.
Reserved (1 byte)	This field is reserved and all 0s.
Sequence Number (4 bytes)	This field ensures that a PE has correct MAC/IP advertisement routes when multiple update messages are received.

HUAWEI

- MAC address migration will be described based on VXLAN in the following sections.

Inclusive Multicast Route

- Inclusive multicast routes are used to process BUM traffic. BUM traffic includes broadcast, multicast, and unknown unicast traffic.
- After a BGP peer relationship is established between PEs, the PEs transmit inclusive multicast routes to each other. An inclusive multicast route carries the RD and RTs of the EVPN instance on the local PE, source IP address (usually the loopback address of the local PE), and Provider Multicast Service Interface (PMSI). The PMSI and RT are carried in route attribute information while the RD and source IP address are carried in NLRI information.

NLRI Format	Field Description
Route Distinguisher (8 bytes)	This field indicates the RD of an EVPN instance.
Ethernet Tag ID (4 bytes)	This field is the VLAN ID configured on the device.
IP Address Length (1 byte)	This field indicates the length of the source IP address configured on a PE.
Originating Router's IP Address (4 or 16 bytes)	This field indicates the source IP address configured on a PE.

HUAWEI

PMSI Attribute

- The PMSI attribute carries the tunnel type (ingress replication or MLDP) and tunnel label information used for transmitting multicast packets.
- A PE forwards the multicast traffic that it receives to other PEs in P2MP mode. The PEs can establish a tunnel to transmit multicast traffic through inclusive multicast routes.
- The PMSI attribute is defined in RFC6514. In the EVPN scenario, the Leaf bit in the Flags field is 0 and the tunnel type is 6 (ingress replication).

PMSI Attribute	Field Description
Flags (1 byte)	The last bit of this field is the Leaf bit.
Tunnel Type (1 byte)	The value of this field is typically 6 (ingress replication) in EVPN.
MPLS Label (3 bytes)	This field is an MPLS label.
Tunnel Identifier (variable length)	This field is the IP address of the tunnel end.

HUAWEI

Ethernet Segment Route

- Ethernet segment routes are mainly used for DF election.
- Ethernet segment routes carry the ESI, source IP address, and RD (source IP address:0) of the local PE. PEs connecting to the same CE use Ethernet segment routes to discover each other.

NLRI Format	Field Description
Route Distinguisher (8 bytes)	This field indicates the RD of an EVPN instance.
Ethernet Segment Identifier (10 bytes)	This field is the unique identifier of the connection between local and peer devices.
IP Address Length (1 byte)	This field indicates the IP address length.
Originating Router's IP Address (4 or 16 bytes)	This field indicates the source address configured on a PE.

HUAWEI

ES-Import Route Target

- ES-Import RT is an extended community attribute newly defined in EVPN and carried in Type 4 routes.
- In an EVPN multi-homing scenario, the ES routes advertised by BGP EVPN must carry the ES-Import RT attribute for ES route filtering.
- ES route filtering: ES-Import RT allows only PEs connected to the same site to import Type 4 routes. RTs are used to import Ethernet segment routes to all the PEs connected to the same site.

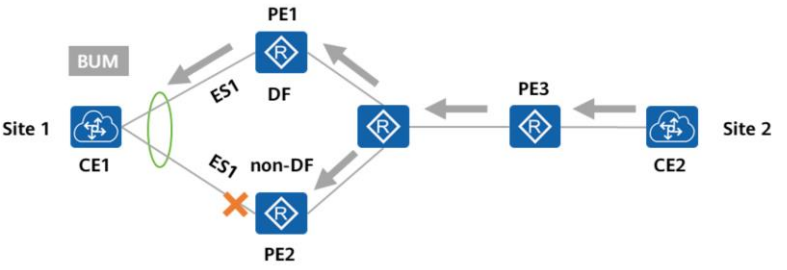
Extended Community Attribute	Field Description
Type (1 byte)	This field has a fixed value of 0x06.
Sub-Type (1 byte)	This field has a fixed value of 0x02.
ES-Import (2 bytes)	The least significant bit of this field is defined as the Single-Active bit. The value 0 indicates the multi-homing scenario.
ES-Import Cont'd (4 bytes)	This field is reserved and all 0s.

HUAWEI

Type 1 Type 2 Type 3 Type 4

DF Election

- In a CE multi-homing scenario, an ES may be configured with multiple Ethernet tags, and only one PE is elected as a designated forwarder (DF). DF election is implemented using Type 4 routes that carry the ES-Import attribute. A DF provides the following functions:
  - Sends multicast and broadcast traffic to the CEs on the specified ES.
  - Floods unknown unicast traffic to CEs.



HUAWEI

- DF election rules:
  - A PE discovers the ES and ESI of the local connection and advertises the Type 4 route carrying the ES-Import.
  - The PE starts the timer. The default value of the timer is 3 seconds, within which ES routes can be received.
  - After the timeout, the PE generates an ordered list. The list contains the IP addresses of all PEs and information about their connections to the ES. The sequence number of the list starts from 0 in ascending order. The sequence number is used to determine the DF.
  - The PE elected as the DF forwards BUM traffic to CEs. When a link fault occurs, the PE withdraws its ES routes, which triggers a re-election process.

New EVPN Routes

- EVPN is not limited to L2VPN applications. With the increase of EVPN route types, more applications, such as L3VPN, are supported.

Route Type	Function	Benefits
(Type 1) Ethernet A-D route	Aliasing MAC address withdrawal in batches All-active indication ESI label advertisement	Loop prevention Fast convergence Load balancing
(Type 2) MAC/IP advertisement route	MAC address learning and advertisement MAC/IP binding MAC address mobility	Policy based on each MAC address ARP suppression Host migration
(Type 3) Inclusive multicast route	Auto-discovery of multicast tunnel endpoints & multicast types	BUM traffic forwarding
(Type 4) Ethernet segment route	ES member auto-discovery DF election	Support for both all-active and single-active modes
(Type 5) IP prefix route	IP prefix advertisement	Support for L3VPN

HUAWEI

IP Prefix Route

- IP prefix routes are used to advertise a host IP address received from an access network or the network segment where the host IP address resides.

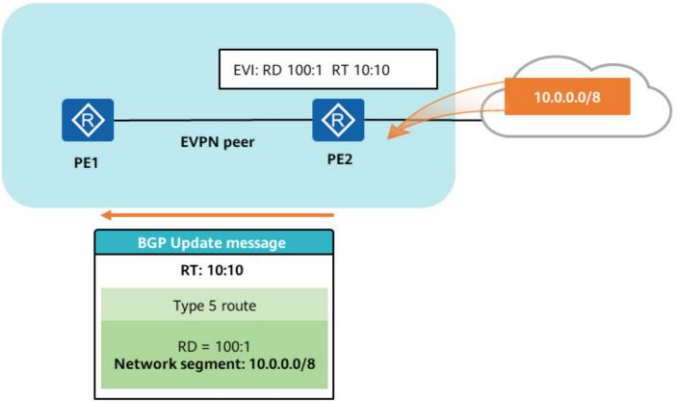
NLRI Format	Field Description
Route Distinguisher (8 bytes)	This field indicates the RD of an EVPN instance.
Ethernet Segment Identifier (10 bytes)	This field uniquely identifies connections between PEs and a CE.
Ethernet Tag ID (4 bytes)	Currently, this field can only be set to 0.
IP Prefix Length (1 byte)	This field indicates the mask length of the IP prefix carried in the route.
IP Prefix (4 or 16 bytes)	This field indicates the IP prefix carried in the route.
GW IP Address (4 or 16 bytes)	This field indicates the default gateway address.
MPLS Label (3 bytes)	This field indicates the label used for Layer 3 service traffic forwarding.

HUAWEI

- For more details about Type 5 routes, see the RFC draft.

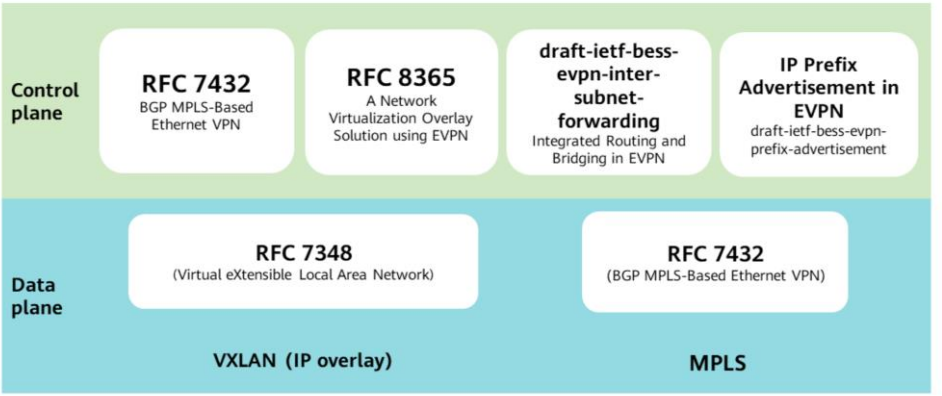
Typical Application Scenarios of IP Prefix Routes

- Type 5 routes are advertised to an EVPN instance to implement interworking between the EVPN instance and external networks.



HUAWEI

EVPN Protocol Standards



HUAWEI

Contents

1. EVPN Background and Terms
2. **EVPN Fundamentals**
   - EVPN Route Overview and Interaction Process
   - EVPN Route Types
   - EVPN Access Principles
3. Inter-AS EVPN
4. Typical EVPN Application Scenarios
5. Basic EVPN Configurations

HUAWEI

EVPN Access Overview

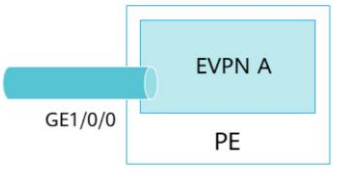
- Multiple Ethernet VPN instances (EVIs) can be configured on PEs at the edge of an EVPN, with each EVI connecting to one or more user networks. EVPN allows user network access in various service modes, as described in the following table.

Service Mode	Application Scenario
Port-based mode	The physical interface connected to a user network is directly bound to a common EVI. This service mode is used to carry only Layer 2 services.
VLAN-based mode	The physical interface connected to a user network is divided into different sub-interfaces. Each sub-interface is bound to a specific EVI. One EVI is required per user. This service mode is used to carry Layer 2 or Layer 3 services.
VLAN bundle mode	Users are divided based on VLANs. Each VLAN is bound to a specific EVI. This service mode is used to carry Layer 2 or Layer 3 services.
VLAN-aware bundle mode	Users are divided based on VLANs. The VLANs are bound to the same EVI. This service mode is used to carry Layer 2 or Layer 3 services.

HUAWEI

Port-based Mode

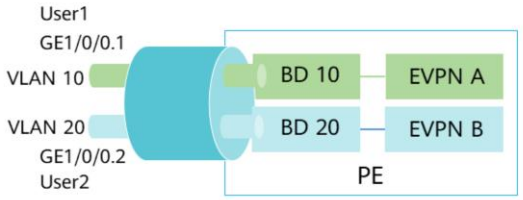
- In port-based mode, an interface is used only by a single user. Specifically, the physical interface connected to a user network is bound to a common EVI and has no sub-interfaces created. This service mode is used only to carry Layer 2 services.




```
[~PE] interface GigabitEthernet 1/0/0
[*PE-GigabitEthernet1/0/0] evpn binding vpn-instance evpna
[*PE-GigabitEthernet1/0/0] commit
[~PE-GigabitEthernet1/0/0] quit
```

VLAN-based Mode

- In VLAN-based mode, the physical interfaces connected to user networks each have different sub-interfaces created. Each sub-interface is associated with a unique VLAN and added to a specific BD, and each BD is bound to a specific EVI. This service mode is used to carry Layer 2 or Layer 3 services. In this example, User1 and User2 access the network through different sub-interfaces.



EVPN A configuration example:

```
[~PE] bridge-domain 10
[*PE-bd10] evpn binding vpn-instance evpna
[*PE-bd10] quit
[*PE] interface GigabitEthernet 1/0/0.1 mode l2
[*PE-G1/0/0.1] encapsulation dot1q vid 10
[*PE-G1/0/0.1] bridge-domain 10
[*PE-G1/0/0.1] commit
```

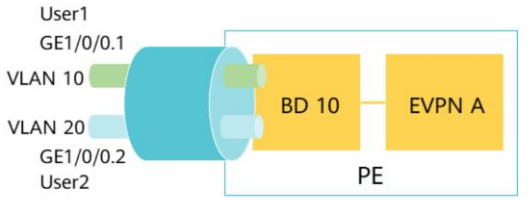
User1's service packets are tagged with VLAN 10 and transmitted through GE1/0/0.1.
User2's service packets are tagged with VLAN 20 and transmitted through GE1/0/0.2.

HUAWEI

- In this service mode, the sub-interface, VLAN, BD, and EVI are exclusively used by a user to access the network, and a separate MAC forwarding table is used on the forwarding plane for each user. Although this mode effectively ensures service isolation, it consumes a large amount of EVI resources because each user requires one EVI.

VLAN Bundle Mode

- In VLAN bundle mode, an EVI connects to multiple users who are divided by VLAN, and the EVI is bound to a BD. In this service mode, the users connected to the same EVI share a MAC forwarding table, requiring each user on the network to have a unique MAC address. This service mode is used to carry Layer 2 or Layer 3 services.



User1's service packets are tagged with VLAN 10 and transmitted through GE1/0/0.1.
User2's service packets are tagged with VLAN 20 and transmitted through GE1/0/0.2.

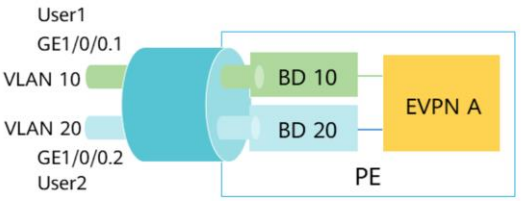
EVPN A configuration example:

```
[~PE] bridge-domain 10
[*PE-bd10] evpn binding vpn-instance evpna
[*PE-bd10] quit
[*PE] interface GigabitEthernet 1/0/0.1 mode l2
[*PE-G1/0/0.1] encapsulation dot1q vid 10
[*PE-G1/0/0.1] bridge-domain 10
[*PE-G1/0/0.1] commit
```


HUAWEI

VLAN-Aware Bundle Mode

- In VLAN-aware bundle mode, an EVI connects to multiple users divided by VLAN. Additionally, the EVI can be bound to multiple BDs. In this service mode, users connected to the same EVI use separate forwarding entries. During traffic forwarding, the system uses the BD tag carried in packets to locate the corresponding BD MAC forwarding table and searches the table for a forwarding entry based on a MAC address.



EVPN A configuration example:

```
[~PE] bridge-domain 10
[*PE-bd10] evpn binding vpn-instance evpna bd-tag 10
[*PE-bd10] quit
[*PE] interface GigabitEthernet 1/0/0.1 mode l2
[*PE-G1/0/0.1] encapsulation dot1q vid 10
[*PE-G1/0/0.1] bridge-domain 10
[*PE-G1/0/0.1] commit
```

User1's service packets are tagged with VLAN 10 and transmitted through GE1/0/0.1.
User2's service packets are tagged with VLAN 20 and transmitted through GE1/0/0.2.

HUAWEI

- When EVPN peers send routes to each other, a BD tag is encapsulated into the Ethernet Tag ID field of Ethernet A-D route packets, MAC/IP advertisement route packets, and inclusive multicast route packets.

Contents

1. EVPN Background and Terms
2. EVPN Fundamentals
3. **Inter-AS EVPN**
4. Typical EVPN Application Scenarios
5. Basic EVPN Configurations

HUAWEI

Inter-AS EVPN Overview

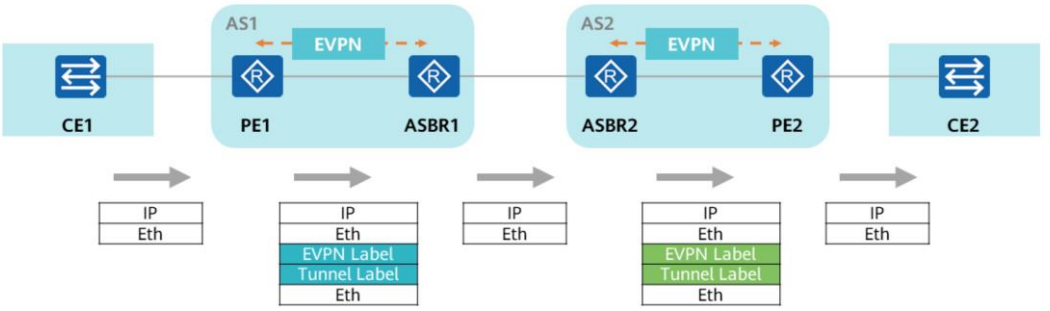
- Typically, an EVPN architecture runs in an autonomous system (AS). Routing information of any EVPN can be flooded only in the local AS. In some complex scenarios, an enterprise may have multiple ASs. In this case, the existing EVPN architecture needs to be extended to provide an inter-AS EVPN.
- Inter-AS EVPN can be implemented in the following ways, similar to inter-AS MPLS VPN:
  - Option A
  - Option B
  - Option C

HUAWEI

- This course describes inter-AS EVPN L3VPN.

Inter-AS EVPN: Option A

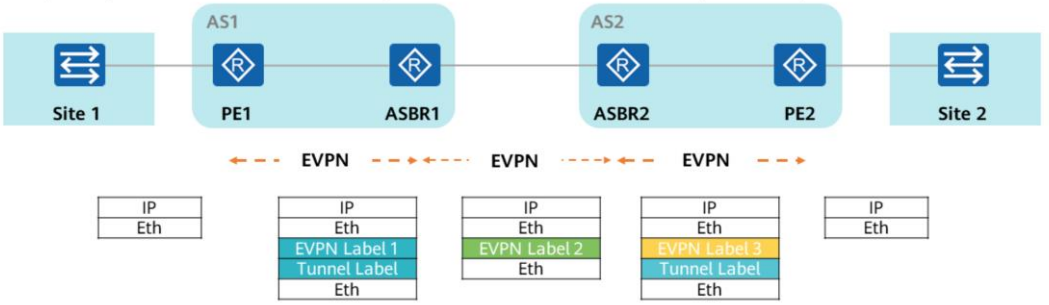
- In inter-AS EVPN Option A, EVPN peer relationships need to be established between PEs and ASBRs, and EVPN does not need to run between ASBRs.
- Multiple sub-interfaces are created between ASBRs and bound to VPN instances to transmit IP routes.
- In Option A, intra-AS traffic has two labels, and inter-AS traffic has no label.



HUAWEI

Inter-AS EVPN: Option B

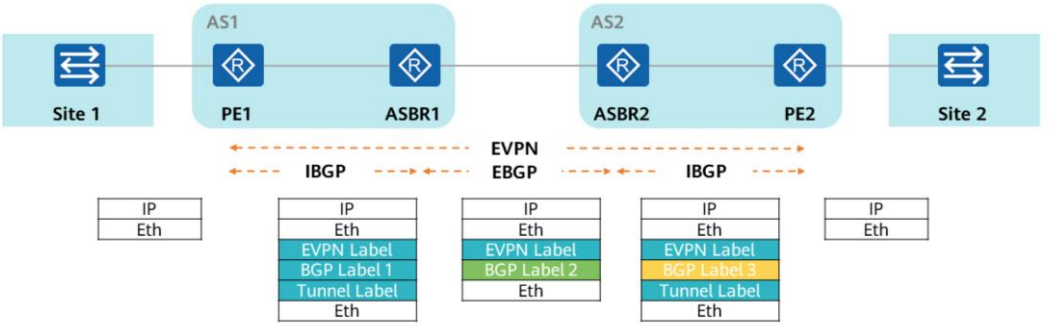
- In inter-AS EVPN Option B, EVPN peer relationships need to be established between PEs and ASBRs, and between ASBRs.
- After receiving a MAC/IP advertisement route, an ASBR changes the next hop to itself, allocates a new label to the route, and sends the route to the ASBR in another AS.
- In Option B, intra-AS traffic has two labels, and inter-AS traffic has one label (EVPN label).



HUAWEI

Inter-AS EVPN: Option C

- In inter-AS EVPN Option C, an E2E BGP LSP needs to be established between the PEs at different sites (Site 1 and Site 2 in this example).
- PE1 and PE2 establish an EVPN peer relationship. ASBRs are unaware of MAC/IP advertisement routes.
- In Option C, intra-AS traffic has three labels, and inter-AS traffic has two labels (EVPN label and BGP label).

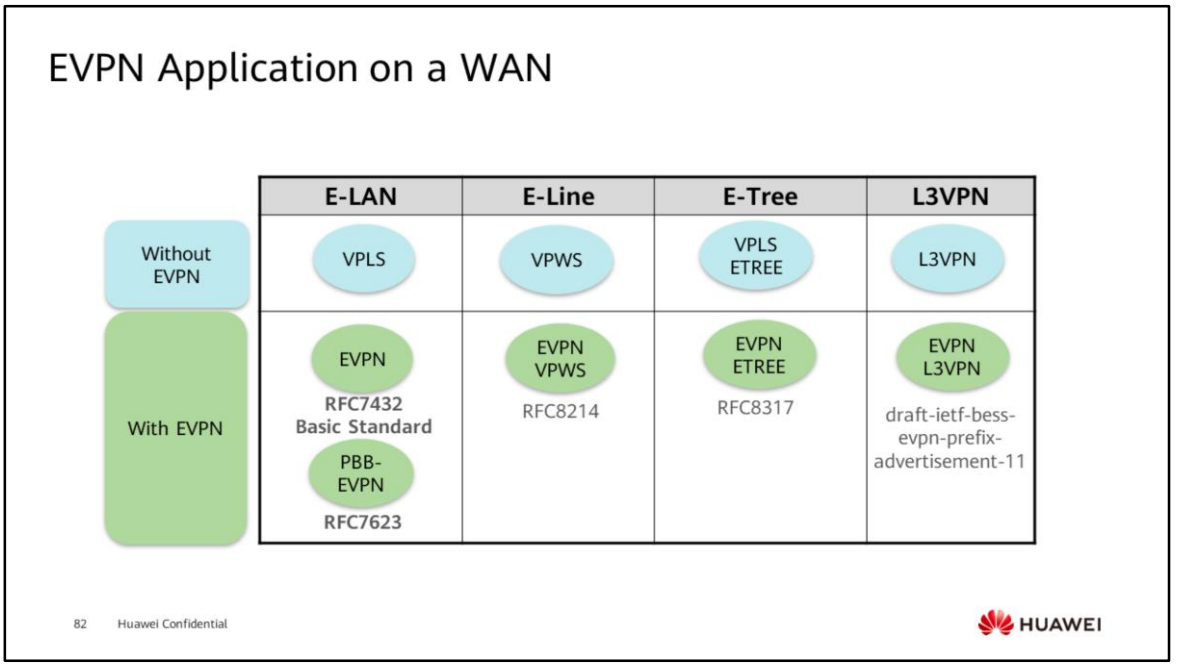


HUAWEI

Contents

1. EVPN Background and Terms
2. EVPN Fundamentals
3. Inter-AS EVPN
4. **Typical EVPN Application Scenarios**
5. Basic EVPN Configurations

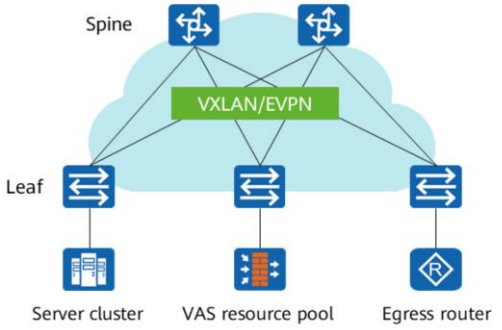
HUAWEI



- E-Line, E-Tree, and E-LAN are three types of Ethernet virtual connection (EVC), which are point-to-point EVC, multipoint-to-multipoint EVC, and rooted-multipoint EVC.
    - E-Line: A point-to-point EVC associates two user-network interfaces (UNIs).
    - E-LAN: A multipoint-to-multipoint EVC can associate two or more UNIs. Users or carriers can add any number of 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 directly communicate with the root UNI in the EVC. Two leaf UNIs cannot communicate with each other directly.

EVPN Application in a DC

- In a cloud DC, the EVPN Network Virtualization Overlay (NVO) solution (RFC8365) can be used.
- It is recommended that the data plane use VXLAN encapsulation and the control plane use EVPN to construct a flexible DC overlay network.

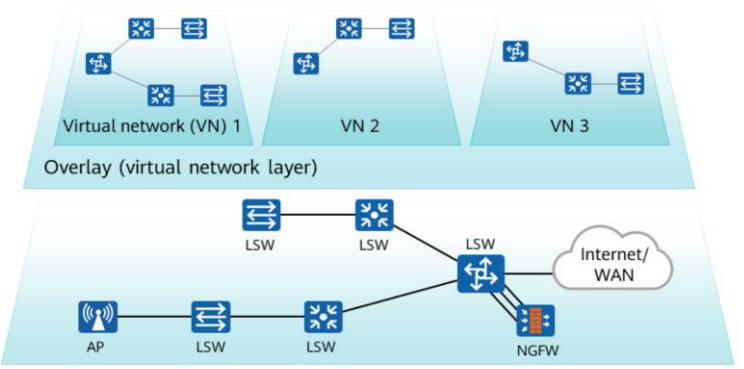


- All services in the DC are carried over the VXLAN overlay network.
- The underlay network consisting of spine and leaf nodes is responsible for high-speed forwarding.

HUAWEI

EVPN Application on a Campus Network

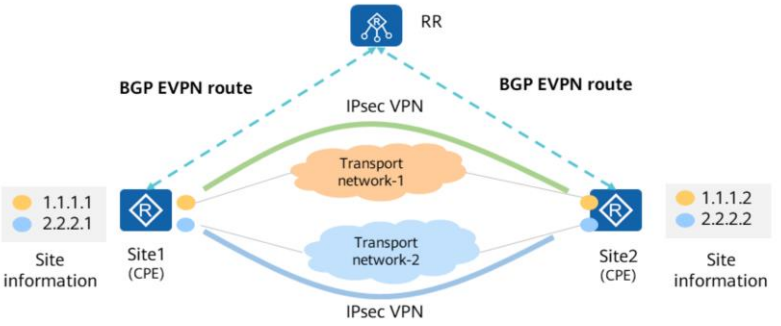
- Similar to the cloud DC solution, the campus network virtualization solution can also use EVPN NVO (RFC8365).
- On different underlying networks, VXLAN encapsulation is used together with EVPN on the control plane to build a flexible campus overlay network.



HUAWEI

EVPN Application in SD-WAN

- SD-WAN is a next-generation enterprise branch interconnection solution that supports features such as intelligent traffic steering, zero touch provisioning (ZTP), and visualization.
- In the SD-WAN solution, EVPN can be deployed between the route reflector (RR) and CPE to transmit SD-WAN overlay VPN routes on the control plane. The data plane uses IPsec VPN to build a secure forwarding channel.



HUAWEI

Contents

1. EVPN Background and Terms
2. EVPN Fundamentals
3. Inter-AS EVPN
4. Typical EVPN Application Scenarios
5. **Basic EVPN Configurations**
   - Configuring an EVPN to Carry Layer 2 Services
   - Configuring an EVPN to Carry Layer 3 Services

HUAWEI

Configuring BD EVPN Functions

1. Configure an EVPN instance in BD mode.

```
[Huawei] evpn vpn-instance evpna bd-mode
[Huawei-evpn-instance-evpna] route-distinguisher 100:1
[Huawei-evpn-instance-evpna] vpn-target 1:1 [ both | export-extcommunity | import-extcommunity ]
```

An EVPN instance named **evpna** is created, the EVPN instance view is displayed, and the RD and RT are set to 100:1 and 1:1, respectively.

2. Configure an EVPN source address.

```
[Huawei] evpn source-address ip-address
```

The EVPN source address is part of EVPN route information and can be used to identify a PE on an EVPN. Configuring EVPN source addresses is a mandatory task for EVPN configuration.

3. Configure an ESI.

```
[Huawei-Eth-Trunk10] esi
```

Perform this configuration if a VLAN is used to access an EVPN. For details about other configurations, see the product manual of the corresponding product.

HUAWEI

- This example describes the configuration commands on Huawei NE20E-S2 (V800R012C10) series routers. For the configuration commands of other models, see the corresponding product manuals.

Configuring a BD and Binding an EVPN Instance to the BD

1. Configure an EVPN instance.

```
[Huawei] bridge-domain 10
[Huawei-bd10] evpn binding vpn-instance evpna [ bd-tag bd-tag ]
```

In MPLS forwarding, bind a BD to the EVPN instance named **evpna**. In this example, the BD ID is 10. By specifying different **bd-tag** values, you can bind multiple BDs with different VLANs to the same EVPN instance and isolate services in these BDs.

2. Add a Layer 2 sub-interface to the BD so that the sub-interface can transmit data packets in this BD.

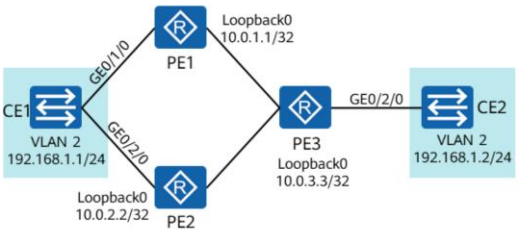
```
[Huawei] interface Eth-Trunk10.1 mode l2
[Huawei-Eth-Trunk10.1] encapsulation dot1q vid 2
[Huawei-Eth-Trunk10.1] rewrite pop single
[Huawei-Eth-Trunk10.1] bridge-domain 10
```

The Layer 2 sub-interface is added to a BD so that it can transmit data packets in this BD. In this example, the sub-interface ID is Eth-Trunk10.1.

The **rewrite pop** { **single** | **double** } command removes single or double VLAN tags from received packets. For single-tagged packets that a Layer 2 sub-interface receives, specify **single** to remove the tags from these packets. If the encapsulation type of packets has been set to QinQ in the previous step, specify **double** to remove double VLAN tags from the received packets.

HUAWEI

Example for Configuring Eth-Trunk Sub-interfaces to Access a BD EVPN (Dual-Homing Active-Active)



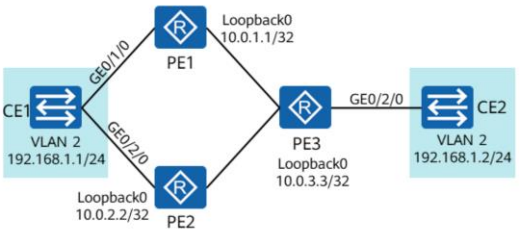
- EVPN needs to be configured on the network to implement Layer 2 communication.
  - CE1 is dual-homed to PE1 and PE2 and works in active-active mode through an E-Trunk.
  - The ESI of CE1 is 0000.1111.2222.1111.1111.
  - IP addresses of CE1 and CE2 are 192.168.1.1/24 and 192.168.1.2/24 respectively.
  - The EVPN instance name is **evpna**.
  - On PE1, PE2, and PE3, the RDs of the EVPN instance **evpna** are 100:1, 200:1, and 300:1 respectively and the RTs are 1:1.

- Configuration roadmap
  - Configure an IGP on the backbone network to allow the PEs to communicate with each other. (The configuration details are not provided here.)
  - Configure basic MPLS functions and enable MPLS LDP to establish LDP LSPs on the backbone network. (The configuration details are not provided here.)
  - Configure a BD EVPN instance on PEs.
  - Configure an ESI and E-Trunk to implement dual-homing active-active networking.
  - Configure local-remote fast reroute (FRR) for MAC routes.
  - Establish BGP EVPN peer relationships.
  - Configure a CE to access the PEs through an Eth-Trunk interface.
  - Verify the configuration.

HUAWEI

- The features required in an EVPN dual-homing scenario, such as fast convergence, split horizon, and DF election, all become invalid in a single-homing scenario. In such a scenario, configuring an ESI is optional on a single-homing PE.

Example for Configuring Eth-Trunk Sub-interfaces to Access a BD EVPN (1)



- **Configure a BD EVPN instance on PEs.**
- Configure an ESI and E-Trunk to implement dual-homing active-active networking.
- Configure local-remote FRR for MAC routes.
- Establish BGP EVPN peer relationships.
- Configure a CE to access the PEs through the Eth-Trunk interface.
- Verify the configuration.

The following uses PE1 as an example to configure an EVPN instance.

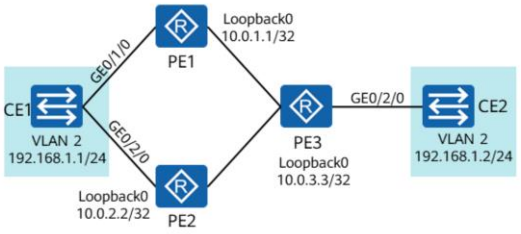
```
[~PE1] evpn vpn-instance evpna bd-mode
[*PE1-evpn-instance-evpna] route-distinguisher 100:1
[*PE1-evpn-instance-evpna] vpn-target 1:1
[*PE1-evpn-instance-evpna] quit
[*PE1] bridge-domain 10
[*PE1-bd10] evpn binding vpn-instance evpna
[*PE1-bd10] quit
[*PE1] commit
```

The following uses PE1 as an example to configure a source address.

```
[~PE1] evpn source-address 1.1.1.1
[*PE1] commit
```


HUAWEI

Example for Configuring Eth-Trunk Sub-interfaces to Access a BD EVPN (2)



- Configure a BD EVPN instance on PEs.
- **Configure an ESI and E-Trunk to implement dual-homing active-active networking.**
- Configure local-remote FRR for MAC routes.
- Establish BGP EVPN peer relationships.
- Configure a CE to access the PEs through the Eth-Trunk interface.
- Verify the configuration.

The following uses PE1 as an example to configure an ESI and E-Trunk for dual-homing active-active networking.

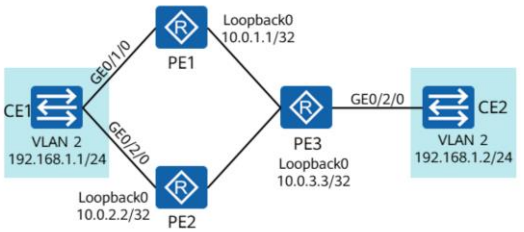
```
[~PE1] lacp e-trunk system-id 00e0-fc00-0000
[*PE1] lacp e-trunk priority 1
[*PE1] e-trunk 1
[*PE1-e-trunk-1] peer-address 2.2.2.2 source-address 1.1.1.1
[*PE1-e-trunk-1] quit
[*PE1] interface eth-trunk 10
[*PE1-Eth-Trunk10] mode lacp-static
[*PE1-Eth-Trunk10] e-trunk 1
[*PE1-Eth-Trunk10] e-trunk mode force-master
[*PE1-Eth-Trunk10] esi 0000.1111.2222.1111.1111
[*PE1-Eth-Trunk10] quit
[*PE1] interface eth-trunk 10.1 mode l2
[*PE1-Eth-Trunk10.1] encapsulation dot1q vid 2
[*PE1-Eth-Trunk10.1] rewrite pop single
[*PE1-Eth-Trunk10.1] bridge-domain 10
[*PE1-Eth-Trunk10.1] quit
[*PE1] interface gigabitethernet 0/1/0
[*PE1-GigabitEthernet0/1/0] eth-trunk 10
[*PE1-GigabitEthernet0/1/0] quit
[*PE1] commit
```

The E-Trunk system ID of PE2 must be the same as that of PE1.

HUAWEI

- For details about the E-Trunk configuration, see the corresponding product documentation.

Example for Configuring Eth-Trunk Sub-interfaces to Access a BD EVPN (3)



- Configure a BD EVPN instance on PEs.
- Configure an ESI and E-Trunk to implement dual-homing active-active networking.
- **Configure local-remote FRR for MAC routes.**
- Establish BGP EVPN peer relationships.
- Configure a CE to access the PEs through the Eth-Trunk interface.
- Verify the configuration.

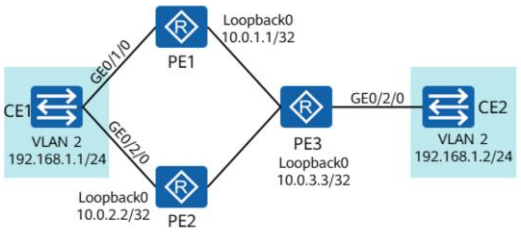
The following uses PE1 as an example to configure local-remote FRR for MAC routes.

```
[~PE1] evpn
[*PE1-evpn] vlan-extend private enable
[*PE1-evpn] vlan-extend redirect enable
[*PE1-evpn] local-remote frr enable
[*PE1-evpn] quit
[*PE1] commit
```

In CE multi-homing scenarios, MAC route redirection enables all PEs connected to the same CE to redirect the outbound interfaces of MAC routes destined for the CE to local AC interfaces. This function shortens forwarding paths and improves forwarding efficiency.
If the AC interface on one of the PEs fails, the outbound interface of the associated MAC route can be quickly redirected so that traffic can be forwarded through another active PE, improving reliability.

HUAWEI

Example for Configuring Eth-Trunk Sub-interfaces to Access a BD EVPN (4)



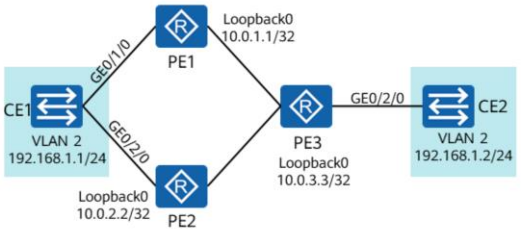
- Configure a BD EVPN instance on PEs.
- Configure an ESI and E-Trunk to implement dual-homing active-active networking.
- Configure local-remote FRR for MAC routes.
- **Establish BGP EVPN peer relationships.**
- Configure a CE to access the PEs through the Eth-Trunk interface.
- Verify the configuration.

The following uses PE1 as an example to establish BGP EVPN peer relationships.

```
[~PE1] bgp 100
[*PE1-bgp] peer 10.0.2.2 as-number 100
[*PE1-bgp] peer 10.0.2.2 connect-interface loopback 1
[*PE1-bgp] peer 10.0.3.3 as-number 100
[*PE1-bgp] peer 10.0.3.3 connect-interface loopback 1
[*PE1-bgp] l2vpn-family evpn
[*PE1-bgp-af-evpn] peer 10.0.2.2 enable
[*PE1-bgp-af-evpn] peer 10.0.3.3 enable
[*PE1-bgp-af-evpn] quit
[*PE1-bgp] quit
[*PE1] commit
```


HUAWEI

Example for Configuring Eth-Trunk Sub-interfaces to Access a BD EVPN (5)



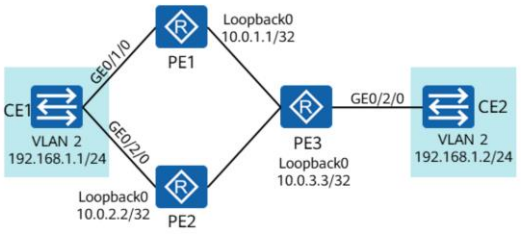
- Configure a BD EVPN instance on PEs.
- Configure an ESI and E-Trunk to implement dual-homing active-active networking.
- Configure local-remote FRR for MAC routes.
- Establish BGP EVPN peer relationships.
- **Configure a CE to access the PEs through the Eth-Trunk interface.**
- Verify the configuration.

The following uses CE1 as an example to configure it to access the PEs through the Eth-Trunk interface.

```
[~CE1] VLAN 2
[*CE1-vlan2] quit
[*CE1] interface Eth-Trunk10
[*CE1-Eth-Trunk10] portswitch
[*CE1-Eth-Trunk10] port link-type trunk
[*CE1-Eth-Trunk10] port trunk allow-pass VLAN 2
[*CE1-Eth-Trunk10] mode lacp-static
[*CE1-Eth-Trunk10] quit
[*CE1] interface gigabitethernet0/1/0
[*CE1-GigabitEthernet0/1/0] eth-trunk 10
[*CE1-GigabitEthernet0/1/0] quit
[*CE1] interface gigabitethernet0/2/0
[*CE1-GigabitEthernet0/2/0] eth-trunk 10
[*CE1-GigabitEthernet0/2/0] quit
[*CE1] commit
```


HUAWEI

Verifying the Configuration (1)



Check the BGP EVPN peer relationship on PE3.

```
[~PE3]display bgp evpn peer
 BGP local router ID : 10.0.3.3
 Local AS number : 100
 Total number of peers : 2    Peers in established state : 2
  Peer            AS  State          PrefRcv
  10.0.1.1        100 Established        4
  10.0.2.2        100 Established        4
```

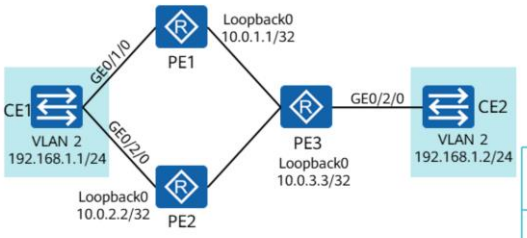
Check Type 3 routes on PE3.

```
<PE3>display bgp evpn  all routing-table inclusive-route
 Local AS number : 100
 EVPN address family:
 Number of Inclusive Multicast Routes: 3
 Route Distinguisher: 100:1
      Network(EthTagId/IpAddrLen/OriginalIp) NextHop
 *>i  0:32:10.0.1.1                          10.0.1.1
 Route Distinguisher: 200:1
      Network(EthTagId/IpAddrLen/OriginalIp) NextHop
 *>i  0:32:10.0.2.2                          10.0.2.2
 Route Distinguisher: 300:1
      Network(EthTagId/IpAddrLen/OriginalIp) NextHop
 *>   0:32:10.0.3.3                          127.0.0.1
 EVPN-Instance evpna:
 Number of Inclusive Multicast Routes: 3
      Network(EthTagId/IpAddrLen/OriginalIp) NextHop
 *>i  0:32:10.0.1.1                          10.0.1.1
 *>i  0:32:10.0.2.2                          10.0.2.2
 *>   0:32:10.0.3.3                          127.0.0.1
```

PE3 receives Type 3 routes from PE1 and PE2 to forward BUM traffic.

HUAWEI

Verifying the Configuration (2)



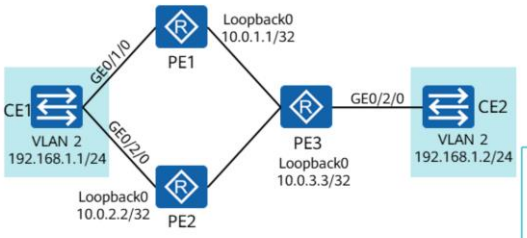
Check Type 4 routes on PE3.

```
<PE3>display bgp evpn  all routing-table es-route
 Local AS number : 100
 EVPN address family:
 Number of ES Routes: 3
 Route Distinguisher: 10.0.1.1:0
      Network(ESI)                          NextHop
 *>i  0000.1111.2222.1111.1111              10.0.1.1
 Route Distinguisher: 10.0.2.2:0
      Network(ESI)                          NextHop
 *>i  0000.1111.2222.1111.1111              10.0.2.2
 Route Distinguisher: 10.0.3.3:0
      Network(ESI)                          NextHop
 *>   014c.1fcc.f95d.e30a.3100              127.0.0.1
 EVPN-Instance evpna:
 Number of ES Routes: 1
      Network(ESI)                          NextHop
 *>   014c.1fcc.f95d.e30a.3100              127.0.0.1
```

Type 4 routes sent by PE1 and PE2 carry the same ESI.

HUAWEI

Verifying the Configuration (3)



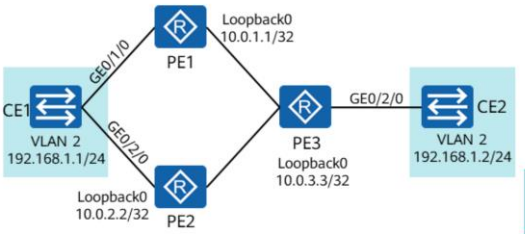
Check Type 2 routes on PE3.

```
<PE3>display bgp evpn  all routing-table mac-route
 Local AS number : 100
 BGP Local router ID is 10.0.3.3
 EVPN address family:
 Number of Mac Routes: 2
 Route Distinguisher: 100:1
  Network(EthTagId/MacAddrLen/MacAddr/IpAddrLen/IpAddr)   NextHop
 *>i   0:48:4c1f-cccf-6675:0:0.0.0.0                      10.0.1.1
 Route Distinguisher: 300:1
  Network(EthTagId/MacAddrLen/MacAddr/IpAddrLen/IpAddr)   NextHop
 *>    0:48:4c1f-ccf9-5de3:0:0.0.0.0                      0.0.0.0

 EVPN-Instance evpna:
 Number of Mac Routes: 2
     Network(EthTagId/MacAddrLen/MacAddr/IpAddrLen/IpAddr)   NextHop
 *>i   0:48:4c1f-cccf-6675:0:0.0.0.0                      10.0.1.1
 *>    0:48:4c1f-ccf9-5de3:0:0.0.0.0                      0.0.0.0
```

PE1 sends the MAC address of CE1 to PE3.

HUAWEI

Verifying the Configuration (4)



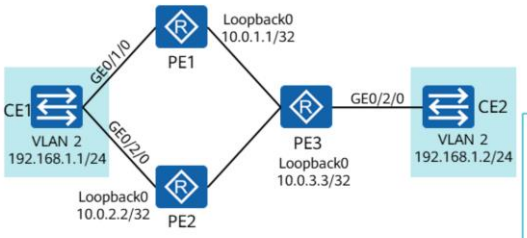
Check Type 1 routes on PE3.

```
<PE3>display bgp evpn  all routing-table ad-route
 Local AS number : 100
 BGP Local router ID is 10.0.3.3
EVPN address family:
 Number of A-D Routes: 5
 Route Distinguisher: 100:1
      Network(ESI/EthTagId)                    NextHop
*>i   0000.1111.2222.1111.1111:0               10.0.1.1
 Route Distinguisher: 200:1
      Network(ESI/EthTagId)                    NextHop
*>i   0000.1111.2222.1111.1111:0               10.0.2.2
 EVPN-Instance evpna:
 Number of A-D Routes: 4
      Network(ESI/EthTagId)                    NextHop
*>i   0000.1111.2222.1111.1111:0               10.0.1.1
  i                                            10.0.2.2
*>    014c.1fcc.f95d.e30a.3100:0               127.0.0.1
```

PE1 and PE2 advertise Type 1 routes to implement split horizon.

HUAWEI

Verifying the Configuration (5)



Check the ARP table on CE2 and ping CE1 from CE2.

```
<CE2>display arp
IP ADDRESS      MAC ADDRESS     INTERFACE
--------------------------------------------------------------------------------
192.168.1.2     4c1f-ccf9-5de3  Vlanif2          #MAC address of CE2
192.168.1.1     4c1f-cccf-6675  Eth-Trunk10      #MAC address of CE1
<CE2>ping 192.168.1.1
  PING 192.168.1.1: 56  data bytes, press CTRL_C to break
    Reply from 192.168.1.1: bytes=56 Sequence=1 ttl=255 time=10 ms
    Reply from 192.168.1.1: bytes=56 Sequence=2 ttl=255 time=50 ms
```

Check the MAC address table of **evpna** on PE3.

```
<PE3>display evpn mac routing-table evpn-instance evpna verbose
MAC Flag: D - download to fib
--------------------------------------------------------------------------------EVPN
name: evpna
MACs: 2        Entries: 3
MAC Address: 4c1f-cccf-6675
  Protocol: BGP                 VLAN/BD: 10
    State: Active               Age: 00h13m01s
    Type: Dynamic               Label: 48002
 IndirectID: 0x1000080          PeerIP: 10.0.1.1
    Flag: D
  TunnelID: 0x0000000001004c4b82     Interface: Ethernet1/0/0
MAC Address: 4c1f-cccf-6675
  Protocol: BGP                 VLAN/BD: 10
    State: Active               Age: 00h13m01s
    Type: Dynamic               Label: 48002
 IndirectID: 0x100007D          PeerIP: 10.0.2.2
    Flag: D
  TunnelID: 0x0000000001004c4b81     Interface: Ethernet1/0/1
```

MAC address of CE1

Label allocated by PE1 to CE1

HUAWEI

Contents

1. EVPN Background and Terms
2. EVPN Fundamentals
3. Inter-AS EVPN
4. Typical EVPN Application Scenarios
5. **Basic EVPN Configurations**
   - Configuring an EVPN to Carry Layer 2 Services
   - Configuring an EVPN to Carry Layer 3 Services

HUAWEI

Configuring an EVPN to Carry Layer 3 Services

1. Configure an L3VPN instance.

```
[Huawei] ip vpn-instance vpna
[Huawei-vpn-instance-vpna] route-distinguisher route-distinguisher
[Huawei-vpn-instance-vpna] vpn-target vpn-target &<1-8> [ both | export-extcommunity | import-extcommunity ] evpn
```

Create an L3VPN instance named **vpna**, enter the L3VPN instance view, and configure an RD and an RT. The keyword **evpn** needs to be appended to the RT.

2. Enable EVPN to generate and advertise IP prefix routes and IRB routes.

```
[Huawei-vpn-instance-vpna] evpn mpls routing-enable
```

In an EVPN L3VPN scenario, when MPLS or SR-MPLS tunnels are used as public network tunnels, enable EVPN to generate and advertise IP prefix routes and IRB routes so that the local device can advertise IP prefix routes and IRB routes to EVPN peers. If non-MPLS tunnels, such as VXLAN and SRv6 tunnels, are used as public network tunnels, you do not need to run this command.

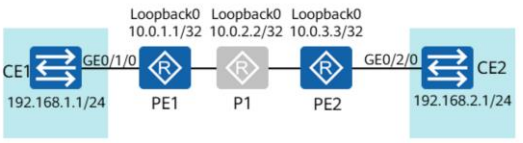
3. Configure IP prefix route advertisement.

```
[Huawei-bgp-vpna] advertise l2vpn evpn
```

By default, locally leaked routes in a VPN instance are not sent to an EVPN instance. To support mutual access between different VPN instances, enable the function of advertising locally leaked routes from a VPN instance to an EVPN instance so that these routes can be sent to the remote device through a BGP EVPN peer relationship.

HUAWEI

Example for Configuring an EVPN to Carry Layer 3 Services: Requirement Description

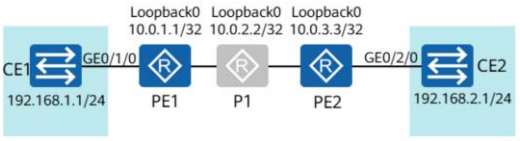


- Configure EVPN on the network to implement Layer 3 communication.
  - The gateway address of CE1 is 192.168.1.254/24.
  - The gateway address of CE2 is 192.168.2.254/24.
  - The L3VPN instance name is **vpna**.
  - On PE1 and PE2, the RDs of the L3VPN instance **vpna** are 100:1 and 200:1 respectively and the RTs are 1:1.

- Configuration roadmap
  - Configure an IGP on the backbone network to allow the PEs to communicate. (The configuration details are not provided here.)
  - Configure basic MPLS functions and enable MPLS LDP to establish MPLS LSPs on the backbone network. (The configuration details are not provided here.)
  - Configure an L3VPN instance on PEs.
  - Establish BGP EVPN peer relationships.
  - Configure CEs to access PEs and advertise routes.
  - Verify the configuration.

HUAWEI

Example for Configuring an EVPN to Carry Layer 3 Services (1)



- Configure an IGP on the backbone network to allow the PEs to communicate. (The configuration details are not provided here.)
- Configure basic MPLS functions and enable MPLS LDP to establish MPLS LSPs on the backbone network.
- **Configure an L3VPN instance on PEs.**
- **Establish BGP EVPN peer relationships.**
- Configure CEs to access PEs and advertise routes.
- Verify the configuration.

The following uses PE1 as an example to configure an L3VPN instance.

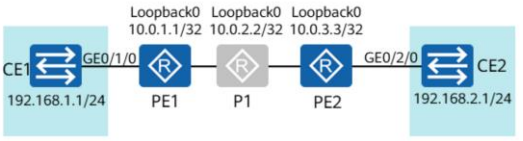
```
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] route-distinguisher 100:1
[*PE1-vpn-instance-vpna-af-ipv4] vpn-target 1:1 both evpn
[*PE1-vpn-instance-vpna-af-ipv4] evpn mpls routing-enable
[*PE1-vpn-instance-vpna-af-ipv4] quit
[*PE1-vpn-instance-vpna] quit
[*PE1] commit
```

The following uses PE1 as an example to establish BGP EVPN peer relationships.

```
[~PE1] bgp 100
[*PE1-bgp] peer 10.0.3.3 as-number 100
[*PE1-bgp] peer 10.0.3.3 connect-interface loopback 0
[*PE1-bgp] l2vpn-family evpn
[*PE1-bgp-af-evpn] peer 10.0.3.3 enable
[*PE1-bgp-af-evpn] quit
[*PE1-bgp] quit
[*PE1] commit
```


HUAWEI

Example for Configuring an EVPN to Carry Layer 3 Services (2)



- Configure an IGP on the backbone network to allow the PEs to communicate. (The configuration details are not provided here.)
- Configure basic MPLS functions and enable MPLS LDP to establish MPLS LSPs on the backbone network.
- Configure an L3VPN instance on PEs.
- Establish BGP EVPN peer relationships.
- **Configure CEs to access PEs and advertise routes.**
- Verify the configuration.

The following uses PE1 as an example to configure CEs to access PEs.

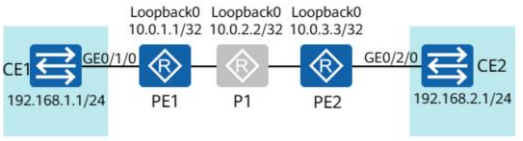
```
[~PE1] interface GigabitEthernet 0/1/0
[*PE1-GigabitEthernet0/1/0] ip binding vpn-instance vpna
[*PE1-GigabitEthernet0/1/0] ip address 192.168.1.254
255.255.255.0
[*PE1-GigabitEthernet0/1/0] quit
[*PE1] commit
```

The following uses PE1 as an example to configure CE-side route advertisement.

```
[~PE1] bgp 100
[~PE1-bgp] ipv4-family vpn-instance vpna
[~PE1-bgp-vpna] import-route direct
[~PE1-bgp-vpna] advertise l2vpn evpn
[~PE1-bgp-vpna] quit
[~PE1-bgp] quit
[~PE1] commit
```


HUAWEI

Verifying the Configuration (1)



- Configure an IGP on the backbone network to allow the PEs to communicate. (The configuration details are not provided here.)
- Configure basic MPLS functions and enable MPLS LDP to establish MPLS LSPs on the backbone network.
- Configure an L3VPN instance on PEs.
- Establish BGP EVPN peer relationships.
- Configure CEs to access PEs and advertise routes.
- **Verify the configuration.**

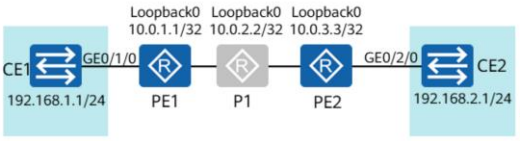
Check Type 5 routes on PE1.

```
<PE1>display bgp evpn all routing-table prefix-route
 Local AS number : 100
 BGP Local router ID is 10.0.1.1
 EVPN address family:
 Number of Ip Prefix Routes: 2
 Route Distinguisher: 100:1
      Network(EthTagId/IpPrefix/IpPrefixLen)     NextHop
 *>    0:192.168.1.0:24                          0.0.0.0
 Route Distinguisher: 200:1
      Network(EthTagId/IpPrefix/IpPrefixLen)     NextHop
 *>i   0:192.168.2.0:24                          10.0.3.3
```

This is the Type 5 route advertised by PE3.

HUAWEI

Verifying the Configuration (2)



- Configure an IGP on the backbone network to allow the PEs to communicate. (The configuration details are not provided here.)
- Configure basic MPLS functions and enable MPLS LDP to establish MPLS LSPs on the backbone network.
- Configure an L3VPN instance on PEs.
- Establish BGP EVPN peer relationships.
- Configure CEs to access PEs and advertise routes.
- **Verify the configuration.**

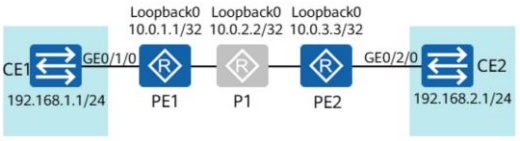
Check detailed information about Type 5 routes on PE1.

```
[~PE1]display bgp evpn all routing-table prefix-route  0:192.168.2.0:24
 BGP local router ID : 10.0.1.1
 Local AS number : 100
 Total routes of Route Distinguisher(200:1): 1
 BGP routing table entry information of 0:192.168.2.0:24:
 Label information (Received/Applied): 48005/NULL
 From: 10.0.3.3 (10.0.13.3)
 Route Duration: 0d01h29m51s
 Relay IP Nexthop: 10.0.13.3
 Relay Tunnel Out-Interface: LDP LSP    #The outer tunnel is an LDP
tunnel.
 Original nexthop: 10.0.3.3
 Qos information : 0x0
 Ext-Community: RT <1 : 1>
 AS-path Nil, origin incomplete,  MED 0, localpref 100, pref-val 0, valid, intern
al, best, select, pre 255, IGP cost 1
 Route Type: 5 (Ip Prefix Route)
 Ethernet Tag ID: 0, IP Prefix/Len: 192.168.2.0/24, ESI:
0000.0000.0000.0000.000
0, GW IP Address: 0.0.0.0
 Not advertised to any peer yet
```

This label is allocated by PE3.

HUAWEI

Verifying the Configuration (3)



- Configure an IGP on the backbone network to allow the PEs to communicate. (The configuration details are not provided here.)
- Configure basic MPLS functions and enable MPLS LDP to establish LDP LSPs on the backbone network.
- Configure an L3VPN instance on PEs.
- Establish BGP EVPN peer relationships.
- Configure CEs to access PEs and advertise routes.
- **Verify the configuration.**

Check the routing table of **vpna** on PE1.

```
<PE1>display ip routing-table vpn-instance vpna
------------------------------------------------------------
Routing Table : vpna
Destination/Mask    Proto   NextHop         Interface
192.168.1.0/24      Direct  192.168.1.254   GigabitEthernet0/1/0
192.168.2.0/24      IBGP    10.0.3.3        GigabitEthernet0/2/0
```

Ping 192.168.2.1 from CE1.

```
<CE1>ping 192.168.2.1
  PING 192.168.2.1: 56  data bytes, press CTRL_C to break
    Reply from 192.168.2.1: bytes=56 Sequence=1 ttl=126 time=30 ms
    Reply from 192.168.2.1: bytes=56 Sequence=2 ttl=126 time=10 ms
    Reply from 192.168.2.1: bytes=56 Sequence=3 ttl=126 time=50 ms
    Reply from 192.168.2.1: bytes=56 Sequence=4 ttl=126 time=40 ms
    Reply from 192.168.2.1: bytes=56 Sequence=5 ttl=126 time=50 ms
  --- 192.168.2.1 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 10/36/50 ms
```


HUAWEI

Quiz

1. (Single-answer question) Which of the following route types is used by EVPN to implement fast convergence? ( )
   A. Type 1
   B. Type 2
   C. Type 3
   D. Type 4
2. (Single-answer question) Which of the following types of routes carries the MPLS label used for forwarding unicast data? ( )
   A. Type 1
   B. Type 2
   C. Type 3
   D. Type 4

HUAWEI

1. A
2. B

Summary

- EVPN uses BGP extensions to implement MAC address learning and advertisement on the control plane instead of the data plane. EVPN allows a device to manage MAC addresses in the same way as it manages routes, implementing load balancing between EVPN routes with the same destination MAC address but different next hops. In addition, EVPN supports the deployment of BGP RRs, significantly reducing network complexity.
- EVPN uses the labels carried in Type 3 routes to guide BUM traffic forwarding, uses the labels carried in Type 2 routes to guide unicast traffic forwarding, uses Type 1 routes to implement split horizon and fast convergence, and uses Type 4 routes to implement DF election and automatic ES member discovery. EVPN also supports the advertisement of IP prefix routes through Type 5 routes.
- With the enrichment of application scenarios and extension of protocols, EVPN can be used in various scenarios, including WAN, DC, and campus network scenarios.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

IPv6 Routing

HUAWEI

Foreword

- With the advent of the Internet of Everything (IoE), IPv4 address space is running out. It is imperative that we replace IPv4 with IPv6. How do we enable reachability between nodes on an IPv6 network? Similar to an IPv4 network, an IPv6 network also supports static routes and dynamic routing protocols.
- IPv6 static routes and IPv4 static routes are configured in a similar way. To support IPv6 networks, the Internet Engineering Task Force (IETF) defines Open Shortest Path First version 3 (OSPFv3) and extends Intermediate System to Intermediate System (IS-IS) and Border Gateway Protocol (BGP).
- This course describes the concepts and configurations of IPv6 static routes and common IPv6 dynamic routing protocols, including OSPFv3, IS-IS (IPv6), and BGP4+.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Configure IPv6 static routes.
  - Analyze the differences between OSPFv3 and OSPFv2.
  - Configure basic OSPFv3 functions.
  - Describe the IS-IS extensions for IPv6.
  - Configure basic IS-IS (IPv6) functions.
  - Describe the BGP extensions for IPv6.
  - Configure basic BGP4+ functions.

HUAWEI

Contents

1. **IPv6 Static Routes**
2. OSPFv3 Implementation and Configurations
3. IS-IS (IPv6) Implementation and Configurations
4. BGP4+ Implementation and Configurations

HUAWEI

IPv6 Static Routes

- Similar to IPv4 static routes, IPv6 static routes are configured by the administrator and are applicable to IPv6 networks with simple structures.
- When creating an IPv6 static route, you can specify both the outbound interface and next hop, or specify only the outbound interface or next hop.
  - Point-to-point (P2P) interface: specify the outbound interface.
  - Broadcast interface: specify the next hop.
- Load balancing and backup:
  - Setting the same preference value for IPv6 static routes to the same destination implements load balancing among these routes, whereas setting different preference values for such routes implements route backup.

HUAWEI

- IPv6 static routes and IPv4 static routes differ mainly in destination and next-hop IP addresses. IPv6 static routes use IPv6 addresses, whereas IPv4 static routes use IPv4 addresses.

Basic Commands for Configuring IPv6 Static Routes

1. Configure an IPv6 static route on the public network:

   [Huawei] **ipv6 route-static** *dest-ipv6-address prefix-length* { *interface-type interface-number* [*nexthop-ipv6-address*] | *nexthop-ipv6-address* | **vpn-instance** *vpn-destination-name nexthop-ipv6-address* } [**preference** *preference*][**permanent** | **inherit-cost**] [**description** *text*]

2. Configure an IPv6 static route in a VPN instance:

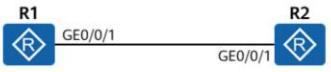
   [Huawei] **ipv6 route-static vpn-instance** *vpn-instance-name dest-ipv6-address prefix-length* { [*interface-type interface-number* [*nexthop-ipv6-address*]] | *nexthop-ipv6-address* [**public**] | **vpn-instance** *vpn-destination-name nexthop-ipv6-address* } [**preference** *preference*] [**permanent** | **inherit-cost**] [**description** *text*]

- Precautions:
  - If no preference is set for a static route, the static route uses the default preference 60.
  - If both the destination address and prefix length are set to all zeros (:: 0), the default route (::/0) is configured.
  - If permanent is specified in a command of the undo form, permanent advertisement is disabled for the IPv6 static route, but this route is not deleted.

HUAWEI

- [Huawei] **ipv6 route-static** *dest-ipv6-address prefix-length* { *interface-type interface-number* [*nexthop-ipv6-address*] | *nexthop-ipv6-address* | **vpn-instance** *vpn-destination-name nexthop-ipv6-address* } [**preference** *preference*][**permanent** | **inherit-cost**] [**description** *text*]
  - **preference** *preference*: specifies a preference value for the route. The value is an integer ranging from 1 to 255. The default value is 60.
  - **permanent**: enables the function of permanently advertising the IPv6 static route.
  - **inherit-cost**: enables the static route to inherit the cost of the recursive route.
  - **description** *text*: specifies a description for the static route. The value is a string of 1 to 80 characters and can contain spaces.
- [Huawei] **ipv6 route-static vpn-instance** *vpn-instance-name dest-ipv6-address prefix-length* { [*interface-type interface-number* [*nexthop-ipv6-address*]] | *nexthop-ipv6-address* [**public**] | **vpn-instance** *vpn-destination-name nexthop-ipv6-address* } [**preference** *preference*] [**permanent** | **inherit-cost**] [**description** *text*]
  - **public**: indicates that *nexthop-ipv6-address* is a public network address instead of an address in the source VPN instance.

Example for Configuring IPv6 Static Routes



Device	Interface	IPv6 Address
R1	GE0/0/1	2001:DB8:2345:12::1/64
	Loopback0	2001:DB8:2345:1::1/128
R2	GE0/0/1	2001:DB8:2345:12::2/64
	Loopback0	2001:DB8:2345:2::2/128

Scenario description:

- A company has an IPv6 network deployed for service testing. In the initial deployment phase, static routes are used to implement interconnection and communication over the IPv6 network.
- Specifically, static routes are used to enable communication between loopback0 interfaces of R1 and R2.

1. Assign an IPv6 address to each router interface. (The configuration details are not provided here.)

2. Configure IPv6 static routes.

```
[R1] ipv6 route-static 2001:DB8:2345:2::2 128 2001:DB8:2345:12::2
```

```
[R2] ipv6 route-static 2001:DB8:2345:1::1 128 2001:DB8:2345:12::1
```

3. Test the IPv6 network connectivity.

```
[R1] ping ipv6 2001:DB8:2345:2::2
  PING 2001:DB8:2345:2::2 : 56  data bytes, press CTRL_C to break
    Reply from 2001:DB8:2345:2::2
    bytes=56 Sequence=1  hop limit=64  time = 90 ms
    Reply from 2001:DB8:2345:2::2
    bytes=56 Sequence=2  hop limit=64  time = 20 ms
    Reply from 2001:DB8:2345:2::2
    bytes=56 Sequence=3  hop limit=64  time = 20 ms
    Reply from 2001:DB8:2345:2::2
    bytes=56 Sequence=4  hop limit=64  time = 20 ms
    Reply from 2001:DB8:2345:2::2
    bytes=56 Sequence=5  hop limit=64  time = 20 ms
  --- 2001:DB8:2345:2::2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 20/34/90 ms
```


HUAWEI

Contents

1. IPv6 Static Routes
2. **OSPFv3 Implementation and Configurations**
   - OSPFv3 Implementation
   - Basic OSPFv3 Configurations
3. IS-IS (IPv6) Implementation and Configurations
4. BGP4+ Implementation and Configurations

HUAWEI

OSPFv3 Overview

- OSPF is a link-state Interior Gateway Protocol (IGP) defined by the IETF. Currently, OSPF version 2 (OSPFv2) is intended for IPv4, and OSPF version 3 (OSPFv3) is intended for IPv6.
- OSPFv3 is mainly developed as a routing protocol independent of any specific network layer. The internal routing information of OSPFv3 has been redesigned to serve this purpose.

OSPFv2 is an IGP running over IPv4, whereas OSPFv3 is an IGP running over IPv6. They are not compatible.

HUAWEI

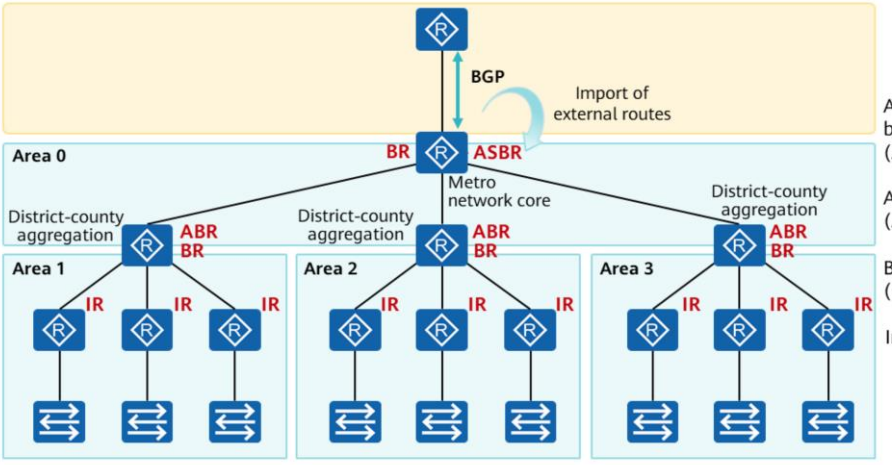
Similarities Between OSPFv3 and OSPFv2

- OSPF's basic running mechanisms remain unchanged, including:
  - Basic concepts:
    - Area division and router types
    - Parameters that affect route calculation: preference and cost
    - Supported network types: broadcast, non-broadcast multiple access (NBMA), P2P, and point-to-multipoint (P2MP)
    - Packet types: Hello, database description (DD), link state request (LSR), link state update (LSU), and link state acknowledgment (LSAck)
  - Working mechanisms:
    - Establishment of neighbor relationships and transition of the neighbor relationship status
    - Election of the Designated router (DR) and backup DR (BDR)
    - Link state advertisement (LSA) flooding
    - Route calculation process

HUAWEI

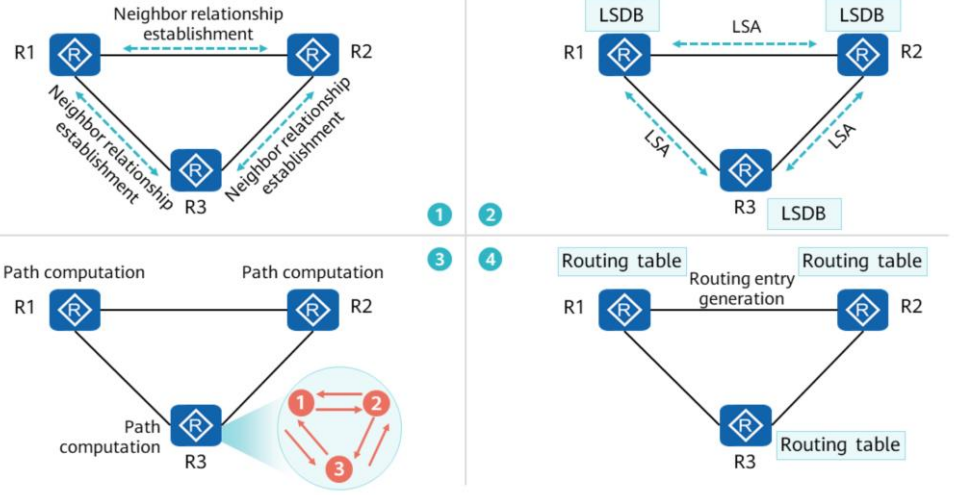
- Similarities also include support for special areas, support for virtual links, and multi-process support.
- For details, see the "HCIP-Datacom-Core Technology" course.

OSPFv3 Topology and Router Types



HUAWEI

OSPFv3 and OSPFv2 Have Similar Working Mechanisms



HUAWEI

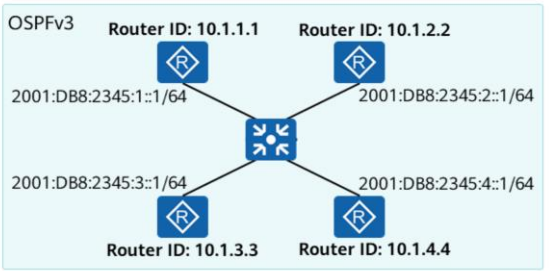
Differences Between OSPFv3 and OSPFv2

- OSPFv3 runs and performs topology calculation based on links instead of network segments.
- OSPFv3 supports multiple instances per link.
- OSPFv3 packets and LSAs do not contain IP addresses, and the packet and LSA formats are reconstructed.
  - OSPFv3 packets and router-LSAs/network-LSAs do not contain IP addresses.
  - OSPFv3 LSAs define the flooding scopes of LSAs.
  - OSPFv3 has new types of LSAs defined to carry IPv6 addresses and prefixes.
  - OSPFv3 neighbors are identified only by router IDs, not by IP addresses.

HUAWEI

Unique Neighbor ID: Router ID

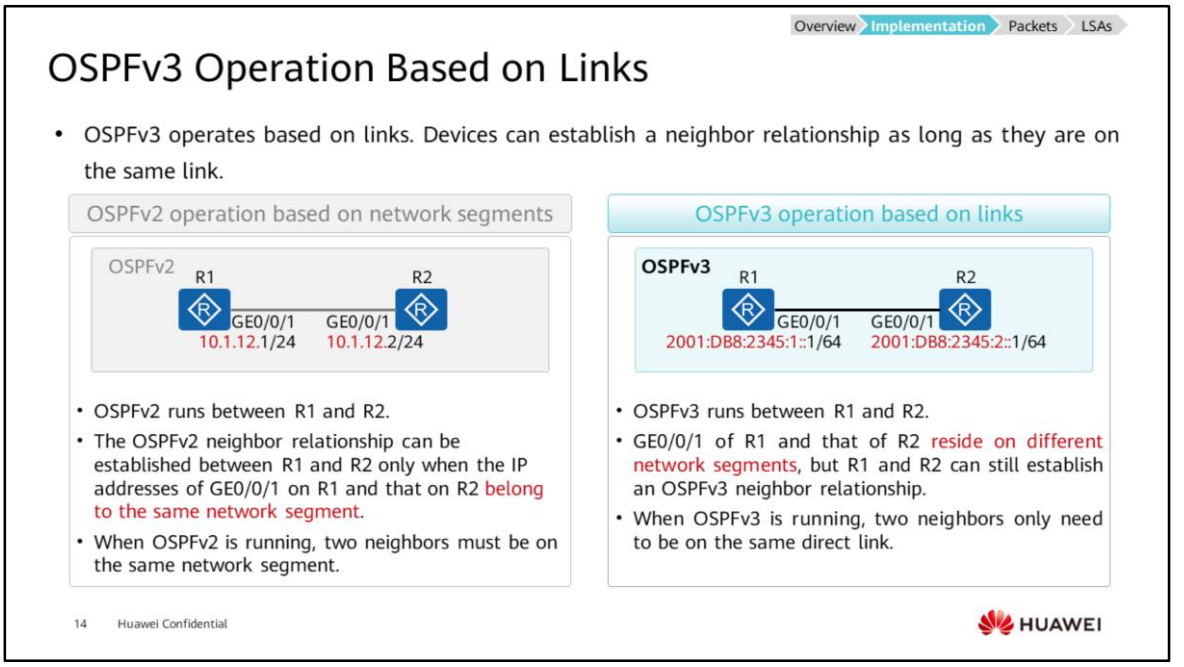
- OSPFv3 uses router IDs to identify network devices.



- A router ID uniquely identifies an OSPFv3 device in an AS. If no router ID is configured, an OSPFv3 process cannot run.
- When you configure a router ID for a device, ensure that the router ID in an AS is unique.
- A router ID is a 32-bit local identifier, which is not related to IPv6 addresses and is expressed in dotted decimal notation.

HUAWEI

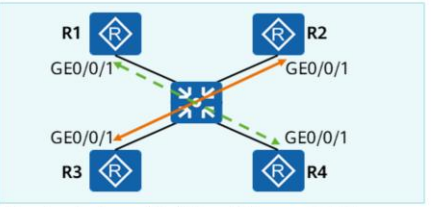
- On broadcast, NBMA, P2P, and P2MP networks, OSPFv2 uses IPv4 interface addresses to identify neighbors. On virtual-link networks, however, OSPFv2 uses router IDs to identify neighbors.



- IPv6 emphasizes the link concept. Multiple IPv6 prefixes that indicate different IP subnets can be allocated to the same link. Different from IPv4, IPv6 allows two nodes on the same link to communicate even if they do not have the same IPv6 prefix. This greatly changes the OSPF behavior.
- In OSPFv3, the concepts "link" and "prefix" are frequently used, which however are independent of each other. The terms "network" and "subnet" used in OSPFv2 should be replaced with the term "link" when OSPFv3 is discussed.

Multi-Instance Support by Links

- An OSPFv3 physical interface can be bound to multiple instances, which are identified by different instance IDs. A single physical link can run multiple OSPFv3 instances.
- These OSPFv3 instances can each establish a neighbor relationship with the device at the remote end of the link, send packets to the remote device, and do not interfere with each other, thereby fully sharing resources on the same link.

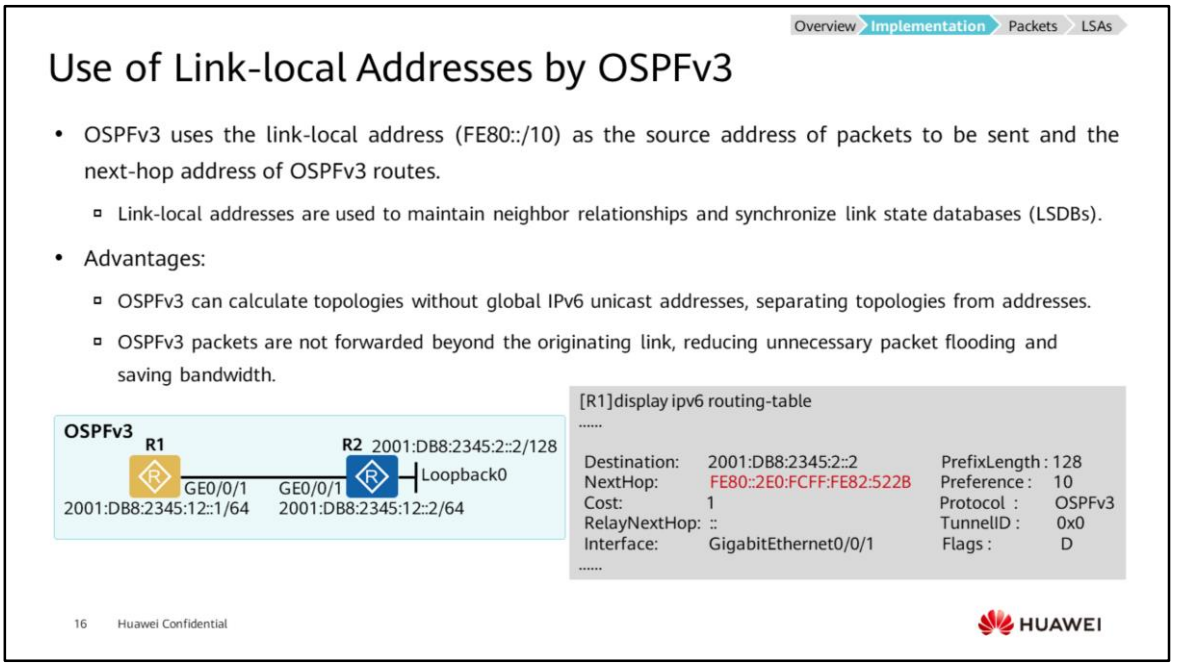


By using instance IDs, R2 and R3 can establish an OSPFv3 neighbor relationship, and R1 and R4 can also establish an OSPFv3 neighbor relationship.

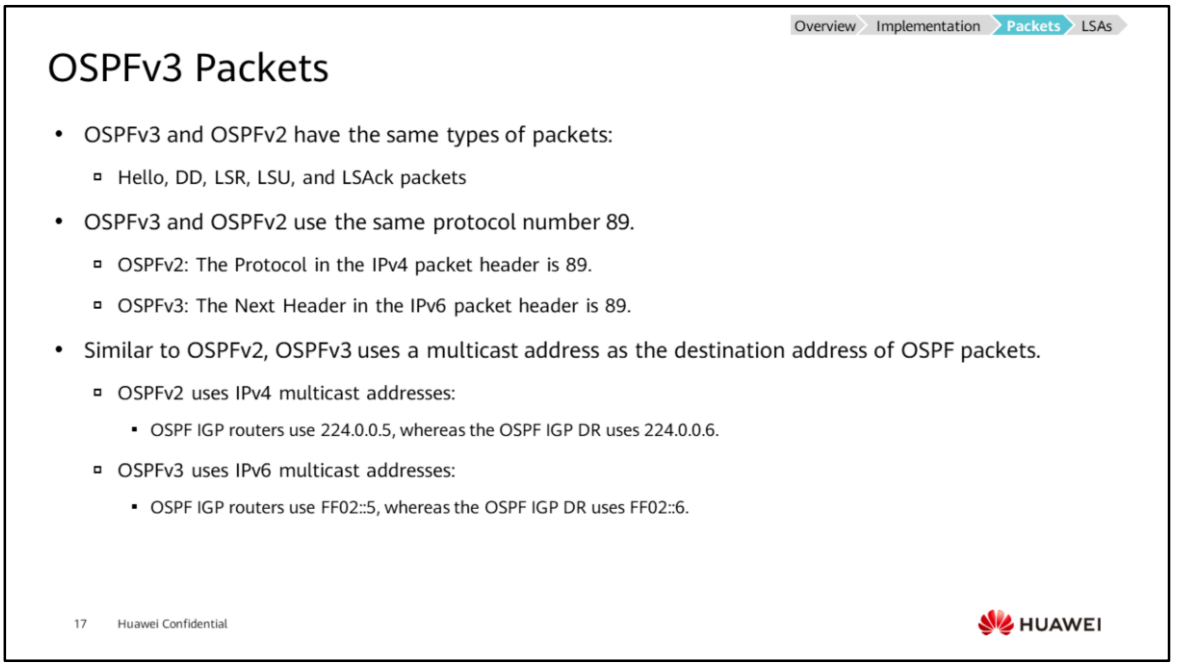
Instance 1
Instance 2

HUAWEI

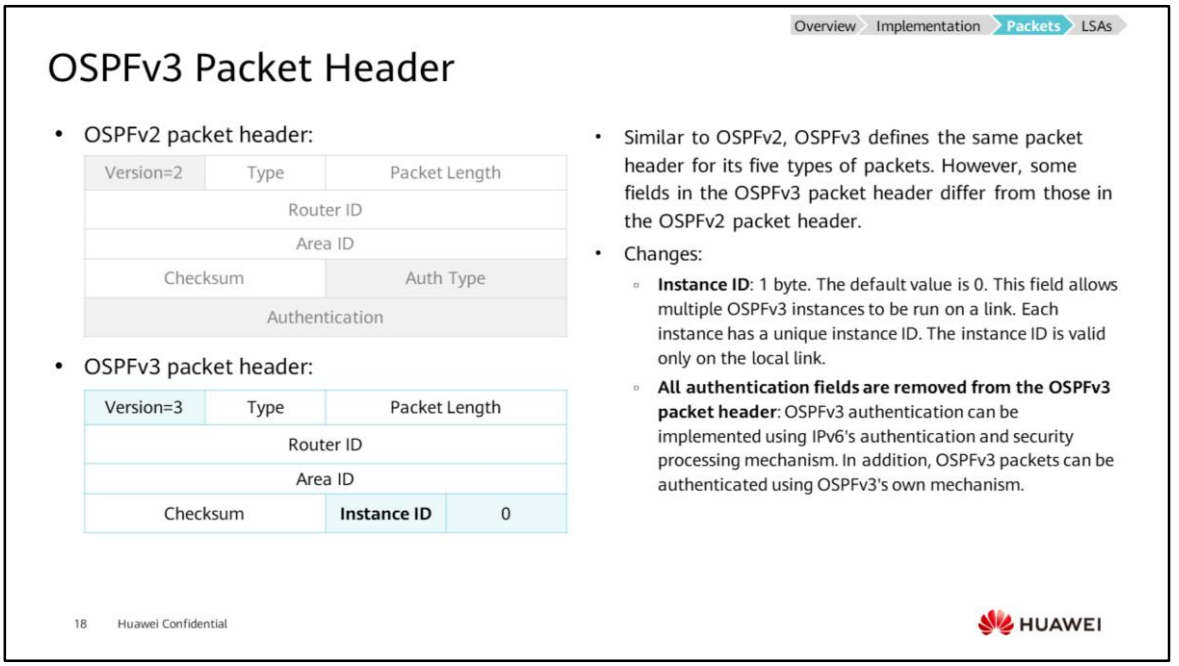
- In multi-instance, each instance is differentiated by adding a specific instance ID to the OSPFv3 packet header. If an instance is assigned a specific instance ID, the OSPFv3 packets that do not match the instance ID are discarded.



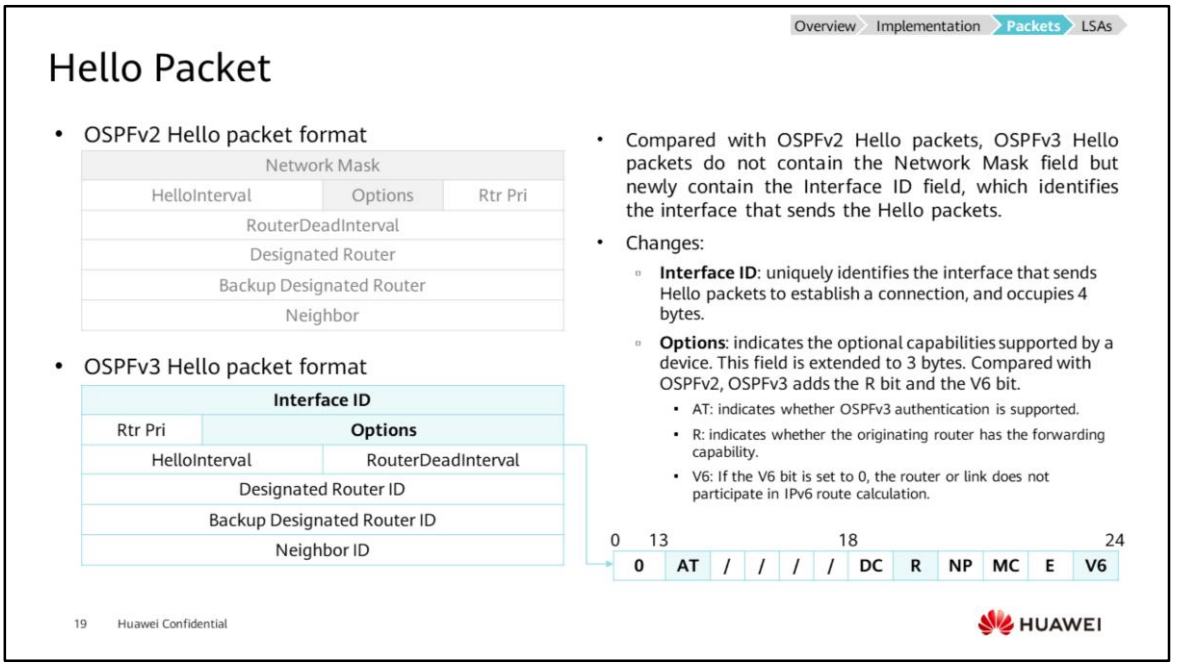
- IPv6 implements neighbor discovery and automatic configuration using link-local addresses. Routers running IPv6 do not forward IPv6 packets whose destination addresses are link-local addresses. Such packets are valid only on the local link.
- OSPFv3 is a routing protocol running on IPv6 and uses link-local addresses to send OSPFv3 packets.
    - OSPFv3 assumes that each router has been assigned a link-local address on each link. All OSPFv3 interfaces except virtual-link interfaces use the associated link-local addresses as the source addresses to send OSPFv3 packets.
    - A router learns the link-local addresses of all the other routers attached to the same link and uses these addresses as the next-hop addresses to forward packets.
    - Note: Description of link-local addresses is only contained in link-LSAs (new type of LSA supported in OSPFv3).
- Note: On a virtual link, the global unicast address or a site's local address must be used as the source address of OSPFv3 packets.



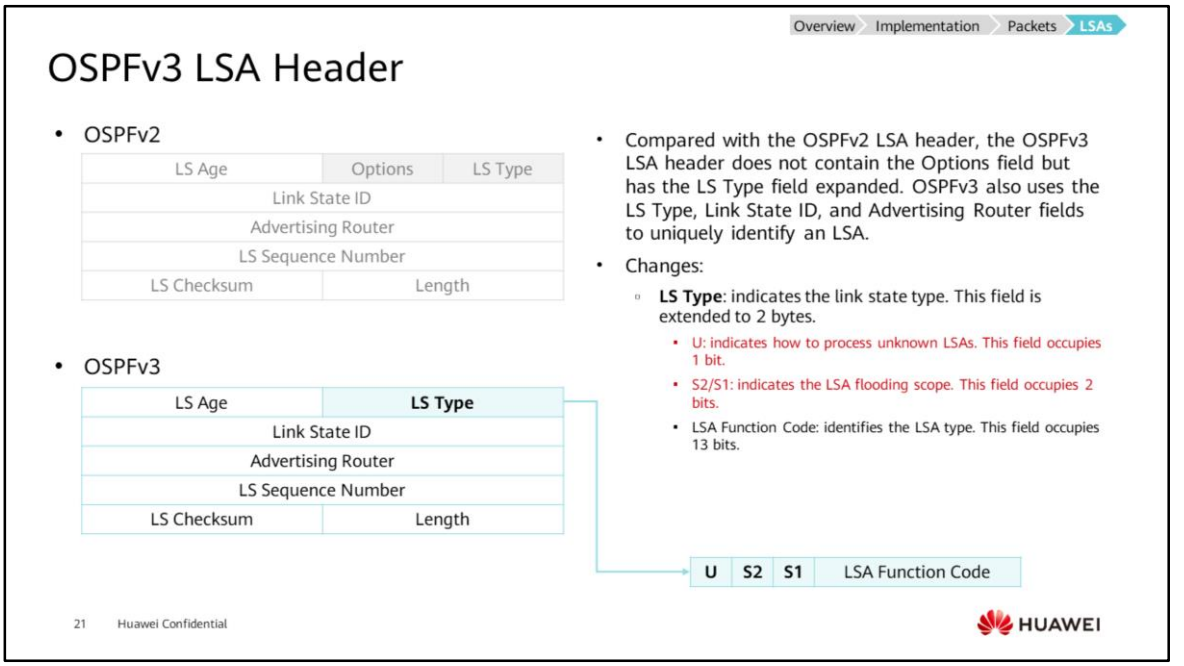
- OSPFv3 packets have the following functions:
  - Hello packet: Hello packets are sent periodically to discover, establish, and maintain OSPFv3 neighbor relationships.
  - DD packet: A DD packet describes the summary of a local LSDB and is used for LSDB synchronization between two devices.
  - LSR packet: An LSR packet is used to request the required LSAs from a neighbor. An OSPFv3 device sends LSR packets to its neighbor only after DD packets have been successfully exchanged between them.
  - LSU packet: An LSU packet is sent to a neighbor to provide required LSAs.
  - LSAck packet: An LSAck packet is used to acknowledge the received LSAs.



- Version: indicates the OSPF version, and occupies 1 byte. For OSPFv3, the value is 3.
- Type: indicates the type of an OSPFv3 packet and occupies 1 byte. The following types are available:
  - 1: Hello packet
  - 2: DD packet
  - 3: LSR packet
  - 4: LSU packet
  - 5: LSAck packet
- Packet length: indicates the total length of an OSPFv3 packet, including the packet header. The field occupies 2 bytes.
- Router ID: indicates the router ID of the router that originates the packet, and occupies 4 bytes.
- Area ID: indicates the area in which the packet is sent, and occupies 4 bytes.
- Checksum: indicates the standard 16-bit IPv6 checksum and occupies 2 bytes.
- 0: Occupying 1 byte, this field is reserved and must be set to 0.



- Rot Pri: indicates the router's router priority, which is used for DR election. This field occupies 1 byte, and the default value is 1. If it is set to 0, the router cannot participate in DR or BDR election.
- Options: indicates the optional capabilities supported by the router and occupies 3 bytes.
    - AT: indicates whether OSPFv3 authentication is supported. This option occupies 1 bit. If the AT bit is 1, an authentication tail field containing authentication information is added to the OSPFv3 packet.
    - DC: indicates whether the capability of processing demand circuits is supported. This option occupies 1 bit.
    - R: indicates whether the originator is a valid router. This option occupies 1 bit.
    - NP: indicates whether the area to which the originating router interface belongs is a not-so-stubby area (NSSA). This option occupies 1 bit.
    - MC: indicates whether multicast data packets can be forwarded. This option occupies 1 bit.
    - E: indicates whether external routes are supported. This option occupies 1 bit.
    - V6: indicates whether the router or link can participate in route calculation. This option occupies 1 bit. If it is set to 0, the router or link does not participate in IPv6 route calculation.



- LS Age: indicates the time elapsed since the LSA was generated, in seconds. This field occupies 2 bytes. The value of this field continually increases regardless of whether the LSA is transmitted over a link or saved in an LSDB.
- LS Type: indicates the LSA type. This field occupies 2 bytes. The high-order three bits of this field identify generic properties of the LSA, whereas the remaining bits identify the LSA's specific function.
    - The U-bit indicates how to process an unknown LSA, that is, how a router that does not recognize an LSA's function code should process this LSA.
        - 0: The LSA is treated as if it had the link-local flooding scope.
        - 1: The LSA is stored and flooded as if its type had been understood.
    - The S2 and S1 bits indicate the flooding scope of the LSA.
        - S2 S1 = 0 0: link-local flooding scope. The LSA is flooded only on the originating link.
        - S2 S1 = 0 1: area flooding scope. The LSA is flooded to all routers in the originating area.
        - S2 S1 = 1 0: AS flooding scope. The LSA is flooded to all routers in the local AS.
        - S2 S1 = 1 1: reserved.

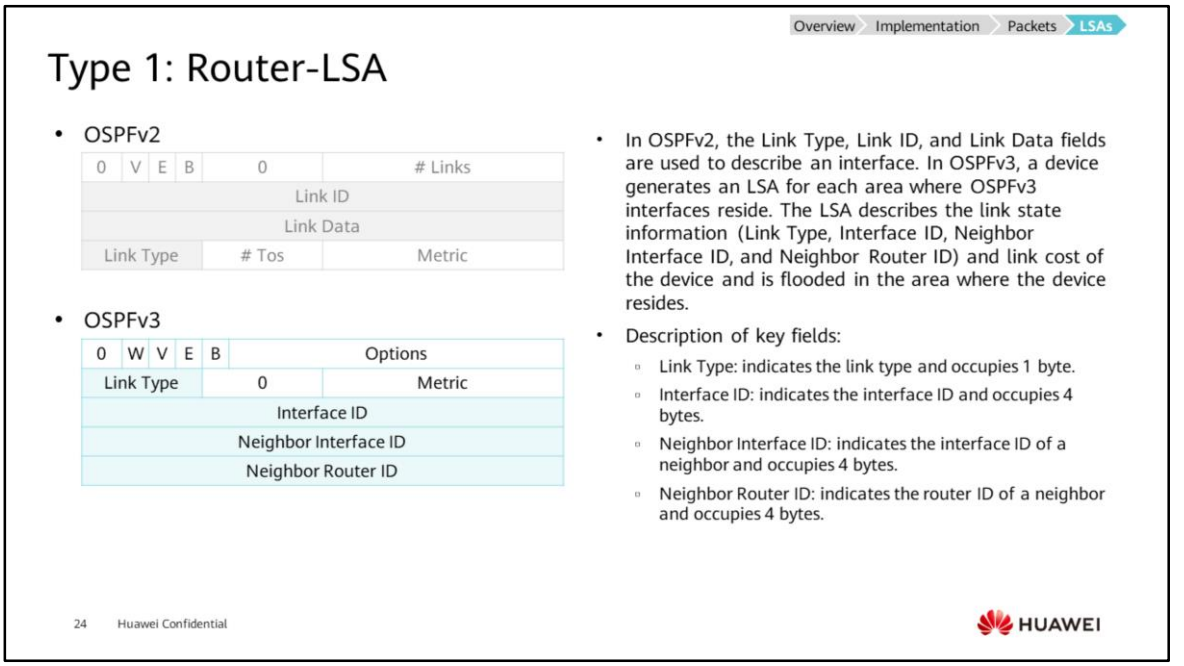
OSPFv3 LSA Types

- Compared with OSPFv2, OSPFv3 LSAs have similar names but slightly different functions.
- OSPFv3 supports two new types of LSAs: link-LSA and intra-area-prefix-LSA.

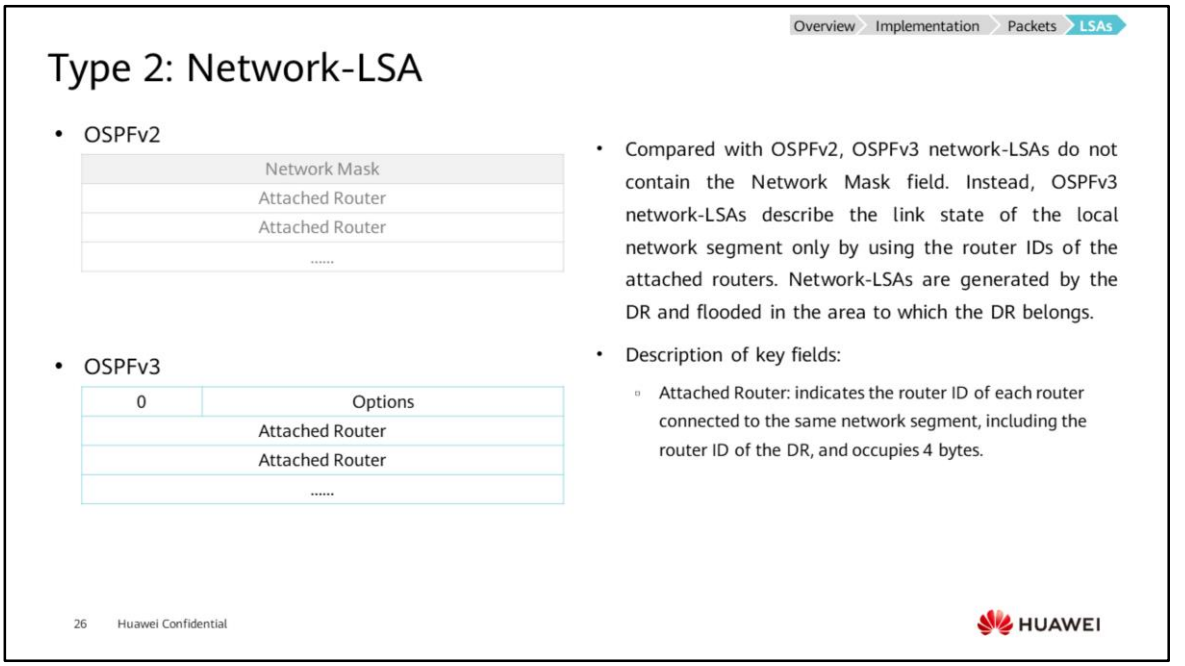
OSPFv2 LSAs		OSPFv3 LSAs		Similarities and Differences
Type	Name	Type	Name	
1	Router-LSA	0x2001	Router-LSA	The functions are similar, but the OSPFv3 LSAs no longer describe address information. Instead, they only describe topology structures.
2	Network-LSA	0x2002	Network-LSA	
3	Network-summary-LSA	0x2003	Inter-area-prefix-LSA	The functions are similar, but the names are different.
4	ASBR-summary-LSA	0x2004	Inter-area-router-LSA	
5	AS-external-LSA	0x4005	AS-external-LSA	The functions and names are the same.
7	NSSA-LSA	0x2007	NSSA-LSA	
		0x0008	Link-LSA	New in OSPFv3
		0x2009	Intra-area-prefix-LSA	

HUAWEI

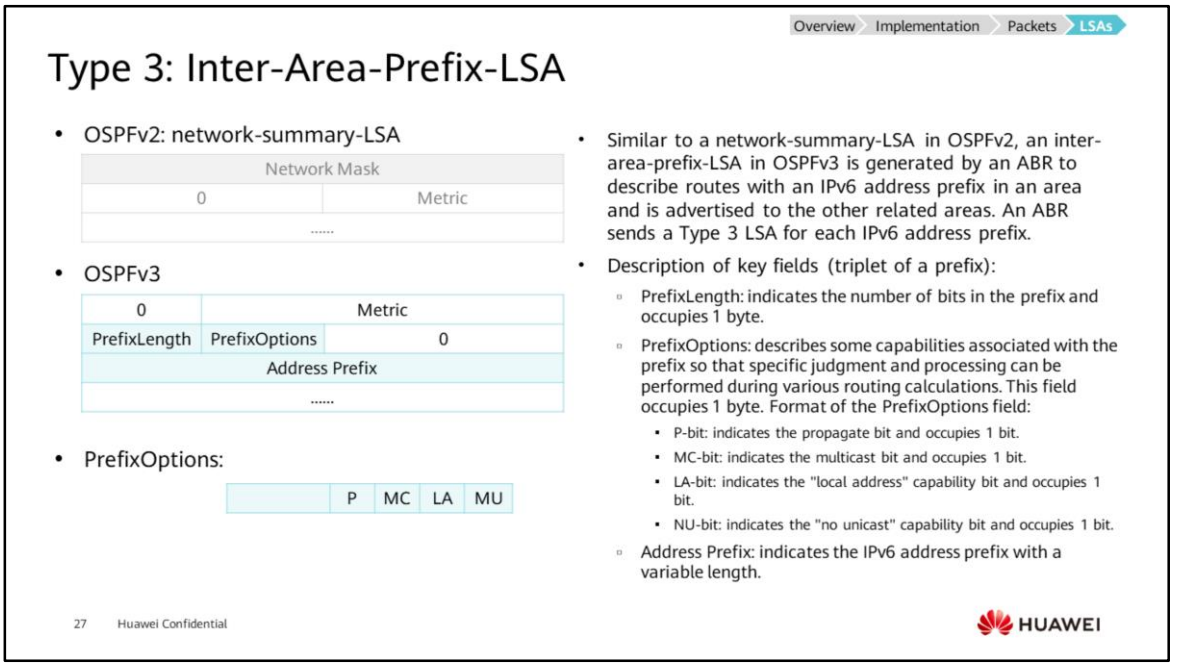
- As shown in the figure, the U-bit in the LS Type field of the OSPFv3 LSA header is 0 by default. Except the Type 5 and Type 8 LSAs, the other types of LSAs all have the area flooding scope (S2 S1 = 0 1).
  - Link-local flooding scope: LSAs, including link-LSAs, are flooded only on the local link.
  - Area flooding scope: The following types of LSAs are flooded in a single OSPF area: router-LSA, network-LSA, inter-area-prefix-LSA, inter-area-router-LSA, NSSA-LSA, and intra-area-prefix-LSA.
  - AS flooding scope: LSAs, including AS-external-LSAs, are flooded in an entire routing domain (AS).



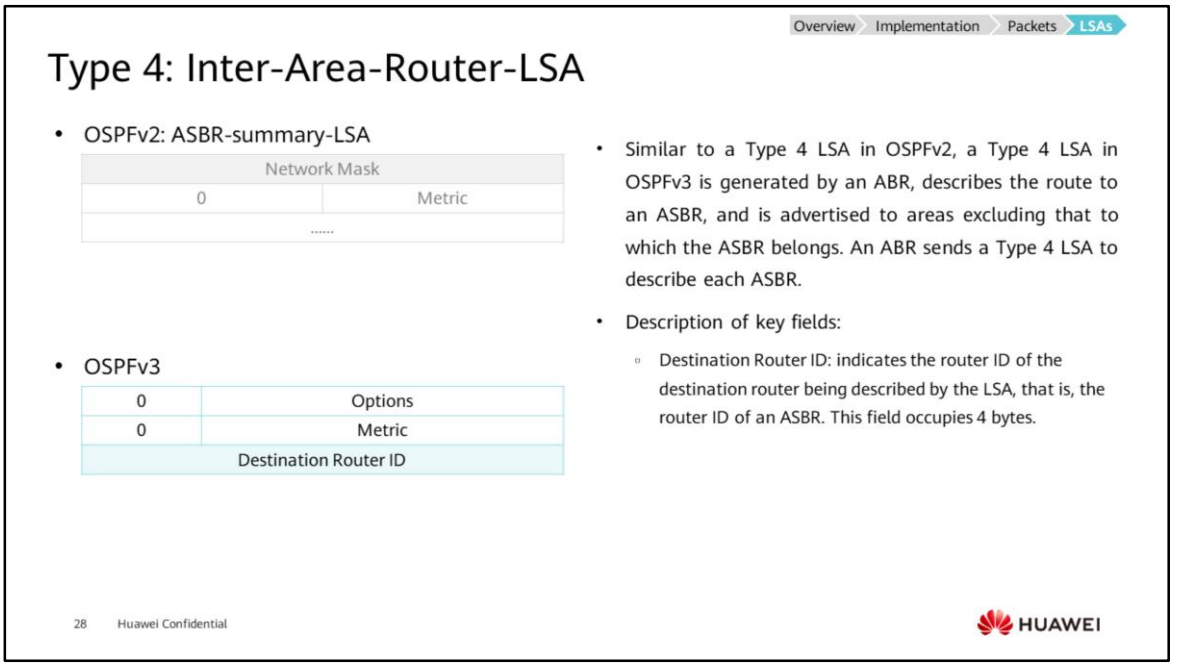
- The fields in an OSPFv3 router-LSA are described as follows:
  - W: wildcard receiver. The value 1 indicates that the router supports multicast routes.
  - V: virtual link. The value 1 indicates that the router that generates the LSA is at one end of the virtual link.
  - E: external. The value 1 indicates that the router that generates the LSA is an ASBR.
  - B: border. The value 1 indicates that the router that generates the LSA is an ABR.
  - Options: indicates the optional capabilities supported by the router and occupies 3 bytes.
    - DC: indicates whether the capability of processing demand circuits is supported. This option occupies 1 bit.
    - R: indicates whether the originator is a valid router. This option occupies 1 bit.
    - NP: indicates whether the area to which the originating router interface belongs is a not-so-stubby area (NSSA). This option occupies 1 bit.



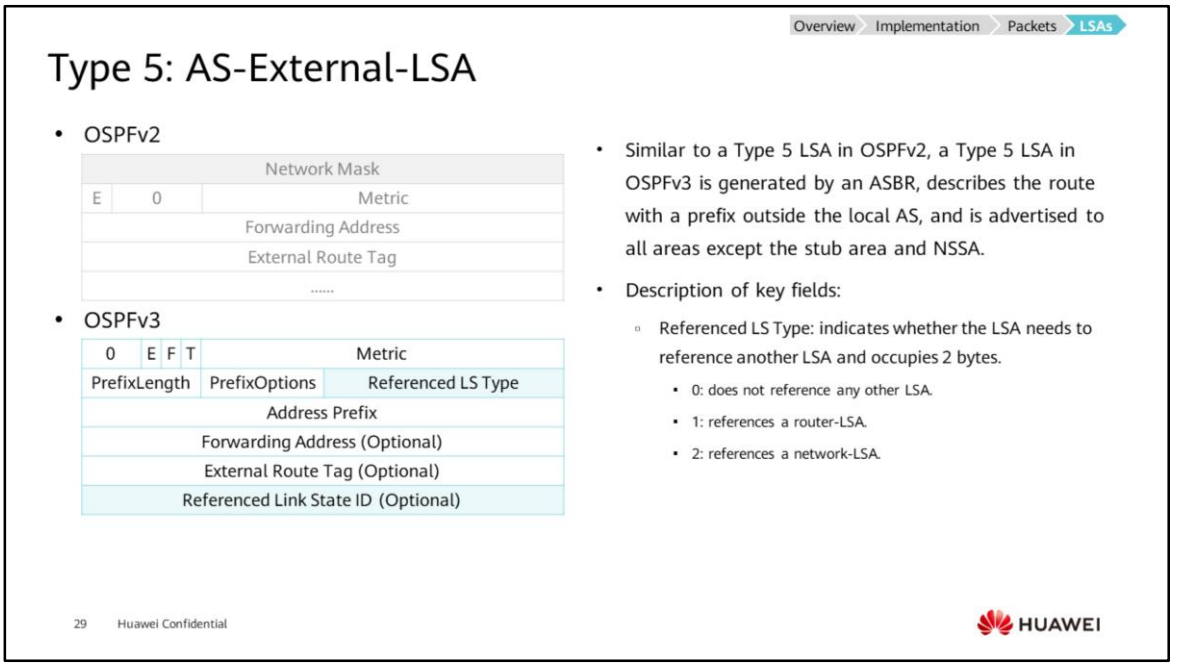
- The fields in an OSPFv3 network-LSA are described as follows:
  - Options: same as the Options field in a router-LSA.



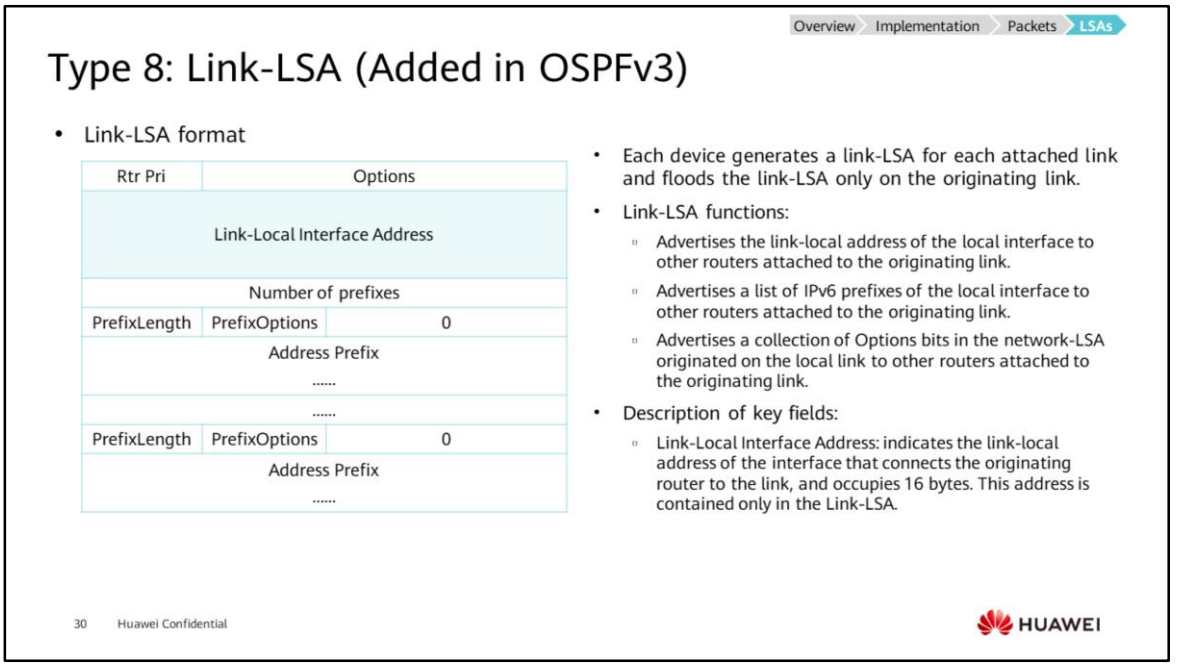
- The fields in an OSPFv3 inter-area-prefix-LSA are described as follows:
  - Metric: indicates the cost of the route to the destination address and occupies 3 bytes.
  - PrefixOptions: Each prefix advertised by an LSA has its own PrefixOptions field.
    - P-bit: propagate bit. This bit needs to be set to 1 if the prefix of an NSSA needs to be advertised by an ABR.
    - MC-bit: multicast bit. If this bit is set to 1, the prefix is used for multicast route calculation. Otherwise, the prefix is not used for multicast route calculation.
    - LA-bit: local address capability bit. If this bit is set to 1, the prefix is an interface address of the router.
    - NU-bit: no unicast capability bit. If this bit is set to 1, the prefix is not used for IPv6 unicast route calculation.
- Note: The prefix length of the default route is 0. An ABR can also originate an inter-area Type 3 LSA to advertise a default route to a stub area.



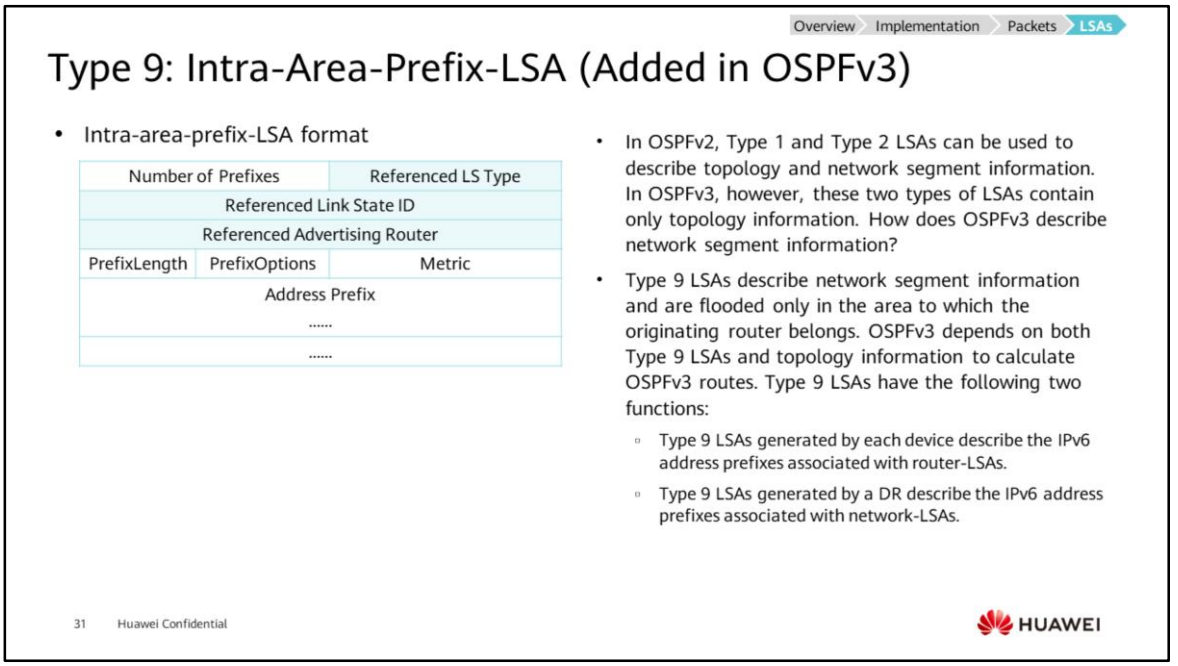
- The fields in an OSPFv3 inter-area-router-LSA are described as follows:
  - Options: This field describes the optional capabilities supported by the destination router instead of those supported by the source router. Therefore, the value of this field should equal that of the Options field in the router-LSA generated by the destination router.
  - Metric: indicates the cost of the route to the destination address and occupies 3 bytes.



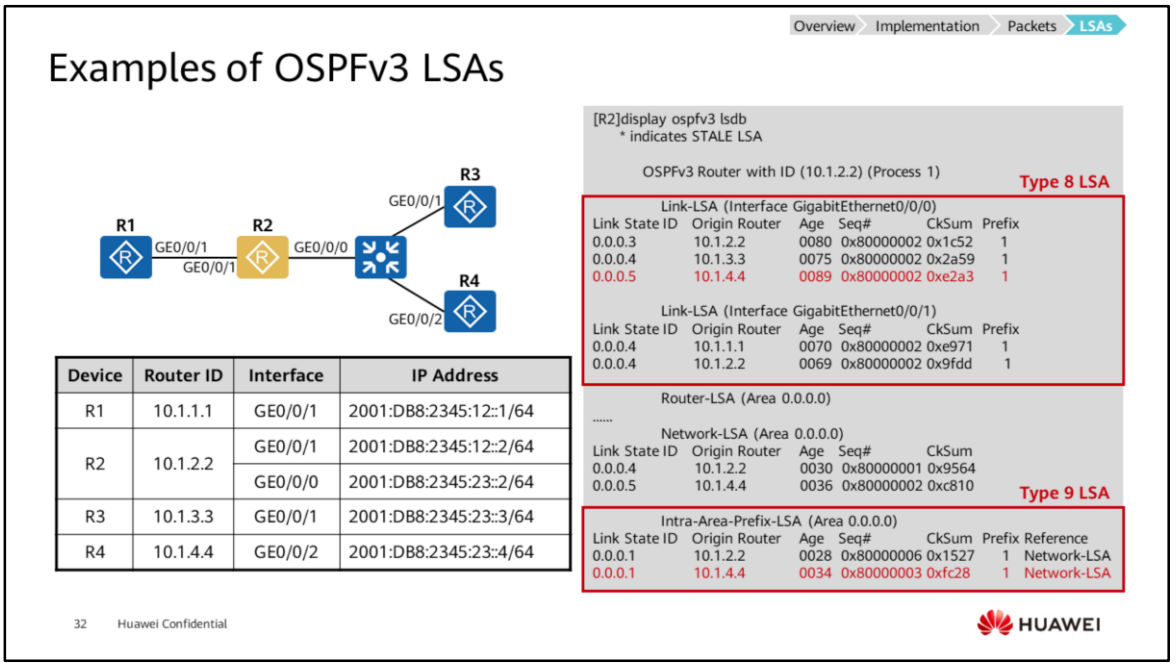
- The fields in an OSPFv3 AS-external-LSA are described as follows:
  - Bit E: indicates the cost type of an AS external route and occupies 1 bit.
    - The value 1 indicates the cost of a Type 2 external route. This cost does not increase during route transmission.
    - The value 0 indicates the cost of a Type 1 external route. This cost increases during route transmission.
  - Bit F: occupies 1 bit. The value 1 indicates that the Forwarding Address field (optional) is included.
  - Bit T: occupies 1 bit. The value 1 indicates that the External Route Tag field (optional) is included.
  - Metric: indicates the cost of the route to the destination address and occupies 3 bytes.
  - PrefixLength, PrefixOptions, and Address Prefix are triplets that describe a prefix and have the same meanings as those in an inter-area-prefix-LSA.
  - Forwarding Address: is an optional 128-bit IPv6 address and occupies 4 bytes. This field is included if bit F is 1. In this case, a data packet needs to be forwarded to this address before reaching its destination.
  - External Route Tag: an optional flag, which occupies 4 bytes. It can be used for communication between ASBRs. In a typical scenario where each of two ASBRs imports an AS external route, the imported routes can be tagged differently to facilitate route filtering.
  - Referenced Link State ID: occupies 4 bytes. This field is included if the Referenced LS Type field is not 0, indicating the link state ID of the referenced LSA.



- The fields in an OSPFv3 link-LSA are described as follows:
  - Rtr Pri: indicates the router priority of the interface attaching the originating router to the link and occupies 1 byte.
  - Options: indicates a collection of Options bits that the router sets in the network-LSA and occupies 3 bytes.
  - Number of Prefixes: indicates the number of IPv6 address prefixes carried in the LSA, and occupies 4 bytes.
  - PrefixLength, PrefixOptions, and Address Prefix are triplets that describe a prefix and have the same meanings as those in an inter-area-prefix-LSA.



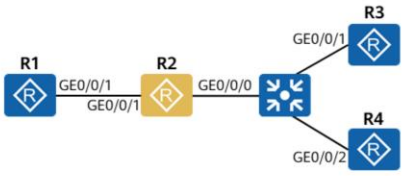
- The fields in an OSPFv3 intra-area-prefix-LSA are described as follows:
  - Number of Prefixes: indicates the number of IPv6 address prefixes carried in the LSA, and occupies 4 bytes. If necessary, prefixes can be carried in multiple intra-area-prefix-LSAs to limit the size of each Type 9 LSA.
  - Referenced LS Type: indicates whether the LSA references a router-LSA or a network-LSA, and occupies 4 bytes.
    - Type=1: A router-LSA is referenced.
    - Type=2: A network-LSA is referenced.
  - Referenced Link State ID: 4 bytes.
    - If the LSA references a router-LSA, this field is set to 0.
    - If the LSA references a network-LSA, this field is set to the interface ID of the DR on the attached link.
  - Referenced Advertising Router: 4 bytes.
    - If the LSA references a router-LSA, this field is set to the router ID of the associated router.
    - If the LSA references a network-LSA, this field is set to the router ID of the DR on the attached link.
  - PrefixLength, PrefixOptions, and Address Prefix are triplets that describe a prefix and have the same meanings as those in an inter-area-prefix-LSA.
  - Metric: indicates the prefix cost and occupies 2 bytes. This field has the same unit as the interface cost of a router-LSA.
- In OSPFv3, when a link or its prefix changes, the attached router sends an intra-area-prefix-LSA, which however does not trigger SPF calculation.



- As shown in the figure, R1, R2, R3, and R4 run OSPFv3 and are all deployed in the backbone area.
- After the network becomes stable, check the LSDB of R2. The command output shows information about the following types of LSAs: router-LSA (Type 1), network-LSA (Type 2), Link-LSA (Type 8), and intra-area-prefix-LSA (Type 9).

Overview > Implementation > Packets > LSAs

Link-LSA Example



Device	Router ID	Interface	IP Address
R1	10.1.1.1	GE0/0/1	2001:DB8:2345:12::1/64
R2	10.1.2.2	GE0/0/1	2001:DB8:2345:12::2/64
		GE0/0/0	2001:DB8:2345:23::2/64
R3	10.1.3.3	GE0/0/1	2001:DB8:2345:23::3/64
R4	10.1.4.4	GE0/0/2	2001:DB8:2345:23::4/64

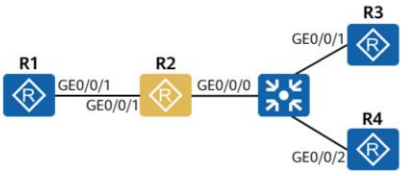
- Check the link-LSA with the Link State ID of 0.0.0.5 on R2. The command output is as follows:

```
[R2]display ospfv3 lsdb link 0.0.0.5

        OSPFv3 Router with ID (10.1.2.2) (Process 1)

        Link-LSA (Interface GigabitEthernet0/0/0)

 LS Age: 446
 LS Type: Link-LSA
 Link State ID: 0.0.0.5
 Originating Router: 10.1.4.4
 LS Seq Number: 0x80000002
 Retransmit Count: 0
 Checksum: 0xE2A3
 Length: 56                                          Type 8 LSA
 Priority: 1
 Options: 0x000013 (-|R|-|-|E|V6)
 Link-Local Address: FE80::2E0:FCFF:FEAC:6C5   Link-local address
 Number of Prefixes: 1

 Prefix: 2001:DB8:2345:23::/64   IPv6 address prefix of the network segment
  Prefix Options: 0 (-|-|-|-|-)      where the physical interface resides
```


HUAWEI

- The command output is described as follows:
  - LS Age: aging time of the LSA.
  - LS Type: LSA type, which can be any of the following:
    - Router-LSA, Network-LSA, Inter-Area-Prefix-LSA, Inter-Area-Router-LSA, AS-external-LSA, NSSA-LSA, Link-LSA, or Intra-Area-Prefix-LSA
  - Link State ID: link state ID in the LSA header.
  - Originating Router: router that generates the LSA.
  - LS Seq Number: sequence number of the LSA. This field is carried in the LSA header.
  - Checksum: checksum of the LSA.
  - Length: length of the LSA.
  - Priority: priority of the interface attached to the link.
  - Options: optional capabilities of the link.
  - Link-Local Address: link-local address.
  - Number of Prefixes: number of IPv6 prefixes contained in the LSA.
  - Prefix: IPv6 prefix.
  - Prefix Options: optional capabilities associated with the prefix.

Intra-Area-Prefix-LSA Example



Device	Router ID	Interface	IP Address
R1	10.1.1.1	GE0/0/1	2001:DB8:2345:12::1/64
R2	10.1.2.2	GE0/0/1	2001:DB8:2345:12::2/64
		GE0/0/0	2001:DB8:2345:23::2/64
R3	10.1.3.3	GE0/0/1	2001:DB8:2345:23::3/64
R4	10.1.4.4	GE0/0/2	2001:DB8:2345:23::4/64

- Check the intra-area-prefix-LSA on R2. The command output is as follows:

```
[R2]display ospfv3 lsdb intra-prefix

        OSPFv3 Router with ID (10.1.2.2) (Process 1)

            Intra-Area-Prefix-LSA (Area 0.0.0.0)
......

  LS Age: 556
  LS Type: Intra-Area-Prefix-LSA
  Link State ID: 0.0.0.1
  Originating Router: 10.1.4.4
  LS Seq Number: 0x80000003
  Retransmit Count: 0
  Checksum: 0xFC28
  Length: 44
  Number of Prefixes: 1
  Referenced LS Type: 0x2002
  Referenced Link State ID: 0.0.0.5
  Referenced Originating Router: 10.1.4.4

  Prefix: 2001:DB8:2345:23::/64
    Prefix Options: 0 (-|-|-|-|-)
    Metric: 0
```

Type 9 LSA

Reference the network-LSA generated by the DR (R4)

IPv6 address prefix of the transit network segment to which the local router is directly connected. This LSA is generated by the DR.

HUAWEI

Contents

1. IPv6 Static Routes
2. **OSPFv3 Implementation and Configurations**
   - OSPFv3 Implementation
   - Basic OSPFv3 Configurations
3. IS-IS (IPv6) Implementation and Configurations
4. BGP4+ Implementation and Configurations

HUAWEI

Basic OSPFv3 Configuration Commands (1)

1. Start OSPFv3.

   [Huawei] **ospfv3** [*process-id*] [**vpn-instance** *vpn-instance-name*]

   An OSPFv3 process is created and runs, and (optional) the process is bound to a VPN instance.

   [Huawei-ospfv3-1] **router-id** *router-id*

   A router ID is configured for the local device in this OSPFv3 process.
   Note: If no router ID is configured, the OSPFv3 process cannot run.

2. Enable OSPFv3 on an interface.

   [Huawei-GigabitEthernet0/0/1] **ospfv3** *process-id* **area** *area-id* [**instance** *instance-id*]

   The interface is enabled in the specified OSPFv3 process, and the area to which the interface belongs is specified. You can also specify the ID of the instance to which the interface belongs.
   Note: Before running this command, you must create an OSPFv3 process and enable the IPv6 function.

HUAWEI

- The configuration commands and methods of OSPFv3 are similar to those of OSPFv2. For details, see the "HCIP-Datacom-Core Technology" course.

Basic OSPFv3 Configuration Commands (2)

3. (Optional) Configure an OSPFv3 network type for the interface.

   [Huawei-GigabitEthernet0/0/1] **ospfv3 network-type** { **broadcast** | **nbma** | **p2mp** [**non-broadcast**] | **p2p** } [**instance** *instance-id*]

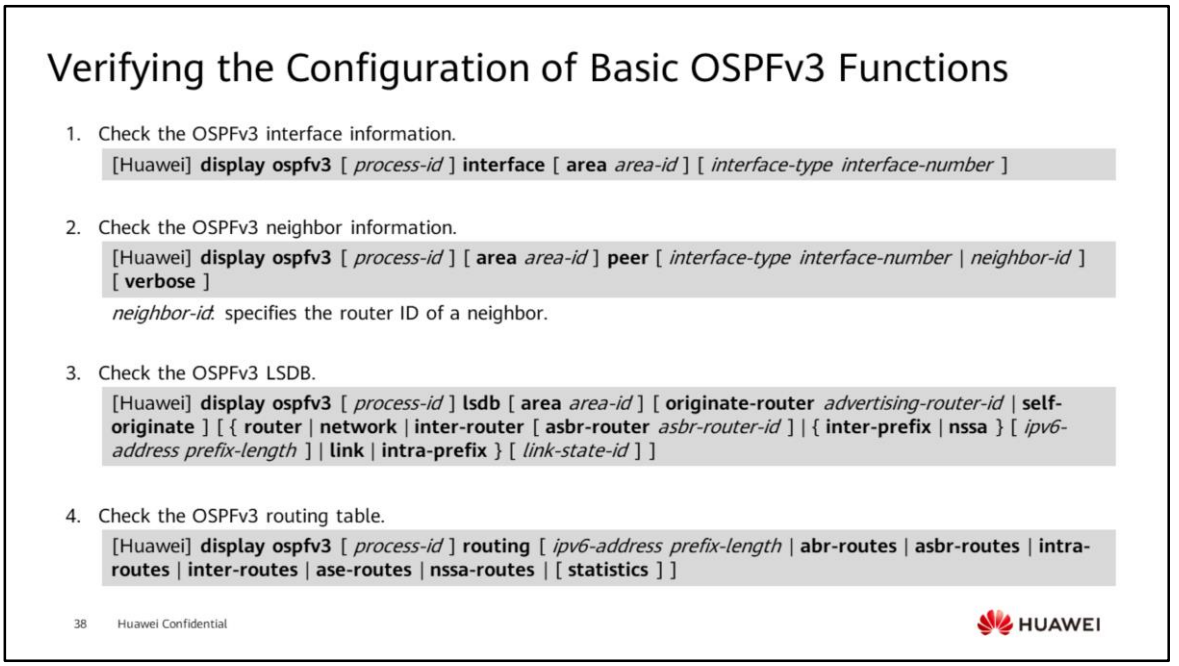
   By default, the OSPFv3 network type of a physical interface depends on the data link layer encapsulation of the interface. The default network type of an Ethernet interface is broadcast, whereas that of a serial interface (encapsulated with PPP or HDLC) is P2P.

4. Enter the OSPFv3 area view.

   [Huawei-ospfv3-1] **area** *area-id*

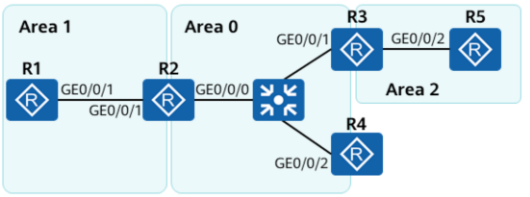
   The specified area ID can be a decimal integer or in the IPv4 address format. Regardless of the specified format, the area ID is displayed as an IPv4 address.

HUAWEI



- [Huawei] **display ospfv3** [*process-id*] **lsdb** [**area** *area-id*] [**originate-router** *advertising-router-id* | **self-originate**] [{ **router** | **network** | **inter-router** [**asbr-router** *asbr-router-id*] | { **inter-prefix** | **nssa** } [*ipv6-address prefix-length*] | **link** | **intra-prefix** | **grace** } [*link-state-id*]]
  - *process-id:* specifies the ID of an OSPFv3 process. The value is an integer ranging from 1 to 65535.
  - **area** *area-id:* specifies the ID of an area. The area ID can be a decimal integer or in the IPv4 address format. For a decimal integer, the value ranges from 0 to 4294967295. For the IPv4 address format, the value is in dotted decimal notation.
  - **external:** displays information about AS-external LSAs in the LSDB.
  - **inter-prefix:** displays information about inter-area-prefix LSAs in the LSDB.
  - **inter-router:** displays information about inter-area-router LSAs in the LSDB.
  - **intra-prefix:** displays information about intra-area-prefix LSAs in the LSDB.
  - **nssa:** displays information about NSSA LSAs in the LSDB.
  - **link:** displays information about link-LSAs in the LSDB.
  - **network:** displays information about network-LSAs in the LSDB.
  - **router:** displays information about router-LSAs in the LSDB.
  - *link-state-id:* specifies a link state ID. The value is in dotted decimal notation.

Example for Configuring OSPF IPv4/IPv6 Dual Stack

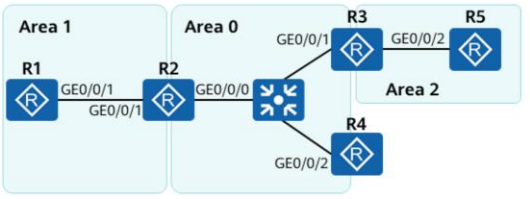


Device	Router ID	Interface	IPv4 Address	IPv6 Address
R1	10.1.1.1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
R2	10.1.2.2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
		GE0/0/0	10.1.23.2/24	2001:DB8:2345:23::2/64
R3	10.1.3.3	GE0/0/1	10.1.23.3/24	2001:DB8:2345:23::3/64
		GE0/0/2	10.1.35.3/24	2001:DB8:2345:35::3/64
R4	10.1.4.4	GE0/0/2	10.1.23.4/24	2001:DB8:2345:23::4/64
R5	10.1.5.5	GE0/0/2	10.1.35.5/24	2001:DB8:2345:35::5/64

- Scenario description:
  - A company has OSPFv2 deployed over an IPv4 network to implement interconnection and communication. To ensure future service development, the company has an IPv6 network deployed for service testing. OSPFv3 runs on the IPv6 network to implement IPv6 interconnection and communication.
  - Each router runs both OSPFv2 and OSPFv3, and the entire AS is divided into three areas. After the configuration is complete, each router should learn the IPv4 and IPv6 routes to all network segments in the AS.

HUAWEI

Deploying the IPv4 Network (1)



Device	Router ID	Interface	IPv4 Address	IPv6 Address
R1	10.1.1.1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
R2	10.1.2.2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
		GE0/0/0	10.1.23.2/24	2001:DB8:2345:23::2/64
R3	10.1.3.3	GE0/0/1	10.1.23.3/24	2001:DB8:2345:23::3/64
		GE0/0/2	10.1.35.3/24	2001:DB8:2345:35::3/64
R4	10.1.4.4	GE0/0/2	10.1.23.4/24	2001:DB8:2345:23::4/64
R5	10.1.5.5	GE0/0/2	10.1.35.5/24	2001:DB8:2345:35::5/64

1. Assign an IPv4 address to each router interface. (The configuration details are not provided here.)

2. Configure basic OSPF functions.

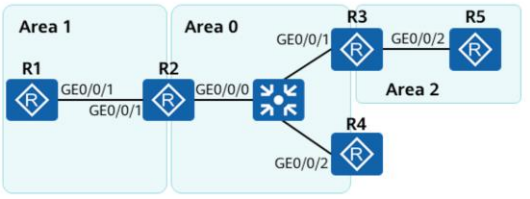
```
[R1] ospf 1 router-id 10.1.1.1
[R1-ospf-1] area 1
[R1-ospf-1-area-0.0.0.1] network 10.1.12.1 0.0.0.0
```

```
[R2] ospf 1 router-id 10.1.2.2
[R2-ospf-1] area 1
[R2-ospf-1-area-0.0.0.1] network 10.1.12.2 0.0.0.0
[R2-ospf-1-area-0.0.0.1] quit
[R2-ospf-1] area 0
[R2-ospf-1-area-0.0.0.0] network 10.1.23.2 0.0.0.0
```

```
[R3] ospf 1 router-id 10.1.3.3
[R3-ospf-1] area 0
[R3-ospf-1-area-0.0.0.0] network 10.1.23.3 0.0.0.0
[R3-ospf-1-area-0.0.0.0] quit
[R3-ospf-1] area 2
[R3-ospf-1-area-0.0.0.2] network 10.1.35.3 0.0.0.0
```


HUAWEI

Deploying the IPv4 Network (2)



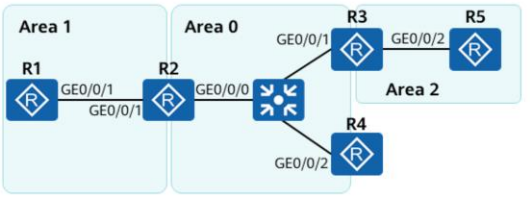
Device	Router ID	Interface	IPv4 Address	IPv6 Address
R1	10.1.1.1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
R2	10.1.2.2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
		GE0/0/0	10.1.23.2/24	2001:DB8:2345:23::2/64
R3	10.1.3.3	GE0/0/1	10.1.23.3/24	2001:DB8:2345:23::3/64
		GE0/0/2	10.1.35.3/24	2001:DB8:2345:35::3/64
R4	10.1.4.4	GE0/0/2	10.1.23.4/24	2001:DB8:2345:23::4/64
R5	10.1.5.5	GE0/0/2	10.1.35.5/24	2001:DB8:2345:35::5/64

```
[R4] ospf 1 router-id 10.1.4.4
[R4-ospf-1] area 0
[R4-ospf-1-area-0.0.0.0] network 10.1.23.4 0.0.0.0
```

```
[R5] ospf 1 router-id 10.1.5.5
[R5-ospf-1] area 2
[R5-ospf-1-area-0.0.0.2] network 10.1.35.5 0.0.0.0
```


HUAWEI

Deploying the IPv6 Network (1)



Device	Router ID	Interface	IPv4 Address	IPv6 Address
R1	10.1.1.1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
R2	10.1.2.2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
		GE0/0/0	10.1.23.2/24	2001:DB8:2345:23::2/64
R3	10.1.3.3	GE0/0/1	10.1.23.3/24	2001:DB8:2345:23::3/64
		GE0/0/2	10.1.35.3/24	2001:DB8:2345:35::3/64
R4	10.1.4.4	GE0/0/2	10.1.23.4/24	2001:DB8:2345:23::4/64
R5	10.1.5.5	GE0/0/2	10.1.35.5/24	2001:DB8:2345:35::5/64

1. Enable IPv6 globally and on interfaces, and assign IPv6 addresses to interfaces on each router. (The configuration details are not provided here.)

2. Enable OSPFv3.

```
[R1] ospfv3 1
[R1-ospfv3-1] router-id 10.1.1.1
```

```
[R2] ospfv3 1
[R2-ospfv3-1] router-id 10.1.2.2
```

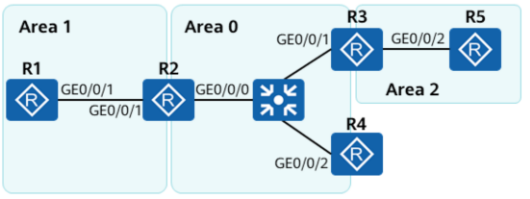
```
[R3] ospfv3 1
[R3-ospfv3-1] router-id 10.1.3.3
```

```
[R4] ospfv3 1
[R4-ospfv3-1] router-id 10.1.4.4
```

```
[R5] ospfv3 1
[R5-ospfv3-1] router-id 10.1.5.5
```


HUAWEI

Deploying the IPv6 Network (2)



Device	Router ID	Interface	IPv4 Address	IPv6 Address
R1	10.1.1.1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
R2	10.1.2.2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
		GE0/0/0	10.1.23.2/24	2001:DB8:2345:23::2/64
R3	10.1.3.3	GE0/0/1	10.1.23.3/24	2001:DB8:2345:23::3/64
		GE0/0/2	10.1.35.3/24	2001:DB8:2345:35::3/64
R4	10.1.4.4	GE0/0/2	10.1.23.4/24	2001:DB8:2345:23::4/64
R5	10.1.5.5	GE0/0/2	10.1.35.5/24	2001:DB8:2345:35::5/64

3. Enable OSPFv3 on interfaces.

```
[R1] interface gigabitethernet 0/0/1
[R1-GigabitEthernet0/0/1] ospfv3 1 area 1
```

```
[R2] interface gigabitethernet 0/0/1
[R2-GigabitEthernet0/0/1] ospfv3 1 area 1
[R2-GigabitEthernet0/0/1] quit
[R2] interface gigabitethernet 0/0/0
[R2-GigabitEthernet0/0/0] ospfv3 1 area 0
```

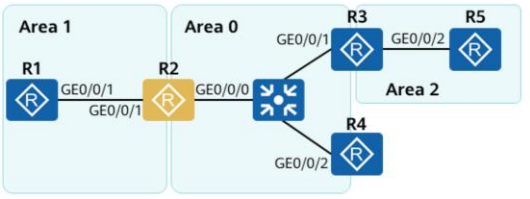
```
[R3] interface gigabitethernet 0/0/1
[R3-GigabitEthernet0/0/1] ospfv3 1 area 0
[R3-GigabitEthernet0/0/1] quit
[R3] interface gigabitethernet 0/0/2
[R3-GigabitEthernet0/0/2] ospfv3 1 area 2
```

```
[R4] interface gigabitethernet 0/0/2
[R4-GigabitEthernet0/0/2] ospfv3 1 area 0
```

```
[R5] interface gigabitethernet 0/0/2
[R5-GigabitEthernet0/0/2] ospfv3 1 area 2
```


HUAWEI

Checking Neighbor Information on the OSPFv3 Network



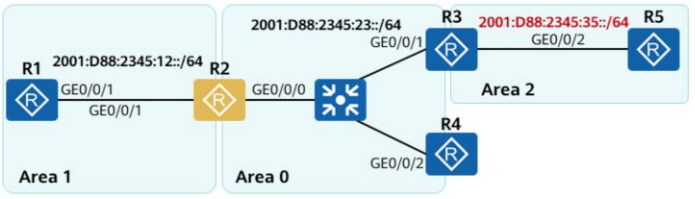

```
[R2]display ospfv3 peer
OSPFv3 Process (1)
OSPFv3 Area (0.0.0.0)
Neighbor ID     Pri  State           Dead Time Interface  Instance ID
10.1.3.3          1  Full/Backup     00:00:31  GE0/0/0    0
10.1.4.4          1  Full/DR         00:00:36  GE0/0/0    0
OSPFv3 Area (0.0.0.1)
Neighbor ID     Pri  State           Dead Time Interface  Instance ID
10.1.1.1          1  Full/Backup     00:00:37  GE0/0/1    0
```

If no instance ID is specified, the default value 0 is used.
The instance IDs of the directly connected interfaces on the two devices must be the same. Otherwise, the OSPFv3 neighbor relationship cannot be established.

HUAWEI

- You can run the **display ospf peer** command to check OSPFv2 neighbor information.
- By comparing OSPFv2 neighbor information and OSPFv3 neighbor information, you will find that the elected DR and BDR are the same, indicating that the DR election modes between OSPFv2 and OSPFv3 are the same.

Checking the Routing Information on the OSPFv3 Network




```
[R2]display ospfv3 routing

Codes : E2 - Type 2 External, E1 - Type 1 External, IA - Inter-Area, N - NSSA, U - Uninstalled

OSPFv3 Process (1)
     Destination                 Metric    Next-hop
     2001:DB8:2345:12::/64       1         directly connected, GigabitEthernet0/0/1
     2001:DB8:2345:23::/64       1         directly connected, GigabitEthernet0/0/0
IA 2001:DB8:2345:35::/64         2         via FE80::2E0:FCFF:FE38:50A0, GigabitEthernet0/0/0
```

In OSPFv3, the next-hop address of a route is a link-local address. The next-hop address of this route is the link-local address of R3's GE0/0/1.

HUAWEI

- You can run the **display ospf routing** command to check the routing information on the OSPFv2 network.
- By comparing OSPFv2 routing information and OSPFv3 routing information, you will find that the paths to the same network segment are the same, indicating that the route calculation methods between OSPFv2 and OSPFv3 are the same.

Checking the LSDB on the OSPFv3 Network

```
[R2]display ospfv3 lsdb
* indicates STALE LSA
        OSPFv3 Router with ID (10.1.2.2) (Process 1)
            Link-LSA (Interface GigabitEthernet0/0/0)
Link State ID Origin Router    Age  Seq#         CkSum  Prefix
0.0.0.3       10.1.2.2         0862 0x80000003   0x486d  1
0.0.0.4       10.1.3.3         0854 0x80000003   0xc512  1
0.0.0.5       10.1.4.4         0840 0x80000003   0x4e3c  1
            Link-LSA (Interface GigabitEthernet0/0/1)
Link State ID Origin Router    Age  Seq#         CkSum  Prefix
0.0.0.4       10.1.1.1         0878 0x80000003   0x9d8f  1
0.0.0.4       10.1.2.2         0868 0x80000003   0xcbf8  1
            Router-LSA (Area 0.0.0.0)
Link State ID Origin Router    Age  Seq#         CkSum  Link
0.0.0.0       10.1.2.2         0567 0x8000000f   0xd411  1
0.0.0.0       10.1.3.3         0576 0x8000000e   0xd70c  1
0.0.0.0       10.1.4.4         0577 0x8000000a   0xdd08  1
            Network-LSA (Area 0.0.0.0)
Link State ID Origin Router    Age  Seq#         CkSum
0.0.0.5       10.1.4.4         0577 0x80000006   0xc014
            Inter-Area-Prefix-LSA (Area 0.0.0.0)
Link State ID Origin Router    Age  Seq#         CkSum
0.0.0.1       10.1.2.2         0862 0x80000003   0xceb3
0.0.0.1       10.1.3.3         0782 0x80000004   0x3824
            Intra-Area-Prefix-LSA (Area 0.0.0.0)
Link State ID Origin Router    Age  Seq#         CkSum  Prefix Reference
0.0.0.1       10.1.4.4         0576 0x8000000a   0xee2f   1    Network-LSA
```

```
            Router-LSA (Area 0.0.0.1)
Link State ID Origin Router    Age  Seq#         CkSum  Link
0.0.0.0       10.1.1.1         0819 0x80000007 0x9460   1
0.0.0.0       10.1.2.2         0827 0x80000007 0x8a67   1
            Network-LSA (Area 0.0.0.1)
Link State ID Origin Router    Age  Seq#         CkSum
0.0.0.4       10.1.2.2         0828 0x80000003 0x9166
            Inter-Area-Prefix-LSA (Area 0.0.0.1)
Link State ID Origin Router    Age  Seq#         CkSum
0.0.0.1       10.1.2.2         0570 0x80000001 0x510f
0.0.0.2       10.1.2.2         0596 0x80000001 0xfb76
            Intra-Area-Prefix-LSA (Area 0.0.0.1)
Link State ID Origin Router    Age  Seq#         CkSum  Prefix  Reference
0.0.0.1       10.1.2.2         0826 0x80000004 0x1925   1     Network-LSA
```

- The LSDB on R2 contains the Type 1, Type 2, Type 3, Type 8, and Type 9 LSAs.

HUAWEI

- You can run the **display ospf lsdb** command to check the OSPFv2 LSDB. You will find that the LSDB contains the Type 1, Type 2, and Type 3 LSAs.

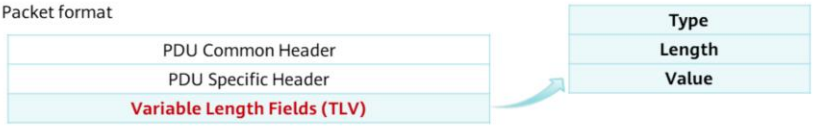
Contents

1. IPv6 Static Routes
2. OSPFv3 Implementation and Configurations
3. **IS-IS (IPv6) Implementation and Configurations**
4. BGP4+ Implementation and Configurations

HUAWEI

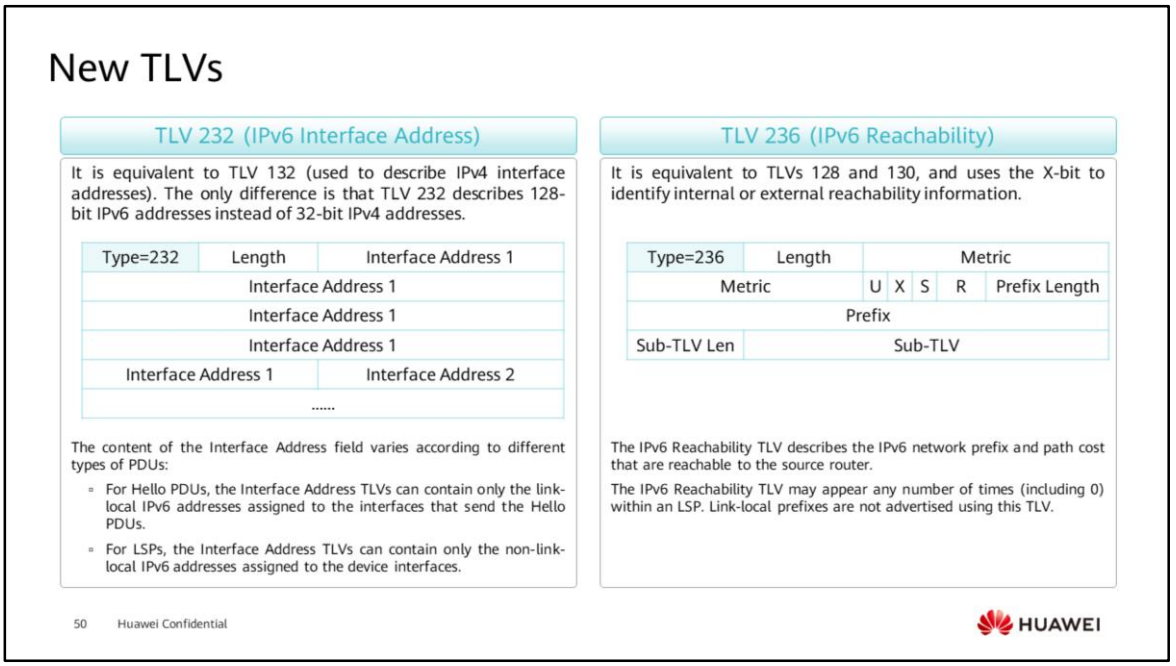
IS-IS (IPv6) Overview

- IS-IS is a link-state dynamic routing protocol, which was initially designed for the OSI network. Later, to support IPv4 routing, extensions were made to apply IS-IS to IPv4 networks. This is called integrated IS-IS.
- IS-IS packets are classified into Hello protocol data units (PDUs), link state PDUs (LSPs), and sequence number PDUs (SNPs).
  - Packet format



  - The Variable Length Fields (TLV) part consists of multiple TLV triplets. Using the TLV structure to construct packets improves the flexibility and scalability of IS-IS. To support new features, only new TLVs need to be added.
- To support the processing and calculation of IPv6 routes, two TLVs and one network layer protocol identifier (NLPID) are added to IS-IS.

HUAWEI



- Fields in TLV 232 (IPv6 Interface Address) are described as follows:
  - Type: indicates the TLV type and occupies 8 bits. The value is 232 (0xE8).
  - Length: indicates the length of the Value field in the TLV and occupies 8 bits.
  - Interface Address: indicates a 128-bit IPv6 address.
- Fields in TLV 236 (IPv6 Reachability) are described as follows:
  - Type: indicates the TLV type and occupies 8 bits. The value is 236 (0xEC).
  - Length: indicates the length of the Value field in the TLV and occupies 8 bits.
  - Metric: a 32-bit field indicating the cost.
  - U: up/down bit. This 1-bit field indicates whether the prefix is advertised from a higher level to a lower level.
  - X: external original bit. This 1-bit field indicates whether the prefix was imported into IS-IS from another routing protocol.
  - S: sub-TLV present bit. This 1-bit field is optional.
  - R: This 5-bit field is reserved.
  - Prefix Length: indicate the prefix length and occupies 8 bits.
  - Prefix: indicates an IPv6 address prefix.
  - Sub-TLV Length: indicates the length of a sub-TLV and occupies 8 bits. If the S-bit is set to 1, this field is included.
  - Sub-TLV: If the S-bit is set to 1, this field is included.

New NLPID in TLV 129

- To support the processing and calculation of IPv6 routes, IS-IS adds an NLPID to TLV 129.
- TLV 129 (Protocol Supported)

Type=129	Length	NLPID	NLPID
.....			

Field description:

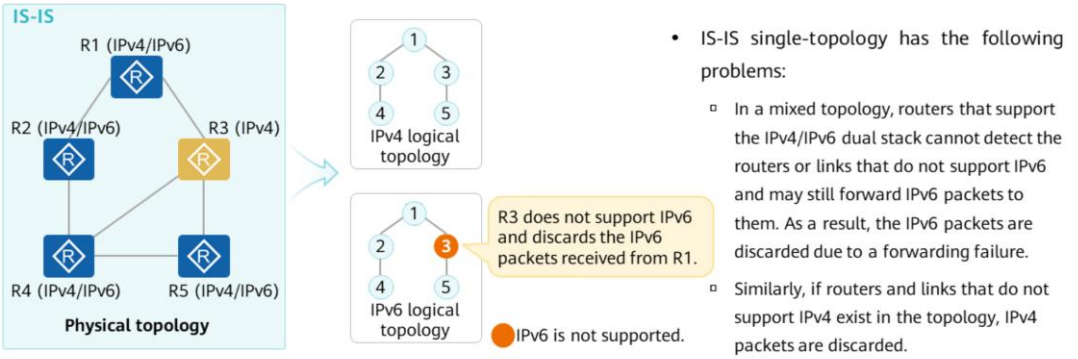
- Type: indicates the TLV type and occupies 8 bits. The value is 129 (0x81), indicating the Protocol Supported TLV.
- Length: indicates the length of the Value field in the TLV and occupies 8 bits.
- NLPID: indicates the network layer protocol identifier and occupies 8 bits. If IPv4 is supported, the value is 204 (0xCC). If IPv6 is supported, the value is 142 (0x8E).

- If IS-IS supports IPv6, the IPv6 routing information that it advertises must carry the NLPID.

HUAWEI

IS-IS Multi-Topology Background

- By default, on an IS-IS network, a mix of IPv4 and IPv6 topologies is considered as an integrated topology, and IS-IS calculates the same SPT for IPv4 and IPv6.



- IS-IS single-topology has the following problems:
  - In a mixed topology, routers that support the IPv4/IPv6 dual stack cannot detect the routers or links that do not support IPv6 and may still forward IPv6 packets to them. As a result, the IPv6 packets are discarded due to a forwarding failure.
  - Similarly, if routers and links that do not support IPv4 exist in the topology, IPv4 packets are discarded.

HUAWEI

- IS-IS single-topology has the following disadvantages:
  - Network deployment is not suitable for topology separation.
  - To maintain the same topology, each interface must run both IS-IS (IPv4) and IS-IS (IPv6), which is not flexible.
  - IPv4 areas cannot be used to connect different IPv6 areas. That is, IPv4 networks cannot be used to address IPv6 network isolation.

IS-IS Multi-Topology Overview

- The IS-IS multi-topology (MT) feature indicates that multiple independent IP topologies run in an IS-IS domain. For example, IPv4 and IPv6 topologies are not considered as a single integrated topology. This ensures that IS-IS routes are calculated for both IPv4 and IPv6 networks. Based on the IP versions supported by links, separate SPF calculation is performed in each topology to shield networks from each other.
- IS-IS MT implementation
  - Topology establishment: Neighbor relationships are established through packet exchange to set up multiple topologies.
  - SPF calculation: SPF calculation is performed in each topology.

HUAWEI

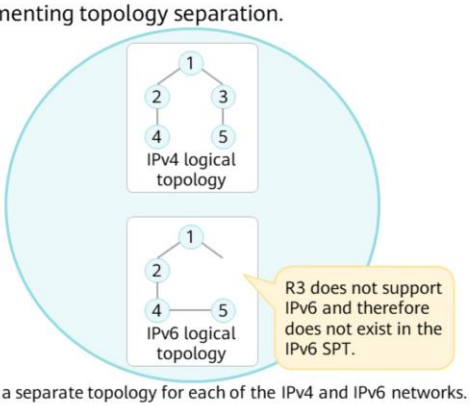
- The IS-IS MT feature can overcome the disadvantages of IS-IS single topology.

IS-IS MT Implementation

- IS-IS defines new TLVs, which contain MT information, that is, information about the topologies to which interfaces belong. The advertisement of MT information enables the network to perform separate SPF calculations for different topologies, implementing topology separation.
- Multi-Topology TLV 229

Type=229	Length	O	A	R	MT ID
.....					

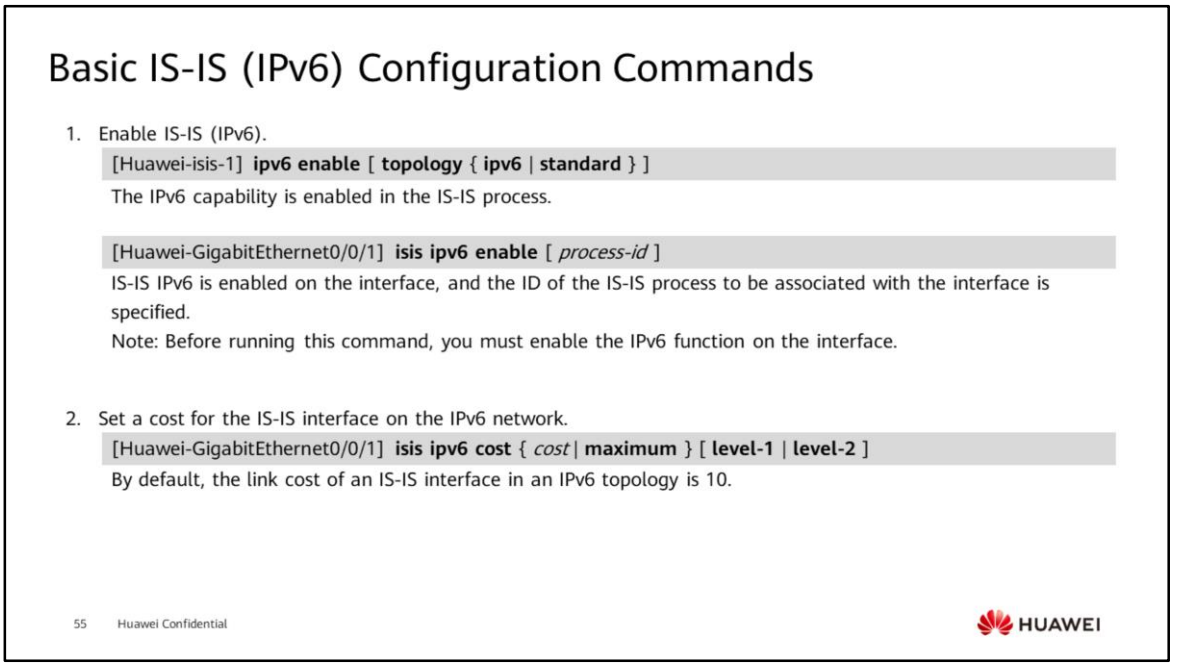
  - Key fields are described as follows:
    - Type: indicates the TLV type and occupies 8 bits. The value is 229 (0xE5), indicating support for MT.
    - O: Overload bit, which occupies 1 bit.
    - A: Attach bit, which occupies 1 bit.
    - MT ID: indicates the topology to which the interface belongs, and occupies 12 bits.



IS-IS calculates a separate topology for each of the IPv4 and IPv6 networks.

HUAWEI

- To support MT, IS-IS defines multiple types of TLVs, including Multi-Topology TLV, MT Intermediate Systems TLV, Multi-Topology Reachable IPv4 Prefixes TLV, and Multi-Topology Reachable IPv6 Prefixes TLV. This course focuses on the Multi-Topology TLV and does not elaborate on the other ones.
- Multi-Topology TLV:
  - This TLV is contained only in IIH PDUs and fragment zero LSPs.
  - Reserved MT IDs:
    - MT ID=0, equivalent to the standard IPv4 topology.
    - MT ID=2, reserved for the IPv6 routing topology.



- The basic configuration commands and methods of IS-IS (IPv6) are the same as those of IS-IS (IPv4). For details, see the "HCIP-Datacom-Core Technology" course.
- [Huawei-isis-1] **ipv6 enable [topology { ipv6 | standard }]**
  - **topology:** sets a network topology type.
  - **ipv6:** sets the topology type to IPv6. That is, the IPv6 capability for the IS-IS process is enabled in an IPv6 topology. Links on the network can be configured as IPv4 or IPv6 links. SPF calculation is performed separately in IPv4 and IPv6 topologies.
  - **standard:** sets the topology type to standard. That is, the IPv6 capability for the IS-IS process is enabled in an integrated topology. A network administrator must ensure that all links on the network support the same topology type. By default, the standard type is used when the IPv6 capability is enabled for an IS-IS process.

Verifying the Configuration of Basic IS-IS (IPv6) Functions

1. Check information about the interface with IS-IS (IPv6) enabled.

   [Huawei] **display isis interface** *interface-type interface-number* [**verbose**]

2. Check IS-IS (IPv6) neighbor information.

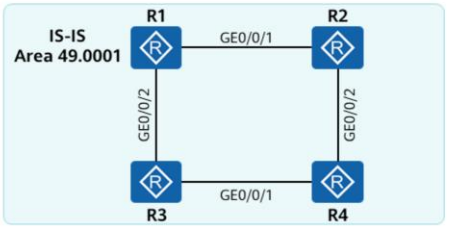
   [Huawei] **display isis** *process-id* **peer** [**verbose**]

3. Check IS-IS (IPv6) routing information.

   [Huawei] **display isis route** [*process-id* | **vpn-instance** *vpn-instance-name*] **ipv6** [**verbose** | [**level-1** | **level-2**] | *ipv6-address* [*prefix-length*]]

HUAWEI

Example for Configuring IS-IS IPv4/IPv6 Dual Stack

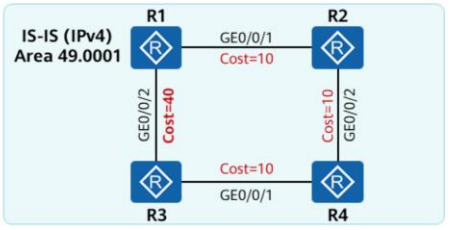


Device	Interface	IPv4 Address	IPv6 Address
R1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
	GE0/0/2	10.1.13.1/24	2001:DB8:2345:13::1/64
R2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
	GE0/0/2	10.1.24.2/24	2001:DB8:2345:24::2/64
R3	GE0/0/1	10.1.34.3/24	2001:DB8:2345:34::3/64
	GE0/0/2	10.1.13.3/24	2001:DB8:2345:13::3/64
R4	GE0/0/1	10.1.34.4/24	2001:DB8:2345:34::4/64
	GE0/0/2	10.1.24.4/24	2001:DB8:2345:24::4/64

- Scenario description:
  - A company has IS-IS deployed over an IPv4 network to implement IPv4 network interconnection. To ensure future service development, the company has an IPv6 network deployed for service testing. In this case, the IPv6 function needs to be enabled in the IS-IS process on the network.
  - As device performance varies, link cost modification is required to ensure that IPv4 and IPv6 services are routed through different paths, and IS-IS MT is required to ensure that a separate topology is calculated for the IPv6 network.
  - All routers run IS-IS and are all Level-2 routers. The entire network is in area 49.0001. After the configuration is complete, each router should learn the IPv4 and IPv6 routes to all network segments in the AS.

HUAWEI

Deploying the IPv4 Network



Device	Interface	IPv4 Address	IPv6 Address
R1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
	GE0/0/2	10.1.13.1/24	2001:DB8:2345:13::1/64
R2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
	GE0/0/2	10.1.24.2/24	2001:DB8:2345:24::2/64
R3	GE0/0/1	10.1.34.3/24	2001:DB8:2345:34::3/64
	GE0/0/2	10.1.13.3/24	2001:DB8:2345:13::3/64
R4	GE0/0/1	10.1.34.4/24	2001:DB8:2345:34::4/64
	GE0/0/2	10.1.24.4/24	2001:DB8:2345:24::4/64

1. Assign an IPv4 address to each router interface. (The configuration details are not provided here.)

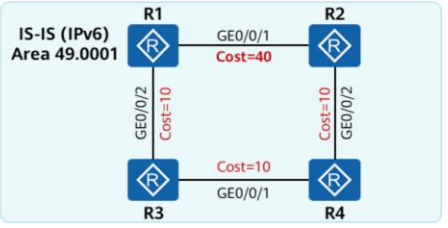
2. Configure basic IS-IS functions.

```
[R1] isis 1
[R1-isis-1] is-level level-2
[R1-isis-1] network-entity 49.0001.0000.0000.0001.00
#
[R1] interface gigabitethernet 0/0/1
[R1-GigabitEthernet0/0/1] isis enable 1
[R1-GigabitEthernet0/0/1] quit
[R1] interface gigabitethernet 0/0/2
[R1-GigabitEthernet0/0/2] isis enable 1
[R1-GigabitEthernet0/0/2] isis cost 40 level-2
```

The configurations of R2, R3, and R4 are similar to the configuration of R1, and are not provided here.
The network entity titles (NETs) are as follows:
R2: 49.0001.0000.0000.0002.00
R3: 49.0001.0000.0000.0003.00
R4: 49.0001.0000.0000.0004.00

HUAWEI

Deploying the IPv6 Network



Device	Interface	IPv4 Address	IPv6 Address
R1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
	GE0/0/2	10.1.13.1/24	2001:DB8:2345:13::1/64
R2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
	GE0/0/2	10.1.24.2/24	2001:DB8:2345:24::2/64
R3	GE0/0/1	10.1.34.3/24	2001:DB8:2345:34::3/64
	GE0/0/2	10.1.13.3/24	2001:DB8:2345:13::3/64
R4	GE0/0/1	10.1.34.4/24	2001:DB8:2345:34::4/64
	GE0/0/2	10.1.24.4/24	2001:DB8:2345:24::4/64

1. Assign an IPv6 address to each router interface. (The configuration details are not provided here.)

2. Enable IS-IS (IPv6).

```
[R1] isis 1
[R1-isis-1] ipv6 enable topology ipv6
#
[R1] interface gigabitethernet 0/0/1
[R1-GigabitEthernet0/0/1] isis ipv6 enable 1
[R1-GigabitEthernet0/0/1] isis ipv6 cost 40 level-2
[R1-GigabitEthernet0/0/1] quit
[R1] interface gigabitethernet 0/0/2
[R1-GigabitEthernet0/0/2] isis ipv6 enable 1
```

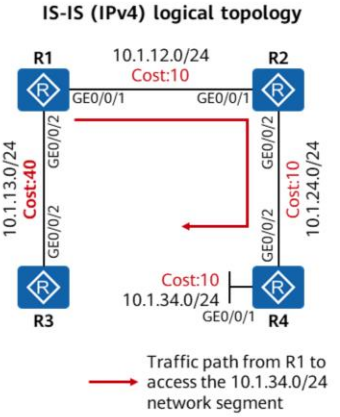
The configurations of R2, R3, and R4 are similar to the configuration of R1, and are not provided here.

HUAWEI

Check the Routing Information on the IS-IS (IPv4) Network

- The following example uses R1 as the root to calculate the SPT. The IS-IS (IPv4) logical topology is shown in the right figure.

```
[R1]dis ip routing-table protocol isis
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Public routing table : ISIS
         Destinations : 2        Routes : 2

ISIS routing table status : <Active>
         Destinations : 2        Routes : 2

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

      10.1.24.0/24  ISIS-L2 15   20        D     10.1.12.2       GigabitEthernet0/0/1
      10.1.34.0/24  ISIS-L2 15   30        D     10.1.12.2       GigabitEthernet0/0/1

ISIS routing table status : <Inactive>
         Destinations : 0        Routes : 0
```



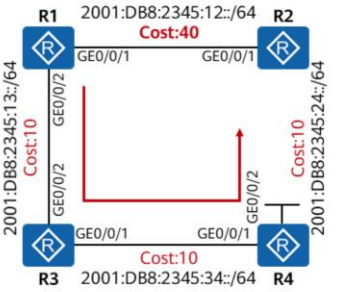
HUAWEI

Check the Routing Information on the IS-IS (IPv6) Network

- The following example uses R1 as the root to calculate the SPT. The IS-IS (IPv6) logical topology is shown in the right figure.

```
[R1]dis ipv6 routing-table protocol isis
Public Routing Table : ISIS
Summary Count : 2

ISIS Routing Table's Status : < Active >
Summary Count : 2

 Destination :              2001:DB8:2345:24::          PrefixLength : 64
 NextHop :                  FE80::2E0:FCFF:FE90:3D3A    Preference   : 15
 Cost :                     30                          Protocol     : ISIS-L2
 RelayNextHop :             ::                          TunnelID     : 0x0
 Interface :   GigabitEthernet0/0/2                     Flags        : D

 Destination :              2001:DB8:2345:34::          PrefixLength : 64
 NextHop :   FE80::2E0:FCFF:FE90:3D3A                   Preference   : 15
 Cost :                     20                          Protocol     : ISIS-L2
 RelayNextHop :             ::                          TunnelID     : 0x0
 Interface :   GigabitEthernet0/0/2                     Flags        : D

ISIS Routing Table's Status : < Inactive >
Summary Count : 0
```

IS-IS (IPv6) logical topology



Traffic path from R1 to access the 2001:DB8:2345:24::/64 network segment

HUAWEI

Contents

1. IPv6 Static Routes
2. OSPFv3 Implementation and Configurations
3. IS-IS (IPv6) Implementation and Configurations
4. **BGP4+ Implementation and Configurations**

HUAWEI

BGP4+ Overview

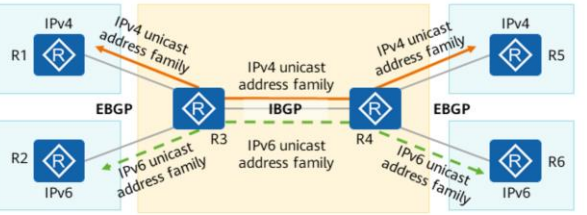
- Traditional BGP-4 can manage only IPv4 unicast routing information. Multiprotocol Extensions for BGP (MP-BGP) can carry routing information for multiple network layer protocols. Currently, MP-BGP uses extended attributes and address families to support IPv6, multicast, and VPN, without changing the messaging and routing mechanisms of BGP-4.
- MP-BGP for IPv6 unicast networks is called BGP4+. BGP4+ establishes an independent topology for an IPv6 unicast network and stores routing information in an independent routing table. This ensures that routing information on an IPv4 unicast network is isolated from that on an IPv6 unicast network.

HUAWEI

- To support IPv6, BGP needs to map IPv6 routing information to the NLRI attributes.

Address Families That MP-BGP Supports

- MP-BGP uses address families to differentiate network layer protocols. To enable BGP peers to exchange different types of routing information, activating the peers and advertising BGP routes need to be performed in the correct address families.
- MP-BGP supports the following address families:
  - IPv4 unicast address family
  - IPv4 multicast address family
  - IPv6 unicast address family
  - VPNv4 address family
  - VPNv6 address family
  - ...



Peers on the IPv4 network must be activated in the IPv4 unicast address family, whereas peers on the IPv6 network must be activated in the IPv6 unicast address family.

HUAWEI

BGP Path Attributes

- BGP Update messages are used to transmit routing information between peers as well as advertise and withdraw routes.
- Update message format:

Withdrawn Routes Length (2 bytes)
Withdrawn Routes (*N* bytes)
Total Path Attribute Length (2 bytes)
Path Attributes (*N* bytes)
Network Layer Reachability Information (*N* bytes)

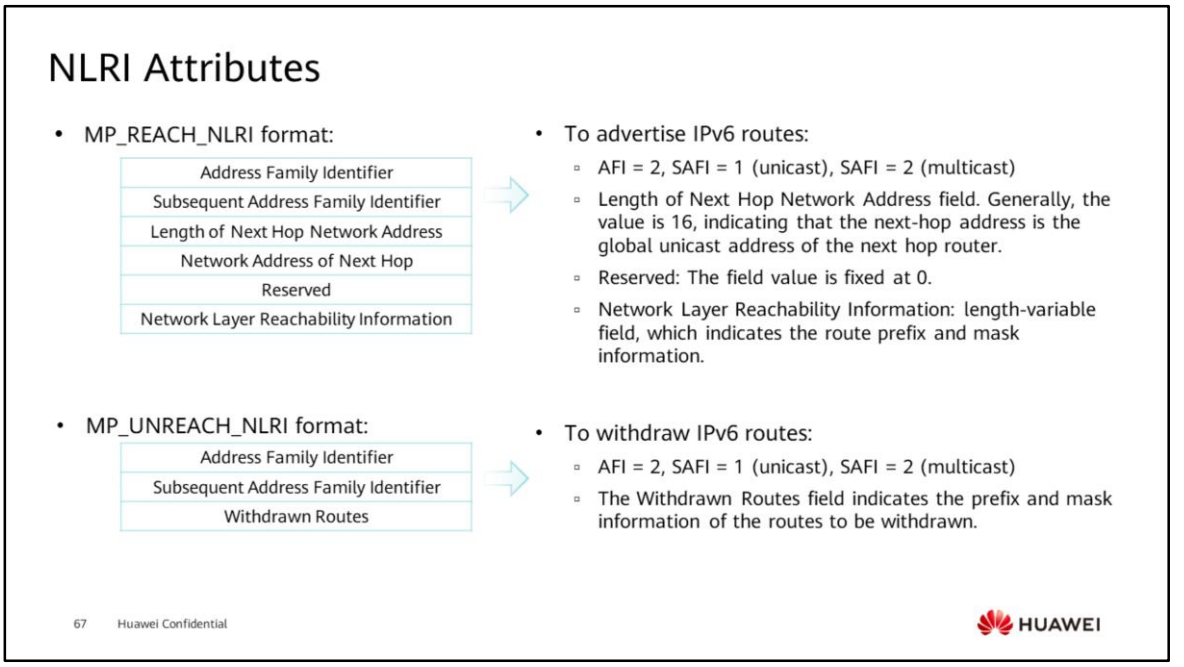
Attr. Type		Attr. Length	Attr. Value
Flags	Type Code		

14: MP_REACH_NLRI
15: MP_UNREACH_NLRI

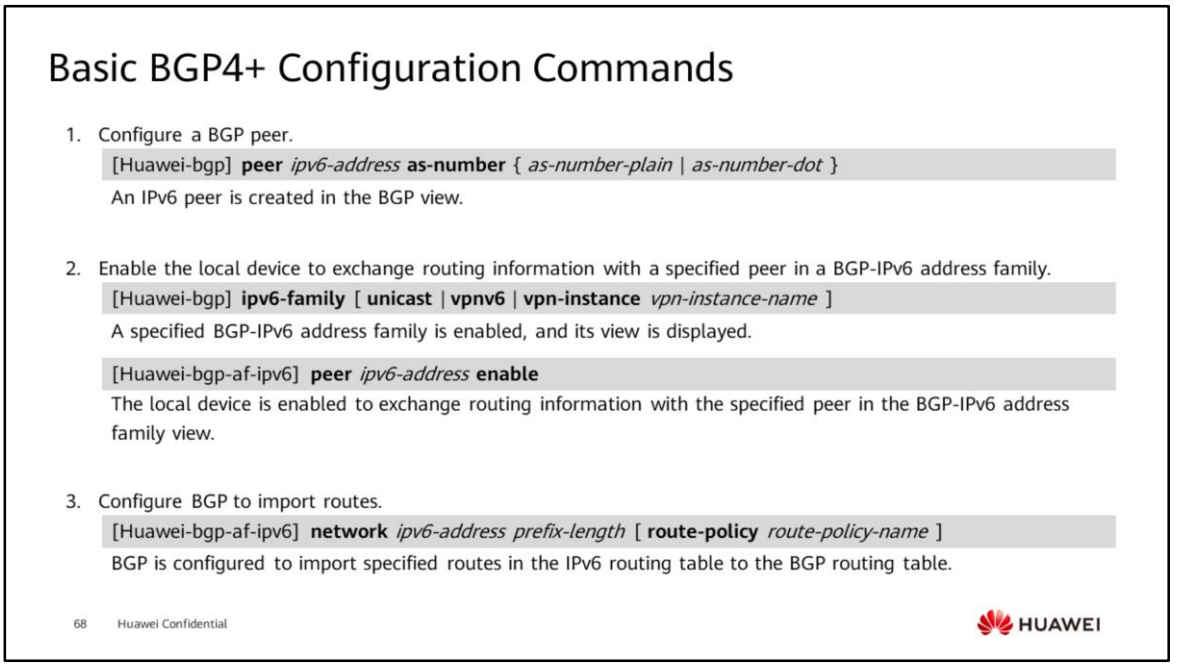
- BGP4+ introduces two NLRI attributes:
  - MP_REACH_NLRI: indicates the multiprotocol reachable NLRI, which is used to advertise reachable routes and next hop information.
  - MP_UNREACH_NLRI: indicates the multiprotocol unreachable NLRI, which is used to withdraw unreachable routes.

HUAWEI

- Update message:
  - An Update message can be used to advertise multiple routes with the same path attribute. These routes are stored in the NLRI attribute. An Update message can also carry multiple unreachable routes, which are stored in the Withdrawn Routes field, to instruct peers to withdraw these routes.



- Fields in the MP_REACH_NLRI attribute are described as follows:
  - Address Family Information: consists of a 2-byte address family identifier (AFI) and a 1-byte subsequent address family identifier (SAFI).
  - Length of Next Hop Network Address: indicates the length of the next-hop address and occupies 1 byte. Generally, the value is 16.
  - Network Address of Next Hop: The length is variable and depends on the preceding field. Generally, the value is a global unicast address.
  - Reserved: 1 byte. The value must be 0.
  - Network Layer Reachability Information: contains information about the routes with the same attributes. The value 0 indicates the default route.
- Fields in the MP_UNREACH_NLRI field are described as follows:
  - Withdrawn Routes: indicates the routes to be withdrawn. The value is in the <mask length, route prefix> format. If the mask length is 0, the associated route is a default route.



- The basic configuration commands and methods of BGP4+ are the same as those of BGP. For details, see the "HCIP-Datacom-Core Technology" course.
- [Huawei-bgp] **ipv6-family** [**unicast** | **vpnv6** | **vpn-instance** *vpn-instance-name*]
  - **unicast**: enters the IPv6 unicast address family view.
  - **vpnv6**: enters the BGP-VPNv6 address family view.
  - **vpn-instance** *vpn-instance-name*: associates a specified VPN instance with the IPv6 address family and enters the BGP-VPN instance IPv6 address family view. The value is a string of 1 to 31 case-sensitive characters. If spaces are used, the string must start and end with double quotation marks (").
- [Huawei-bgp-af-ipv6] **network** *ipv6-address prefix-length* [**route-policy** *route-policy-name*]
  - *ipv6-address*: specifies the IPv6 address advertised by BGP. The value is a 32-digit hexadecimal number, in the format of *X:X:X:X:X:X:X:X*.
  - *prefix-length*: specifies the prefix length of the IPv6 address advertised by BGP. The value is an integer ranging from 0 to 128.
  - **route-policy** *route-policy-name*: specifies the name of the route-policy applied to route advertisement. The value is a string of 1 to 40 case-sensitive characters. If spaces are used, the string must start and end with double quotation marks (").

Verifying the Configuration of Basic BGP4+ Functions

1. Check BGP4+ peer information.

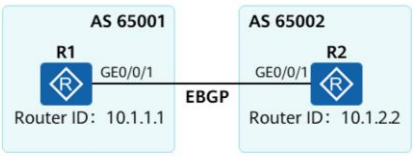
   [Huawei] **display bgp ipv6 peer** *ipv6-address* [**verbose**]

2. Check BGP4+ routing information.

   [Huawei] **display bgp ipv6 routing-table**

HUAWEI

Example for Configuring BGP IPv4/IPv6 Dual Stack

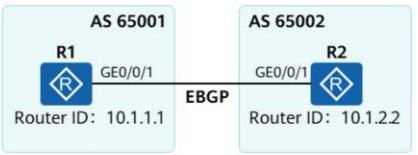


Device	Interface	IPv4 Address	IPv6 Address
R1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
	Loopback0	10.1.1.1/32	2001:DB8:2345:1::1/128
R2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
	Loopback0	10.1.2.2/32	2001:DB8:2345:2::2/128

- Scenario description:
  - Two branches of a company communicate with each other through BGP. To ensure future service development, an IPv6 network is deployed in each branch for service testing. In this case, BGP4+ needs to be deployed to implement IPv6 network interworking.

HUAWEI

Deploying the IPv4 Network



Device	Interface	IPv4 Address	IPv6 Address
R1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
	Loopback0	10.1.1.1/32	2001:DB8:2345:1::1/128
R2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
	Loopback0	10.1.2.2/32	2001:DB8:2345:2::2/128

1. Assign an IPv4 address to each router interface. (The configuration details are not provided here.)

2. Configure basic BGP functions: establish an EBGP peer relationship.

```
[R1] bgp 65001
[R1-bgp] router-id 10.1.1.1
[R1-bgp] peer 10.1.12.2 as-number 65002
```

```
[R2] bgp 65002
[R2-bgp] router-id 10.1.2.2
[R2-bgp] peer 10.1.12.1 as-number 65001
```

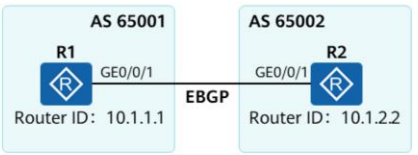
3. Configure BGP to advertise IPv4 routes.

```
[R1-bgp] iPv4-family unicast
[R1-bgp-af-ipv4] network 10.1.1.1 32
```

```
[R2-bgp] iPv4-family unicast
[R2-bgp-af-ipv4] network 10.1.2.2 32
```


HUAWEI

Deploying the IPv6 Network



Device	Interface	IPv4 Address	IPv6 Address
R1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
	Loopback0	10.1.1.1/32	2001:DB8:2345:1::1/128
R2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
	Loopback0	10.1.2.2/32	2001:DB8:2345:2::2/128

1. Assign an IPv6 address to each router interface. (The configuration details are not provided here.)

2. Configure basic BGP4+ functions: establish an EBGP peer relationship and enable the local device to exchange routing information with the EBGP peer.

```
[R1] bgp 65001
[R1-bgp] peer 2001:DB8:2345:12::2 as-number 65002
[R1-bgp] iPv6-family unicast
[R1-bgp-af-ipv6] peer 2001:DB8:2345:12::2 enable
```

```
[R2] bgp 65002
[R2-bgp] peer 2001:DB8:2345:12::1 as-number 65001
[R2-bgp] iPv6-family unicast
[R2-bgp-af-ipv6] peer 2001:DB8:2345:12::1 enable
```

3. Configure BGP4+ to advertise IPv6 routes.

```
[R1-bgp-af-ipv6] network 2001:DB8:2345:1::1 128
```

```
[R2-bgp-af-ipv6] network 2001:DB8:2345:2::2 128
```


HUAWEI

Checking the BGP Peer Information

- Run the **display bgp [ipv6] peer** command to check information about the peers on the IPv4 and IPv6 networks. By comparing the command outputs, you will find that the peer information is basically the same except for the addresses used to establish the peer relationships.

BGP

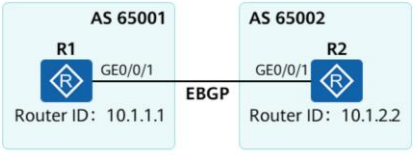
```
[R1]display bgp peer

 BGP local router ID :  10.1.1.1
 Local AS number:       65001
 Total number of peers:  1                 Peers in established state : 1

  Peer                  V          AS  MsgRcvd  MsgSent  OutQ  Up/Down  State        PrefRcv
  10.1.12.2             4       65002       11       12     0  00:06:49  Established      1
```

BGP4+

```
[R1]display bgp ipv6 peer

 BGP local router ID : 10.1.1.1
 Local AS number : 65001
 Total number of peers : 1                 Peers in established state : 1

  Peer                  V          AS  MsgRcvd  MsgSent  OutQ  Up/Down  State        PrefRcv
  2001:DB8:2345:12::2   4       65002       14       15     0  00:07:35  Established      1
```


HUAWEI

Checking the BGP4+ Routing Information



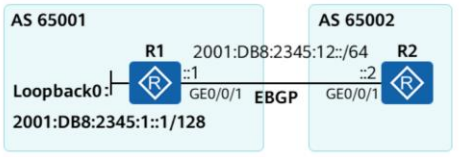
Device	Interface	IPv4 Address	IPv6 Address
R1	GE0/0/1	10.1.12.1/24	2001:DB8:2345:12::1/64
	Loopback0	10.1.1.1/32	2001:DB8:2345:1::1/128
R2	GE0/0/1	10.1.12.2/24	2001:DB8:2345:12::2/64
	Loopback0	10.1.2.2/32	2001:DB8:2345:2::2/128

```
[R1]display bgp ipv6 routing-table

 BGP Local router ID is 10.1.1.1
 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 :          2001:DB8:2345:1::1      PrefixLen :   128
     NextHop :          ::                      LocPrf  :
     MED :              0                       PrefVal :     0
     Label  :
     Path/Ogn :         i
 *>  Network :          2001:DB8:2345:2::2      PrefixLen :   128
     NextHop :          2001:DB8:2345:12::2     LocPrf  :
     MED :              0                       PrefVal :     0
     Label :
     Path/Ogn :         65002  i
```

Check the BGP4+ routing table. The command output shows that a total of two valid routes exist.

HUAWEI

Checking the NLRI Attributes of BGP4+



- By analyzing the Update message sent by R1, you can check information about the MP_REACH_NLRI attribute. In the attribute:
  - Next-hop address of the reachable route: 2001:db8:2345:12::1
  - Prefix and prefix length of the reachable route: 2001:db8:2345:1::1/128

```
UPDATE Message-Border Gateway Protocol
  UPDATE Message
    Marker: 16 bytes
    Length: 85 bytes
    Type: UPDATE Message (2)
    Unfeasible routes length: 0 bytes
    Total path attribute length: 62 bytes
    Path attributes
      ORIGIN: IGP (4 bytes)
      AS_PATH: 65001 (9 bytes)
      MULTI_EXIT_DISC: 0 (7 bytes)
      MP_REACH_NLRI (42 bytes)
        Flags: 0x90 (Optional, Non-transitive, Complete, Extended Length)
        Type code: MP_REACH_NLRI (14)
        Length: 38 bytes
        Address family: IPv6 (2)
        Subsequent address family identifier: Unicast (1)
        Next hop network address (16 bytes)
          Next hop: 2001:db8:2345:12::1 (16)
        Subnetwork points of attachment: 0
        Network layer reachability information (17 bytes)
          2001:db8:2345:1::1/128
            MP Reach NLRI prefix length: 128
            MP Reach NLRI prefix: 2001:db8:2345:1::1
```


HUAWEI

Quiz

1. (Multiple-answer question) Which of the following types of LSAs belong to OSPFv3? ( )
   A. Router-LSA
   B. Network-LSA
   C. Network-summary-LSA
   D. Link-LSA
2. (True or false) IS-IS can calculate the SPTs for IPv4 and IPv6 networks separately. ( )
   A. True
   B. False

HUAWEI

1. ABD
2. A

Summary

- IPv6 static routes can meet network interconnection requirements. However, as the network scale expands, dynamic routing protocols need to be leveraged. Dynamic routing protocols are required to support IPv6 and carry IPv6 addresses for route advertisement.
- To support IPv6 networks, the IETF enhanced and improved OSPFv2 and finally developed the new protocol, OSPFv3.
- IS-IS features high scalability and supports IPv6 through new TLVs.
- Similarly, BGP supports IPv6 by using its own extended path attributes and multi-protocol address families.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

IPv6 Transition Technologies

HUAWEI

Foreword

- As shortage of IPv4 addresses becomes severer, IPv4-to-IPv6 transition has become the focus of network development. Although IPv6 is the ultimate solution to this problem, IPv4 networks will be retained for a long period of time to meet service continuity requirements. During the IPv4-to-IPv6 transition, it is inevitable that IPv4 and IPv6 addresses coexist on networks.
- This course describes the transition technologies used when IPv4 and IPv6 coexist, including IPv6 over IPv4 tunneling, IPv6 Provider Edge (6PE), IPv6 VPN Provider Edge (6VPE), IPv4 over IPv6 tunneling, Network Address Translation IPv6-to-IPv4 (NAT64), and IVI.

HUAWEI

Objectives

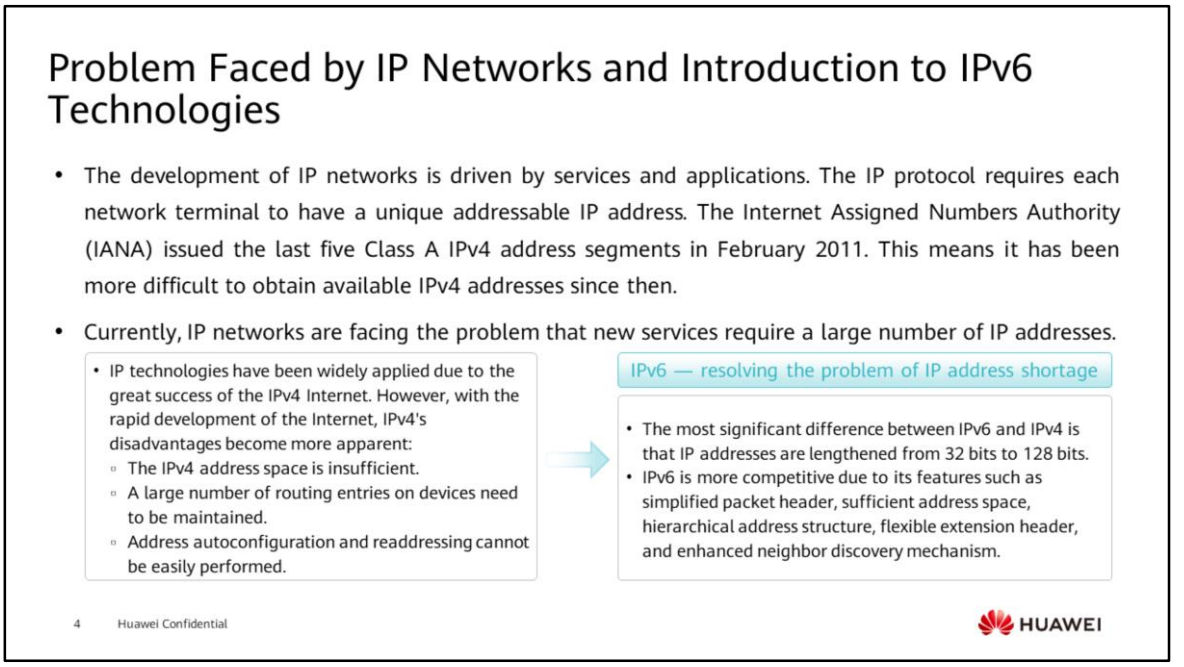
- On completion of this course, you will be able to:
  - Master the basics of IPv6, such as classification and obtaining of IPv6 addresses.
  - Describe IPv6 transition technologies and their implementations, and complete the configuration.

HUAWEI

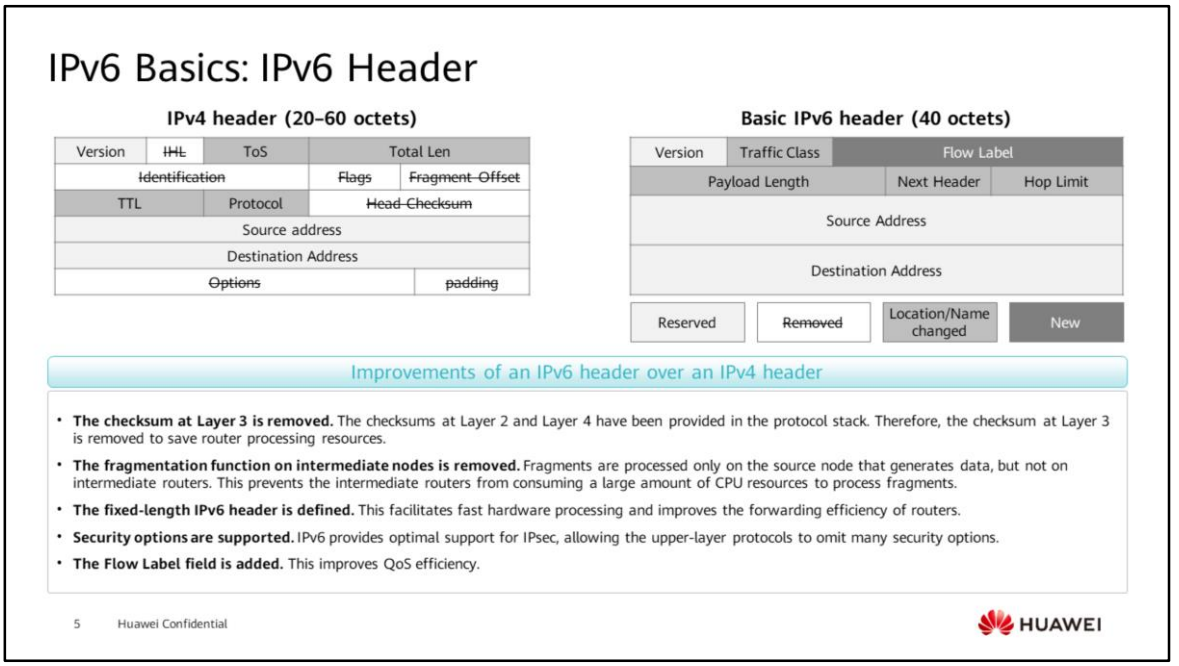
Contents

1. **IPv6 Basics**
2. IPv6 Transition Technologies

HUAWEI

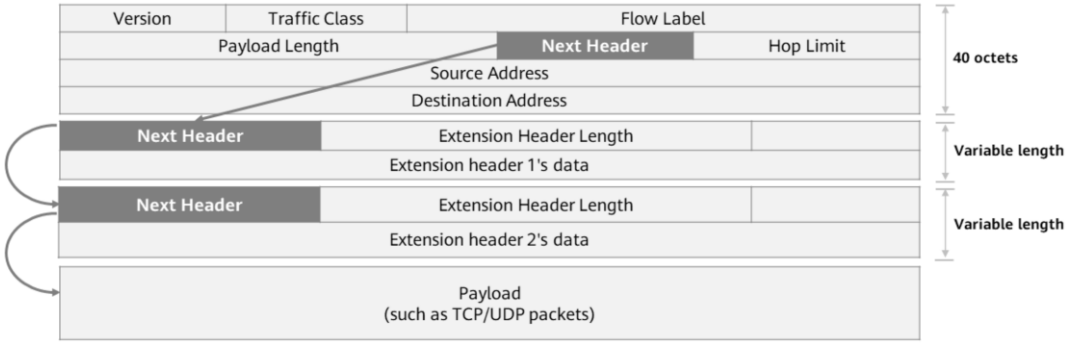


- Disadvantages of IPv4:
  - **The IPv4 address space is insufficient.** IPv4 addresses are 32 bits long. Theoretically, 4.3 billion IPv4 addresses can be provided. However, the number of IPv4 addresses that are actually available cannot reach this number due to various address assignment reasons. Additionally, IPv4 address resources are unevenly assigned. IPv4 addresses in U.S. account for almost half of all addresses, leaving insufficient addresses for Europe and even fewer for the Asia-Pacific region. The shortage of IPv4 addresses limits further development of mobile IP and broadband technologies that require an increasing number of IP addresses.
  - **A large number of routing entries on devices need to be maintained.** At the initial stage of IPv4 development, many discontinuous IPv4 addresses are assigned. As a result, routes cannot be effectively summarized. The increasingly large routing table consumes a lot of memory resources and affects forwarding efficiency. Device vendors must continually upgrade devices to improve route addressing and forwarding performance.
  - **Address autoconfiguration and readdressing cannot be easily performed.** IPv4 addresses are only 32 bits long, and they are unevenly assigned. As a result, IP addresses need to be reassigned during network expansion or reconstruction, increasing the maintenance workload.
- Advantages of IPv6:
  - Easy to deploy.
  - Compatible with various applications.
  - Smooth transition from IPv4 networks to IPv6 networks.



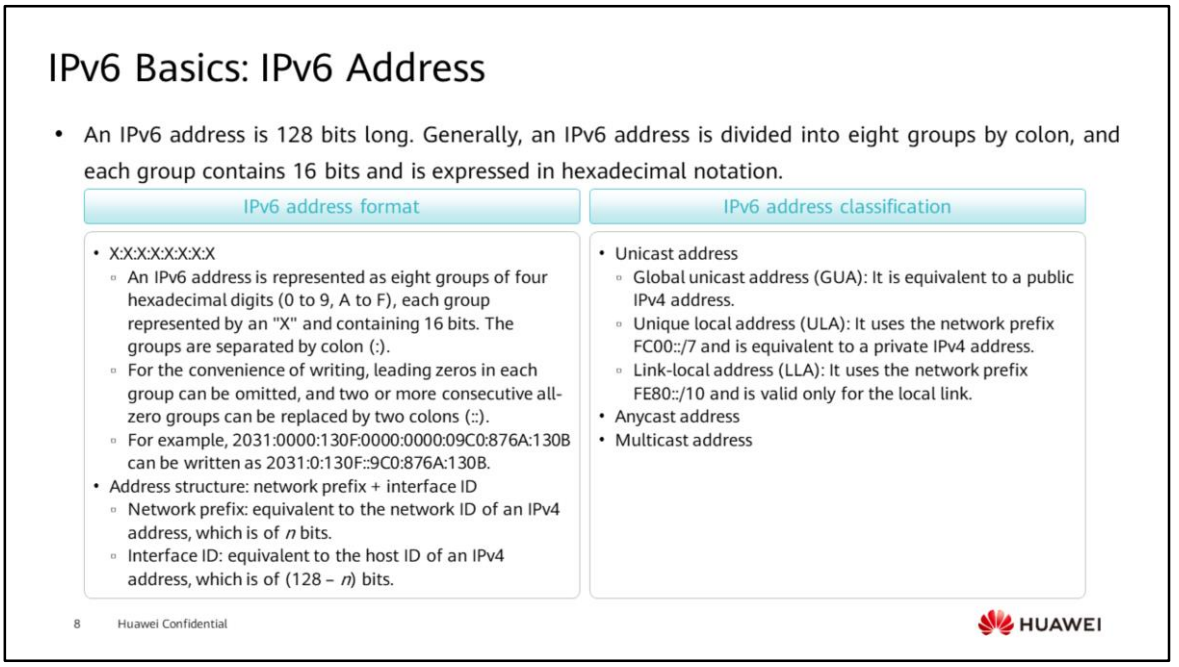
- The IPv6 header is also called fixed header, which contains eight fields in 40 octets. The fields are Version, Traffic Class, Flow Label, Payload Length, Next Header, Hop Limit, Source Address, and Destination Address.
    - **Version:** This field indicates the version of IP and its value is 6. The 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:** An IPv4 header does not contain this field. It is a new field. It is used by a source to label sequences of packets for which the label requests special handling by IPv6 routers. The length is 20 bits. Generally, the flow label, together with the source and destination IPv6 addresses, can determine a flow.
    - **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. This field is 16 bits long and can specify a maximum length of 65535 octets for the payload. If the payload size exceeds 65535 octets, 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.

IPv6 Basics: IPv6 Extension Header



- **Next Header**: 8 bits long. This field is similar to the Next Header field in the basic IPv6 header, indicating the next extension header (if any) or upper-layer protocol type.
- **Extension Header Length**: 8 bits long. This field indicates the extension header length excluding the Next Header field.
- **Extension Header Data**: variable length. This field includes a series of options and the padding field.

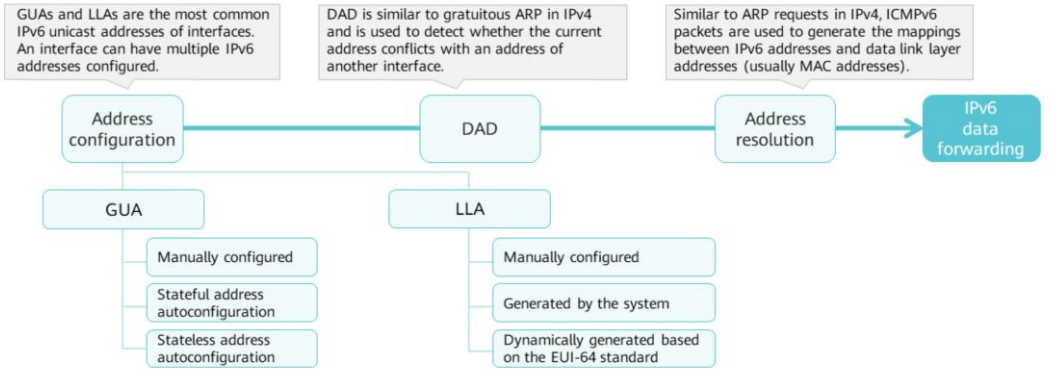
HUAWEI



- Global unicast address (GUA): It is also known as 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.
- Unique local address (ULA): It 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.
- Link-local address (LLA): It is another type of IPv6 address with a limited application scope.

IPv6 Basics: Service Process of IPv6 Unicast Addresses

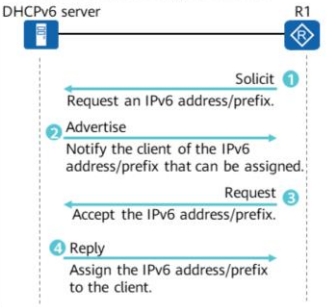
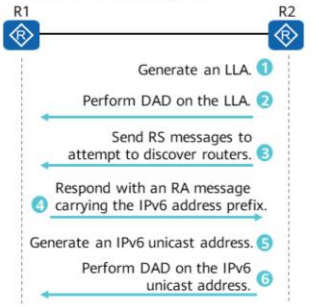
- Before sending an IPv6 packet, an interface undergoes address configuration, duplicate address detection (DAD), and address resolution. The Neighbor Discovery Protocol (NDP) plays a significant role in this process.



HUAWEI

IPv6 Basics: Configuration Methods of GUAs

- Similar to IPv4 address assignment, IPv6 address assignment can be classified into manual configuration, stateful address autoconfiguration, and stateless address autoconfiguration.

Manual configuration	Stateful address autoconfiguration (DHCPv6)	Stateless address autoconfiguration (SLAAC)
• An IPv6 GUA can be: ▫ Generated in EUI-64 format: An IPv6 GUA in EUI-64 format contains a manually configured prefix and an interface ID that is automatically generated by an interface. ▫ Configured manually.	• Implemented through DHCPv6:  	• Implemented through NDP: 

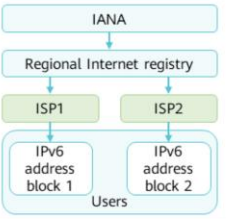
HUAWEI

IPv6 Address Application

- There are two types of IPv6 address space: Provider-Aggregatable (PA) address space and Provider-Independent (PI) address space.

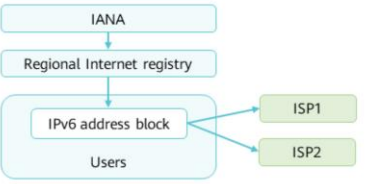
PA address space

- The PA address space is a block of IP addresses assigned by a regional Internet registry to an Internet service provider (ISP).
- An ISP can assign addresses in an address block based on user requirements. PA addresses are provided to users together with services such as private line and Internet egress. This mode is cost-effective and applies to small and midsize enterprises. However, the assigned addresses are changed if a different ISP is used.

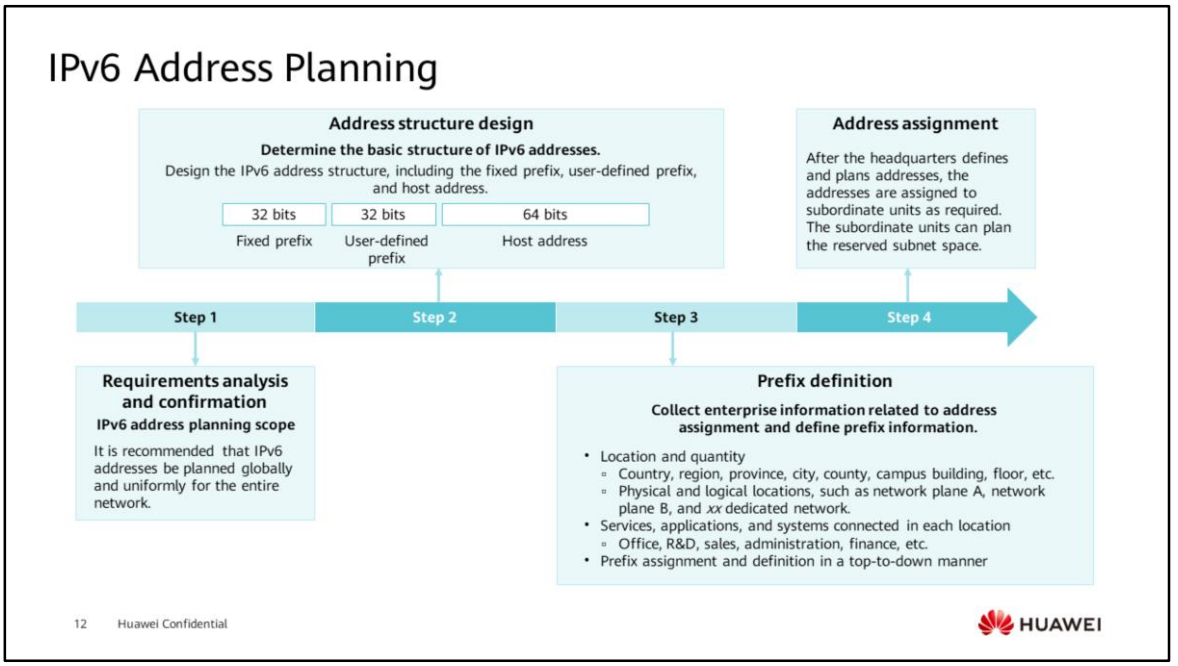


PI address space

- The PI address space is a block of IP addresses assigned by a regional Internet registry directly to end users. Users who obtain such addresses must contact an ISP to obtain the Internet access service and advertise the routes destined for related address segments on the Internet.
- PI addresses allow users to connect to multiple ISPs and plan addresses flexibly. However, the cost is high. This mode applies to medium- and large-sized enterprises.



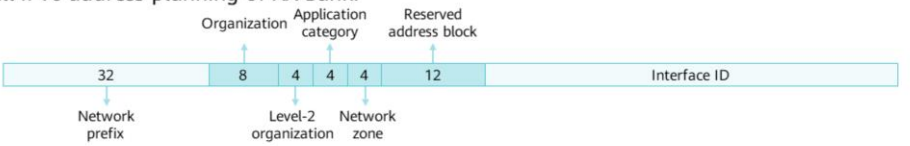
HUAWEI



- General rules for IPv6 address planning:
  - **Hierarchical design:** The large address space of IPv6 imposes higher requirements on route summarization. Specifically, network address fragments need to be reduced and route summarization needs to be enhanced to improve routing efficiency. The hierarchical design helps reduce the size of routing tables, ease IP address expansion and configuration, and facilitate fault locating, management, and capacity planning. In the hierarchical design, an IPv6 address is divided into several independent fields. Each field can be planned individually to implement route summarization. For example, a user-defined prefix can be divided into multiple fields, with each field representing a different meaning.
  - **Semantics:** Different fields are defined and assigned with meanings such as service type and physical location, facilitating O&M and fault locating.
  - **Security:** Services with shared attributes have the same security requirements. Mutual access between services needs to be controlled. Services with shared attributes are assigned addresses in the same address space, facilitating security design and policy management.
  - **Continuity:** It is recommended that IPv6 addresses in or between address segments be contiguous. This avoids wasting addresses and facilitates address management and summarization. This also eases route summarization, reduces the size of routing tables, and improves routing efficiency.
  - **Extensibility:** Some IP addresses should be reserved for future network expansion.

Example of IPv6 Address Planning

- Overall IPv6 address planning of *XX* Bank:



- In this example, the IPv6 prefix assigned to the bank is XXXX:XXXX::/32.

Field	Level-2 Organization
0000	HQ
0001	Branch A
0010	
0011	
...	...

Field	Application Category
0000	Production
0001	Office
0010	Development and test
0011	Out-of-band management of network devices
0100	Network device interconnection
0101	Video surveillance
...	...

Application Category	Field	Network Zone
Production	0000	Production terminal
	0001	External DMZ
	0010	E-banking web zone
	0011	E-banking app zone
	0100	Production service area
	0101	Management service zone
	0110	Core service zone
	...	...

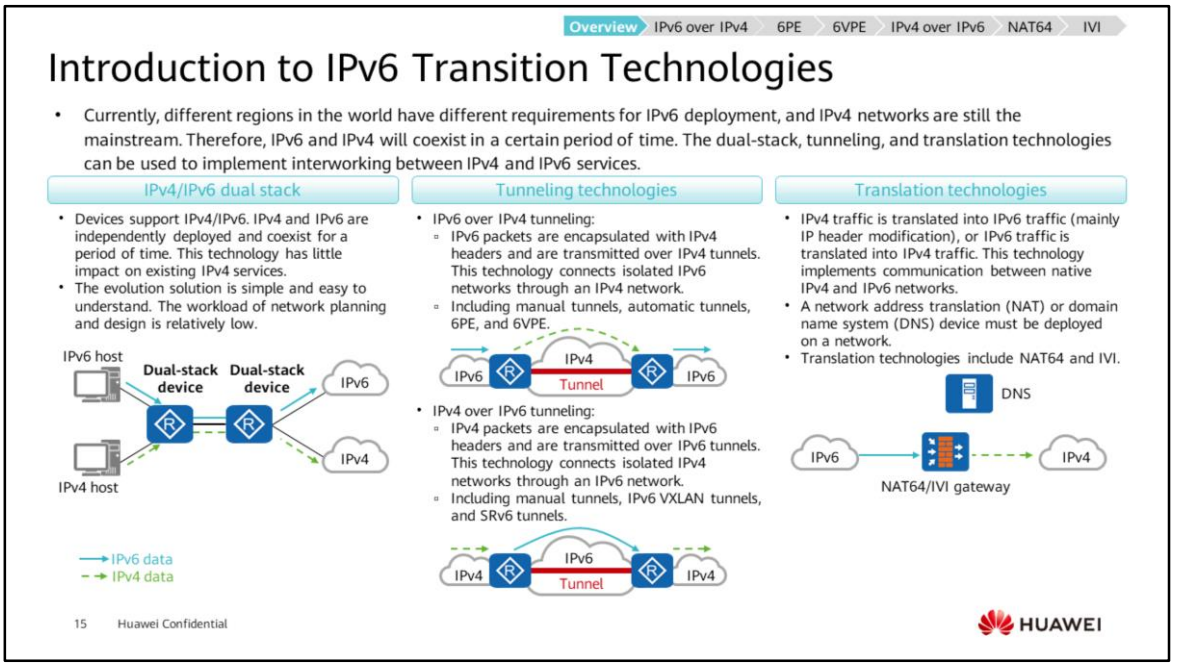
- HQ - production - production service zone: XXXX:XXXX:0000:4000::/64
- HQ - production - core service zone: XXXX:XXXX:0000:6000::/64

HUAWEI

Contents

1. IPv6 Basics
2. **IPv6 Transition Technologies**

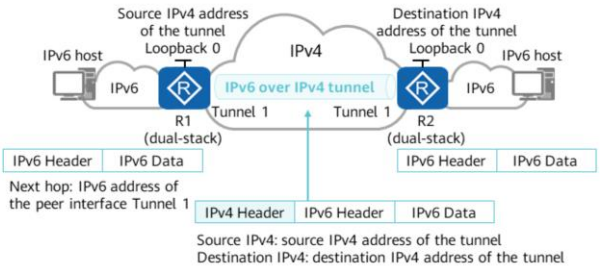
HUAWEI



- IPv4 address exhaustion fuels the urgency to transition to IPv6, but this requires existing IPv4 devices to be replaced as these devices are incompatible with IPv6 networks. The main issue is that replacing a large number of IPv4 devices will incur huge costs and interrupt services on the live networks. Therefore, the transition from IPv4 to IPv6 must be a gradual process. During the early stage of IPv4-to-IPv6 transition, IPv6 networks are scattered across a large number of IPv4 networks. Therefore, IPv6 transition technologies are required to implement IPv6 service interworking.
- Note: The VXLAN and SRv6 technologies will be described in detail in subsequent courses.

Tunneling Technology: IPv6 over IPv4

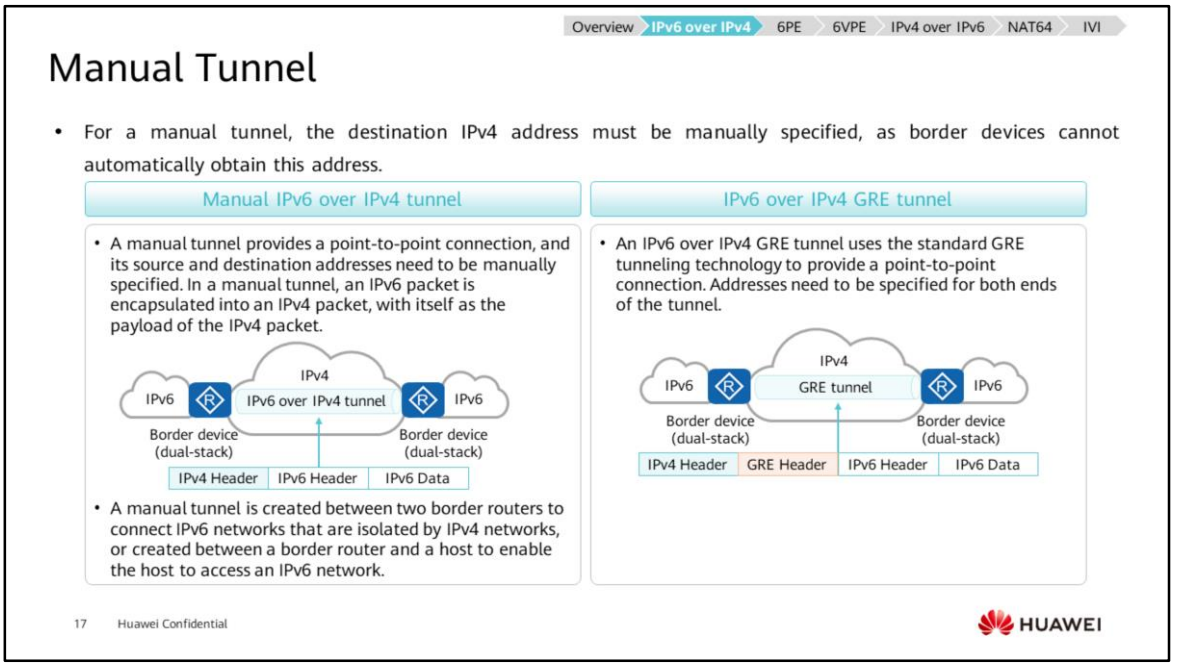
- IPv6 over IPv4 tunneling enables interconnection between isolated IPv6 networks over an IPv4 network.
- Fundamentals of IPv6 over IPv4 tunneling:



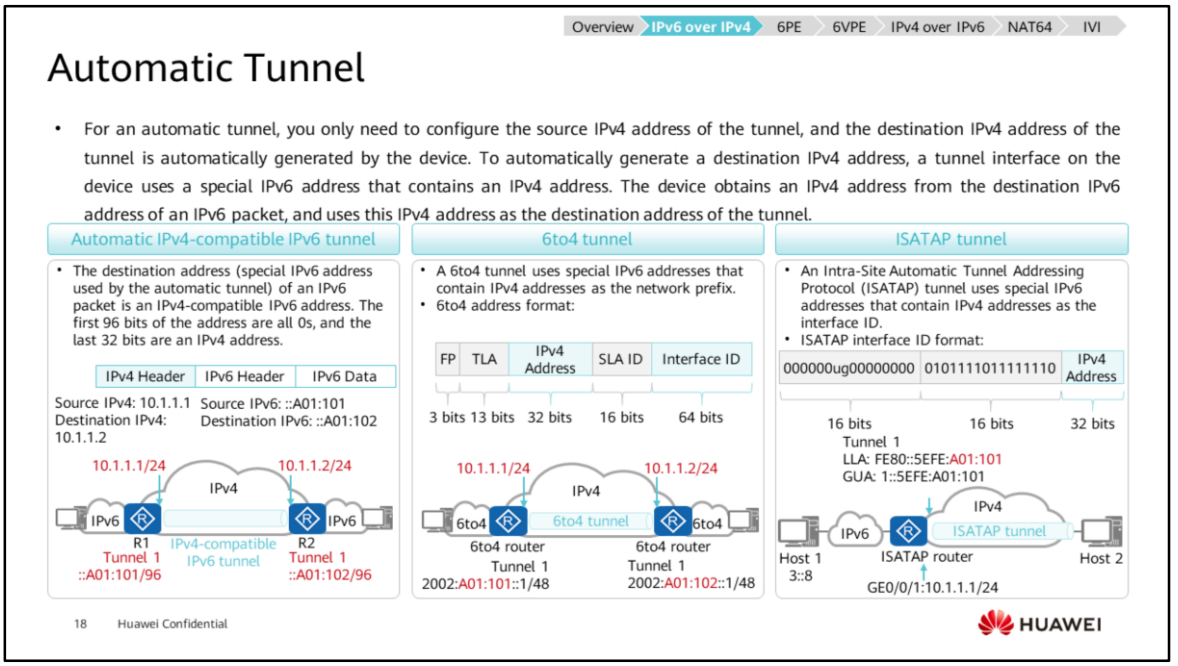
1. IPv4/IPv6 dual stack is enabled on border devices (R1 and R2), and an IPv6 over IPv4 tunnel is configured.
2. When receiving an IPv6 packet from the connected IPv6 network, R1 encapsulates it into an IPv4 packet by attaching an IPv4 header to the IPv6 packet. This is done only if the IPv6 packet is not destined for R1 and has a tunnel interface as the next hop.
3. R1 transmits the resulting IPv4 packet to R2 over the IPv4 network.
4. R2 decapsulates the IPv4 packet by removing the IPv4 header, and sends the resulting IPv6 packet to the destination IPv6 network.

- Classification of IPv6 over IPv4 tunnels:
  - For an IPv6 over IPv4 tunnel, the source IPv4 address must be manually specified, and the destination IPv4 address can be manually or automatically determined. From this perspective, IPv6 over IPv4 tunnels are classified into manual tunnels and automatic tunnels.

HUAWEI



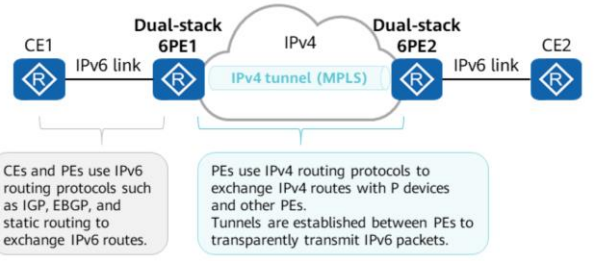
- Manual IPv6 over IPv4 tunnel:
  - Devices at both ends of a tunnel must support the IPv4/IPv6 dual stack. Other devices only need to support a single protocol stack.
  - The source and destination addresses of a manual tunnel are manually configured on devices. If a border device needs to establish tunnels with several devices, manually configuring multiple tunnels on the device is complex. Therefore, manual tunnels apply when two border routers connect only two IPv6 networks.
  - Tunnel forwarding mechanism: When receiving a packet from an IPv6 network, a border device searches its routing table for an entry matching the destination address of the IPv6 packet. If the outbound interface is a virtual tunnel interface, the border device encapsulates the IPv6 packet into an IPv4 packet based on the source and destination IPv4 addresses configured on the interface. The resulting IPv4 packet is then forwarded to the remote end of the tunnel over an IPv4 network. When receiving this IPv4 packet, the remote end decapsulates the packet to obtain the original IPv6 packet and processes it using the IPv6 protocol stack.
- IPv6 over IPv4 GRE tunnel:
  - Tunnel forwarding mechanism: Same as that of the manual IPv6 over IPv4 tunnel.



- The data forwarding process of an automatic IPv4-compatible IPv6 tunnel is as follows:
  - When receiving an IPv6 packet, R1 searches for an IPv6 route destined for ::A01:102 and finds that the next hop of the route is a virtual tunnel interface.
  - R1 encapsulates the IPv6 packet into an IPv4 packet. The source address of the IPv4 packet is the source IPv4 address 10.1.1.1 of the tunnel, and the destination IPv4 address is the last 32 bits (10.1.1.2) of the IPv4-compatible IPv6 address ::A01:102.
  - R1 sends the resulting IPv4 packet out from its tunnel interface. Then, the packet is routed to the destination node R2 at 10.1.1.2 over the IPv4 network. When receiving this packet, R2 decapsulates the packet to obtain the original IPv6 packet, and processes the IPv6 packet using the IPv6 protocol stack.
  - The response packet returned by R2 is processed in a similar way as the IPv6 packet sent by R1.
- 6to4 tunnel:
  - The network prefix of a 6to4 address is 64 bits long.
    - The first 48 bits (2002: a.b.c.d) are determined by the IPv4 address assigned to a router and cannot be changed.
    - The last 16 digits (SLA) are defined by users.

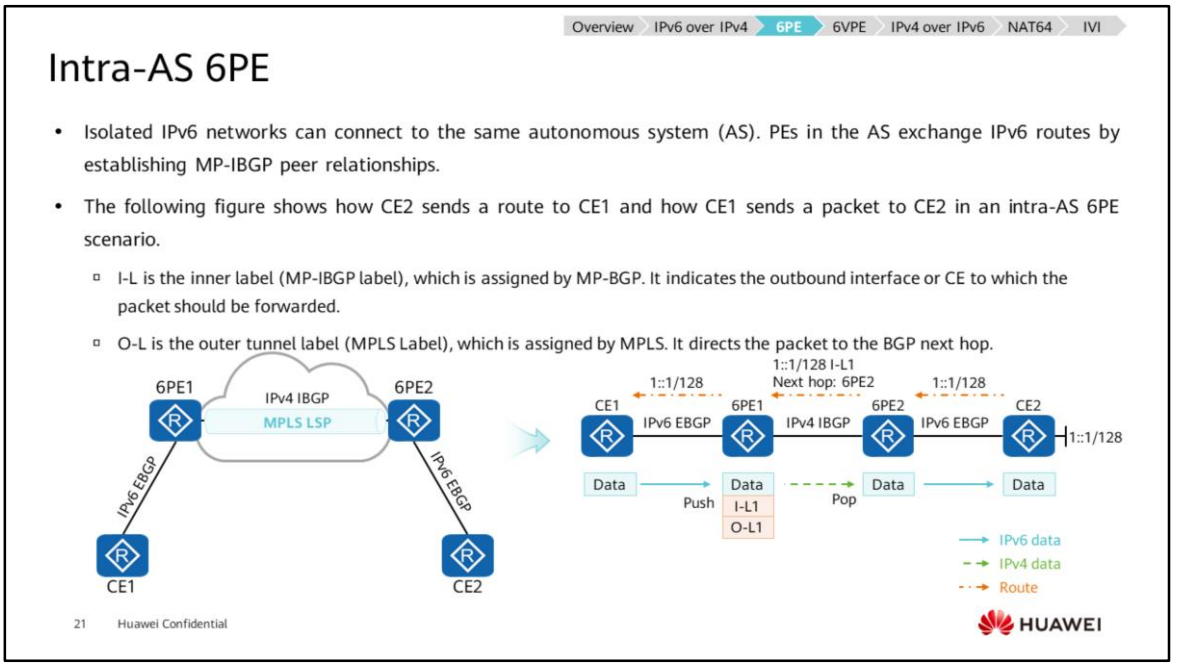
Tunneling Technology: 6PE

- To provide IPv6 services for users, carriers do not need to build new IPv6 backbone networks; instead, they can use existing IPv4 networks to connect isolated IPv6 sites. IPv6 Provider Edge (6PE) is designed based on this concept.
- The 6PE solution connects isolated IPv6 networks using multiprotocol label switching (MPLS) tunnels over IPv4 networks. 6PE implements IPv4/IPv6 dual stack on provider edges (PEs) of Internet service providers (ISPs), and uses the Multiprotocol Extensions for Border Gateway Protocol (MP-BGP) to assign labels to IPv6 routes. In this way, isolated IPv6 networks are connected through IPv4 tunnels between PEs.



HUAWEI

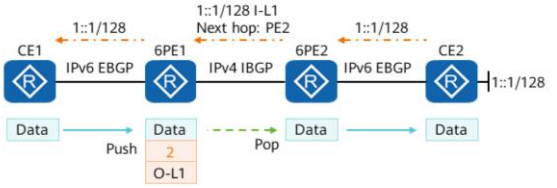
- Advantages of 6PE:
  - **Easy maintenance:** All configurations are performed on PEs. IPv6 services are carried over the existing IPv4 networks, simplifying network maintenance. Additionally, users on IPv6 networks are unaware of the IPv4 networks.
  - **Low network construction costs:** Carriers can provide IPv6 services to users over existing MPLS networks without upgrading the networks. 6PE devices can also provide various types of services, such as IPv6 VPN and IPv4 VPN.



- The route transmission process in an intra-AS 6PE scenario is as follows:
  1. CE2 sends an intra-AS IPv6 route to its EBGP peer 6PE2.
  2. Upon receipt, 6PE2 changes the next hop of the IPv6 route to itself, and assigns an inner label to the IPv6 route. Then, 6PE2 sends the labeled IPv6 route to its IBGP peer 6PE1.
  3. When receiving the labeled IPv6 route, 6PE1 recurses the route to a tunnel, and adds the route to the local forwarding table. Then, 6PE1 changes the next hop of the IPv6 route to itself, removes the label from the route, and sends the route to its EBGP peer CE1.
- The packet transmission process in an intra-AS 6PE scenario is as follows:
  1. CE1 sends an ordinary IPv6 packet to 6PE1 over an IPv6 link on the public network.
  2. Upon receipt of the IPv6 packet, 6PE1 looks up the destination address of the packet in its forwarding table, and encapsulates the packet with inner and outer labels. Then, 6PE1 sends the resulting IPv6 packet to 6PE2 over a public network tunnel.
  3. When receiving the IPv6 packet, 6PE2 removes the inner and outer labels and forwards the resulting IPv6 packet to CE2 based on the destination address over an IPv6 link.

Explicit Null Label Shared by 6PE Routes

- In a 6PE networking, by default, each 6PE route to be sent to 6PE peers is assigned a label. This will cause a large number of label resources to be consumed if many 6PE routes need to be sent.
- After 6PE routes are configured to share the same explicit null label, all 6PE routes to be sent to 6PE peers share the explicit null label 2. The number of required labels is irrelevant to the number of 6PE routes to be sent. This greatly saves label resources on 6PE devices.
- Explicit null label 2 is a special label that must be popped out on the egress PE, and packets must be forwarded based on IPv6.

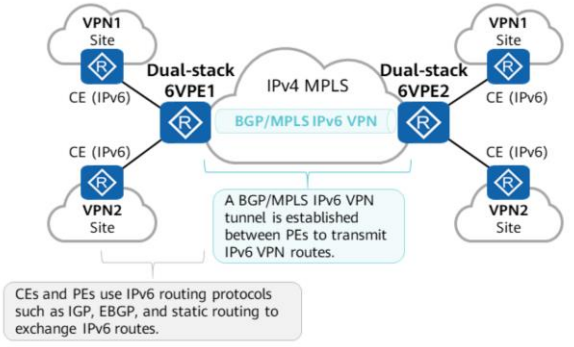


HUAWEI

- When 6PE routes are configured to share the same explicit null label on 6PE2, 6PE2 advertises 6PE routes with an explicit null label to 6PE1 without applying for labels for the routes.
- When forwarding data to 6PE2, 6PE1 adds two labels to the data. The outer label is distributed by LDP pointing to 6PE2, and the inner label is an explicit null label distributed by MP-BGP.
- When the IPv6 data packet arrives at 6PE2, 6PE2 pops out the explicit null label and forwards the packet to CE2.

Tunneling Technology: 6VPE

- All IPv6 services connected through 6PE run in the same VPN and cannot be logically isolated. Therefore, 6PE can be used only for open and unprotected IPv6 network interconnection. To isolate the connected IPv6 services logically, that is, to implement IPv6 VPNs, use the IPv6 VPN Provider Edge (6VPE) technology.
- 6VPE is an extension of BGP/MPLS IPv6 VPN. It transmits IPv6 VPN services over an IPv4 MPLS backbone network.



- 6VPE uses MP-BGP to advertise VPNv6 routes on an IPv4 MPLS backbone network, uses MPLS to assign labels to identify IPv6 packets, and uses tunneling mechanisms such as LSP and MPLS TE to transmit private network data on the backbone network.
- If the backbone network is an IPv4 network, IPv4 addresses are used to establish VPNv6 peer relationships between PEs to transmit IPv6 VPN routes. For these routes, IPv4 tunnels on the backbone network can be selected to transmit IPv6 VPN services.
- Except the routing protocols running between PEs and CEs, implementation of 6VPE is the same as that of IPv4 VPN.

HUAWEI

- In 6VPE, IPv6 routing protocols run between PEs and CEs. The following IPv6 routing protocols can be used to provide IPv6 VPN services:
  - BGP4+
  - Static IPv6 routing
  - OSPFv3
  - IS-IS for IPv6

Configuring 6VPE (1)

1. Configure a VPN instance.

```
[HUAWEI] ip vpn-instance vpn-instance-name
```

```
[HUAWEI-vpn-instance-vpna] ipv6-family
[HUAWEI-vpn-instance-vpna-af-ipv6] route-distinguisher route-distinguisher
[HUAWEI-vpn-instance-vpna-af-ipv6] vpn-target vpn-target [ both | export-extcommunity | import-extcommunity ]
```

A route distinguisher (RD) and VPN target extended community are configured for the VPN instance IPv6 address family.

2. Bind the VPN instance to an interface.

```
[HUAWEI-GigabitEthernet0/0/1] ip binding vpn-instance vpn-instance-name
[HUAWEI-GigabitEthernet0/0/1] ipv6 enable
[HUAWEI-GigabitEthernet0/0/1] ipv6 address { ipv6-address prefix-length | ipv6-address|prefix-length }
```


HUAWEI

- Note: This course uses a Huawei NetEngine 8000 series router as an example to describe how to configure 6VPE.
- Command: <HUAWEI>**system-view [immediately]**
  - **immediately:** indicates that the configuration takes effect immediately.

Configuring 6VPE (2)

3. Configure MP-IBGP between PEs to transmit VPNv6 routes.

```
[HUAWEI] bgp as-number
[HUAWEI-bgp] peer ipv4-address as-number as-number
[HUAWEI-bgp] peer ipv4-address connect-interface interface-number
```

The remote PE is configured as an IBGP peer in the BGP view.

```
[HUAWEI-bgp] ipv6-family vpnv6
[HUAWEI-bgp-af-ipv6] peer ipv4-address enable
```

The BGP VPNv6 address family view is enabled and displayed, and the ability to exchange VPNv6 routes with peers is enabled in this view.

HUAWEI

Configuring 6VPE (3)

4. Configure route exchange between PEs and CEs. (A static route is used as an example. The configuration on CEs is the same as that of an ordinary IPv6 static route, and is not mentioned here.)

[PE] **ipv6 route-static vpn-instance** *vpn-instance-name dest-ipv6-address prefix-length* { *interface-type interface-number* [*nexthop-ipv6-address*] | **vpn-instance** *vpn-destination-name nexthop-ipv6-address* | *nexthop-ipv6-address* } [**preference** *preference*]

An IPv6 static route is configured for the VPN instance enabled with the IPv6 address family.

[PE-bgp] **ipv6-family vpn-instance** *vpn-instance-name*
[PE-bgp-bgp6-vpnb] **import-route static** [**med** *med* | **route-policy** *route-policy-name*]

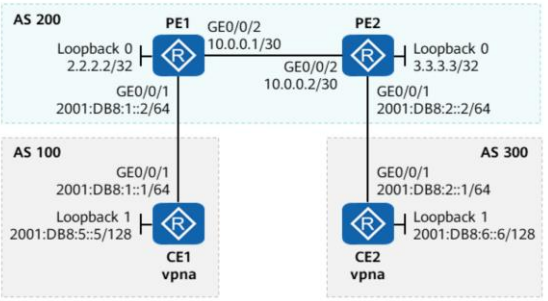
The BGP-VPN instance IPv6 address family view is displayed, and the configured static route is imported to the routing table of the BGP-VPN instance IPv6 address family.

- **med** *med*: specifies the MED of the imported route. The value is an integer in the range from 0 to 4294967295.
- **route-policy** *route-policy-name*: specifies a route-policy for filtering routes and modifying route attributes when routes of other routing protocols are imported.

HUAWEI

- Note: For details about route exchange between PEs and CEs, see the corresponding product documentation.

Example of Configuring 6VPE (1)



- Configuration requirements:
  - Users at different sites desire IPv6 data communication between each other across a public network without having the internal route information known to the public network, and also implement service isolation.
  - As shown in the figure, CE1 and CE2 belong to **vpna**. It is required that 6VPE be configured to allow sites in **vpna** to communicate with each other through an MPLS backbone network, and PEs and CEs exchange static routes.

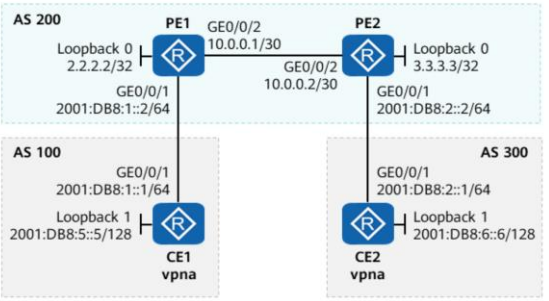
1. Configure IPv4 and IPv6 addresses for interfaces. (The configuration details are not provided here.)
2. Configure IS-IS on PE1 and PE2 so that they can learn routes to each other's interface (loopback 0). (The configuration details are not provided here.)
3. Enable MPLS and MPLS LDP on the backbone network, and enable PEs to establish an LDP LSP.

```
[PE1] mpls lsr-id 2.2.2.2
[PE1] mpls
[PE1-mpls] quit
[PE1] mpls ldp
[PE1-mpls-ldp] quit
[PE1] interface GigabitEthernet0/0/2
[PE1-GigabitEthernet2/0/0] mpls
[PE1-GigabitEthernet2/0/0] mpls ldp
```

The configuration on PE2 is similar to that on PE1, and is not provided here.

HUAWEI

Example of Configuring 6VPE (2)



- Configuration requirements:
  - Users at different sites desire IPv6 data communication between each other across a public network without having the internal route information known to the public network, and also implement service isolation.
  - As shown in the figure, CE1 and CE2 belong to **vpna**. It is required that 6VPE be configured to allow sites in **vpna** to communicate with each other through an MPLS backbone network, and PEs and CEs exchange static routes.

4. Create a VPN instance supporting the IPv6 address family on each PE, and bind the VPN instance to the PE's interface connected to a CE.

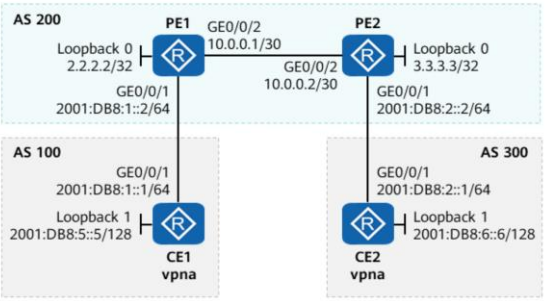
```
[PE1] ip vpn-instance vpna
[PE1-vpn-instance-vpna] ipv6-family
[PE1-vpn-instance-vpna-af-ipv6] route-distinguisher 100:1
[PE1-vpn-instance-vpna-af-ipv6] vpn-target 22:22 both
```

```
[PE1] interface GigabitEthernet0/0/1
[PE1-GigabitEthernet0/0/1] ip binding vpn-instance vpna
[PE1-GigabitEthernet0/0/1] ipv6 enable
[PE1-GigabitEthernet0/0/1] ipv6 address 2001:DB8:1::2 64
```

The configuration on PE2 is similar to that on PE1, and is not provided here.

HUAWEI

Example of Configuring 6VPE (3)



- Configuration requirements:
  - Users at different sites desire IPv6 data communication between each other across a public network without having the internal route information known to the public network, and also implement service isolation.
  - As shown in the figure, CE1 and CE2 belong to **vpna**. It is required that 6VPE be configured to allow sites in **vpna** to communicate with each other through an MPLS backbone network, and PEs and CEs exchange static routes.

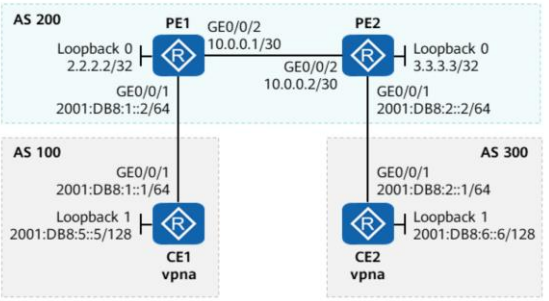
5. Enable PEs to establish a VPNv6 peer relationship.

```
[PE1] bgp 200
[PE1-bgp] peer 3.3.3.3 as-number 200
[PE1-bgp] peer 3.3.3.3 connect-interface loopback 0
[PE1-bgp] ipv6-family vpnv6
[PE1-bgp-af-vpnv6] peer 3.3.3.3 enable
```

```
[PE2] bgp 200
[PE2-bgp] peer 2.2.2.2 as-number 200
[PE2-bgp] peer 2.2.2.2 connect-interface loopback 0
[PE2-bgp] ipv6-family vpnv6
[PE2-bgp-af-vpnv6] peer 2.2.2.2 enable
```


HUAWEI

Example of Configuring 6VPE (4)



- Configuration requirements:
  - Users at different sites desire IPv6 data communication between each other across a public network without having the internal route information known to the public network, and also implement service isolation.
  - As shown in the figure, CE1 and CE2 belong to **vpna**. It is required that 6VPE be configured to allow sites in **vpna** to communicate with each other through an MPLS backbone network, and PEs and CEs exchange static routes.

6. Configure static routes on PEs and CEs, and import the static routes to the routing table of the BGP-VPN instance IPv6 address family on each PE.

```
[PE1] ipv6 route-static vpn-instance vpna 2001:DB8:5:: 64
2001:DB8:1::1
[PE1] bgp 200
[PE1-bgp] ipv6-family vpn-instance vpna
[PE1-bgp6-vpna] import-route static
```

```
[CE1] ipv6 route-static :: 0 2001:DB8:1::2
```

```
[PE2] ipv6 route-static vpn-instance vpna 2001:DB8:6:: 64
2001:DB8:2::1
[PE2] bgp 200
[PE2-bgp] ipv6-family vpn-instance vpna
[PE2-bgp6-vpna] import-route static
```

```
[CE2] ipv6 route-static :: 0 2001:DB8:2::2
```


HUAWEI

Verifying the Configuration

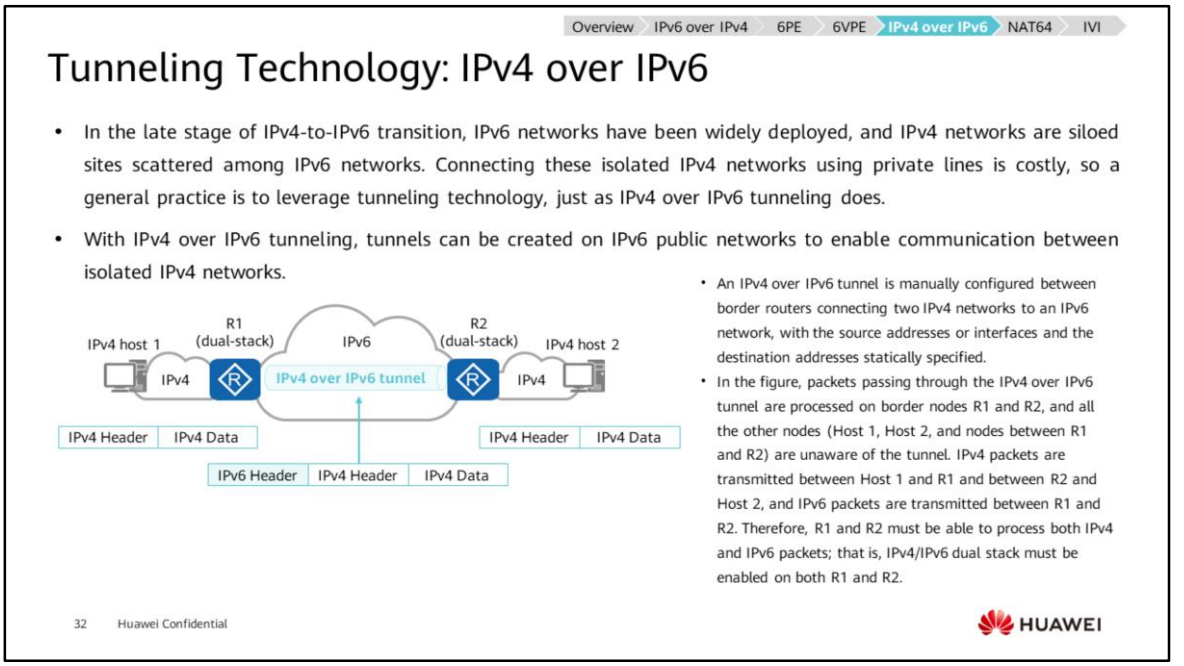
1. On PEs, check whether a VPNv6 peer relationship is established.

```
[PE1]display bgp vpnv6 all peer
 BGP local router ID : 10.0.0.1
 Local AS number : 200
 Total number of peers : 1                 Peers in established state : 1

  Peer            V    AS  MsgRcvd    MsgSent    OutQ    Up/Down       State       PrefRcv
  3.3.3.3         4   200       19         19       0     00:12:25      Established   1
```

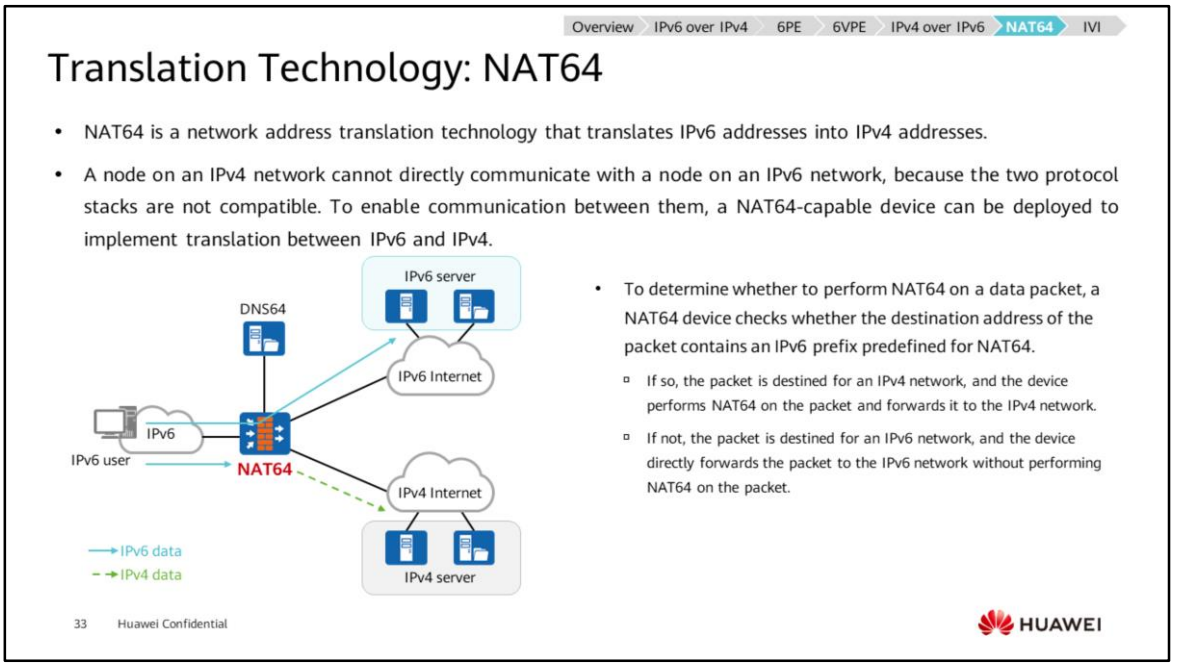
2. Check whether CE1 can ping the loopback 1 interface of CE2.

```
[CE1]ping ipv6 -a 2001:db8:5::5 2001:db8:6::6
  PING 2001:db8:6::6 : 56  data bytes, press CTRL_C to break
    Request time out
    Reply from 2001:DB8:6::6
    bytes=56 Sequence=2  hop limit=62  time = 50 ms
    Reply from 2001:DB8:6::6
    bytes=56 Sequence=3  hop limit=62  time = 40 ms
    Reply from 2001:DB8:6::6
    bytes=56 Sequence=4  hop limit=62  time = 30 ms
    Reply from 2001:DB8:6::6
    bytes=56 Sequence=5  hop limit=62  time = 30 ms

  --- 2001:db8:6::6 ping statistics ---
    5 packet(s) transmitted
    4 packet(s) received
    20.00% packet loss
    round-trip min/avg/max = 30/37/50 ms
```

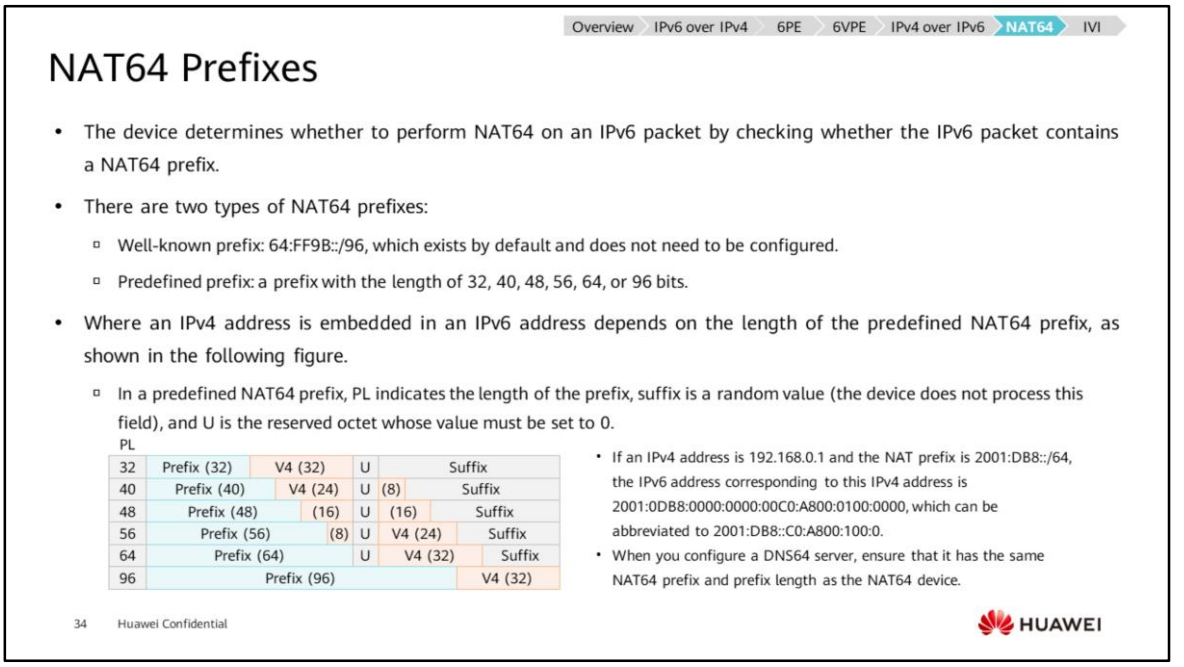

HUAWEI



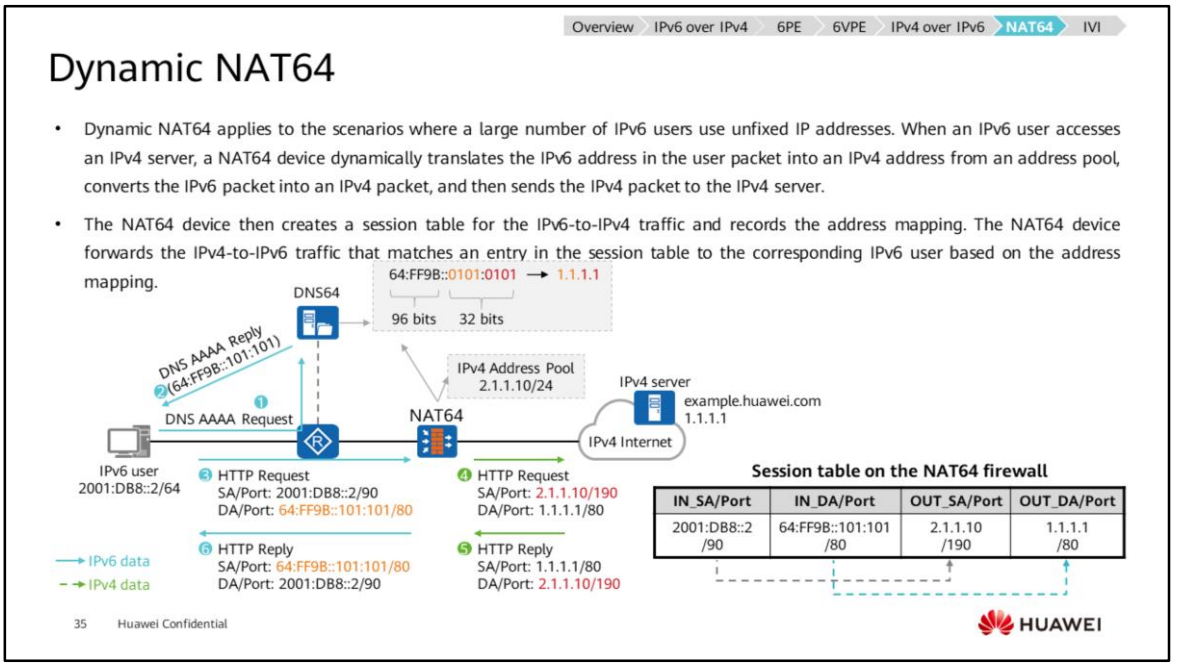
- Packet processing on an IPv4 over IPv6 tunnel:
    1. IPv4 packet forwarding: Host 1 sends an IPv4 packet destined for Host 2 to R1.
    2. Tunnel encapsulation: When receiving the IPv4 packet from Host 1, R1 discovers that the destination address of the IPv4 packet is not its own and the next hop is a tunnel interface, so R1 attaches an IPv6 header to the IPv4 packet. Specifically, in the IPv6 header, R1 encapsulates both its own and R2's IPv6 addresses into the Source Address and Destination Address fields, respectively, sets the Version field to 6 and the Next Header field to 4, and encapsulates the fields that ensure effective transmission of the packet along the tunnel based on the configuration.
    3. Tunnel forwarding: R1 searches its IPv6 routing table for an entry matching the destination address in the IPv6 header, and forwards the IPv6 packet to R2. Other nodes on the IPv6 network are unaware of the tunnel and process the IPv6 packet as an ordinary IPv6 packet.
    4. Tunnel decapsulation: Upon receipt of the IPv6 packet, R2 discovers that the destination address is its own IPv6 address and the encapsulated packet is an IPv4 packet based on the Next Header field, and then decapsulates the IPv6 packet by removing the IPv6 header based on the Version field.
    5. IPv4 packet forwarding: R2 searches its IPv4 routing table for an entry matching the destination address of the IPv4 packet and forwards the packet to Host 2.



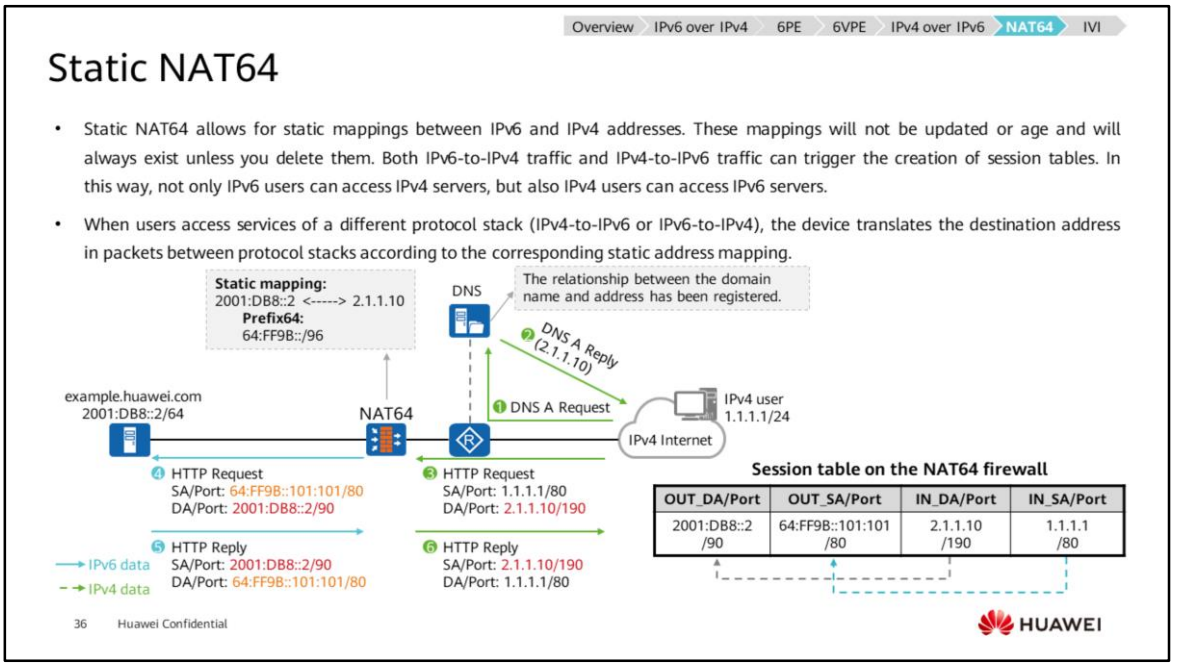
- Scenario where IPv6 users access IPv4 servers:
  - At the early stage of IPv4-to-IPv6 evolution, carriers provide users with single-stack IPv6 access. However, lots of Internet servers still use IPv4 and do not support dual stack. This requires carriers to provide NAT64 devices for single-stack IPv6 users to access IPv4 servers.
  - The carriers only need to assign IPv6 addresses to users, with communication between IPv6 devices implemented based on IPv6 routes. Access from IPv6 users to IPv4 services is implemented through NAT64 devices that support dual stack.
  - There are two NAT64 modes: static and dynamic. Static NAT64 is recommended when a small number of IPv6 users use fixed IP addresses, while dynamic NAT64 is recommended when a large number of IPv6 users use unfixed IP addresses.
- Scenario where IPv4 users access IPv6 servers:
  - At the late stage of IPv4-to-IPv6 evolution, many service providers have begun to provide IPv6 services. However, there are still many single-stack IPv4 users on the Internet. To retain these users, the service providers deploy NAT64 devices, which allow these users to access IPv6 servers.
  - In this scenario, service providers only need to assign IPv6 addresses to servers, without the need to upgrade access devices. NAT64 devices, which must support dual stack, are deployed at the egresses, and communication between IPv6 devices is implemented based on IPv6 routes. IPv4-to-IPv6 access is implemented using the static NAT64 technology.



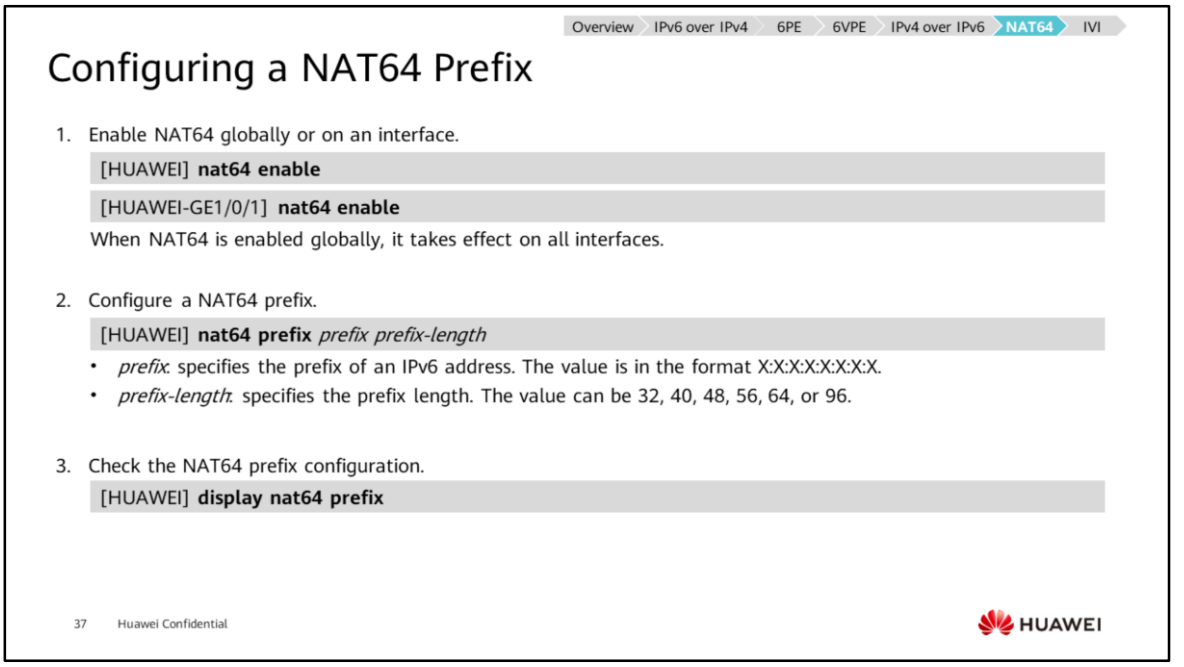
- NAT64 types:
  - PAT-based NAT64: translates both addresses and port numbers by mapping **[IPv6 address, port number]** into **[IPv4 address, port number]**. Multiple IPv6 addresses can be translated into the same IPv4 address. The mappings are differentiated by port number. This mode is commonly used.
  - No-PAT–based NAT64: translates only addresses by mapping **[IPv6 address]** into **[IPv4 address]**. There are one-to-one mapping relationships between IPv6 and IPv4 addresses.



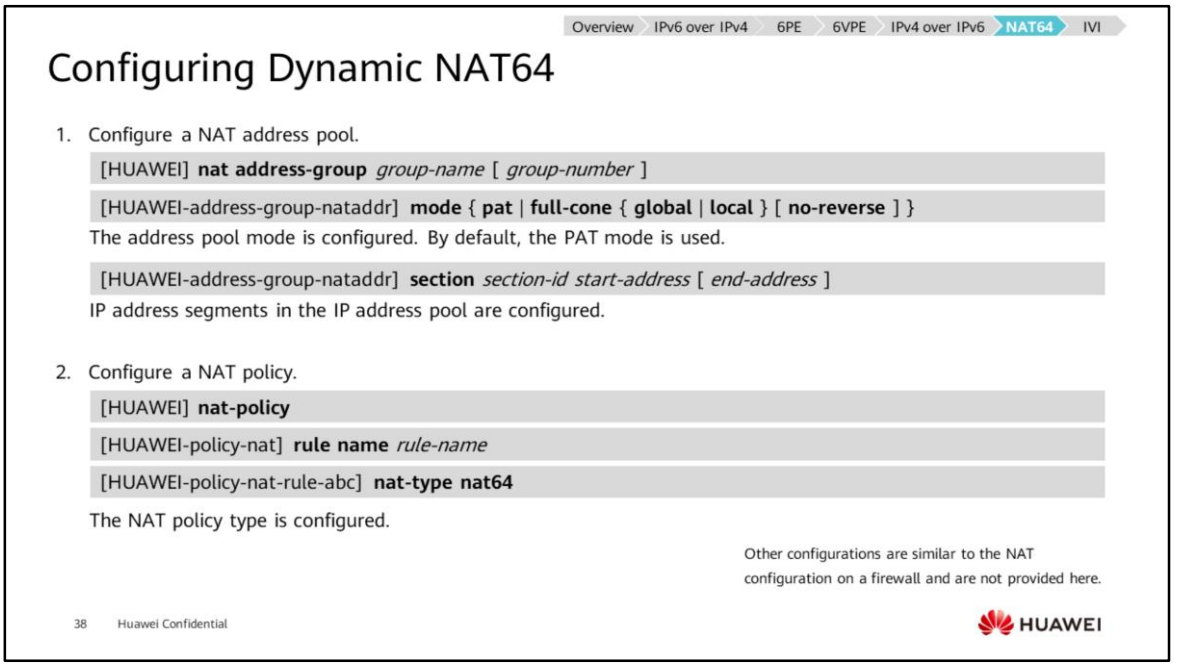
- Dynamic NAT64 mapping process:
  1. A single-stack IPv6 user sends an AAAA DNS request for a remote service (example.huawei.com).
  2. After receiving the request, the DNS64 server parses the AAAA request. If no IPv6 address is found, it sends an A request, and parses the received reply packet to obtain the service IPv4 address (1.1.1.1). The DNS64 server then combines the IPv4 address and the preconfigured NAT64 prefix 64::/n (64:FF9B::/96) into the IPv6 address 64:FF9B::101:101, and sends this IPv6 address to the user.
  3. When receiving the DNS64 reply from the DNS64 server, the user sends a packet with the IPv6 address as the destination address to the remote server.
  4. After receiving the IPv6 packet from the user, the NAT64 device extracts the IPv4 address (1.1.1.1) from the IPv6 packet using the address translation algorithm, changes the destination address of the packet to this IPv4 address, changes the source address of the packet to an IPv4 address (2.1.1.10) from the NAT address pool based on the address mapping configured in the NAT64 policy, and converts the IPv6 packet into an IPv4 packet. The NAT64 device then sends the IPv4 packet to the remote server on the IPv4 network, and generates a session table with the address mapping.
  5. Upon receipt of the IPv4 packet, the IPv4 server returns a reply packet.
  6. After receiving the reply packet from the IPv4 server, the NAT64 device converts the IPv4 packet into an IPv6 packet according to the session table, and sends the IPv6 packet to the IPv6 user.
- DNS AAAA and A records:
  - AAAA record: maps a host name or domain name to an IPv6 address.
  - A record: maps a domain name to an IPv4 address.



- Static NAT64 mapping process:
    1. An IPv4 user sends request A (example.huawei.com) to the DNS server.
    2. After receiving the request, the DNS server parses the request to obtain the IPv4 address (2.1.1.10) corresponding to the domain name, and sends a reply containing the IPv4 address to the user. In this scenario, the mapping between the domain name and IPv4 address has been predefined on the DNS server. If the DNS A request does not contain any IPv4 address, the packet is discarded.
    3. After receiving the DNS reply, the user sends a packet with the obtained IPv4 address as the destination address to the remote server.
    4. Upon receipt of the IPv4 packet, the NAT64 device translates the destination IPv4 address into an IPv6 address (2001:DB8::2) according to the preconfigured static address mapping (based on which a server mapping table is generated), combines the source IPv4 address and the preconfigured NAT64 prefix into a source IPv6 address (64:FF9B::101:101), and converts the IPv4 packet into an IPv6 packet. The NAT64 device then sends the IPv6 packet to the remote server on the IPv6 network, and generates a session table.
    5. Upon receipt of the IPv6 packet, the server returns a reply packet.
    6. After receiving the reply packet from the IPv6 server, the NAT64 device converts the IPv6 packet into an IPv4 packet according to the session table, and sends the IPv4 packet to the IPv4 user.



- NAT64 prefix configuration:
  - A NAT64 prefix must be different from the IPv6 address prefix of any interface on the device. Otherwise, the device considers the packets whose destination IPv6 addresses are on the same network segment as the interfaces to be NAT64 packets, and starts NAT64 processing for these packets.
  - When multiple NAT64 prefixes are configured and dynamic NAT64 is used, all of these NAT64 prefixes can be used for NAT64 translation of IPv6-to-IPv4 traffic. On the other hand, if static NAT64 is used, the device randomly selects one of these prefixes.
- Note: This course uses a Huawei USG6000 series firewall as an example to describe how to configure NAT64.



- Command: **mode { pat | full-cone { global | local } [no-reverse] }**
  - **pat:** specifies the PAT mode.
  - **full-cone:** specifies the 3-tuple NAT mode.
  - **global:** specifies the global 3-tuple NAT mode. The generated server mapping table does not contain security zone parameters and is not subject to restrictions of interzone relationships.
  - **local:** specifies the local 3-tuple NAT mode. The generated server mapping table contains security zone parameters and is subject to restrictions of interzone relationships.
  - **no-reverse:** specifies that no reverse server mapping table will be generated.

Configuring Static NAT64

1. Configure static NAT64.

[HUAWEI] **nat64 static** *ipv6-address ipv4-address* [**route**]

Static NAT64 mapping for a specified IPv6 address is configured.

[HUAWEI] **nat64 static protocol icmp** *ipv6-address ipv4-address* [**route**]

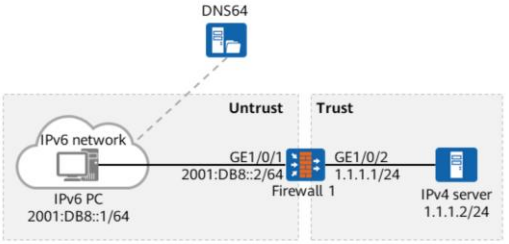
Static NAT64 mapping for ICMP is configured.

[HUAWEI] **nat64 static protocol** { **tcp** | **udp** } *ipv6-address* [*ipv6-port*] *ipv4-address* [*ipv4-port*] [**route**]

Static NAT64 mapping for TCP or UDP is configured.

HUAWEI

Example for Configuring Dynamic NAT64 (1)



- Configuration requirements:
  - Firewall 1 supports the IPv4/IPv6 dual stack and is connected to an IPv4 network and an IPv6 network.
  - The IPv6 network has a large number of PCs using unfixed IP addresses. These PCs need to access the server on the IPv4 network through the domain name example.huawei.com.

1. Configure IPv4 and IPv6 addresses for Firewall 1 and configure security zones. (The configuration details are not provided here.)
2. Configure a security policy.

```
[FW1] security-policy
[FW1-policy-security] rule name sec1
[FW1-policy-security-rule-sec1] source-zone untrust
[FW1-policy-security-rule-sec1] destination-zone trust
[FW1-policy-security-rule-sec1] source-address 2001:db8:: 64
[FW1-policy-security-rule-sec1] action permit
```

3. Enable NAT64 on GE1/0/1 of Firewall 1.

```
[FW1] interface GigabitEthernet 1/0/1
[FW1-GigabitEthernet1/0/1] nat64 enable
```

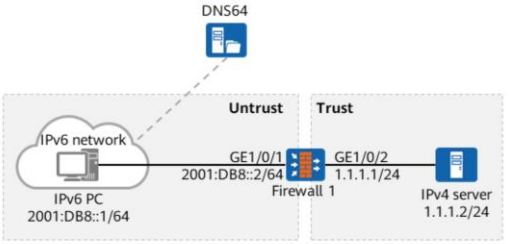
4. Set the NAT64 prefix to 2001:db8:1::/96 (a predefined prefix).

```
[FW1] nat64 prefix 2001:db8:1:: 96
```


HUAWEI

Overview | IPv6 over IPv4 | 6PE | 6VPE | IPv4 over IPv6 | NAT64 | IVI

Example for Configuring Dynamic NAT64 (2)



- Configuration requirements:
  - Firewall 1 supports the IPv4/IPv6 dual stack and is connected to an IPv4 network and an IPv6 network.
  - The IPv6 network has a large number of PCs using unfixed IP addresses. These PCs need to access the server on the IPv4 network through the domain name example.huawei.com.

5. Configure a NAT64 address pool.

```
[FW1] nat address-group pool1
[FW1-address-group-pool1] mode pat
[FW1-address-group-pool1] section 1 1.1.1.6 1.1.1.10
```

6. Configure dynamic NAT64.

```
[FW1] nat-policy
[FW1-policy-nat] rule name nat64
[FW1-policy-nat-rule-nat64] nat-type nat64
[FW1-policy-nat-rule-nat64] source-zone untrust
[FW1-policy-nat-rule-nat64] destination-zone trust
[FW1-policy-nat-rule-nat64] source-address 2001:db8:: 64
[FW1-policy-nat-rule-nat64] action source-nat address-
group pool1
```

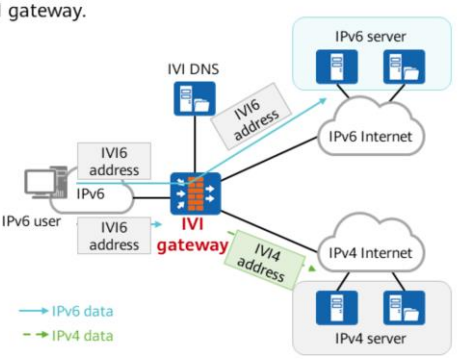

HUAWEI

- The destination address of IPv6 packets sent by the IPv6 PC is the NAT64 address 2001:DB8:1::101:102/96.
- Other configurations:
  - Configure the DNS64 server.
    - Set the IPv6 prefix of the DNS64 server to 2001:DB8:1::/96, which is the NAT64 prefix of Firewall 1.
    - Configure reachable routes from the DNS64 server to the PC and server.
  - Configure the IPv6 address, route, and DNS server for the PC. (The method of configuring an IPv6 address and a route for the PC varies according to the PC operating system. The detailed configuration procedure is not provided here.)
    - Set the PC IPv6 address to 2001:DB8::1/64, which is on the same network segment as GE1/0/1 of Firewall 1.
    - Configure a static route from the PC to the destination network segment 2001:db8:1::/96, with the next hop address set to 2001:db8::2.
    - Configure the DNS64 server IPv6 address on the PC.
  - Configure an IPv4 address for the server. (The configuration method varies according to the server operating system. The detailed configuration procedure is not provided here.)
    - Set the IPv4 address of the server to 1.1.1.2/24, which is on the same network segment as GE1/0/2 of Firewall 1.

Translation Technology: IVI

- The IVI stateless address translation technology translates an IPv4 address segment (IVI4 addresses) reserved by carriers into a special IPv6 address segment (IVI6 addresses embedded with IPv4 addresses).
- Users with IVI6 addresses can directly access the global IPv6 network, and access the global IPv4 network after their IVI6 addresses are translated into IVI4 addresses by an IVI gateway.
- IVI6 address format:

  - The first 32 bits are an IVI prefix, which is an IPv6 address prefix, and the last eight bits are fixed at FF.
  - The 41st to 72nd bits are an embedded IPv4 address.
  - The value of **Suffix** is all 0s.



HUAWEI

- IVI supports communication requests initiated by both IPv6 and IPv4 hosts.
- The following uses access from an IVI6 host to a global IPv4 host as an example:
  1. In this scenario, stateless IPv6 address autoconfiguration cannot be used due to the special IVI6 address format. Therefore, the IVI6 host obtains the IVI6 address, default gateway address, and DNS server address through static configuration or DHCPv6 Options.
  2. The IVI6 host sends an AAAA query request to the dual-stack IVI DNS server. This DNS server stores the IVI4 addresses of IVI servers and their corresponding IVI6 addresses. When receiving the AAAA query request, the IVI DNS server sends an AAAA query request to the target network. If no AAAA record exists, the IVI DNS server sends an A query request, converts the obtained A record into an AAAA record according to the IVI mapping rule, and returns the AAAA record to the IVI6 host.
  3. The IVI6 host sends a data packet. When receiving this data packet, the IVI gateway statelessly converts the packet into an IPv4 packet. During address translation, the IPv4 address embedded in the IVI6 address is extracted and used as the source address in the IPv4 header. During header encapsulation, the Stateless IP/ICMP Translation (SIIT) algorithm is used.
  4. The resulting IPv4 data packet is routed to the IPv4 network, thereby implementing access from the IVI6 host to the IPv4 host.
- IVI restrictions: The IPv6 addresses of hosts and servers must be planned and configured in compliance with the IVI format.

Quiz

1. (Multiple-answer question) Which of the following technologies can be used to carry IPv6 service traffic on an IPv4 WAN? ( )

   A. GRE tunneling

   B. 6PE

   C. NAT64

   D. 6VPE

HUAWEI

1. ABD

Summary

- IPv6 can provide a large number of network address resources, making it a key to implementing Internet of Everything (IoE) and promoting the digital, network-based, and intelligent development of production and life. The IPv6-based next-generation Internet is an inevitable trend of Internet evolution and upgrade.
- In the early stage of IPv4-to-IPv6 transition, IPv4 networks are widely deployed, whereas IPv6 networks are siloed networks scattered around the world. Tunneling technologies can be used to create tunnels over IPv4 networks, so as to connect these siloed IPv6 networks. Address translation technologies can also be used to translate between IPv6 and IPv4 addresses and protocols for bidirectional communication.
- This course mainly describes the fundamentals and applications of transition technologies during IPv6 network evolution, including IPv6 over IPv4 tunneling, 6PE, 6VPE, IPv4 over IPv6 tunneling, NAT64, and IVI.

HUAWEI

Thank you.

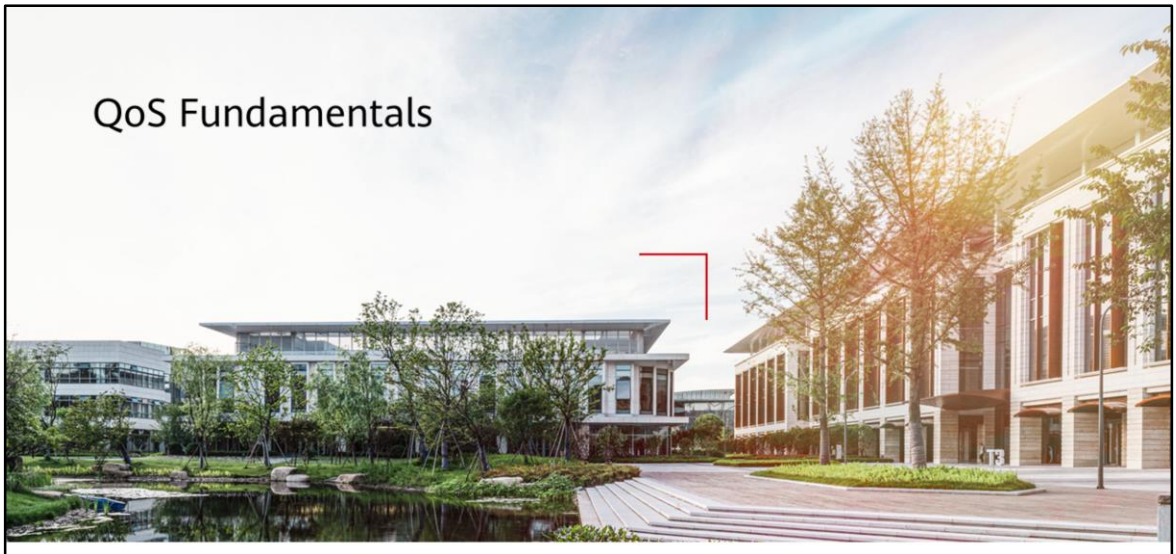
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

QoS Fundamentals



HUAWEI

Foreword

- With continuous development of networks, the network scale and traffic types increase continuously. As a result, Internet traffic increases sharply, network congestion occurs, the forwarding delay increases, and even packet loss occurs. In this case, the service quality deteriorates or even services are unavailable. To deploy real-time and non-real-time services on the IP network, network congestion must be resolved. The commonly used solution is to increase the network bandwidth. However, this solution is not ideal considering the network construction cost.
- Quality of service (QoS) is introduced in this situation. At limited bandwidth, QoS uses a "guaranteed" policy to manage network traffic and provides different priority services for different traffic.
- This course describes QoS fundamentals.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the QoS background.
  - Describe QoS types.
  - Describe the implementation of the QoS DiffServ model.
  - Describe application scenarios of different QoS functions.
  - Describes basic configuration of QoS.
  - Describe HQoS fundamentals.
  - Describe basic configuration of HQoS.

HUAWEI

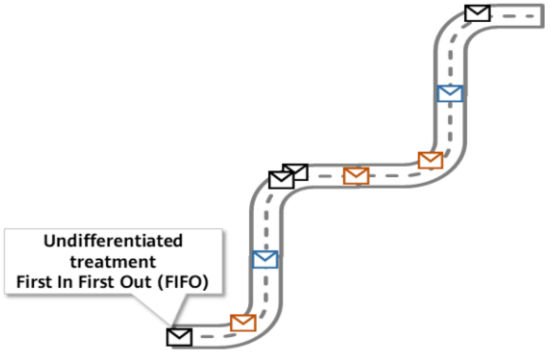
Contents

1. **Introduction to QoS**
2. Traffic Classification and Marking
3. Traffic Limiting Technology
4. Congestion Avoidance Technology
5. Congestion Management Technology
6. Introduction to HQoS

HUAWEI

"Best-Effort" Traditional Network

- When the IP network emerges, there is no QoS guarantee.
- You only know that the packets have been sent out. Whether the packets can be received and when the packets can be received are unknown.

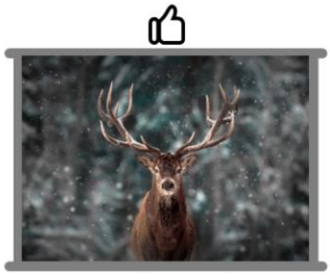


HUAWEI

- On the traditional IP network, each network device handles all packets in an undifferentiated manner and follows the First In First Out (FIFO) rule to transmit packets. The devices transmit packets to the destination in best-effort (BE) mode, but the BE mode cannot ensure the performance such as delay and reliability.

QoS Background

- With continuous technology improvement and fierce product competition, users have increasingly higher requirements on the network quality.

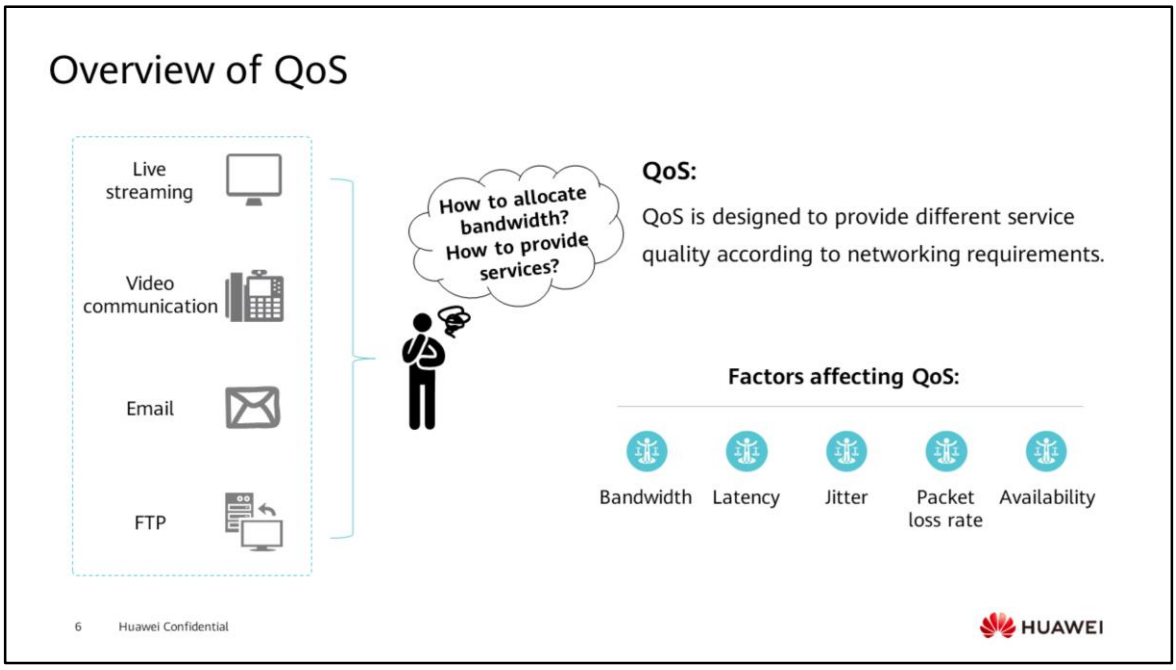


High-definition image quality and high network speed
Good signal quality

Poor image quality, low network speed, and frame freezing
Poor signal quality

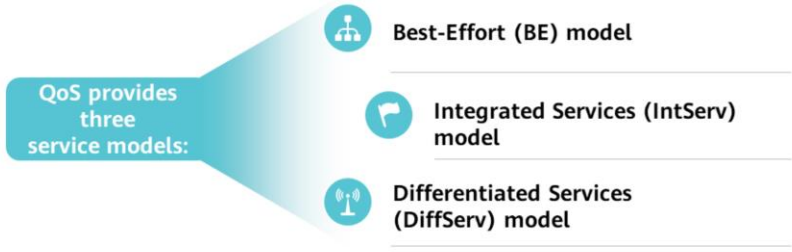
HUAWEI

- With the emergence of new applications on IP networks, new requirements are raised to QoS of IP networks.



- To support voice, video, and data services of different requirements, the network is required to distinguish different communication types before providing corresponding QoS.
    - For example, real-time services such as Voice over IP (VoIP) demand shorter latency. A long latency for packet transmission is unacceptable. Email and the File Transfer Protocol (FTP) services are comparatively insensitive to the latency.
- To support voice, video, and data services of different requirements, the network is required to distinguish different communication types before providing corresponding QoS.
    - The BE mode of traditional IP networks cannot identify and distinguish various communication types on the networks. This distinguishing capability is the premise for providing differentiated services. The BE mode cannot satisfy application requirements, so QoS is introduced.
- What is QoS?

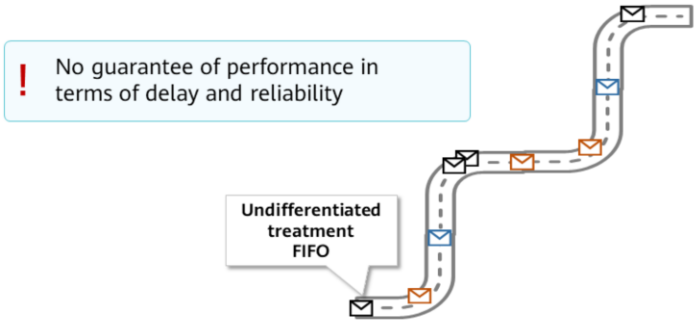
QoS Service Models



HUAWEI

BE Model

- An application can send any number of packets at any time.
- The network then makes the best effort to transmit the packets.

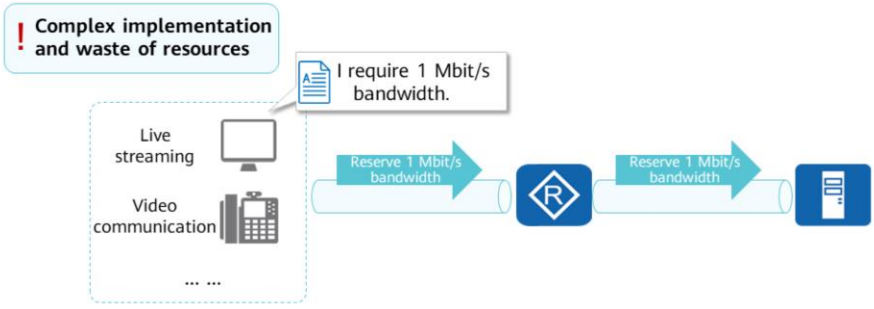


HUAWEI

- The BE model is the simplest service model in which an application can send any number of packets at any time without obtaining approval or notifying the network.
- The network then makes the best effort to transmit the packets but provides no guarantee of performance in terms of delay and reliability.
- The BE model is the default service model for the Internet and applies to various network applications, such as the File Transfer Protocol (FTP) and email. It uses FIFO queues.

IntServ Model

- Before sending packets, an application needs to apply for specific services through signaling.
- After receiving a resource request from an application, the network reserves resources for each information flow by exchanging RSVP signaling information.

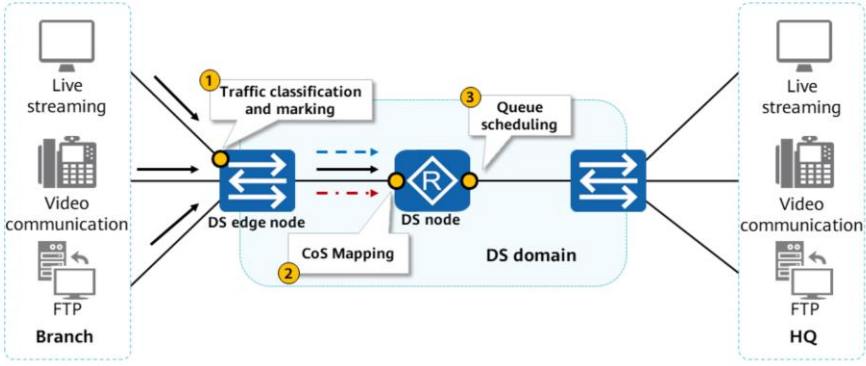


HUAWEI

- The IntServ model is a comprehensive service model to meet various QoS requirements.
- Before sending packets, an application needs to apply for specific services through signaling. This request is sent through RSVP. RSVP applies for network resources for an application before the application starts to send packets.
- Once the network determines to allocate resources to the application, the network maintains a state for each flow (determined by IP addresses, port numbers, and protocol numbers at both ends), and performs packet classification, traffic policing, queuing, and scheduling based on the state. After receiving the acknowledgment message from the network (the application confirms that the network has reserved resources for the packets of the application), the application starts to send packets. As long as packets of the application are controlled within the range described by traffic parameters, the network promises to meet QoS requirements of the application.
- Example: If you want to reserve a vehicle, you need to apply for a service in advance and reserve resources when resources are sufficient.
- However, the vehicle service vendor has to maintain a large number of booking information.
- Disadvantage: The implementation of the IntServ model is complex. When no traffic is transmitted, the bandwidth is still exclusively occupied, and the usage is low. This solution requires that all end-to-end nodes support and run the RSVP protocol.

DiffServ Model

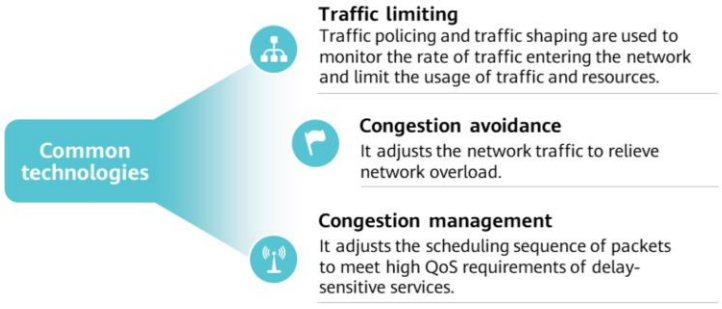
- Traffic on a network is classified into multiple classes, and a corresponding processing behavior is defined for each class, so that the traffic has different forwarding priorities, packet loss rates, and delays.



HUAWEI

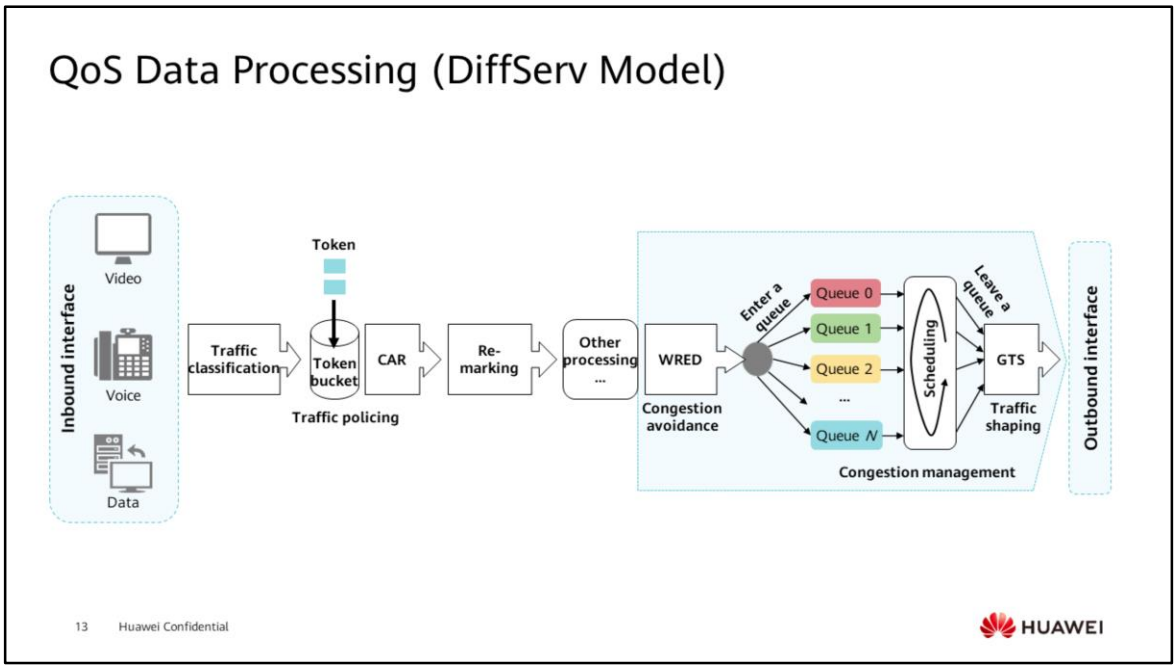
- DiffServ is a multi-service model and can satisfy different QoS requirements. Currently, this model is widely used on IP networks.
- Before sending a packet, the application does not need to notify the network to reserve resources for it. In the DiffServ model, the network does not need to maintain the status of each flow. Instead, it provides specific services based on precedence fields of packets (for example, the DS field in the IP header).
- The DiffServ model classifies network traffic into multiple classes for differentiated processing. To be specific, the DiffServ model implements traffic classification first and allocates different identifiers to different classes of packets. After a network node receives these packets, it simply identifies these identifiers and processes packets based on the actions corresponding to these identifiers.
- There is an analogy between the DiffServ model and train ticket service system. A train ticket marks the service that you book: soft sleeper, hard sleeper, hard seat, or no seat. You get on a train and enjoy the specific service marked in your ticket. On an IP network, an identifier is to a packet as a train ticket is to a passenger.
- In addition to traffic classification and marking, the DiffServ model provides the queuing mechanism. When network congestion occurs on a device, the device buffers packets in queues. The device sends the packets out of queues when network congestion is relieved.

Common QoS Technologies (DiffServ Model)



HUAWEI

- Rate limiting: Traffic policing and traffic shaping monitor the rate of traffic entering the network to limit the traffic and resource usage, providing better services for users.
- Congestion avoidance and congestion management: When congestion occurs on a network, the device determines the sequence in which packets are forwarded according to a certain scheduling policy so key services are processed first. Or, the device proactively adjusts traffic to relieve network overload by discarding packets.



- QoS technology provides the following functions:
  - Traffic classification and marking: identify objects based on certain matching rules, which is the prerequisite for implementing differentiated services. They are usually applied to the inbound direction of an interface.
  - Token bucket: is used to check whether traffic meets packet forwarding conditions.
  - Traffic policing: monitors the volume of specific data traffic that arrives at network devices, and is usually applied to incoming traffic. When the traffic volume exceeds the maximum value, traffic limiting or punishment measures are taken to protect business interests and network resources of service providers.
  - Congestion avoidance: Excessive congestion may damage network resources. Congestion avoidance monitors the usage of network resources. When congestion aggravates, congestion avoidance proactively adjusts traffic to relieve network overload by discarding packets. Congestion avoidance is generally applied to the outbound direction of an interface.
  - Congestion management: is taken to solve the problem of resource competition. Packets are buffered in queues and a scheduling algorithm is used to determine the forwarding sequence of packets. Congestion management is usually applied to the outbound direction of an interface.

Quiz

1. (Multiple-answer question) Which of the following service models are provided by QoS? ( )
   A. DiffServ model
   B. IntServ model
   C. BE model
   D. Network service model

HUAWEI

1. ABC

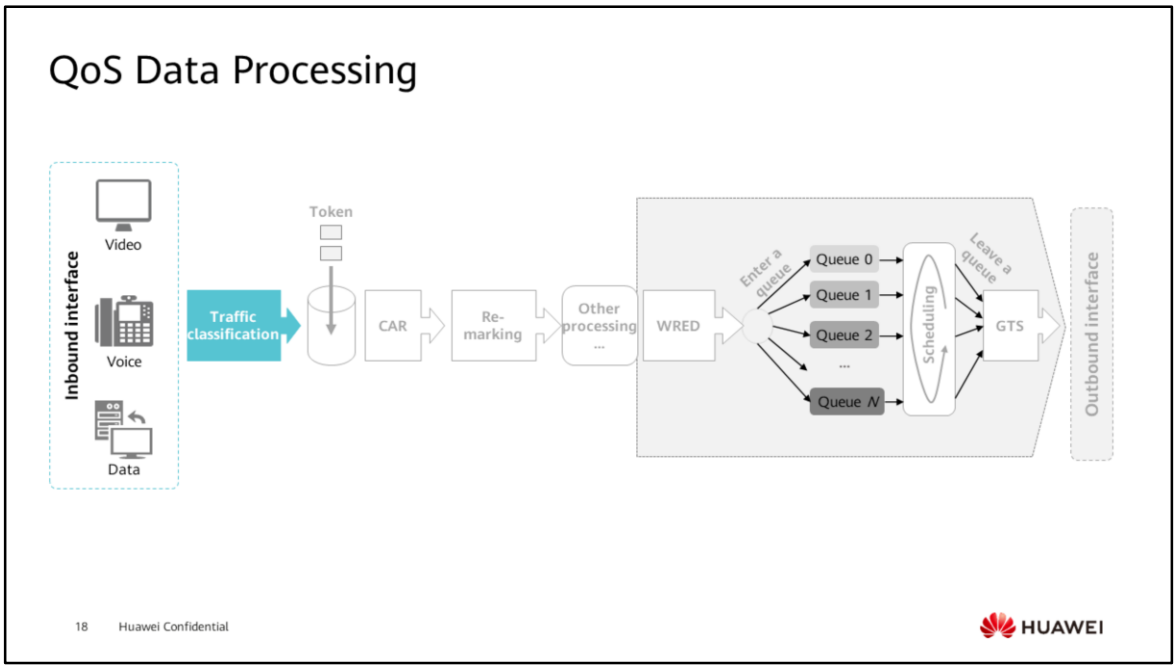
Section Summary

- QoS service models include the DiffServ, IntServ, and BE models.
- The DiffServ model is the most commonly used QoS model. It provides rate limiting, congestion avoidance, and congestion management.

Contents

1. Introduction to QoS
2. **Traffic Classification and Marking**
3. Traffic Limiting Technology
4. Congestion Avoidance Technology
5. Congestion Management Technology
6. Introduction to HQoS

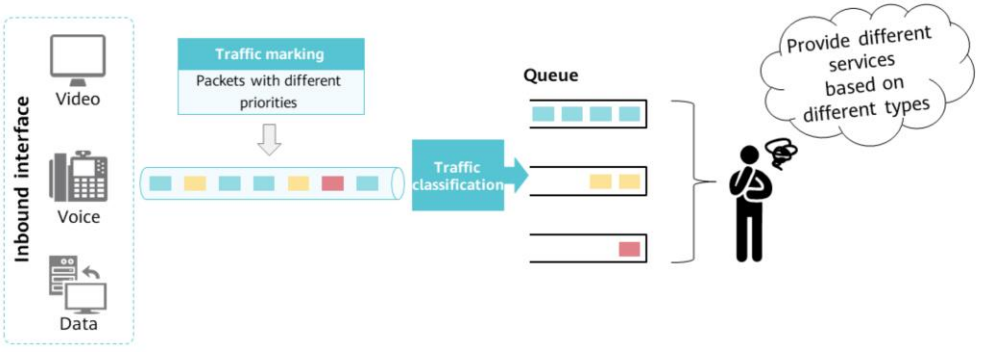
HUAWEI



- Traffic classification and marking: identify objects based on certain matching rules, which is the prerequisite for implementing differentiated services. They are usually applied to the inbound direction of an interface.

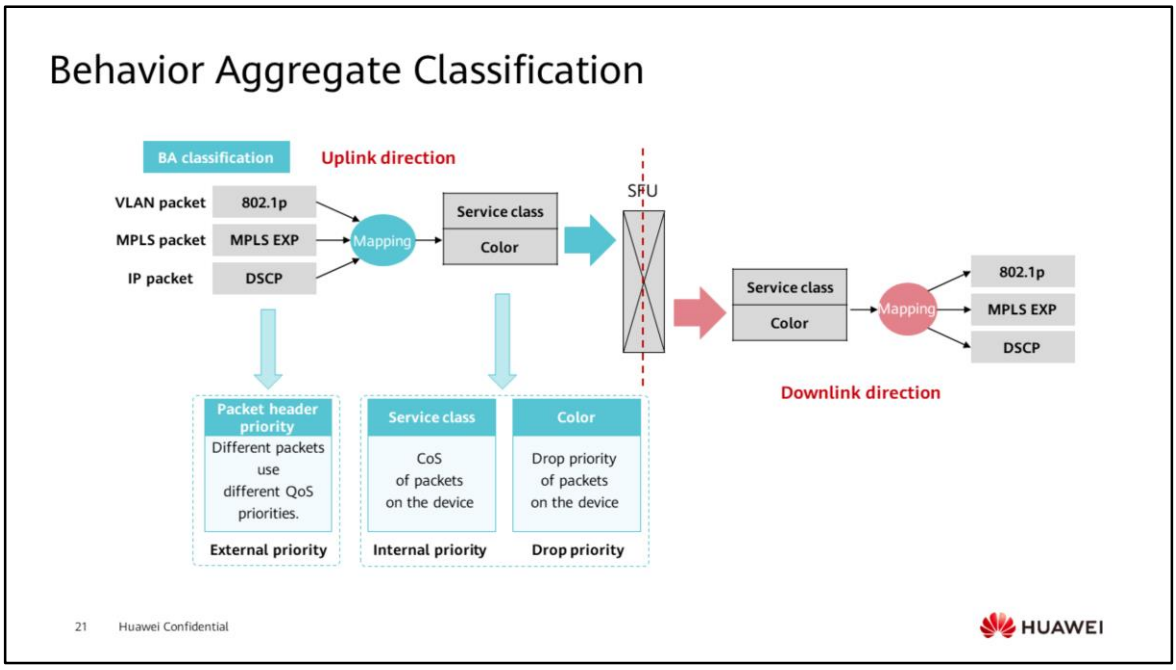
Why Are Traffic Classification and Traffic Marking Required?

- Traffic classification and marking are the basis of QoS and the prerequisite for implementing differentiated services.

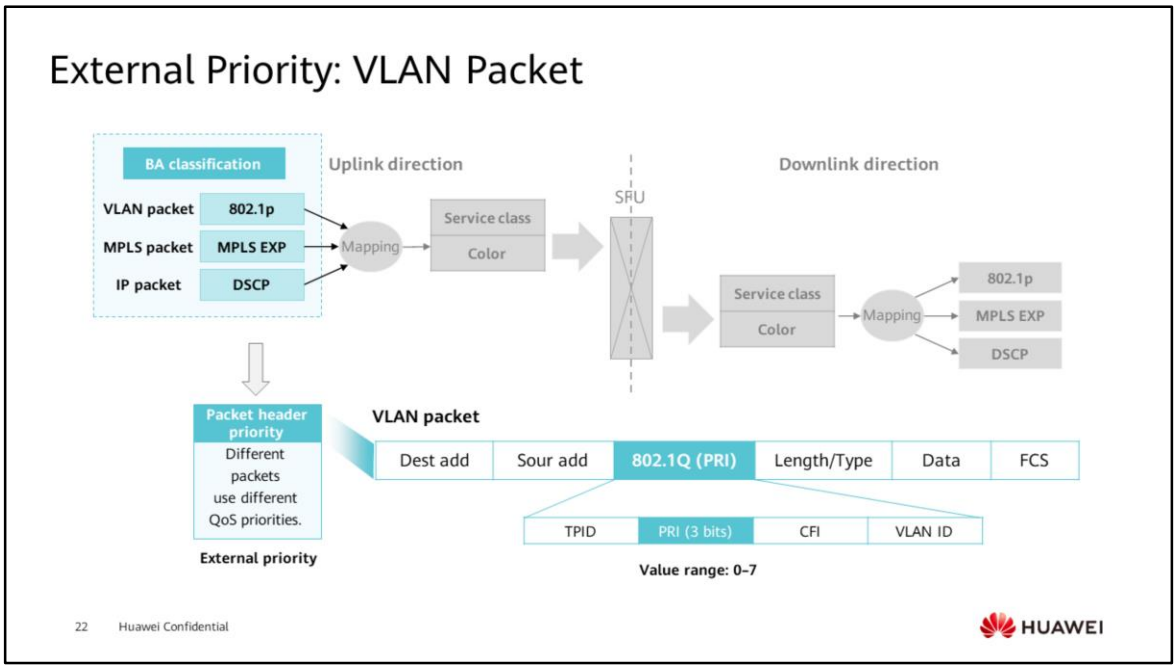


HUAWEI

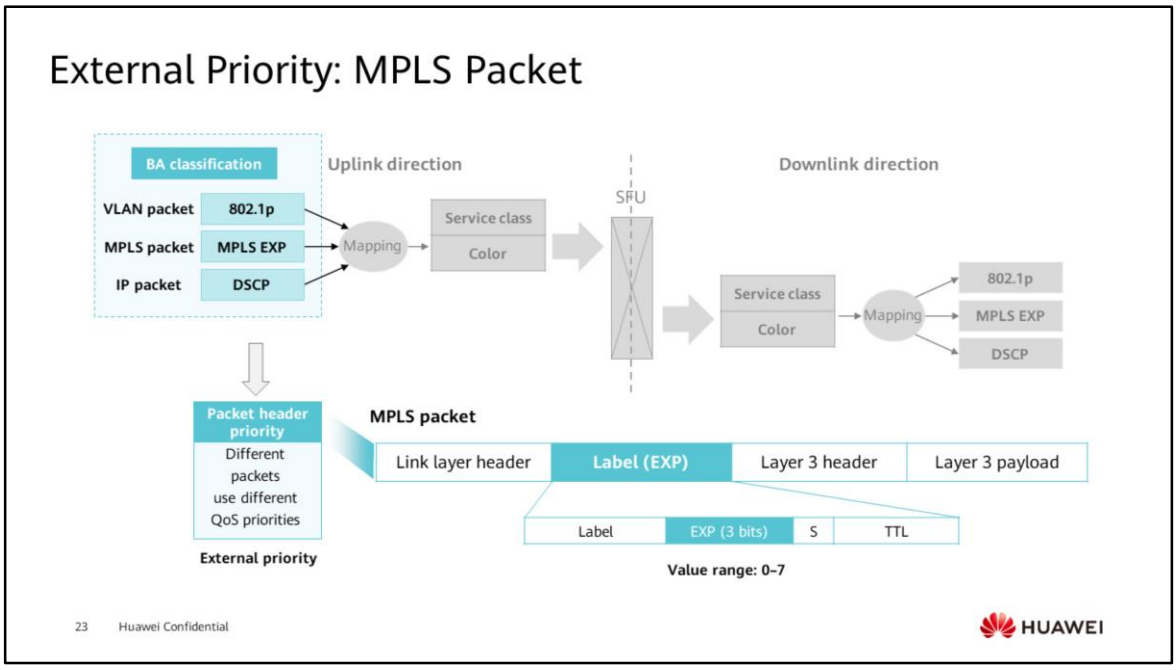
- To implement differentiated services, the traffic entering a DS domain needs to be classified according to certain rules, and then different services are provided for different types of traffic.
- After packets are classified at the DS domain edge, intermediate nodes provide differentiated services for classified packets. The downstream node can accept the classification result of its upstream node or classifies packets based on its own criteria.
- Traffic classification and marking are prerequisites for differentiated services.
  - Traffic classification technology classifies packets into different types, and does not modify the data packets.
  - Marking technology marks packets with different priorities, and modifies the data packets. Marking is classified into internal and external marking.
    - Internal marking
    - Sets the CoS and drop precedence of packets for internal processing on a device so that packets can be placed directly in specific queues.
    - Setting the drop precedence of packets is also called coloring packets. When traffic congestion occurs, packets in the same queue are provided with differentiated buffer services based on colors.



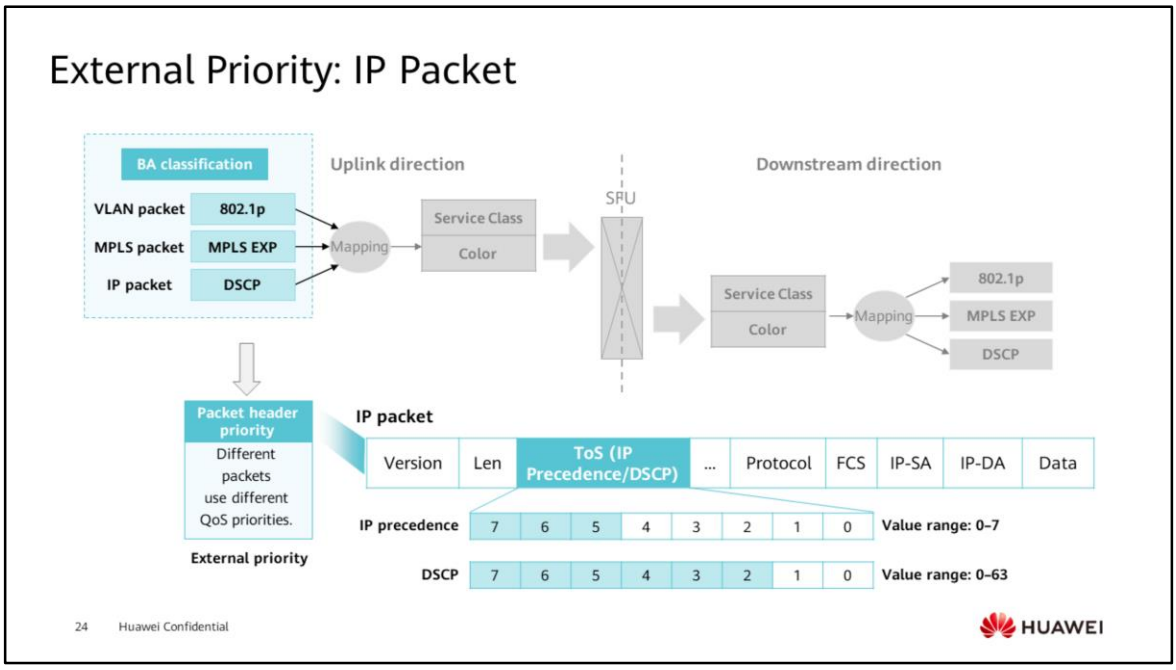
- BA classification allows the device to classify packets based on related values as follows: IP priority or DSCP value of IPv4 packets, TC value of IPv6 packets, EXP value of MPLS packets, and 802.1p value of VLAN packets. It is used to simply identify the traffic that has the specific priority or service classes for mapping between external and internal priorities.
- BA classification confirms that the priority of incoming packets on a device is trusted and mapped to the service class and color based on a priority mapping table. The service class and color of outgoing packets are then mapped back to the priority.
- Packets carry different types of precedence field depending on the network type. For example, packets carry the 802.1p value on a VLAN network, the EXP value on an MPLS network, and the DSCP value on an IP network. To provide differentiated services for different packets, the device maps the QoS priority of incoming packets to the scheduling precedence (also called service class) and drop precedence (also called color), and then performs congestion management based on the service-class and congestion avoidance based on the color. Before forwarding packets out, the device maps the service class and color of the packets back to the QoS priority, which provides a basis for other devices to process the packets.



- Eight service priorities (PRIs) are defined in the VLAN tag of the Ethernet frame header.



- The EXP field in the label is used as the external priority of MPLS packets to differentiate service classes of data traffic.



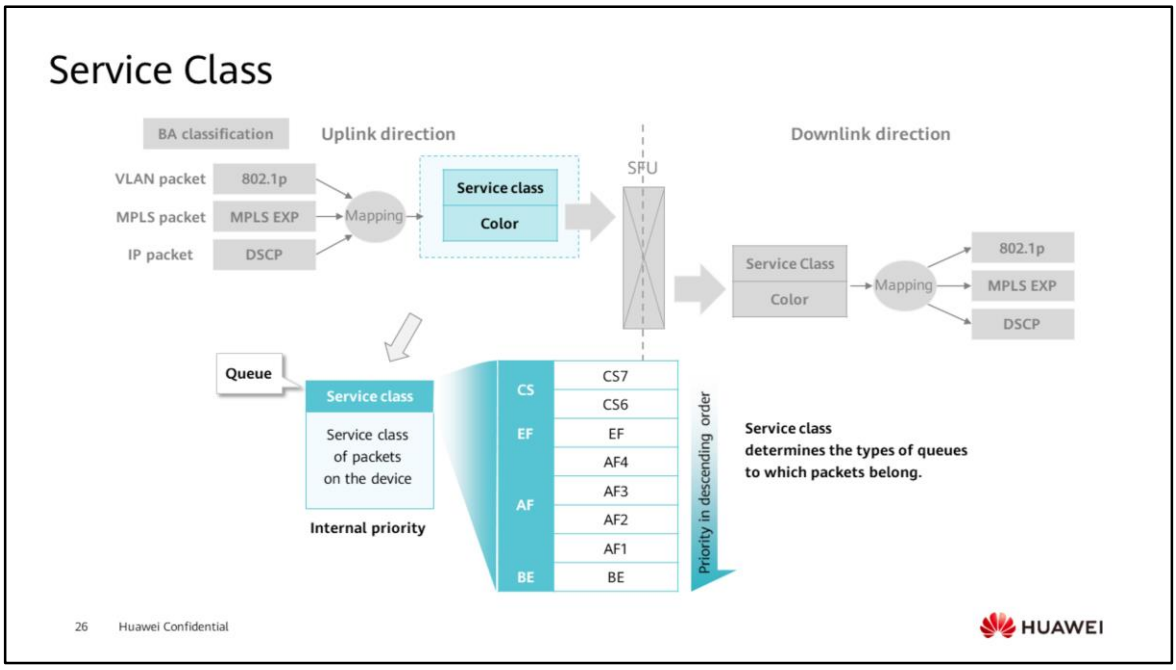
- Eight IP service types are defined in the Precedence field of the ToS field in an IPv4 packet header.
- The ToS field in the IPv4 packet header is redefined as the Differentiated Services (DS) field. That is, the IP Precedence field is extended.

Mapping Between External Priorities

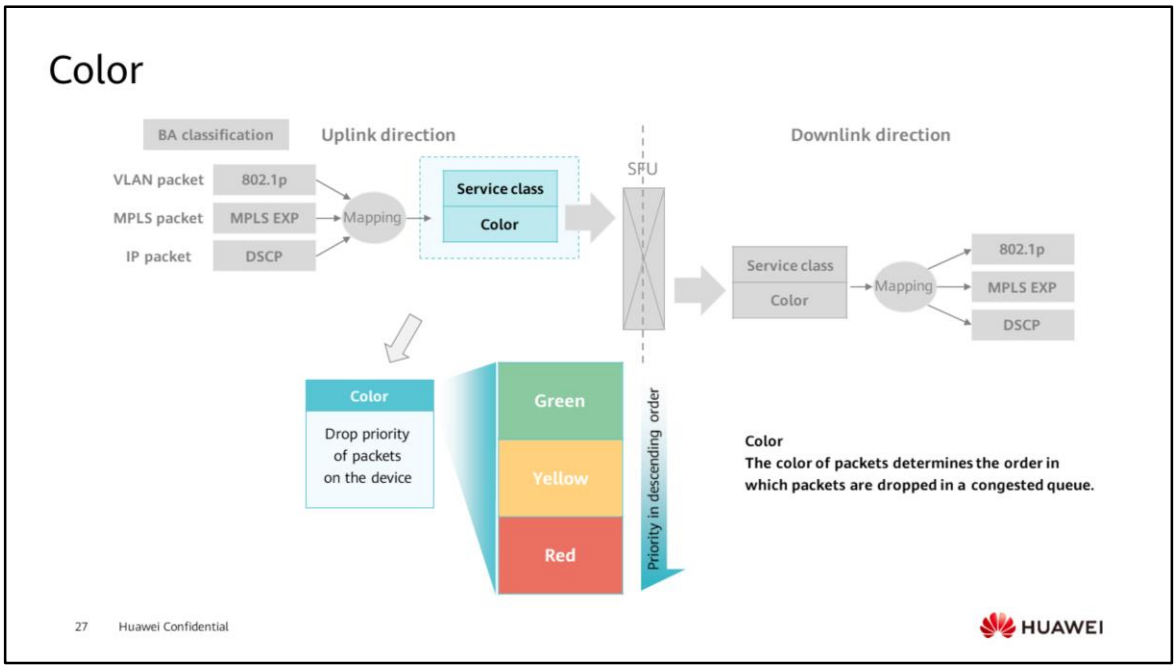
Priority in descending order

802.1P	MPLS Exp	IP-Precedence	DSCP	DSCP Name			
7	7	7	56-63	CS	CS7 (56)		
6	6	6	48-55		CS6 (48)		
5	5	5	40-47	EF	EF (46)		
4	4	4	32-39	AF	AF4	AF41 (34)	AF42 (36)	AF43 (38)
3	3	3	24-31		AF3	AF31 (26)	AF32 (28)	AF33 (30)
2	2	2	16-23		AF2	AF21 (18)	AF22 (20)	AF23 (22)
1	1	1	8-15		AF1	AF11 (10)	AF12 (12)	AF13 (14)
0	0	0	0-7	BE	BE (0)			

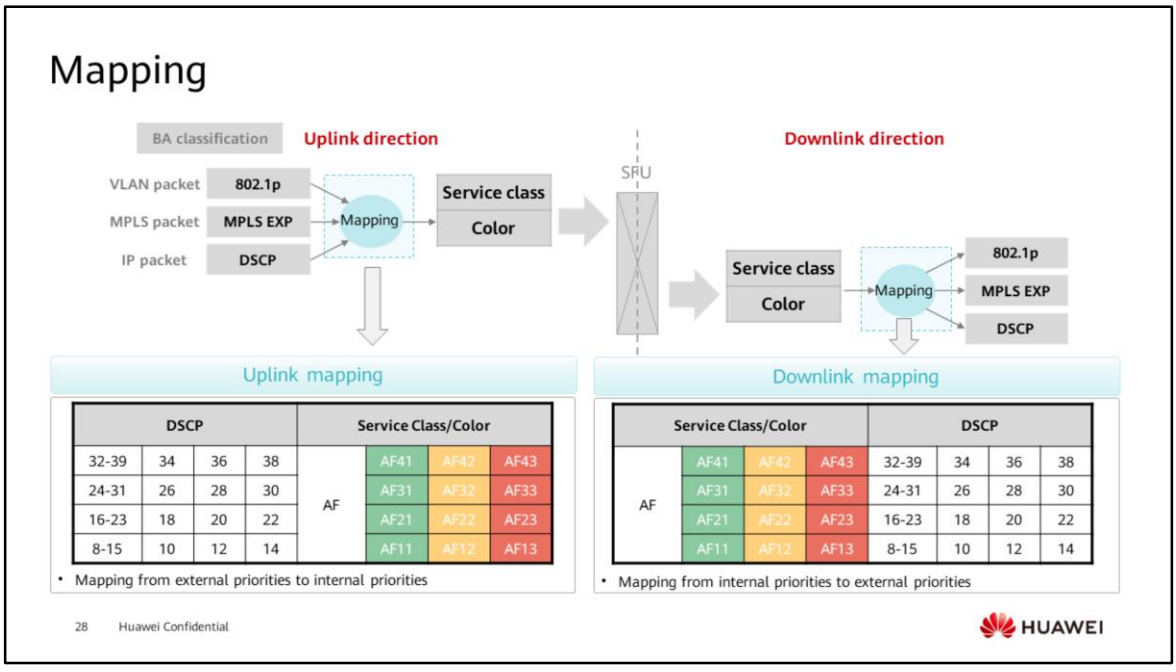
HUAWEI



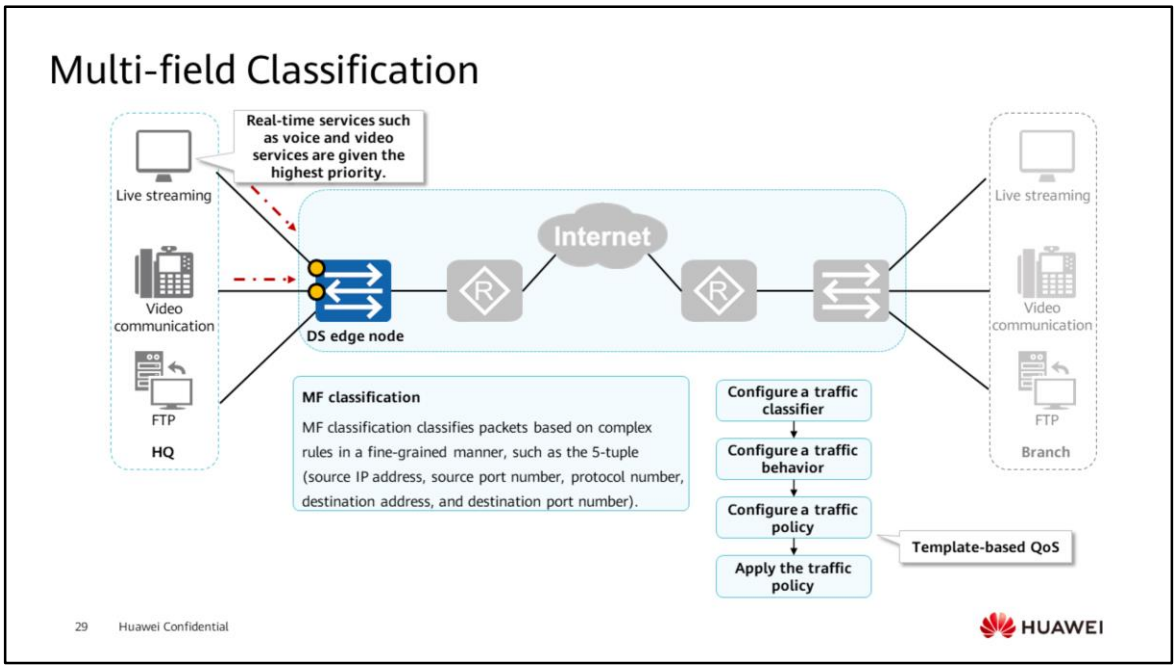
- Service class, that is, queue
- Service classes refer to the internal priorities of packets. Eight service class values are available: class selector 7 (CS7), CS6, expedited forwarding (EF), assured forwarding 4 (AF4), AF3, AF2, AF1, and best-effort (BE). Service classes determine the types of queues to which packets belong.
- The priority of queues with a specific service class is calculated based on scheduling algorithms.
    - If queues with eight service classes all use priority queuing (PQ) scheduling, queues are displayed in descending order of priority: CS7 > CS6 > EF > AF4 > AF3 > AF2 > AF1 > BE.
    - If the BE queue uses PQ scheduling (this configuration is rare on live networks) but all the other seven queues use weighted fair queuing (WFQ) scheduling, the BE queue is of the highest priority.
    - If the queues of eight service classes all use WFQ scheduling, their priorities are the same.



- Color, referring to the drop priority of packets on a device, determines the order in which packets in one queue are dropped when traffic congestion occurs.
- As defined by the Institute of Electrical and Electronics Engineers (IEEE), the color of a packet can be green, yellow, or red.
- Drop priorities are compared based on the configured parameters. For example, if a maximum of 50% of the buffer is configured to store packets colored green, whereas a maximum of 100% of the buffer is configured to store packets colored red, the drop priority of packets colored green is higher than that of packets colored red.



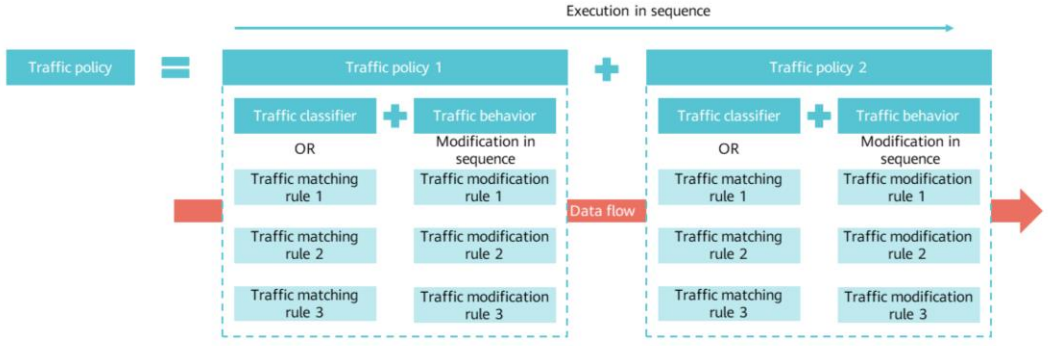
- A device maps the QoS priority to the service class and color for incoming packets and maps the service class and color back to the QoS priority for outgoing packets.



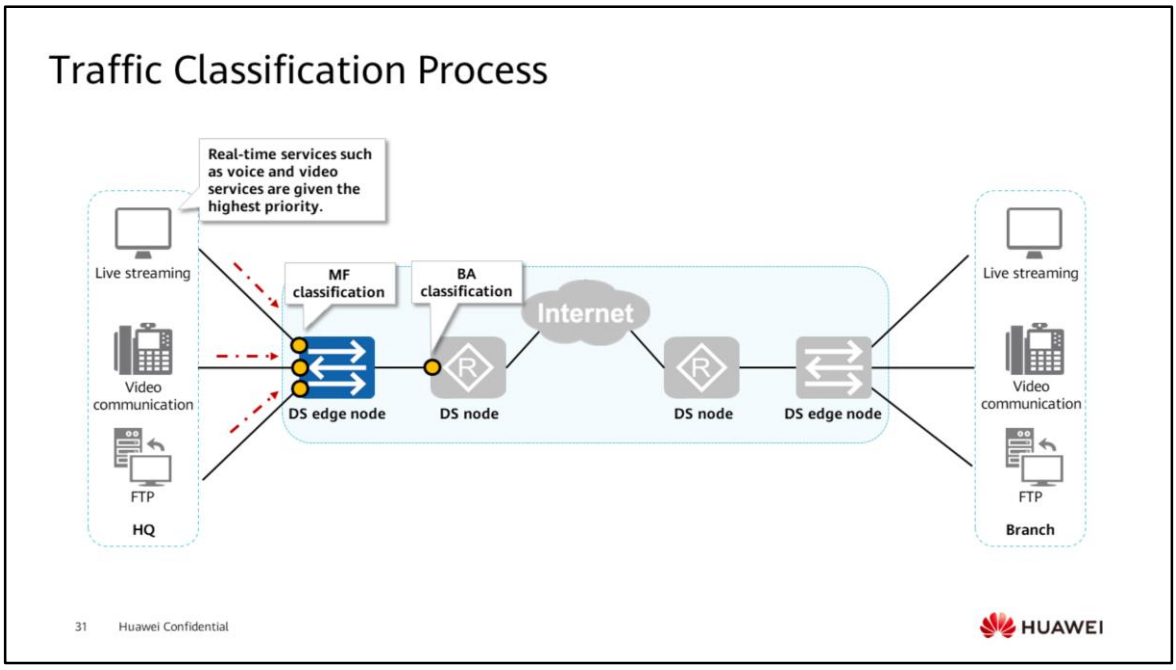
- As networks rapidly develop, services on the Internet become increasingly diversified. Various services share limited network resources. In particular, multiple services use port number 80. Because of this increasing demand, network devices are required to possess a high degree of sensitivity for services, including an in-depth parsing of packets and a comprehensive understanding of any packet field at any layer. This level of sensitivity rises far beyond what BA classification can offer. MF classification can be deployed to help address this sensitivity deficit.
- MF classification classifies packets based on complex rules in a fine-grained manner, such as the 5-tuple (source IP address, source port number, protocol number, destination address, and destination port number).

Traffic Policy Overview

- Modular QoS command line interface (MQC) uses traffic policies.
- A traffic policy is often bound to traffic classifiers and traffic behaviors. A traffic classifier is used to match data packets, and a traffic behavior is used to modify data packets.

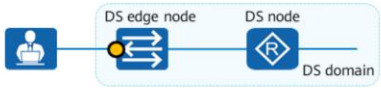


HUAWEI



- Requirement: The highest forwarding priority must be provided for real-time services such as voice and video services.
- Implementation: The DS edge node obtains service traffic such as voice and video traffic through MF classification and maps the traffic to the corresponding priorities. It processes the remaining traffic through BA classification.

Configuring MF Classification



- Typically, the traffic received by the DS edge node is not classified, so complex traffic classification is configured on the DS edge node. The configuration roadmap is as follows:
  - Configure a traffic classifier to match traffic.
  - Configure a traffic behavior to define an action taken on the matched traffic.
  - Bind the traffic classifier and traffic behavior to a traffic policy.
  - Apply the traffic policy to the inbound direction of the interface on the DS edge node.

```
system-view
  traffic classifier [classifier-name]      //Create a traffic classifier.
    if-match [acl| vlan-id| .... ]      //Match traffic based on traffic characteristics.
```

```
system-view
  traffic behavior [behavior-name]      //Create a traffic behavior.
    remark [dscp-name| 8021p-value| EXP| ... ]      //Re-mark the QoS field of traffic.
```

```
system-view
  traffic policy [policy-name]    //Create a traffic policy.
    classifier [classifier-name] behavior [behavior-name]    //Bind the traffic classifier to the traffic behavior.
```

```
system-view
  interface [interface-type interface-num]    //Enter the interface view.
    traffic-policy [policy-name] [inbound | outbound]    //Apply the traffic policy to the inbound direction of an interface.
```


HUAWEI

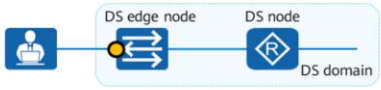
Checking the MF Classification Configuration

- After MF classification is configured, you can run the following commands to check the configuration.

```
system-view
  display traffic classifier user-defined [ classifier-name ]   //Check the traffic classifier configuration.
  display traffic behavior [ system-defined | user-defined ] [ behavior-name ]   //Check the traffic behavior configuration.
  display traffic policy user-defined [ policy-name ] classifier [classifier-name ]   //Check the traffic policy configuration.
  display traffic-policy applied-record [ policy-name ]   //Check the record of the specified traffic policy.
```


HUAWEI

(Optional) Modifying the BA Classification Configuration



- Based on the priority mapping table, BA classification maps data with the specific QoS field to the internal priority.
- The priority mapping table can be modified as required. The roadmap is as follows:
  - Specify the packet priority trusted on an interface.
  - Configure a priority mapping table.

- Specify the packet priority trusted on an

```
system-view
  interface [interface-type interface-num]      //Enter the interface view.
    trust [8021p | dscp]      //Specify the priority to be trusted.
```

```
system-view
  qos map-table [ dot1p-dot1p | dot1p-dscp | dot1p-lp | dscp-dot1p | dscp-dscp | dscp-lp ]      //Enter the priority mapping table view.
    input [input-value1] output [output-value]      //Configure mappings in the priority mapping table.
```


HUAWEI

Checking the Priority Mapping Configuration

- After the priority mapping configuration is modified, you can run the following commands to check the configuration.

```
system-view
  display qos map-table [ dot1p-dot1p | dot1p-dscp | dot1p-lp | dscp-dot1p | dscp-dscp | dscp-lp ]
//Check the mapping between priorities.
```


HUAWEI

Quiz

1. (True or false) MF classification is generally deployed in the inbound direction of the DS edge node.( )
   A. True
   B. False
2. (Multiple-answer question) Which of the following parameters are used to mark the QoS priority of data packets?( )
   A. EXP
   B. 802.1p
   C. DSCP
   D. IP precedence

HUAWEI

1. A
2. ABCD

Section Summary

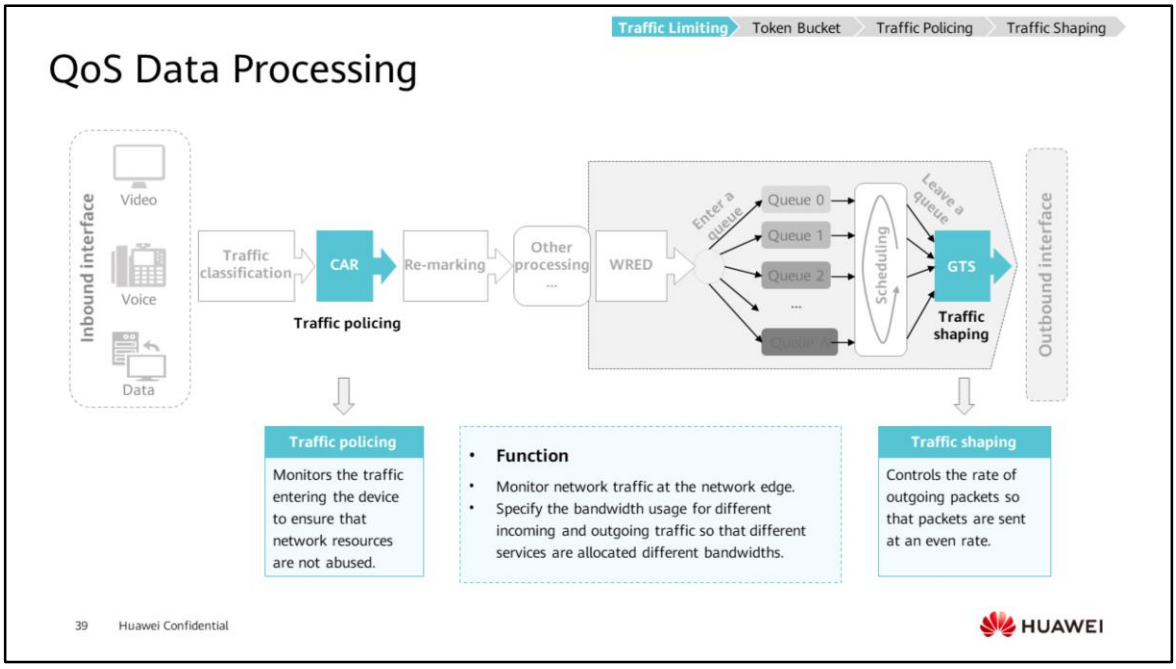
- The DiffServ model must mark packets for differentiating them. Generally, MF classification is used to mark incoming traffic on edge devices in a DS domain, and BA classification is used to mark incoming traffic on devices in a DS domain.
- Tags can be added to multiple types of data packet headers.
  - The Pri bit (802.1p priority) in the VLAN header is used to mark the QoS priority.
  - The EXP bit in the MPLS header is used to mark the QoS priority.
  - The TOS bit (DSCP/IP precedence) in the IP header is used to mark the QoS priority.

HUAWEI

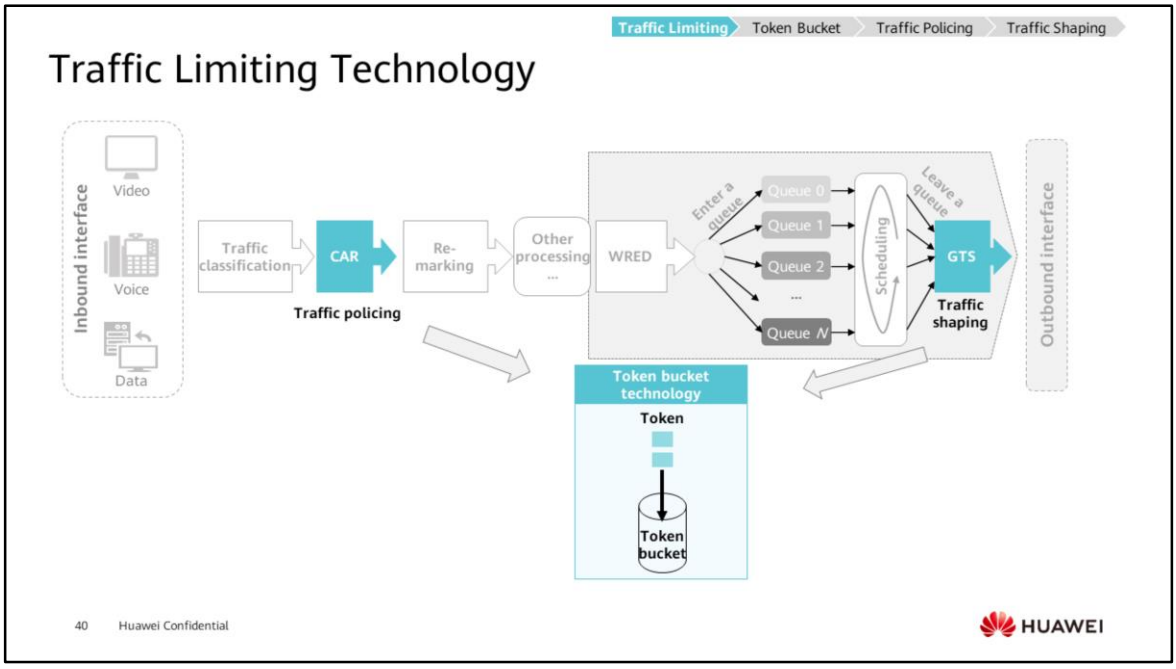
Contents

1. Introduction to QoS
2. Traffic Classification and Marking
3. **Traffic Limiting Technology**
4. Congestion Avoidance Technology
5. Congestion Management Technology
6. Introduction to HQoS

HUAWEI



- This course describes two rate limiting technologies: traffic policing and traffic shaping.
- Traffic policing: If the traffic rate of a connection exceeds the specifications on an interface, traffic policing allows the interface to drop excess packets or re-mark the packet priority to protect network resources and protect carriers' profits. An example of this process is restricting the rate of HTTP packets to 50% of the network bandwidth.
- Traffic shaping: allows the traffic rate to match that on the downstream device. When traffic is transmitted from a high-speed link to a low-speed link or a traffic burst occurs, the inbound interface of the low-speed link is prone to severe data loss. To prevent this problem, traffic shaping must be configured on the outbound interface of the device connecting to the high-speed link.

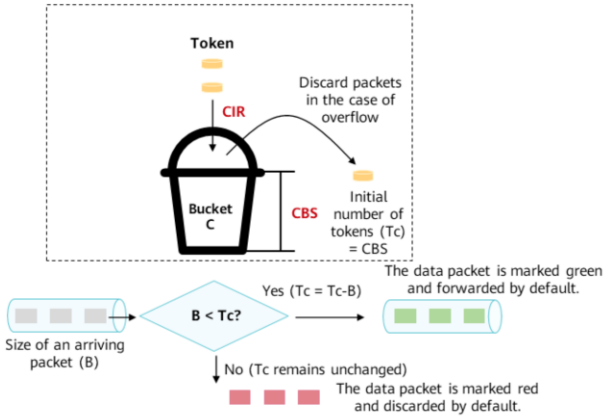


- Both traffic policing and traffic shaping use the token bucket technology.
  - Token bucket: A token bucket is used to check whether traffic meets packet forwarding conditions.

Single-Rate-Single-Bucket Mechanism

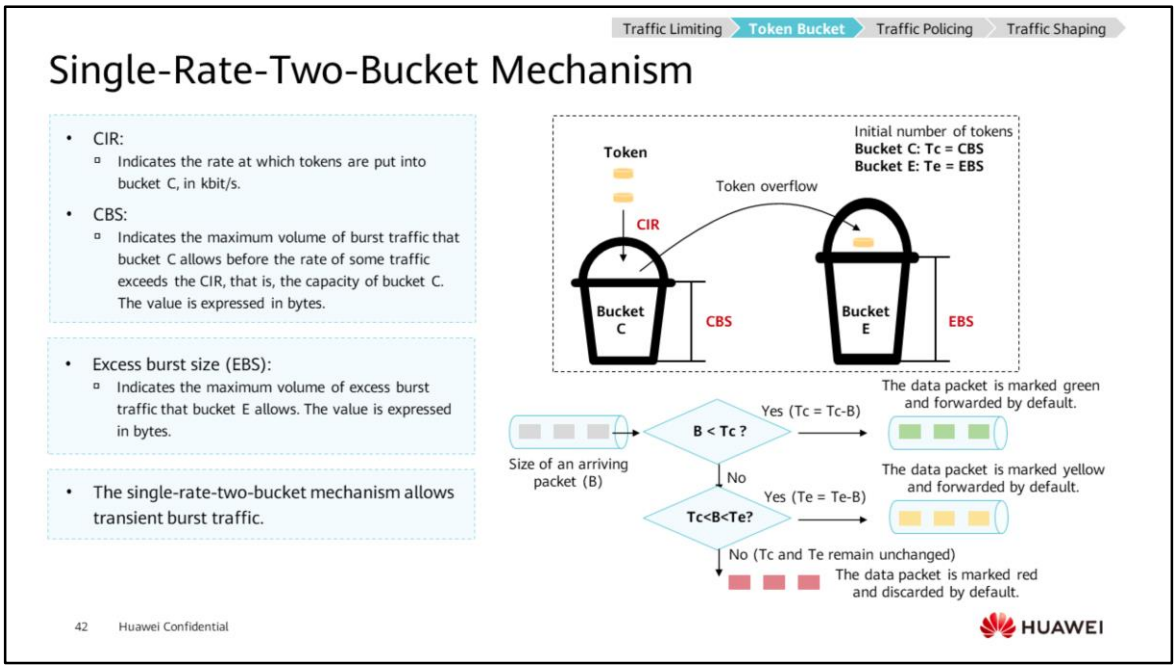
- Committed Information Rate (CIR):
  - indicates the rate at which tokens are put into bucket C, in kbit/s.
- Committed burst size (CBS):
  - indicates the maximum volume of burst traffic that bucket C allows before the rate of some traffic exceeds the CIR, that is, the capacity of bucket C. The value is expressed in bytes.

- The single-rate-single-bucket mechanism does not allow burst traffic. Only committed traffic is allowed.

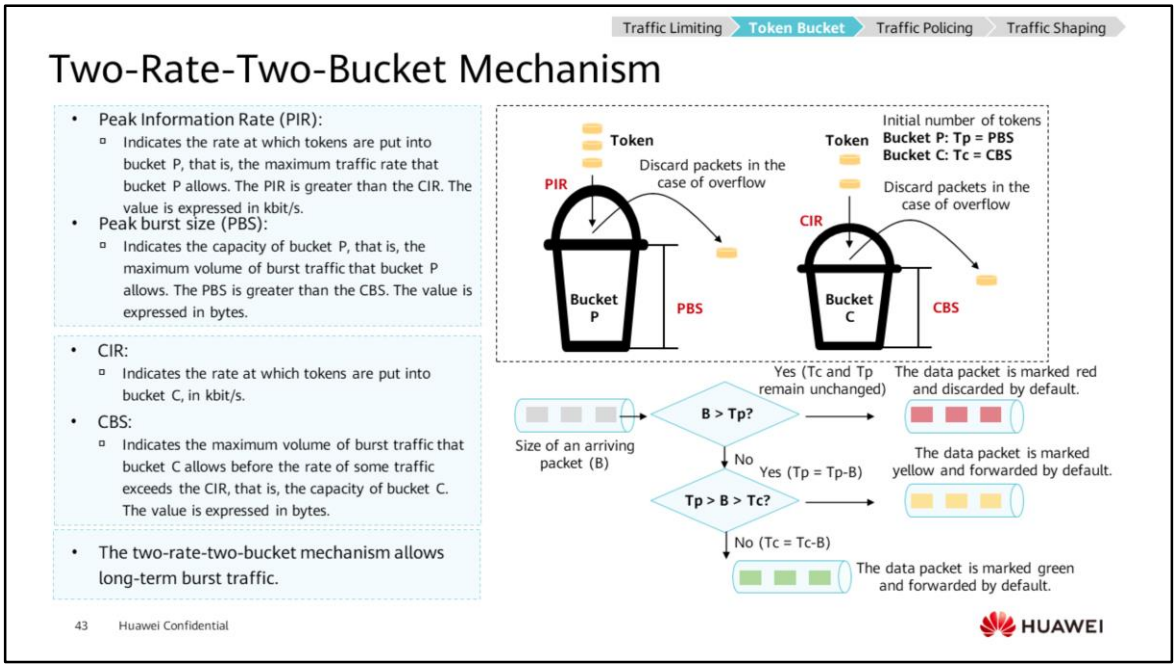


HUAWEI

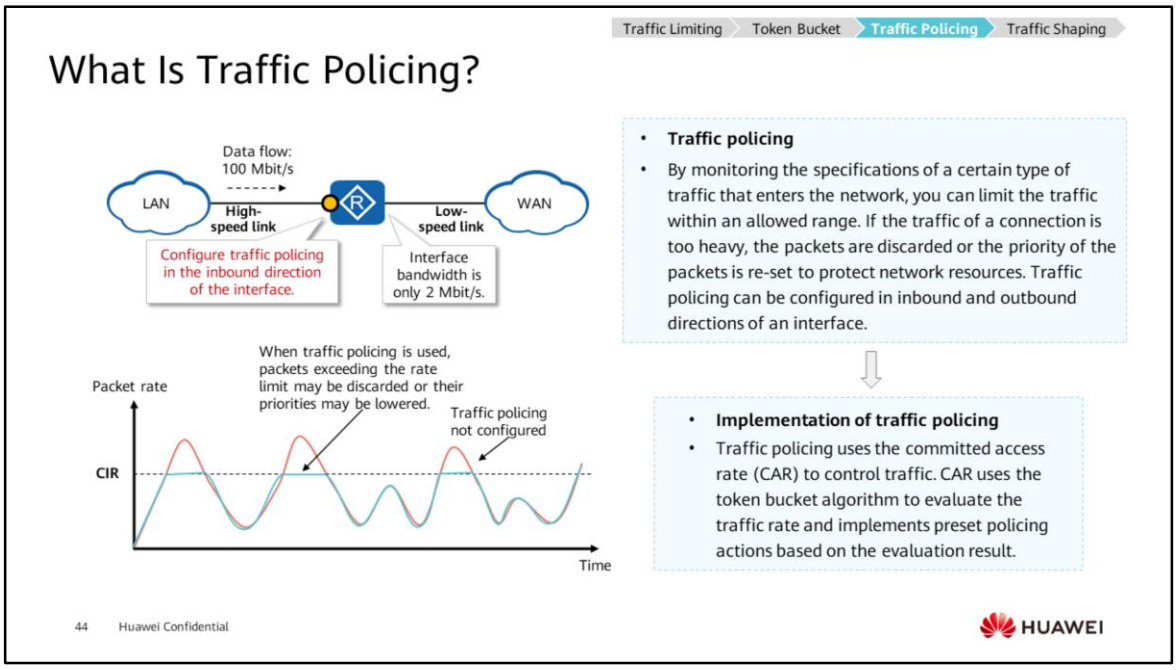
- When a packet arrives, the device compares the packet with the number of tokens in the bucket. If there are sufficient tokens, the packet is forwarded (one token is associated with 1-bit forwarding permission). If there are no enough tokens, the packet is discarded or buffered.
- Tc and Te refer to the numbers of tokens in buckets C and E, respectively. The initial values of Tc and Te are respectively the CBS and EBS.
- In color-blind mode (B indicates the size of an arriving packet):
  - If B is less than or equal to Tc, the packet is marked green, and Tc decreases by B.
  - If B is greater than Tc, the packet is marked red, and Tc remains unchanged.



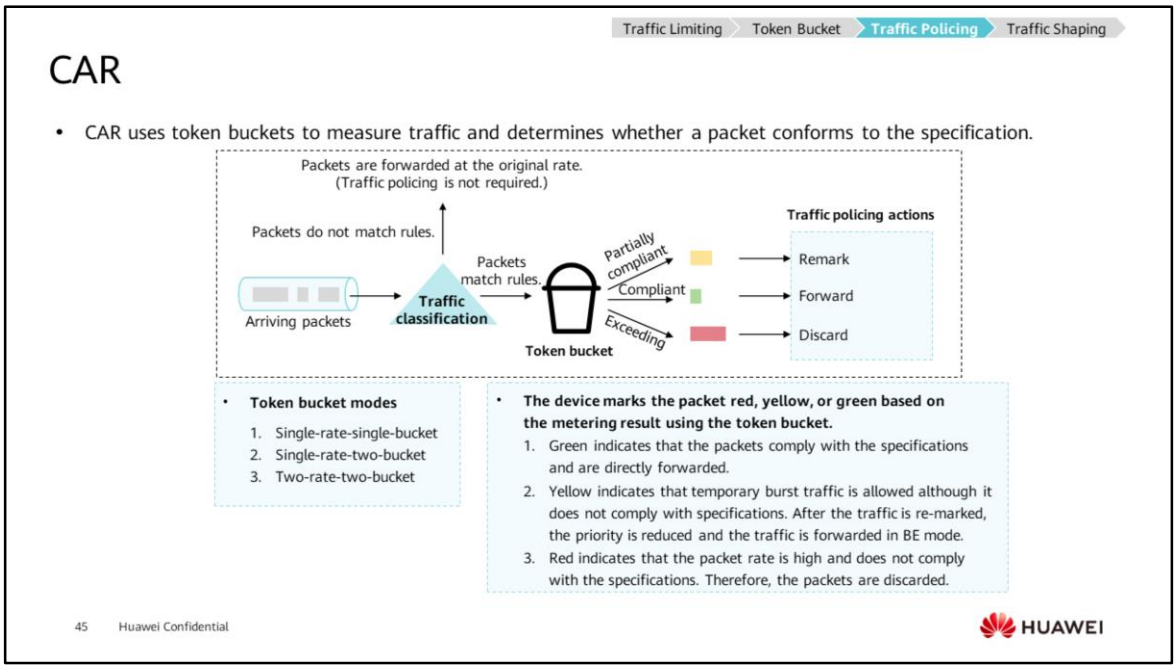
- When a packet arrives, the device compares the packet with the number of tokens in the bucket. If there are sufficient tokens, the packet is forwarded (one token is associated with 1-bit forwarding permission). If there are no enough tokens, the packet is discarded or buffered.
- Tc and Te refer to the numbers of tokens in buckets C and E, respectively. The initial values of Tc and Te are respectively the CBS and EBS.
- In color-blind mode (B indicates the size of an arriving packet):
  - If B is less than or equal to Tc, the packet is marked green and Tc decreases by B.
  - If B is greater than Tc and less than or equal to Te, the packet is marked yellow and Te decreases by B.
  - If B is greater than Te, the packet is marked red, and Tc and Te remain unchanged.



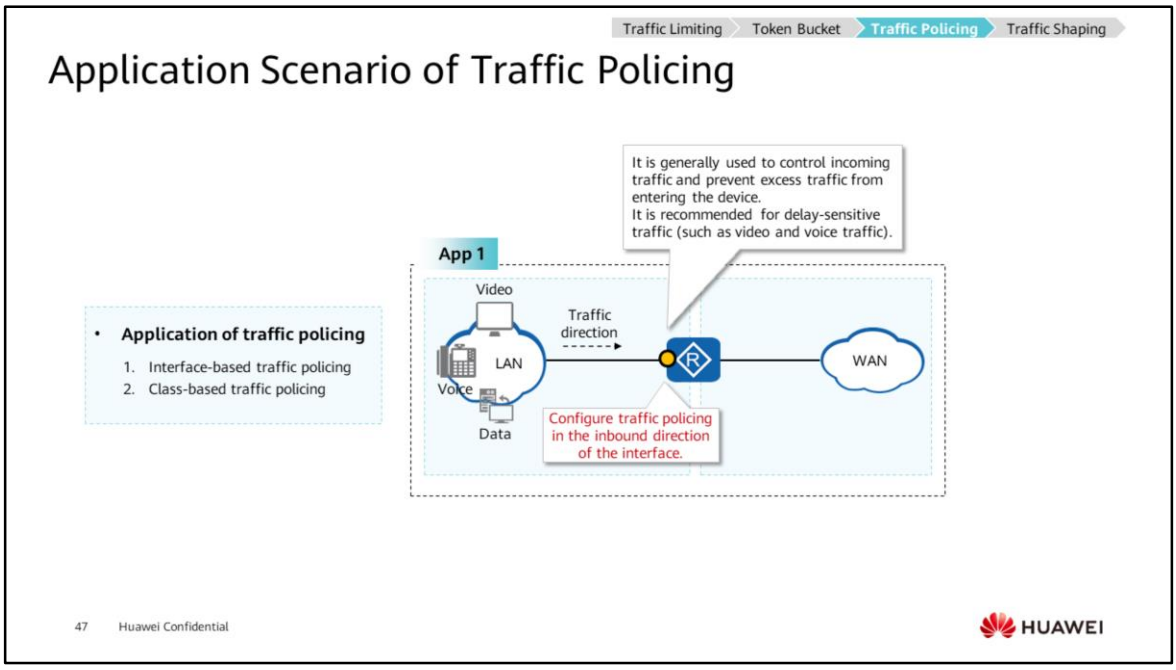
- The two rate three color marker (trTCM) algorithm focuses on the traffic burst rate and checks whether the traffic rate conforms to the specifications. Therefore, traffic is measured based on bucket P and then bucket C.
- Tc and Tp refer to the number of tokens in buckets C and P, respectively. The initial values of Tc and Tp are respectively the CBS and PBS.
- In color-blind mode (B indicates the size of an arriving packet):
  - If B is greater than Tp, the packet is marked red and Tc and Tp remain unchanged.
  - If B is greater than Tc and less than or equal to Tp, the packet is marked yellow and Tp decreases by B.
  - If B is less than or equal to Tc, the packet is marked green, and Tp and Tc decrease by B.



- In the figure:
  - An edge network device connects a wide area network (WAN) and a local area network (LAN). The LAN bandwidth (100 Mbit/s) is higher than the WAN bandwidth (2 Mbit/s).
  - When a LAN user attempts to send a large amount of data to a WAN, the edge network device is prone to traffic congestion. Traffic policing can be configured on the edge network device to restrict the traffic rate, preventing traffic congestion.
- Characteristics of traffic policing:
  - Drops excess traffic over the specifications or re-marks such traffic with a lower priority.
  - Consumes no additional memory resources and brings no delay or jitter.
  - Packet loss may result in packet retransmission.
  - Traffic can be re-marked.

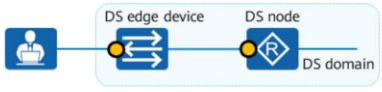


- Traffic policing uses CAR to control traffic. CAR uses token buckets to measure traffic and determines whether a packet conforms to the specification.
- CAR has the following two functions:
  - Rate limiting: Only packets allocated enough tokens are allowed to pass so that the traffic rate is restricted.
  - Traffic classification: Packets are marked internal priorities, such as the service class and drop priority, based on the measurement performed by token buckets.
- CAR process:
  - When a packet arrives, the device matches the packet against matching rules. If the packet matches a rule, the device uses token buckets to meter the traffic rate.
  - The device marks the packet red, yellow, or green based on the metering result using the token bucket. Red indicates that the traffic rate exceeds the specifications. Yellow indicates that the traffic rate exceeds the specifications but is within an allowed range. Green indicates that the traffic rate is conforming to the specifications.
  - The device drops packets marked red, re-marks and forwards packets marked yellow, and forwards packets marked green.



- Voice, video, and data services are transmitted on an enterprise network. When a large amount of traffic enters the network, congestion may occur due to insufficient bandwidth. Different guaranteed bandwidth must be provided for the voice, video, and data services in descending order of priority. In this situation, traffic policing can be configured to provide the highest guaranteed bandwidth for voice packets and lowest guaranteed bandwidth for data packets. This configuration ensures preferential transmission of voice packets during congestion.
- Interface-based traffic policing
    - Interface-based traffic policing controls all traffic that enters an interface and does not identify the packet types. (based on the interface)
- Class-based traffic policing
    - Class-based traffic policing controls the rate of one or more types of packets that enter an interface. (based on traffic classification)

Configuring Interface-based Traffic Policing



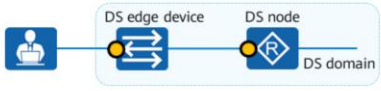
- Typically, traffic policing is performed in the inbound direction of a device. Traffic policing can be deployed on the terminal side or in the inbound direction of an egress device as required. Traffic policing can be configured based on interfaces or MQC.
- The configuration roadmap of interface-based traffic policing is as follows:
  - Set the maximum bandwidth for traffic policing on an interface, select the traffic to be policed, and adjust the behavior to be taken on the excess traffic.

- Configure interface-based traffic policing.

system-view
interface [*interface-type interface-num*] //Enter the interface view.
qos car [inbound | outbound] [acl *acl-number*| *destination-ip-address*| *source-ip-address*] **cir** [*cir-value*] [**pir** *pir-value*] [**cbs** *cbs-value* **pbs** *pbs-value*] //Configure traffic policing for specific traffic in the inbound or outbound direction of an interface. The CIR must be configured. The CIR indicates the maximum committed rate of traffic policing. If the PIR is not configured, it is equal to the CIR. In this case, the traffic rate cannot be higher than the CIR.

HUAWEI

Configuring MQC-based Traffic Policing



- MQC can be used to implement traffic policing in a more refined manner. The device can control traffic and allocate different bandwidths based on the 5-tuple and QoS value of the data packet.
- The configuration roadmap is as follows:
  - Configure a traffic classifier to match traffic.
  - Configure a traffic behavior to define actions taken for packets.
  - Bind the traffic classifier and traffic behavior to a traffic policy.
  - Apply the traffic policy to an interface.

system-view
traffic classifier [*classifier-name*] //Create a traffic classifier.
if-match [*acl*| *vlan-id*| ....] //Match traffic based on traffic characteristics.

system-view
traffic behavior [*behavior-name*] //Create a traffic behavior.
car cir [*cir-value*] [**pir** *pir-value*] [cbs *cbs-value* pbs *pbs-value*] //Set the CIR and PIR.

system-view
traffic policy [*policy-name*] //Create a traffic policy.
classifier [*classifier-name*] **behavior** [*behavior-name*] //Bind the traffic classifier to the traffic behavior.

system-view
interface [*interface-type interface-num*] //Enter the interface view.
traffic-policy [*policy-name*] [inbound | outbound] //Apply the traffic policy to the inbound direction of an interface.

HUAWEI

Checking the Traffic Policing Configuration

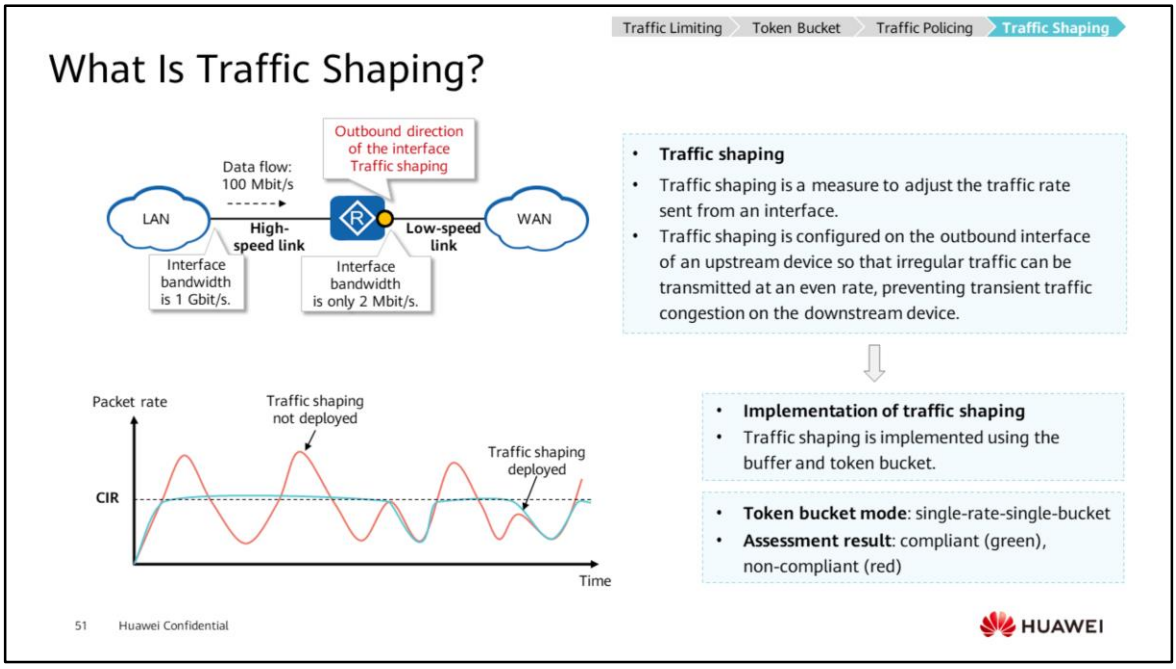
- After interface-based traffic policing is configured, you can run the following commands to check the configuration.

```
system-view
  display qos car statistics interface [interface-type interface-num] [inbound | outbound]    //Check statistics on forwarded and discarded packets on the interface.
```

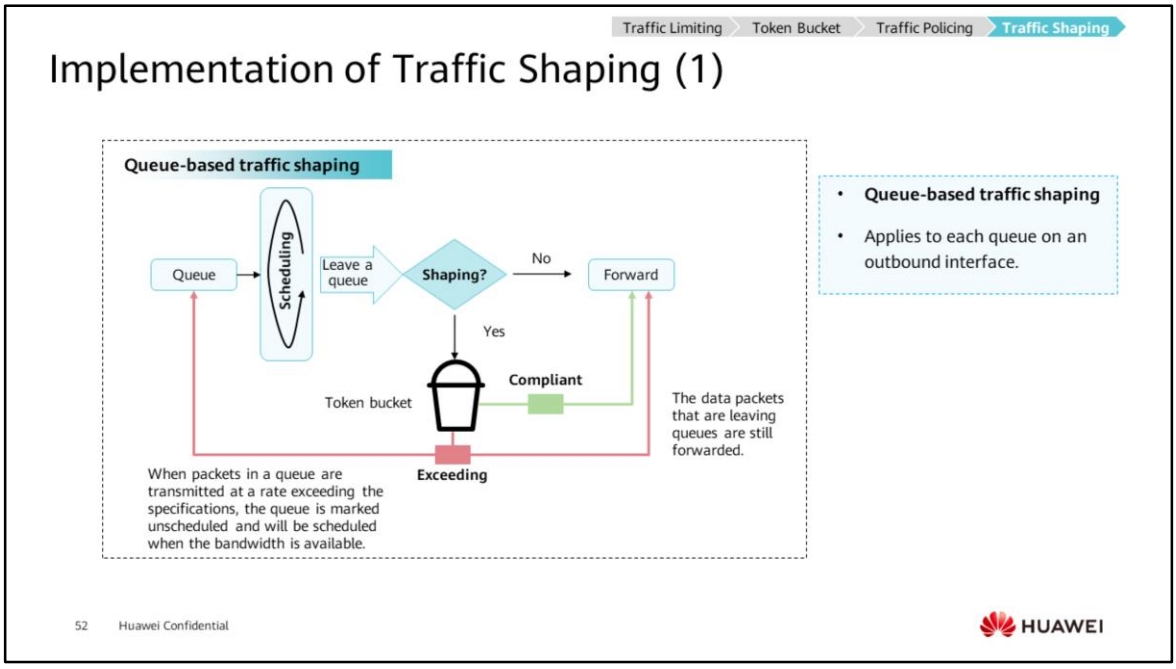
- After MQC-based traffic policing is configured, you can run the following commands to check the configuration.

```
system-view
  display traffic classifier user-defined [ classifier-name ]    //Check the traffic classifier configuration.
  display traffic behavior [ system-defined | user-defined ] [ behavior-name ]    //Check the traffic behavior configuration.
  display traffic policy user-defined [ policy-name ] classifier [classifier-name ]    //Check the traffic policy configuration.
  display traffic-policy applied-record [ policy-name ]    //Check the record of the specified traffic policy.
```


HUAWEI

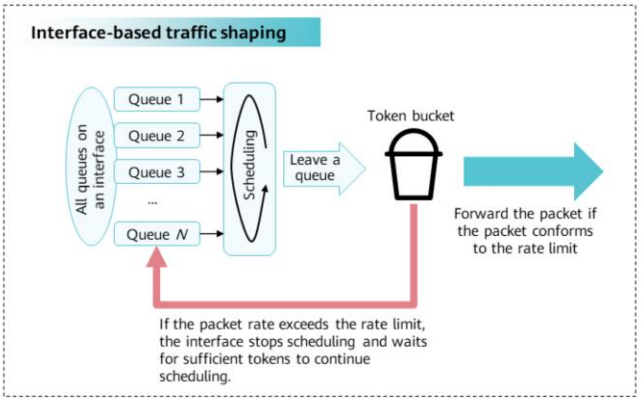


- Generic Traffic Shaping (GTS)
  - When traffic is transmitted from a high-speed link to a low-speed link or a traffic burst occurs, the inbound interface of the low-speed link is prone to severe data loss. To prevent this problem, configure traffic shaping on the outbound interface of the device connecting to the high-speed link.
  - When packets are sent at a high speed, they are cached and then evenly sent through the token bucket.
- Characteristics of traffic shaping:
  - Buffers excess traffic over the specifications.
  - Consumes memory resources for buffering excess traffic and brings delay and jitter.
  - Packet loss rarely occurs, so packets are seldom retransmitted.
  - Traffic re-marking is not supported.
- Token bucket mode: single-rate-single-bucket — The evaluation result can be either green or red.



- When packets leave queues, the packets that do not need to be shaped are forwarded. The packets that need to be shaped are measured against token buckets.
    - If the packet rate conforms to the rate limit, the packet is marked green and forwarded.
    - If the rate of a data packet exceeds the threshold, the data packet is still forwarded. In this case, the status of the queue where the data packet is located is changed to unscheduled, and the queue is scheduled when the token bucket is filled with new tokens. After the queue is marked unscheduled, more packets can be put into the queue, but excess packets over the queue capacity are dropped. Therefore, traffic shaping allows traffic to be sent at an even rate but does not provide zero-packet-loss guarantee.

Implementation of Traffic Shaping (2)



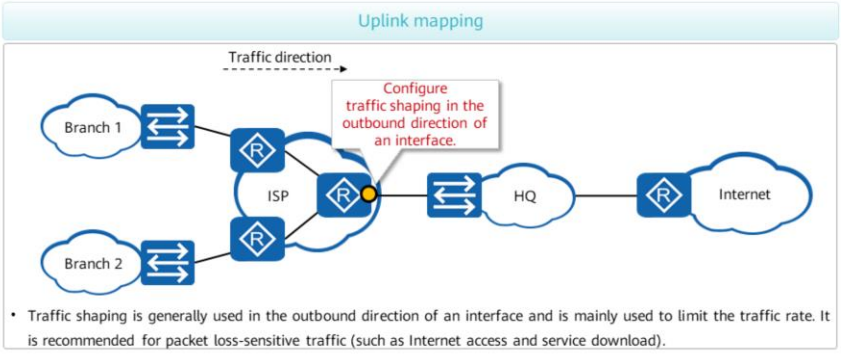
- **Interface-based traffic shaping**
- Limits the total rate of all packets sent by an interface. Traffic shaping is performed on the outbound interface, regardless of packet priorities.

HUAWEI

- When packets leave queues, all queues are measured together against token buckets.
  - If the packet rate conforms to the rate limit, the packet is marked green and forwarded.
  - If the packet rate exceeds the threshold (that is, tokens in the token bucket are insufficient), the packet is marked red. In this case, the interface stops scheduling and continues to schedule the packets when there are sufficient tokens.

Application Scenario of Traffic Shaping

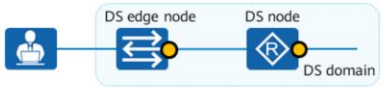
- On an enterprise network, the enterprise headquarters is connected to branches through private lines on an ISP network. Branches connect to the Internet through the headquarters.
- If all branches connect to the Internet at the same time, a large amount of web traffic sent from the headquarters to the Internet causes network congestion. As a result, some web traffic is discarded.. As shown in the figure, to prevent web traffic loss, traffic shaping can be configured before traffic sent from enterprise branches enters the enterprise headquarters.



- Traffic shaping is generally used in the outbound direction of an interface and is mainly used to limit the traffic rate. It is recommended for packet loss-sensitive traffic (such as Internet access and service download).

HUAWEI

Configuring Interface-based Traffic Shaping



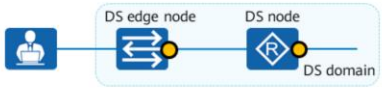
- Traffic shaping can be configured only in the outbound direction of a device. It falls into interface-based, queue-based, and MQC-based traffic shaping.
- Interface-based traffic shaping has a large granularity. The configuration roadmap is as follows:
  - Deploy traffic shaping in the outbound direction of an interface and configure the maximum bandwidth.

- Configure interface-based traffic shaping.

system-view
  interface [*interface-type interface-num*] //Enter the interface view.
    qos gts cir [*cir-value*] [cbs *cbs-value*] //Configure traffic shaping in the outbound direction of an interface. The CIR indicates the maximum traffic shaping rate. You can configure the CBS as required to control the size of the token bucket. The CIR must be configured.

HUAWEI

Configuring Queue-based Traffic Shaping



- To shape packets in each queue on an interface, configure a queue profile and apply it to the interface.
- You can set different traffic shaping parameters for queues with different priorities to provide differentiated services. The configuration roadmap is as follows:
  - Create a queue profile.
  - Configure queue shaping.
  - Apply the queue profile to an interface.

- Create a queue profile and configure queue shaping.

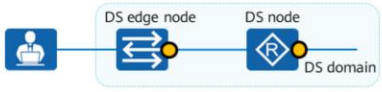
system-view
interface [*interface-type interface-num*] //Enter the interface view.
qos queue-profile [*queue-profile-name*] //Create a queue profile.
queue [*start-queue-index*] **to** [*end-queue-index*] **gts cir** [*cir-value*] [cbs *cbs-value*] //Configure traffic shaping for a specified queue in the outbound direction and set the CIR.

- Apply the queue profile to an interface.

system-view
interface [*interface-type interface-num*] //Enter the interface view.
qos queue-profile [*queue-profile-name*] //Apply the queue profile to the interface.

HUAWEI

Configuring MQC-based Traffic Shaping



- MQC-based traffic policing uses traffic classifiers to implement differentiated services.
- The configuration roadmap is as follows:
  - Configure a traffic classifier to match traffic.
  - Configure a traffic behavior to define an action for packets.
  - Bind the traffic classifier and traffic behavior to a traffic policy.
  - Apply the traffic policy to an interface in the outbound direction.

system-view
  traffic classifier [*classifier-name*] //Create a traffic classifier.
    if-match [*acl*| *vlan-id*| ....] //Match traffic based on traffic characteristics.

system-view
  traffic behavior [*behavior-name*] //Create a traffic behavior.
    gts cir [*cir-value*] | **pct** [*pct-value*] //Configure traffic shaping based on the maximum traffic rate or the percentage of the occupied interface bandwidth.

system-view
  traffic policy [*policy-name*] //Create a traffic policy.
    classifier [*classifier-name*] **behavior** [*behavior-name*] //Bind the traffic classifier to the traffic behavior.

system-view
  interface [*interface-type interface-num*] //Enter the interface view.
    traffic-policy [*policy-name*] [inbound | outbound] //Apply the traffic policy to the interface in the outbound direction.

HUAWEI

Checking the Traffic Shaping Configuration

- After queue-based traffic shaping is configured, you can run the following commands to check the configuration.

```
system-view
  display qos queue-profile [ queue-profile-name ]   //Check the queue profile configuration.
```

- After MQC-based traffic shaping is configured, you can run the following commands to check the configuration.

```
system-view
  display traffic classifier user-defined [ classifier-name ]   //Check the traffic classifier configuration.
  display traffic behavior [ system-defined | user-defined ] [ behavior-name ]   //Check the traffic behavior configuration.
  display traffic policy user-defined [ policy-name ] classifier [classifier-name ]   //Check the traffic policy configuration.
  display traffic-policy applied-record [ policy-name ]   //Check the record of the specified traffic policy.
```


HUAWEI

Quiz

1. (True or false) Traffic shaping caches excess traffic by default, and traffic policing discards excess traffic by default.( )
   A. True
   B. False
2. (Multiple-answer question) How many modes of token buckets are used to measure traffic?( )
   A. Single-rate-single-bucket
   B. Three-rate-two-bucket
   C. Single-rate-two-bucket
   D. Two-rate-two-bucket

HUAWEI

1. A
2. ACD

Section Summary

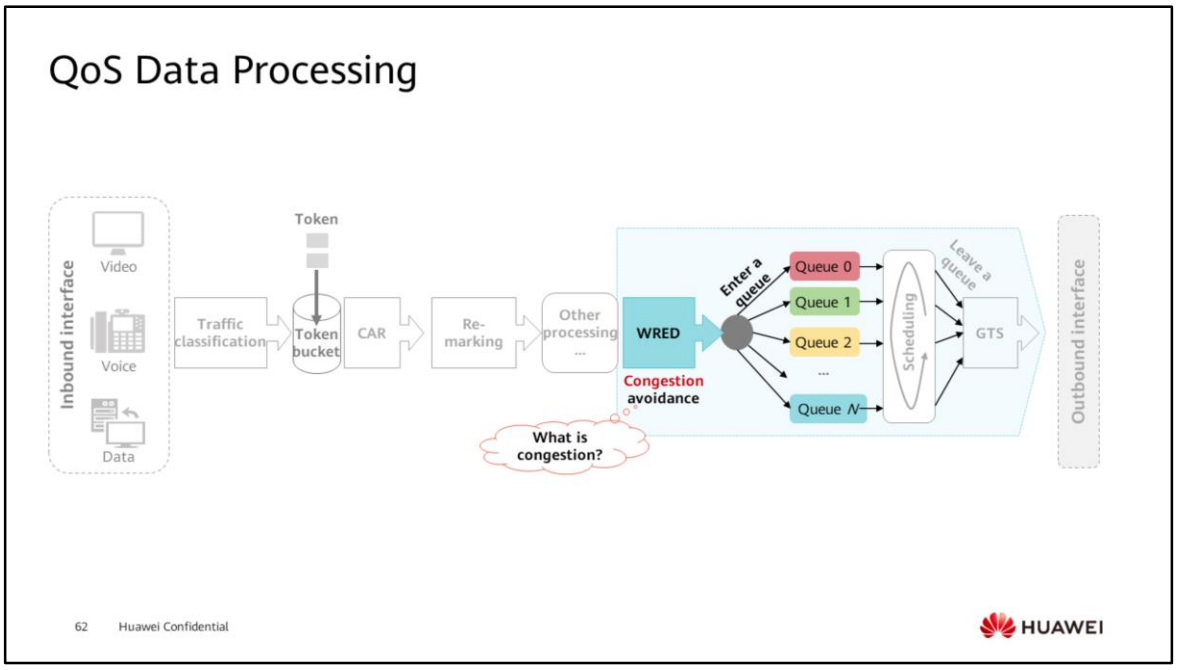
- There are two traffic limiting technologies: traffic policing and traffic shaping.
- Traffic policing discards excess traffic by default. It can be deployed in inbound and outbound directions of a device.
- Traffic shaping caches excess traffic by default. It can be deployed only in the outbound direction of a device.
- The device uses token buckets to measure traffic. There are three modes of token buckets:
  - The single-rate-single-bucket mechanism can be used together with traffic policing and traffic shaping.
  - The single-rate-two-bucket mechanism can be used only with traffic policing, and is mainly used in scenarios where burst traffic occurs occasionally.
  - The two-rate-two-bucket can be used only with traffic policing, and is mainly used in scenarios with long-term burst traffic.

HUAWEI

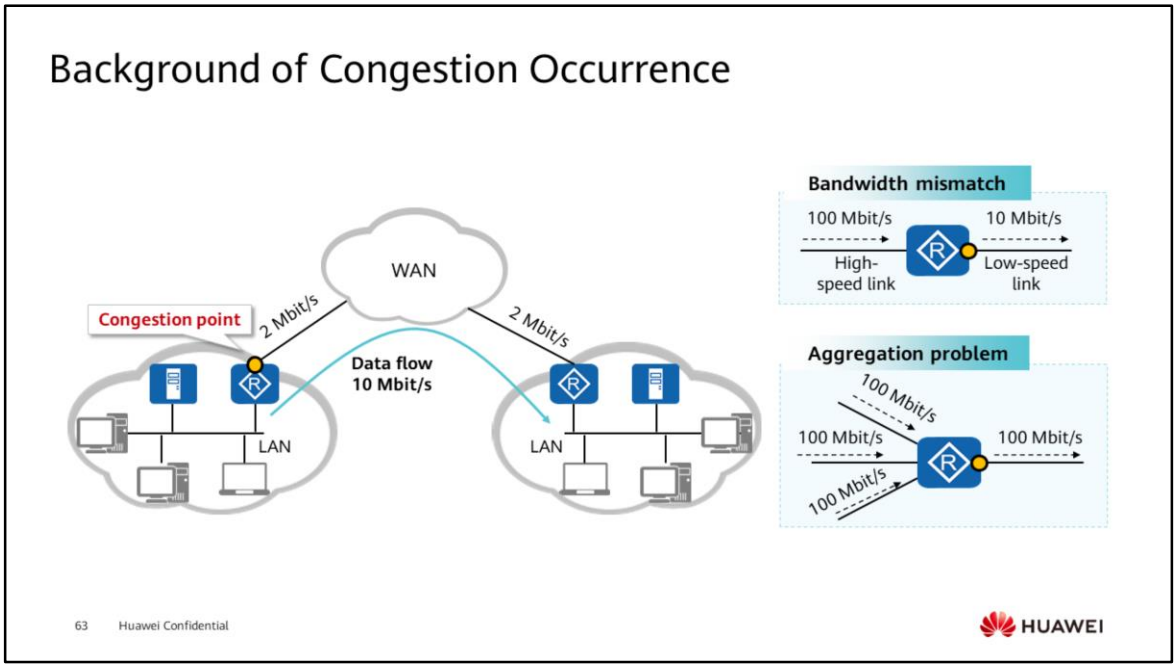
Contents

1. Introduction to QoS
2. Traffic Classification and Marking
3. Traffic Limiting Technology
4. **Congestion Avoidance Technology**
5. Congestion Management Technology
6. Introduction to HQoS

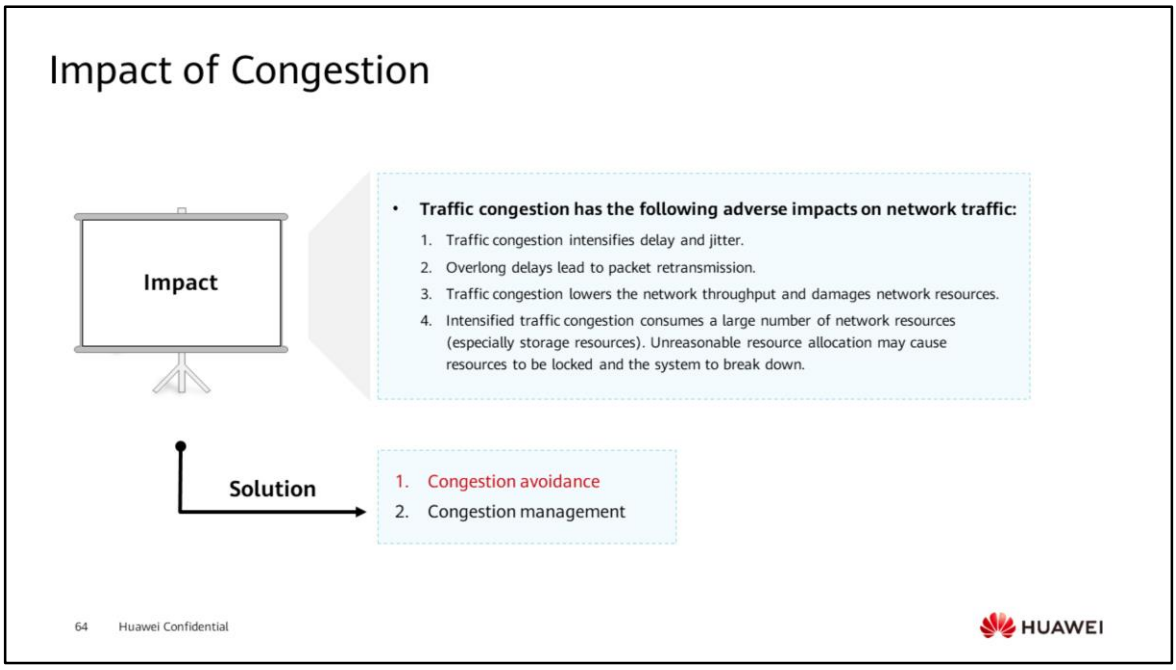
HUAWEI



- Congestion avoidance: Excessive congestion may damage network resources. Congestion avoidance monitors the usage of network resources. When congestion aggravates, congestion avoidance adjusts traffic to relieve network overload by discarding packets. Congestion avoidance is generally applied to the outbound direction of an interface.



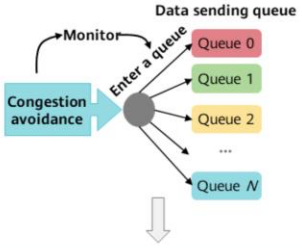
- Traffic congestion occurs when multiple users compete for the same resources (such as the bandwidth and buffer) on the shared network.
  - For example, a user on a LAN sends data to a user on another LAN through a WAN. The WAN bandwidth is lower than the LAN bandwidth. Therefore, data cannot be transmitted at the same rate on the WAN as that on the LAN. Traffic congestion occurs on the router connecting the LAN and WAN.
- Congestion often occurs in the following situations:
  - Traffic rate mismatch: Packets are transmitted to a device through a high-speed link and are forwarded out through a low-speed link.
  - Traffic aggregation: Packets are transmitted from multiple interfaces to a device and are forwarded out through a single interface without enough bandwidth.



- Impact of congestion:
  - Congestion prevents traffic from obtaining resources immediately, which causes service deterioration. However, congestion often occurs in a complex networking environment where packet transmission and provisioning of various services are both required. Therefore, effective methods are required to avoid congestion or prevent congestion from aggravating.
- Solutions:
  - The solutions need to make full use of network resources on the premise of meeting users' requirements for service quality. Congestion management and congestion avoidance are commonly used to relieve traffic congestion.
  - Congestion management provides means to manage and control traffic when traffic congestion occurs.
  - Congestion avoidance is a flow control technique used to relieve network overload. By monitoring the usage of network resources in queues or memory buffer, a device automatically drops packets on the interface that shows a sign of traffic congestion. Congestion avoidance prevents queues from being overflowed due to network overload. The following will introduce congestion avoidance technology.

Congestion Avoidance Technology

- Congestion avoidance is a flow control technique used to relieve network overload. By monitoring the usage of network resources for queues or memory buffers, a device automatically drops packets that shows a sign of traffic congestion.



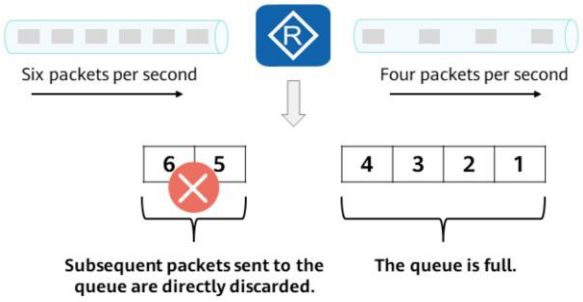
1. Tail drop: traditional processing
2. Random Early Detection (RED)
3. Weighted Random Early Detection (WRED)

- If congestion becomes severe, the packets that do not enter queues are discarded.

HUAWEI

Policy 1: Tail Drop

- When the length of a queue reaches the maximum value, the device enabled with tail drop discards all new packets buffered at the tail of the queue.

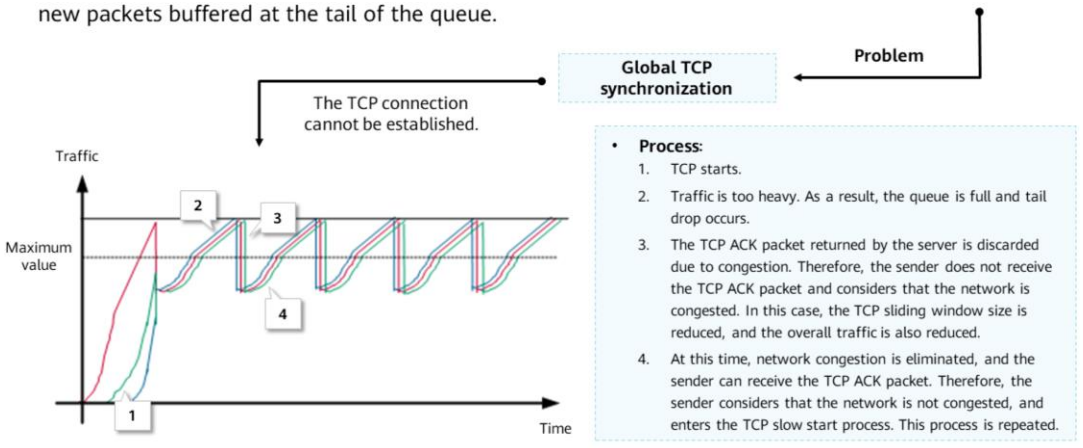


HUAWEI

- Due to the limited length of each queue, when a queue is full, the traditional processing method discards all the packets sent to the queue until the congestion is relieved. This processing method is called tail drop.

Disadvantage 1: Global TCP Synchronization

- When the length of a queue reaches the maximum value, the device enabled with tail drop discards all new packets buffered at the tail of the queue.

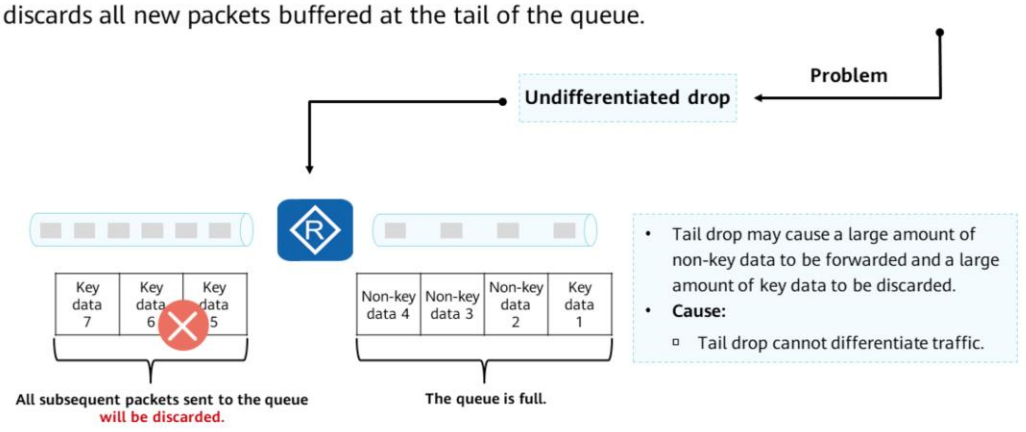


HUAWEI

- As shown in the following figure, three colors indicate three TCP connections.
- Global TCP synchronization:
  - In tail drop mechanism, all newly arrived packets are dropped when congestion occurs, causing all TCP sessions to simultaneously enter the slow start state and the packet transmission to slow down.
  - When packets of multiple TCP connections are discarded in a queue, TCP connections enter the congestion avoidance and slow start state to adjust and reduce traffic. This is called TCP global synchronization. Then all TCP sessions restart their transmission at roughly the same time and then congestion occurs again, causing another burst of packet drops, and all TCP sessions enter the slow start state again. The behavior cycles constantly, severely reducing the network resource usage.

Disadvantage 2: Undifferentiated Drop

- When the length of a queue reaches the maximum value, the device enabled with tail drop discards all new packets buffered at the tail of the queue.



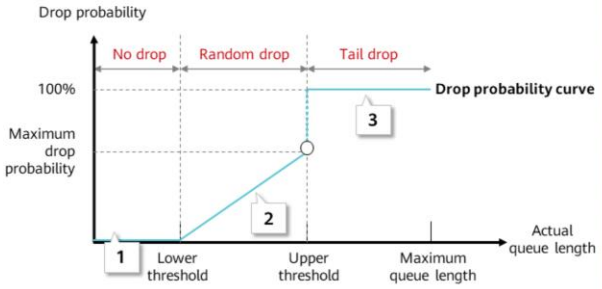
- Tail drop may cause a large amount of non-key data to be forwarded and a large amount of key data to be discarded.
- **Cause:**
  - Tail drop cannot differentiate traffic.

HUAWEI

- Tail drop cannot differentiate services and discard traffic in the same way.

Policy 2: RED

- Random early detection (RED) randomly discards data packets.



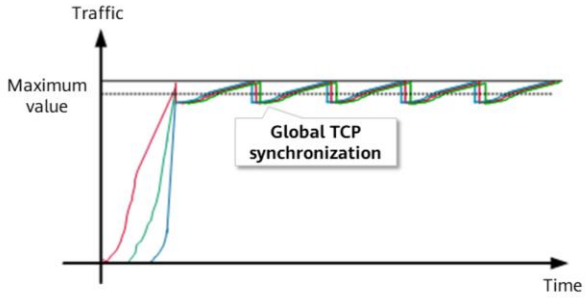
- **Process:**
  1. When the queue length is less than the lower threshold, no packets are discarded.
  2. When the queue length is between the upper threshold and the lower threshold, newly arrived packets are randomly discarded. The longer the queue is, the higher the drop probability is.
  3. All coming packets are discarded if the queue length is greater than the upper threshold.

HUAWEI

- RED defines upper and lower thresholds for the length of each queue:
  - When the queue length is less than the lower threshold, no packets are discarded.
  - When the queue length is greater than the upper drop threshold, all packets are discarded.
  - Coming packets are dropped randomly if the queue length is between upper and lower thresholds. RED generates a random number for each incoming packet and compares it with the drop probability of the current queue. If the random number is greater than the drop probability, the packet is discarded. A longer queue indicates a higher drop probability.

Relieving Global TCP Synchronization

- RED randomly discards packets so that rates of TCP connections are reduced at different times. This prevents global TCP synchronization.



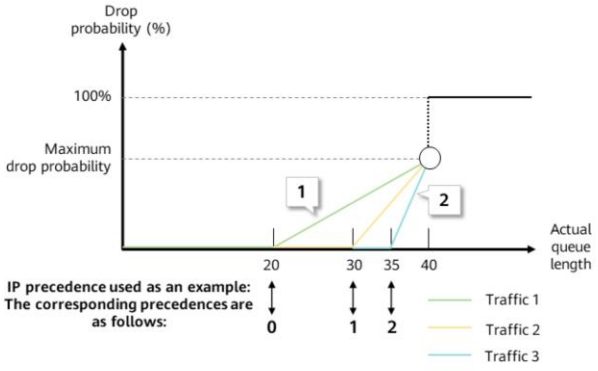
- **Symptom:**
  - Global TCP synchronization may still occur, but the link usage is greatly increased.
- **Disadvantage:**
  - RED cannot distinguish traffic.

HUAWEI

- RED is used to avoid global TCP synchronization that occurs with tail drop. It does this by randomly discarding packets so that the transmission speed of multiple TCP connections is not reduced simultaneously. This results in more stable rates of TCP traffic and other network traffic. — Do not adjust TCP sliding window sizes simultaneously.

Policy 3: WRED

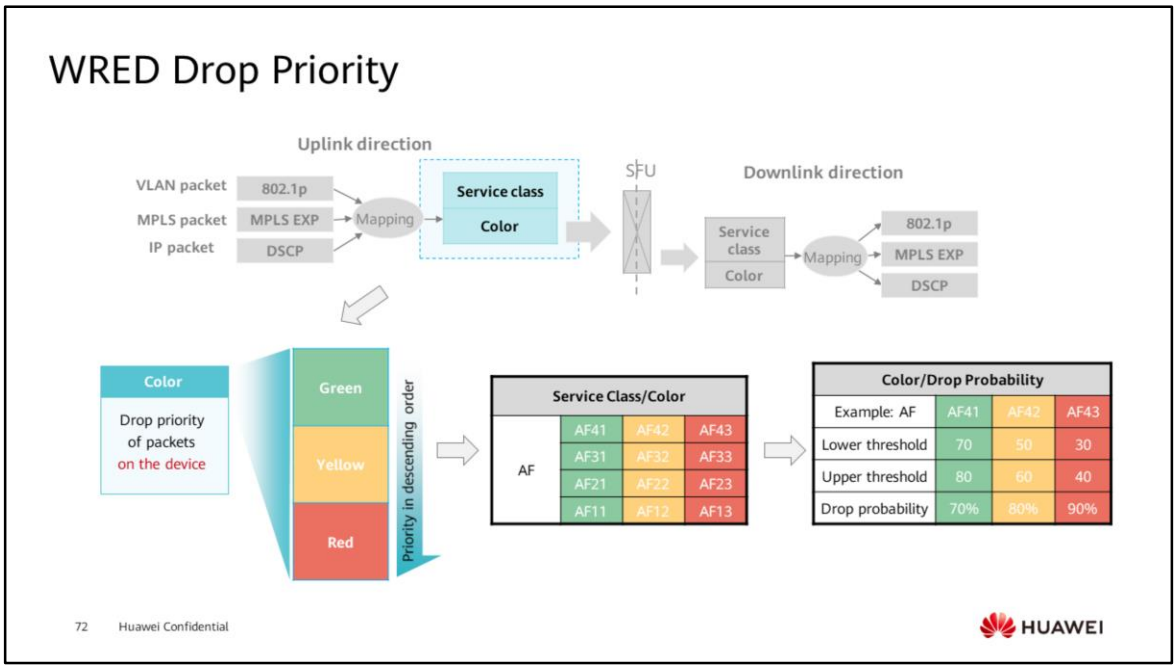
- Weighted Random Early Detection (WRED) sets different drop policies for data packets or queues with different priorities to discard different types of traffic.



- **Example:**
  1. The lower threshold is 20 and the upper threshold is 40 for the traffic whose IP precedence is 0.
  2. The lower threshold is 35 and the upper threshold is 40 for the traffic whose IP precedence is 2. The traffic whose IP precedence is 2 is discarded later than the traffic whose IP precedence is 0.
- **Advantage:**
  1. Do not adjust TCP sliding window sizes simultaneously to avoid global TCP synchronization.
  2. Different traffic is discarded based on weights.

HUAWEI

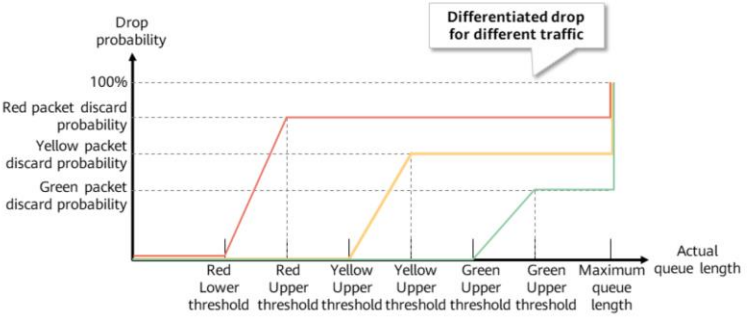
- The device provides WRED based on RED technology.
- WRED discards packets in queues based on DSCP priorities or IP priorities. The upper and lower thresholds, and drop probability can be set for each priority. When the number of packets of a priority reaches the lower threshold, the device starts to discard packets. When the number of packets reaches the upper threshold, the device discards all the packets. As the queue length increases, the packet loss rate increases. The maximum packet loss rate does not exceed the preset packet loss rate. WRED discards packets in queues based on the drop probability, thereby preventing congestion to a certain degree. — Do not adjust TCP sliding window sizes simultaneously.



- Color:
  - The color of packets determines the order in which packets are dropped in a congested queue.
- Application:
  - The WRED lower threshold is recommended to start from 50% and change with the drop priority. The lowest drop probability and highest lower and upper thresholds are recommended for green packets; a medium drop probability and medium lower and upper thresholds are recommended for yellow packets; the highest drop probability and smallest lower and upper thresholds are recommended for red packets.
  - When traffic congestion aggravates, red packets are first dropped due to the smallest lower threshold and high drop probability. As the queue length increases, the device drops green packets at last. If the queue length reaches the upper threshold for red/yellow/green packets, red/yellow/green packets start to be tail dropped.

Curve of the WRED Drop Probability

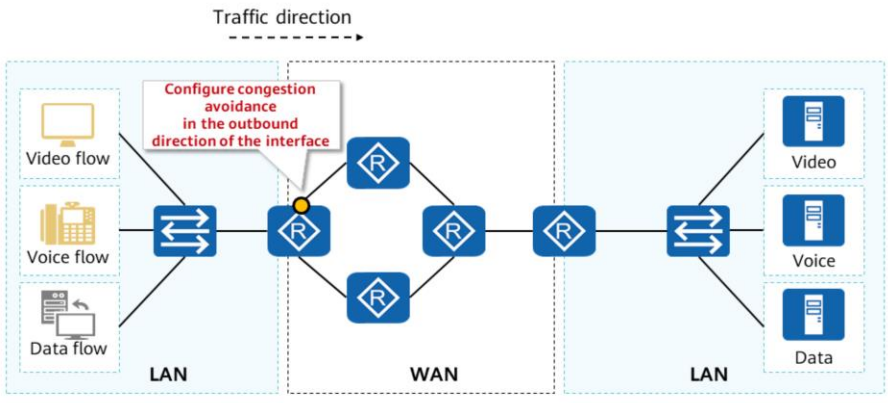
Color/Drop Probability			
Example: AF	AF41	AF42	AF43
Lower threshold	70	50	30
Upper threshold	80	60	40
Drop probability	70%	80%	90%



HUAWEI

- Color:
  - The color of packets determines the order in which packets are dropped in a congested queue.
- Application:
  - The WRED lower threshold is recommended to start from 50% and change with the drop priority. The lowest drop probability and highest lower and upper thresholds are recommended for green packets; a medium drop probability and medium lower and upper thresholds are recommended for yellow packets; the highest drop probability and smallest lower and upper thresholds are recommended for red packets.
  - When traffic congestion aggravates, red packets are first dropped due to the smallest lower threshold and high drop probability. As the queue length increases, the device drops green packets at last. If the queue length reaches the upper threshold for red/yellow/green packets, red/yellow/green packets start to be tail dropped.

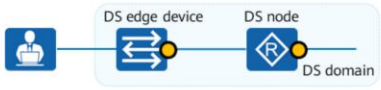
Application of Congestion Avoidance



HUAWEI

- Example:
  - Users in different LANs may upload data to the same server, so data exchanged between users and the server passes the WAN. Because WAN bandwidth is lower than LAN bandwidth, congestion may occur on the edge device between the WAN and LANs. Congestion avoidance can be configured on the edge device to discard low-priority packets such as data packets, reducing network overload and ensuring forwarding of high-priority services.

Configuring Queue-based WRED

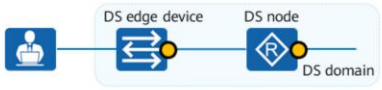


- The device supports WRED based on DSCP priorities or IP priorities. The configuration roadmap is as follows:
  - Configure a drop profile.
  - Configure WRED parameters.
  - Reference the drop profile to a queue profile.
  - Apply the queue profile to the outbound direction of the interface.

system-view
drop-profile [*drop-profile-name*] //Create a drop profile.
wred [dscp | ip-precedence] //Configure a WRED drop profile based on DSCP or IP priorities.
dscp [*dscp-value*] **low-limit** [*low-limit-percentage*] **high-limit** [*high-limit-percentage*] **discard-percentage** [*discard-percentage*] //Configure WRED parameters based on DSCP priorities.
ip-precedence [*ip-precedence-value*] **low-limit** [*low-limit-percentage*] **high-limit** [*high-limit-percentage*] **discard-percentage** [*discard-percentage*] //(Optional) Configure WRED parameters based on IP priorities.
qos queue-profile [*queue-profile-name*] //Enter the queue profile view.
queue [*queue-index*] **drop-profile** [*drop-profile-name*] //Bind the drop profile to the specified queue in the queue profile.
interface [*interface-type interface-num*] //Enter the interface view.
qos queue-profile [*queue-profile-name*] //Apply the queue profile to the interface.

HUAWEI

Configuring MQC to Implement Congestion Avoidance (1)

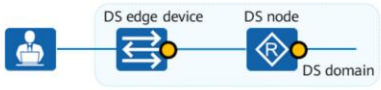


- After a drop profile is bound to a traffic behavior, associate the traffic behavior with the corresponding traffic classifier in the traffic policy and apply the traffic policy to an interface to implement congestion avoidance for traffic matching the traffic classifier. The configuration roadmap is as follows:
  - Configure a drop profile.
  - Configure a traffic classifier and a traffic behavior.
  - Bind the traffic classifier and traffic behavior to a traffic policy.
  - Apply the traffic policy to the outbound direction of the device interface.

system-view
drop-profile [*drop-profile-name*] //Create a drop profile.
wred [dscp | ip-precedence] //Configure a WRED drop profile based on DSCP or IP priorities.
dscp [*dscp-value*] **low-limit** [*low-limit-percentage*] **high-limit** [*high-limit-percentage*] **discard-percentage** [*discard-percentage*] //Configure WRED parameters based on DSCP priorities.
ip-precedence [*ip-precedence-value*] **low-limit** [*low-limit-percentage*] **high-limit** [*high-limit-percentage*] **discard-percentage** [*discard-percentage*] //(Optional) Configure WRED parameters based on IP priorities.

HUAWEI

Configuring MQC to Implement Congestion Avoidance (2)



- After a drop profile is bound to a traffic behavior, associate the traffic behavior with the corresponding traffic classifier in the traffic policy and apply the traffic policy to an interface to implement congestion avoidance for traffic matching the traffic classifier. The configuration roadmap is as follows:
  - Configure a drop profile.
  - Configure a traffic classifier and a traffic behavior.
  - Bind the traffic classifier and traffic behavior to a traffic policy.
  - Apply the traffic policy to the outbound direction of the device interface.

system-view
  traffic classifier [*classifier-name*] //Create a traffic classifier.
    if-match [*acl* | *vlan-id* | ....] //Match traffic based on traffic characteristics.

system-view
  traffic behavior [*behavior-name*] //Create a traffic behavior.
    drop-profile [*drop-profile-name*] //Bind the created drop profile to the traffic behavior.

system-view
  traffic policy [*policy-name*] //Create a traffic policy.
    classifier [*classifier-name*] **behavior** [*behavior-name*] //Bind the traffic classifier to the traffic behavior.

system-view
  interface [*interface-type interface-num*] //Enter the interface view.
    traffic-policy [*policy-name*] **outbound** //Apply the traffic policy to the outbound direction of the interface.

HUAWEI

Checking the Congestion Avoidance Configuration

- Checking the queue-based congestion avoidance configuration

```
system-view
  interface [interface-type interface-num]
    display this    //Check the queue profile bound to the interface.
  qos queue-profile [queue-profile-name]
    display this   //Check the drop profile bound to the queue profile.
  display drop-profile [ drop-profile-name ]   //Check the drop profile configuration.
```

- Checking the MQC-based congestion avoidance configuration

```
system-view
  display traffic classifier user-defined [ classifier-name ]   //Check the traffic classifier configuration.
  display traffic behavior [ system-defined | user-defined ] [ behavior-name ]   //Check the traffic behavior configuration.
  display traffic policy user-defined [ policy-name ] classifier [classifier-name ]    //Check the traffic policy configuration.
  display traffic-policy applied-record [ policy-name ]   //Check the record of the specified traffic policy.
```


HUAWEI

Quiz

1. (Multiple-answer question) Which of the following mechanisms are used by QoS to proactively discard packets?( )
   A. Tail drop
   B. RED
   C. MRED
   D. WRED

HUAWEI

1. ABD

Section Summary

- Congestion avoidance technology cannot avoid congestion, but prevents problems caused by congestion. For example, tail drop causes global TCP synchronization, interface traffic is unstable, and UDP traffic preempts the bandwidth used by TCP traffic.
- RED/WRED randomly discards data packets to prevent problems such as global TCP synchronization.

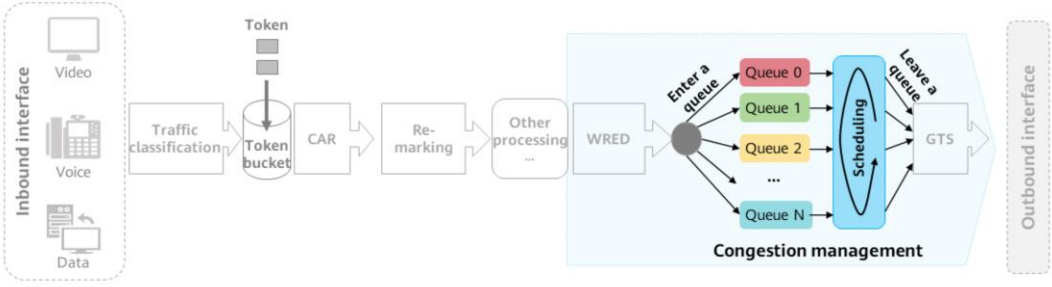
HUAWEI

Contents

1. Introduction to QoS
2. Traffic Classification and Marking
3. Traffic Limiting Technology
4. Congestion Avoidance Technology
5. **Congestion Management Technology**
6. Introduction to HQoS

HUAWEI

QoS Data Processing

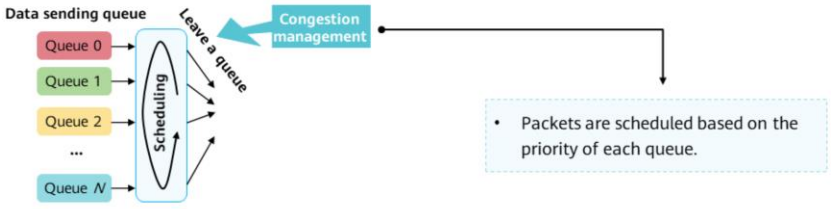


HUAWEI

- Congestion management: It is a measure that must be taken to solve the problem of resource competition. Packets are buffered in queues and a scheduling algorithm is used to determine the forwarding sequence of packets. Congestion management is usually applied to the outbound direction of an interface.

Congestion Management Technology

- Congestion management technology manages and controls different types of service traffic when network congestion occurs.
- It uses queue scheduling technology to handle traffic congestion.



HUAWEI

- Congestion Management Technology
- Congestion management defines a policy that determines the order in which packets are forwarded and specifies drop principles for packets. The queuing technology is used.
- The queue scheduling algorithm determines the order in which packets are leaving a queue and the relationships between queues.
- Queuing technology
- Packets sent from one interface are placed into many queues which are identified with different priorities. The packets are then sent based on the priorities. Different queue scheduling mechanisms are designed for different situations and lead to varying results.

What Is a Queue?

- The queuing technology orders packets in the buffer.

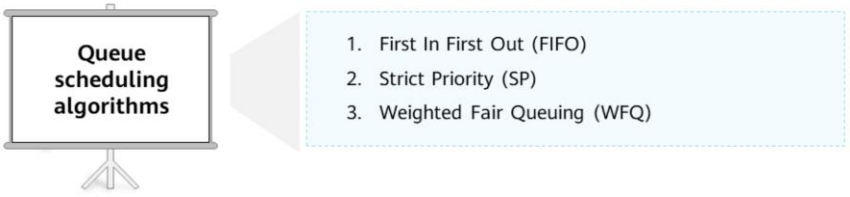
- Each interface has eight downlink queues, which are called class queues (CQs) or port queues.
- They are EF, AF1, AF2, AF3, AF4, BE, CS6, and CS7.

HUAWEI

- What is a queue?
  - The queuing technology orders packets in the buffer. When the packet rate exceeds the interface bandwidth or the bandwidth configured for packets, the packets are buffered in queues and wait to be forwarded.
  - Each interface on the NE20E or NE40E stores eight downlink queues, which are called CQs or port queues. The eight queues are BE, AF1, AF2, AF3, AF4, EF, CS6, and CS7.

Queue Scheduling Algorithms

- Congestion management uses the queuing technology.

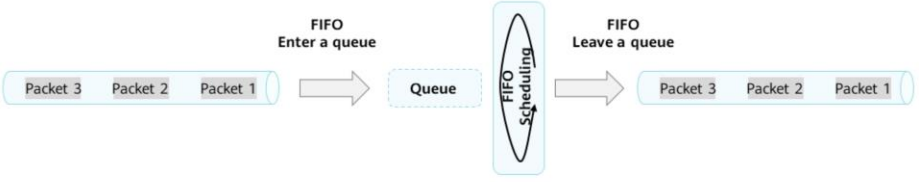


HUAWEI

- Queuing technology places packets sent from one interface into multiple queues with different priorities. These packets are then sent based on the priorities. Different queue scheduling mechanisms are designed for different situations and lead to varying results.

FIFO

- The FIFO mechanism is used to transfer packets in a queue. Resources used to forward packets are allocated based on the arrival order of packets.

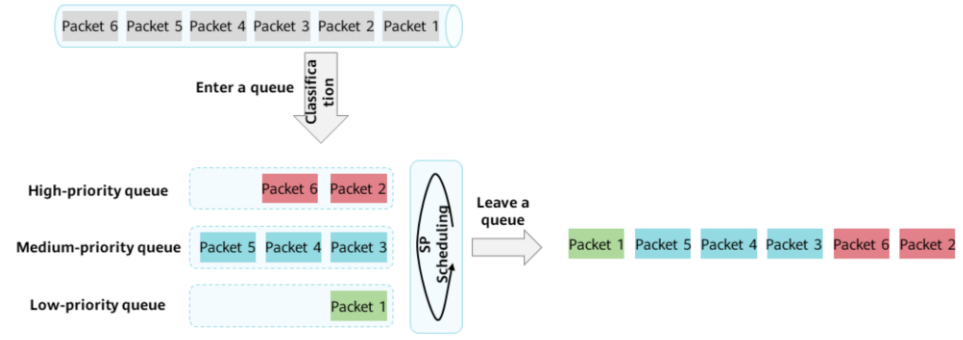


HUAWEI

- FIFO does not classify packets.
- FIFO allows the packets that come earlier to enter the queue first. On the exit of a queue, FIFO allows the packets to leave the queue in the same order as that in which the packets enter the queue.
- Characteristics:
  - Advantage: The implementation mechanism is simple and the processing speed is fast.
  - Disadvantage: Packets with different priorities cannot be processed in differentiated ways.

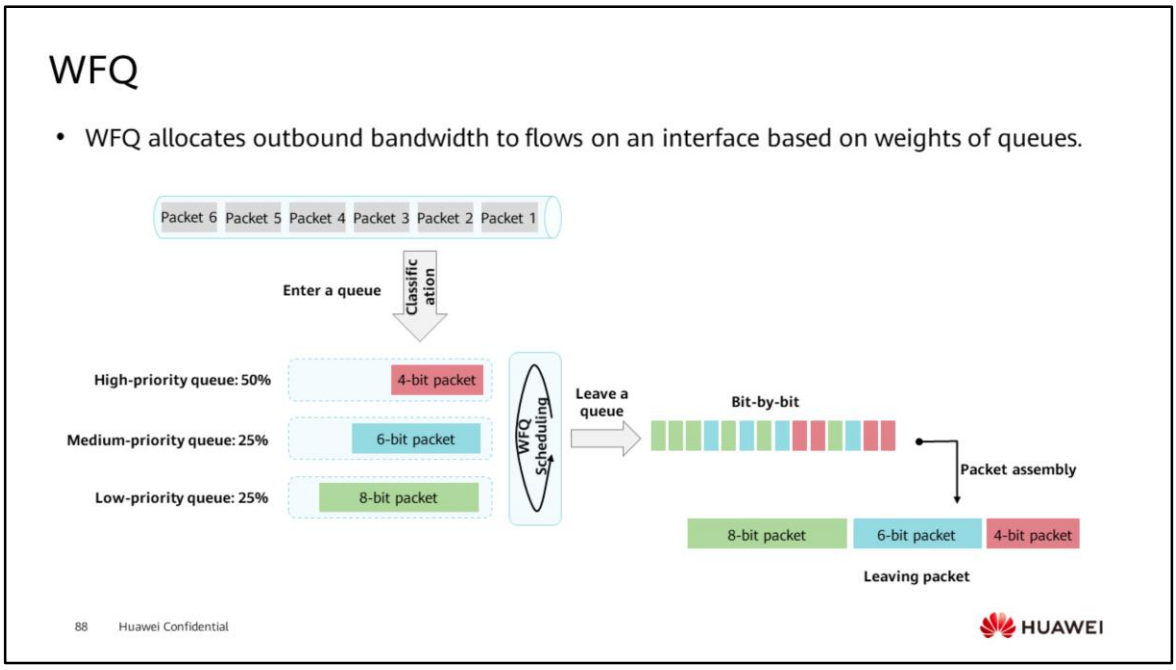
SP

- SP schedules packets strictly based on queue priorities.



HUAWEI

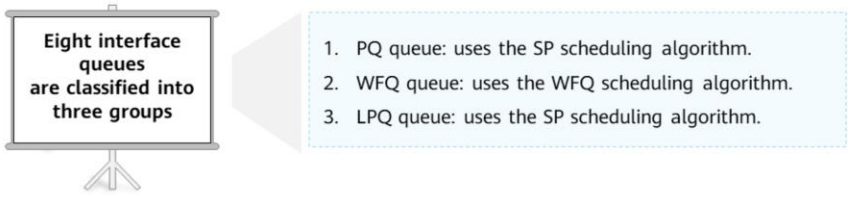
- SP: Packets in queues with a low priority can be scheduled only after all packets in queues with a higher priority are scheduled.
- As shown in the figure, three queues with a high, medium, and low priorities respectively are configured with SP scheduling. The number indicates the order in which packets arrive.
- When packets leave queues, the device forwards the packets in descending order of priority. Packets in the higher-priority queue are forwarded preferentially. If packets in the higher-priority queue come in between packets in the lower-priority queue that is being scheduled, the packets in the high-priority queue are still scheduled preferentially. This implementation ensures that packets in the higher-priority queue are always forwarded preferentially. As long as there are packets in the high-priority queue, no other queue will be served.
- Characteristics:
  - Advantage: High-priority packets are preferentially forwarded.
  - Disadvantage: Low-priority queues may be starved out. That is, when congestion occurs, packets in lower-priority queues are not processed until all the higher-priority queues are empty. As a result, a congested higher-priority queue causes all lower-priority queues to starve out.



- WFQ allocates bandwidths to flows based on weights of queues. In addition, to fairly allocate bandwidths to flows, WFQ schedules packets bit by bit.
- When packets leave queues, the device forwards the packets in descending order of priority. Packets in the higher-priority queue are forwarded preferentially. If packets in the higher-priority queue come in between packets in the lower-priority queue that is being scheduled, the packets in the high-priority queue are still scheduled preferentially. This implementation ensures that packets in the higher-priority queue are always forwarded preferentially. As long as there are packets in the high-priority queue, no other queue will be served.
- Characteristics:
    - Advantages:
        - Packets in different queues are scheduled fairly, and the flow delays have slight differences.
        - If many large and small packets in different queues need to be sent, small packets are scheduled first, reducing the total jitter of each flow.
        - The smaller the weight, the less the allocated bandwidth. Flows with larger weights are allocated higher bandwidth.
    - Disadvantage: Low-latency services cannot be scheduled in a timely manner. User-defined classification rules cannot be implemented.

Queue Scheduling Mode of an Interface

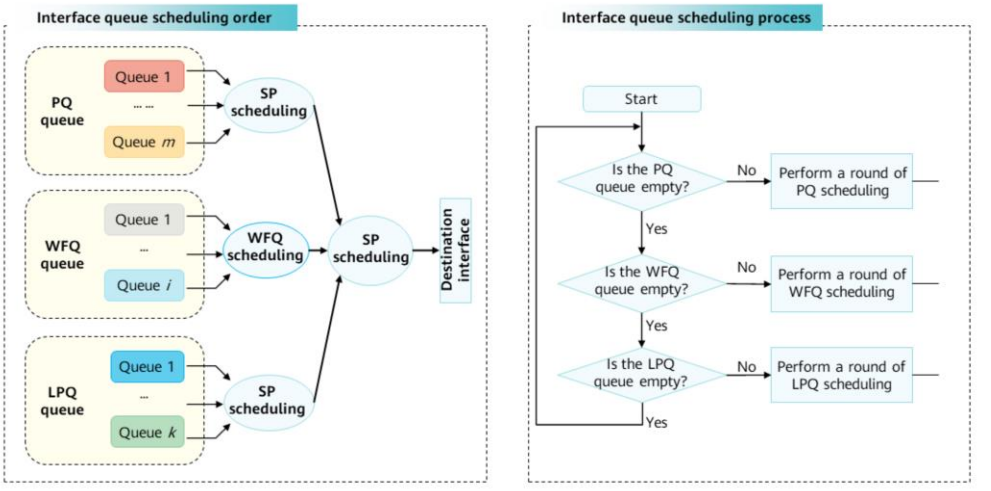
- You can configure SP scheduling or WFQ scheduling for eight queues on an interface.
- Eight queues can be classified into three groups, priority queuing (PQ) queues, WFQ queues, and low priority queuing (LPQ) queues, based on scheduling algorithms.



HUAWEI

- PQ queue
  - SP scheduling applies to PQ queues. Packets in high-priority queues are scheduled preferentially. Therefore, services that are sensitive to delays (such as VoIP) can be configured with high priorities.
  - In PQ queues, however, if the bandwidth of high-priority packets is not restricted, low-priority packets cannot obtain bandwidth and are starved out.
  - Generally, services that are sensitive to delays are put into PQ queues.
- WFQ queue
  - WFQ queues are scheduled based on weights. The WFQ scheduling algorithm can be used to allocate the remaining bandwidth based on weights.
- LPQ queue
  - LPQ is a queue scheduling mechanism that is implemented on a high-speed interface (such as an Ethernet interface). LPQ is not supported on a low-speed interface (such as a serial interface or MP-group interface).
  - SP scheduling applies to LPQ queues. The difference is that when congestion occurs, the PQ queue can preempt the bandwidth of the WFQ queue whereas the LPQ queue cannot. After packets in the PQ and WFQ queues are all scheduled, the remaining bandwidth can be assigned to packets in the LPQ queue.

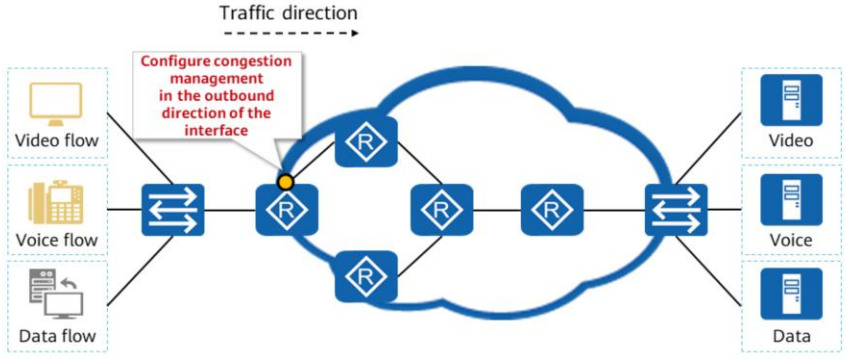
Scheduling Order of Three Types of Queues



HUAWEI

- Interface queue scheduling order:
  - If PQ, WFQ, and LPQ queues use SP scheduling. PQ, WFQ, and LPQ queues are scheduled in sequence.
- Interface queue scheduling process:
  - Packets in PQ queues are preferentially scheduled, and packets in WFQ queues are scheduled only when no packets are buffered in PQ queues. When all PQ queues are empty, WFQ queues start to be scheduled. Packets in PQ queues are preferentially scheduled,
  - and packets in WFQ queues are scheduled only when no packets are buffered in PQ queues. Bandwidths are preferentially allocated to PQ queues to guarantee the PIR of packets in PQ queues.
  - Packets in LPQ queues are scheduled only after all packets in WFQ queues are sent.
- Scheduling result:
  - The PIR of PQ queues is guaranteed first, and the remaining bandwidth is allocated among WFQ queues based on weights.
  - When the PIR of all WFQ queues is guaranteed, the remaining bandwidth is allocated to LPQ queues.

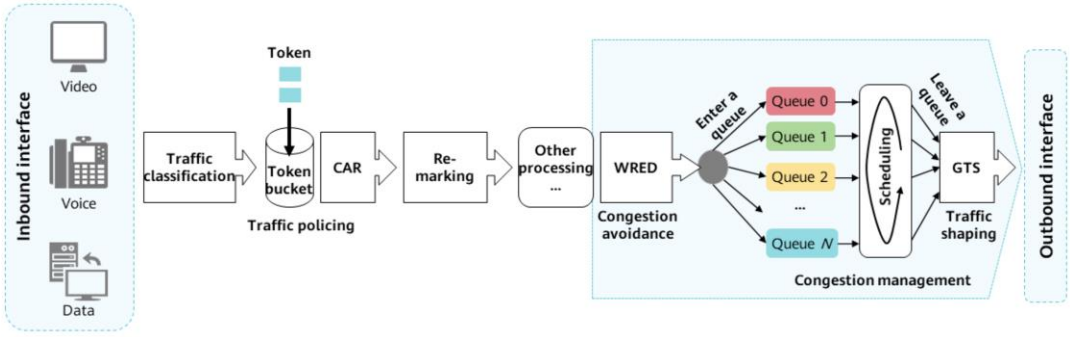
Application of Congestion Management



HUAWEI

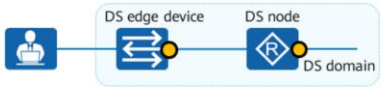
- Example:
  - On a network, when multiple services compete for the same resources (such as the bandwidth and buffer), traffic congestion may occur and high-priority services may be not processed in a timely manner. Packets can be sent to different queues according to the priority mapping result, as shown in the figure. Different scheduling modes are set in the outbound direction to implement differentiated services.

QoS Processing Review



HUAWEI

Configuring Queue-based Congestion Management

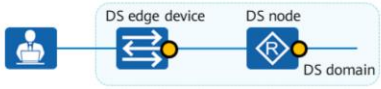


- WAN interfaces support three scheduling modes: PQ, WFQ, and PQ+WFQ. The configuration roadmap is as follows:
  - Create a queue profile.
  - Configure scheduling modes.
  - Apply the queue profile to the interface.

system-view
  qos queue-profile [*queue-profile-name*] //Create a queue profile.
    schedule pq [*queue-index*] | **wfq** [*queue-index*] //Configure scheduling modes for queues on a WAN interface.
  interface [*interface-type interface-num*] //Enter the interface view.
    qos queue-profile [*queue-profile-name*] //Apply the queue profile to the interface.

HUAWEI

Configuring MQC to Implement Congestion Management (1)



- MQC provides three types of queues:
  - Assured Forwarding (AF) queues
  - Expedited Forwarding (EF) or LLQ queues
  - BE queues
- The configuration roadmap is as follows:
  - Configure a traffic classifier and a traffic behavior.
  - Bind the traffic classifier and traffic behavior to a traffic policy.
  - Apply the traffic policy to the outbound direction of the device interface.

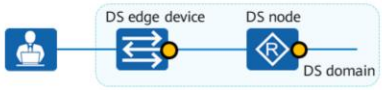
system-view
traffic classifier [*classifier-name*] //Create a traffic classifier.
if-match [*acl*| *vlan-id*| ....] //Match traffic based on traffic characteristics.

system-view
traffic behavior [*behavior-name*] //Create a traffic behavior.
queue af bandwidth [*bandwidth*| **pct** *percentage*] //Configure the minimum bandwidth for AF queues in the traffic behavior.
queue ef bandwidth [*bandwidth*| **pct** *percentage*] //Configure the minimum bandwidth for EF queues in the traffic behavior.
queue llq bandwidth [*bandwidth*| **pct** *percentage*] //Configure the maximum bandwidth for LLQ queues in the traffic behavior.
queue wfq queue-number [*total-queue-number*] //Configure WFQ scheduling parameters for BE queues in the traffic behavior.

HUAWEI

- AF queue: AF queues ensure that service traffic is forwarded when the traffic rate does not exceed the minimum bandwidth.
- EF/LLQ queue: After packets matching certain rules enter EF or LLQ queues, they are scheduled in SP mode. Packets in other queues are scheduled only after all the packets in EF or LLQ queues are scheduled. In addition, EF queues can use the available bandwidth in AF or BE queues. The latency of LLQ queues is lower than that of common EF queues.
- BE queue: The remaining packets that do not enter AF or EF queues enter BE queues. BE queues are scheduled using the WFQ algorithm.
- The total bandwidth used by AF queues and EF queues cannot exceed 100% of the interface bandwidth.
- EF queues are provided with bandwidth preferentially. AF queues share the remaining bandwidth based on their weights.

Configuring MQC to Implement Congestion Management (2)



- MQC provides three types of queues:
  - AF queues
  - EF/LLQ queues
  - BE queues
- The configuration roadmap is as follows:
  - Configure a traffic classifier and a traffic behavior.
  - Bind the traffic classifier and traffic behavior to a traffic policy.
  - Apply the traffic policy to the outbound direction of the device interface.

```
system-view
  traffic policy [policy-name]    //Create a traffic policy.
    classifier [classifier-name] behavior [behavior-name]
//Bind the traffic classifier to the traffic behavior.
```

```
system-view
  interface [interface-type interface-num]    //Enter the
interface view.
    traffic-policy [policy-name] outbound    //Apply the traffic
policy to the outbound direction of the interface.
```


HUAWEI

- AF queue: AF queues ensure that service traffic is forwarded when the traffic rate does not exceed the minimum bandwidth.
- EF/LLQ queue: After packets matching certain rules enter EF or LLQ queues, they are scheduled in SP mode. Packets in other queues are scheduled only after all the packets in EF or LLQ queues are scheduled. In addition, EF queues can use the available bandwidth in AF or BE queues. The latency of LLQ queues is lower than that of common EF queues.
- BE queue: The remaining packets that do not enter AF or EF queues enter BE queues. BE queues are scheduled using the WFQ algorithm.
- The total bandwidth used by AF queues and EF queues cannot exceed 100% of the interface bandwidth.
- EF queues are provided with bandwidth preferentially. AF queues share the remaining bandwidth based on their weights.

Checking the Congestion Management Configuration

- Checking the queue-based congestion management configuration

```
system-view
  interface [interface-type interface-num]
    display this    //Check the queue profile bound to the interface.
  display qos queue-profile [queue-profile-name]   //Check the queue profile configuration.
```

- Checking the traffic classifier-based congestion management configuration

```
system-view
  display traffic classifier user-defined [ classifier-name ]   //Check the traffic classifier configuration.
  display traffic behavior [ system-defined | user-defined ] [ behavior-name ]   //Check the traffic behavior configuration.
  display traffic policy user-defined [ policy-name ] classifier [classifier-name ]   //Check the traffic policy configuration.
  display traffic-policy applied-record [ policy-name ]   //Check the record of the specified traffic policy.
```


HUAWEI

Quiz

1. (Single-answer question) How many queues are there on an interface?( )
   A. 6
   B. 7
   C. 8
   D. 9
2. (Multiple-answer question) Which of the following is a queue scheduling technology?( )
   A. PQ
   B. WFQ
   C. WRED
   D. FIFO

HUAWEI

1. C
2. ABD

Section Summary

- After a data packet enters a queue, the device sends the data packet according to the queue scheduling mechanism.
- Common queue scheduling technologies include FIFO, PQ, and WFQ.
- PQ scheduling is performed before WFQ scheduling and FIFO. Queues scheduled in WFQ mode can transmit data only when queues scheduled in PQ mode have no data to transmit. The queue scheduled in FIFO mode can transmit data only when queues scheduled in PQ and WFQ mode have no data to transmit.

HUAWEI

Contents

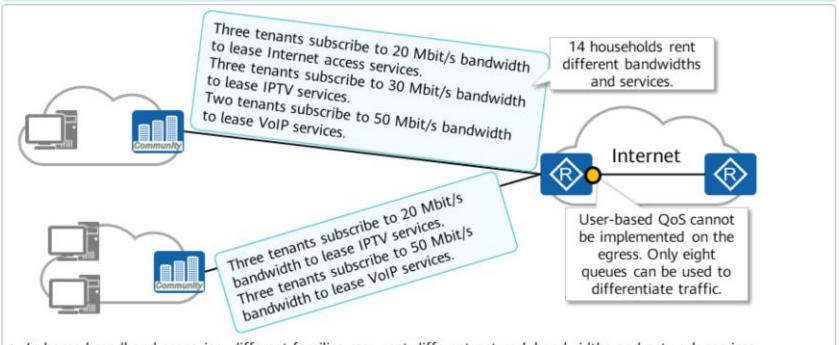
1. Introduction to QoS
2. Traffic Classification and Marking
3. Traffic Limiting Technology
4. Congestion Avoidance Technology
5. Congestion Management Technology
6. **Introduction to HQoS**

HUAWEI

Limitations of QoS

- Traditional QoS distributes a flow into only eight queues for scheduling and control. Therefore, it has great limitations in multi-tenant scenarios.

QoS limitations in home broadband scenarios



- In home broadband scenarios, different families may rent different network bandwidths and network services. Therefore, QoS cannot manage these families in a refined manner.

HUAWEI

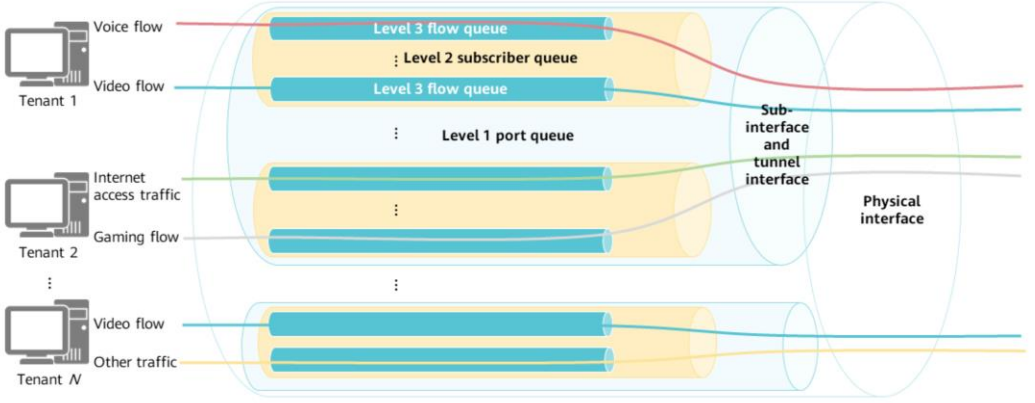
HQoS Overview

- Traditional QoS schedules traffic based on interfaces. An interface can only differentiate service priorities. The traffic of the same priority uses the same interface queue and competes for the same queue resources. Therefore, traditional QoS technology cannot provide differentiated services based on types of traffic and users.
- HQoS meets this requirement by implementing hierarchical scheduling based on multiple levels of queues, differentiating both services and users to provide refined QoS guarantee.
- Different devices provide different HQoS features. This section describes HQoS features supported by the CPE (AR series router).

HUAWEI

Introduction to HQoS Queues

- The CPE supports three-level queues: flow queue (level 3), subscriber queue (level 2), and port queue (level 1).

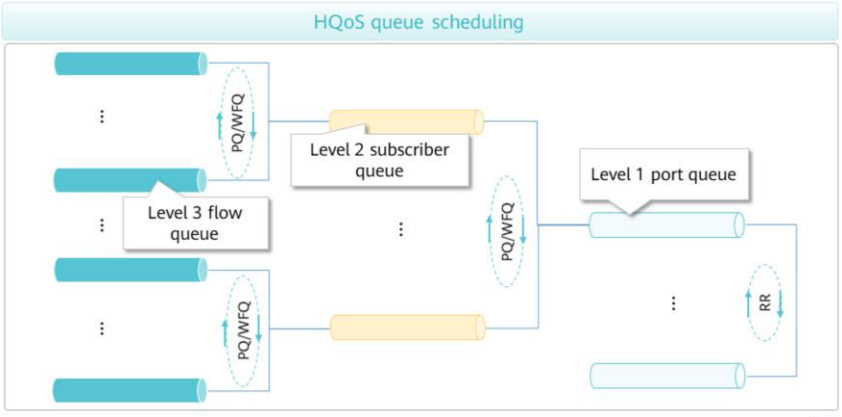


HUAWEI

- Flow queue
  - The same type of services of a user is taken as a service flow. HQoS schedules queues based on service flows. Flow queues correspond to service types and are classified into EF, AF, and BE queues. You can set scheduling modes for flow queues.
- Subscriber queue
  - Services from a user are placed into a subscriber queue. HQoS allows all services in the subscriber queue to share the bandwidth.
- Port queue
  - Each port corresponds to a queue and port queues are scheduled in RR mode. You can configure only interface-based traffic shaping, but cannot configure scheduling modes.

Introduction to HQoS Queue Scheduling

- The flow queue and subscriber queue support PQ scheduling, WFQ scheduling, and PQ+WFQ scheduling. The port queue uses RR scheduling.

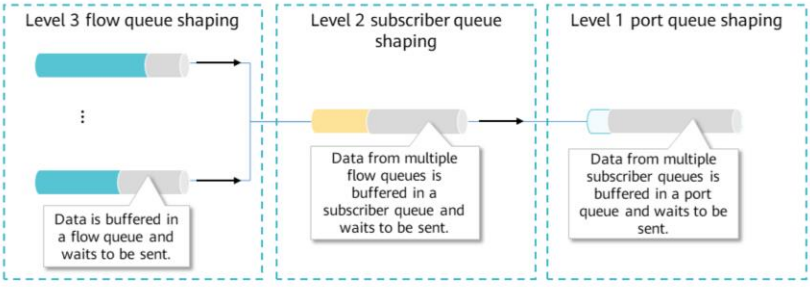


HUAWEI

- HQoS deployment for enterprise users is used as an example. Enterprise users have VoIP, video conference, and data services. Each subscriber queue corresponds to one enterprise user and each flow queue corresponds to a type of services. By deploying HQoS, the device can control the following items:
  - Traffic scheduling among three types of services of a single enterprise user
  - Total bandwidth of three types of services of a single enterprise user
  - Bandwidth allocation between multiple enterprise users
  - Total bandwidth of multiple enterprise users

Introduction to HQoS Traffic Shaping

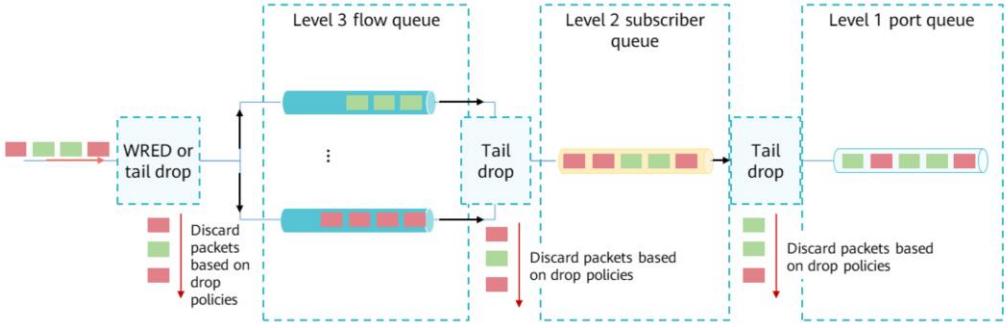
- The HQoS shaper buffers packets and limits the rate of packets. The device supports three levels of shapers, that is, flow queue shaper, subscriber queue shaper, and port queue shaper. After packets enter the device, the device buffers the packets in queues and sends the packets at the limited rate. Shapers can ensure the CIR and limit the maximum rate of packets by using the rate limiting algorithm.



HUAWEI

Introduction to the HQoS Dropper

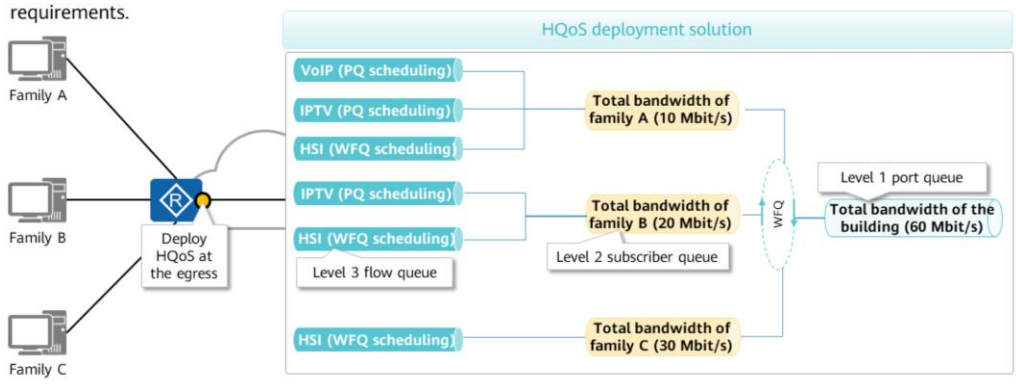
- The HQoS dropper discards packets based on a drop policy before packets are sent to queues.
- The three types of queues supported by HQoS support different drop modes. The port queue and subscriber queue support tail drop; the flow queue supports tail drop and WRED.



HUAWEI

HQoS Application Example

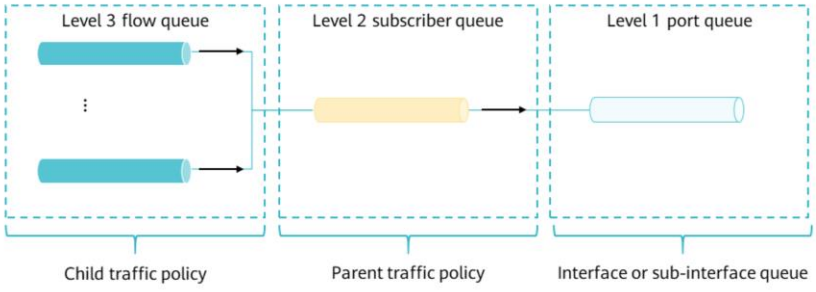
- Assume that there are three families in a building. Family A purchases 10 Mbit/s bandwidth and enables the VoIP, IPTV, and High Speed Internet (HSI) services. Family B purchases 20 Mbit/s bandwidth and enables the IPTV and HSI services. Family C purchases 30 Mbit/s bandwidth and enables only the HSI service. HQoS can meet these requirements.



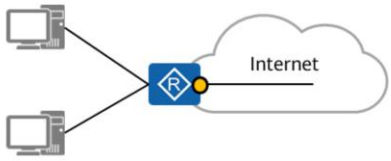
HUAWEI

HQoS Configuration Roadmap

- The HQoS configuration is complex. Generally, the MQC mode is used.
- When HQoS is configured, the policy nesting mode is used.
  - The parent traffic policy differentiates users, and the child traffic policy differentiates traffic.
  - A parent traffic policy can have multiple child traffic policies.
  - A parent traffic policy applies to an interface.



Configuring a Child Traffic Policy



- Child traffic policies are used to differentiate services. You can configure multiple child traffic policies based on services when configuring HQoS.
- The configuration of HQoS child traffic policies is the same as that of common MQC. The configuration roadmap is as follows:
  - Configure a traffic classifier where traffic is matched based on service characteristics.
  - Configure a traffic behavior where the queue scheduling mode and queue bandwidth are defined.
  - Bind the traffic classifier and traffic behavior to a traffic policy.

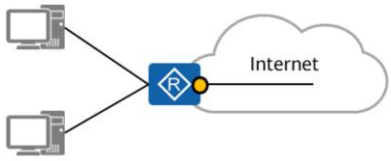
system-view
traffic classifier [*classifier-name*] //Create a traffic classifier.
if-match [*acl* | *vlan-id* | ....] //Match traffic based on service characteristics.

system-view
traffic behavior [*behavior-name*] //Create a traffic behavior.
queue [af | ef | llq] bandwidth [*bandwidth* | **pct** *percentage*] //Configure AF, EF, or LLQ queue parameters in the traffic behavior.
drop-profile [*drop-profile-name*] //Bind the created drop profile to the traffic behavior.

system-view
traffic policy [*policy-name*] //Create a traffic policy.
classifier [*classifier-name*] **behavior** [*behavior-name*] //Bind the traffic classifier to the traffic behavior.

HUAWEI

Configuring a Parent Traffic Policy



- A parent traffic policy is used to differentiate users. When configuring HQoS, you can bind multiple child traffic policies to a parent traffic policy.
- The configuration roadmap is as follows:
  - Configure a traffic classifier to match traffic based on user characteristics.
  - Configure a traffic behavior that needs to invoke a child traffic policy.
  - Bind the traffic classifier and traffic behavior to a traffic policy.

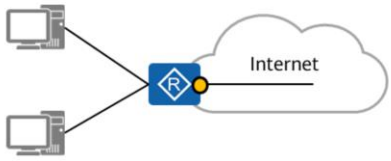
system-view
  traffic classifier [*classifier-name*] //Create a traffic classifier.
    if-match [*acl*| *vlan-id*| ....] //Match traffic based on user characteristics.

system-view
  traffic behavior [*behavior-name*] //Create a traffic behavior.
    queue [af | ef | llq] bandwidth [*bandwidth*| **pct** *percentage*] //(Optional) Configure AF, EF, or LLQ queue parameters in the traffic behavior.
    traffic-policy [*policy-name*] //Bind the sub traffic policy to the traffic behavior.

system-view
  traffic policy [*policy-name*] //Create a parent traffic policy.
    classifier [*classifier-name*] **behavior** [*behavior-name*] //Bind the traffic classifier to the traffic behavior.

HUAWEI

Applying the Parent Traffic Policy




```
system-view
  interface [interface-type interface-num]      //Enter the interface view.
    traffic-policy [policy-name] outbound      //Apply the parent traffic policy to the outbound direction of the interface.
```

- After configuring a parent traffic policy, bind it to an interface or sub-interface.
- If the parent traffic policy is bond to a sub-interface, traffic between different sub-interfaces is sent from the physical interface in polling mode.
- The configuration roadmap is as follows:
  - Apply the parent traffic policy to the outbound direction of the interface.

HUAWEI

Checking the HQoS Configuration

- After configuring HQoS, you can run the following commands to check the configuration.

```
system-view
  display traffic classifier user-defined [ classifier-name ]   //Check the traffic classifier configuration.
  display traffic behavior [ system-defined | user-defined ] [ behavior-name ]   //Check the traffic behavior configuration.
  display traffic policy user-defined [ policy-name ] classifier [classifier-name ]   //Check the traffic policy configuration.
  display traffic-policy applied-record [ policy-name ]   //Check the record of the specified traffic policy.
```


HUAWEI

Quiz

1. (True or false) HQoS cannot distinguish users or services.( )
   A. True
   B. False
2. (Multiple-answer question) What are three types of HQoS queues?( )
   A. Flow queue
   B. Subscriber queue
   C. Data queue
   D. Port queue

HUAWEI

1. B
2. ABD

Section Summary

- HQoS can ensure services with finer granularities.
- HQoS has three levels of queues: flow queue, subscriber queue, and port queue. Traffic shaping can be deployed for the three types of queues. Flow queues are scheduled in PQ+WFQ mode, subscriber queues are scheduled in PQ+WFQ mode, and interface queues are scheduled in RR mode.

HUAWEI

Summary

- QoS is an important means to ensure service quality. Generally, the DiffServ model is used on the live network.
- This model uses rate limiting, congestion avoidance, and congestion management.
- HQoS is used in complex scenarios with finer granularity. Flow queues, subscriber queues, and port queues can be used to distinguish different users and different services of the same user.

HUAWEI

Thank you.

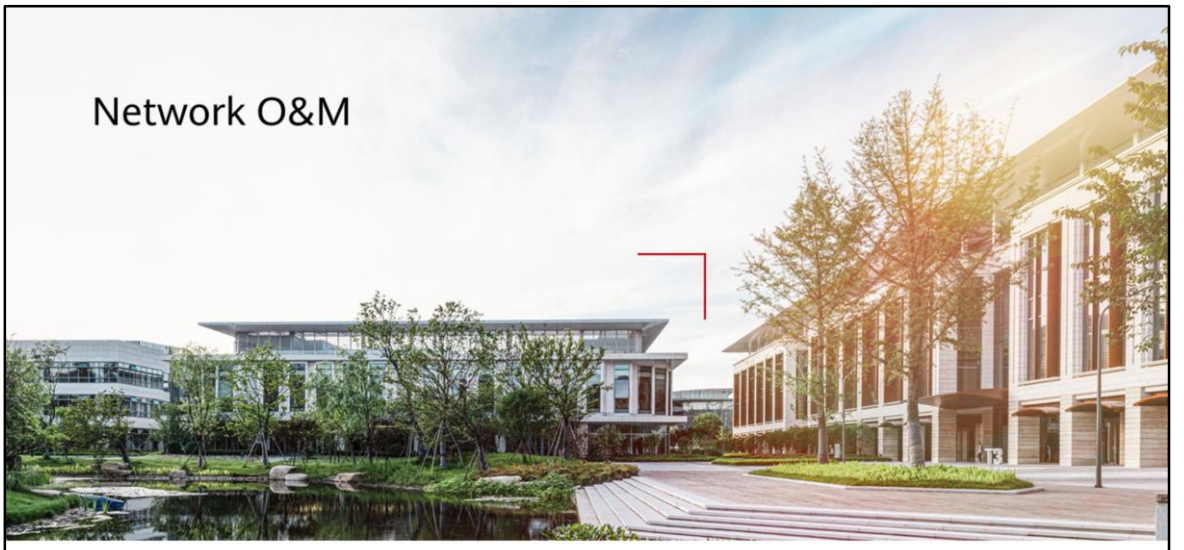
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Network O&M



HUAWEI

Foreword

- Routine maintenance and troubleshooting are required to ensure the proper running of network functions. Routine maintenance is preventive and planned maintenance, and troubleshooting is event-triggered maintenance.
- Good routine maintenance habits help network engineers detect risks in advance. Once an exception or fault occurs on a device, network engineers have to accurately collect events that occur during device running. Routine maintenance and troubleshooting depend on network information collection.
- This course describes the precautions for routine maintenance and common tools for information collection.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Understand check items in routine maintenance.
  - Understand the functions and features of Huawei datacom product information center.
  - Use common maintenance tools.

HUAWEI

Contents

1. **Routine Maintenance**
2. Information Collection Tools

HUAWEI

Overview of Network Maintenance

- The lifecycle of a network involves network planning and design, network implementation, and network maintenance and optimization. Network maintenance involves routine maintenance and troubleshooting.
- Routine maintenance is performed to prevent problems and minimize the possibility of unexpected faults. Fault causes found during the troubleshooting can be used as a reference for routine maintenance.
- Network maintenance is related to not only technology but also management. Routine maintenance does not pose high technical requirements for operators, but poses high requirements on operation standardization. Through routine maintenance, you can obtain various parameters of the network that is working properly, such as network device version, network bandwidth, and network security parameters, which helps you rectify faults.

HUAWEI

- Maintenance is also called "O&M", "operation", or "operation and maintenance."
- Network planning is the starting point of a project. Complete and detailed planning will lay a solid foundation for subsequent project implementation. Specific tasks in network planning are as follows:
    - In the project planning phase, investigate and understand the project background. Properly prepare for project implementation, which ensures the smooth progress of the project.
    - In the project planning phase, the implementation scope of the network project must be specified.
    - Draw up the project budget based on the project objective, project scope, and work content.
    - In the project planning phase, the network design guidelines must be specified to provide guidance and basis for subsequent network design.

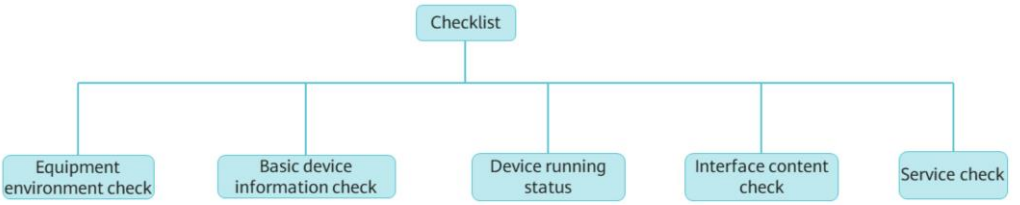
Routine Maintenance: Content and Methods

- Routine maintenance involves the following:
  - Device running environment:
    - The hardware running environment refers to the peripheral environment that covers equipment rooms, power supply system, and heat dissipation system. The hardware running environment is the basis for the proper running of devices.
    - To maintain the device running environment, an engineer has to visit a site in person or uses some professional tools to observe and assess the environment.
  - Device software and hardware running status:
    - The running status of software and hardware is closely related to specific services running on a device. Network engineers must be familiar with common maintenance commands provided by the Versatile Routing Platform (VRP) that Huawei datacom devices.
    - Maintenance personnel can maintain the software and hardware of a device onsite or remotely mainly by running display commands.
- Either of the following methods can be used to perform routine maintenance:
  - Onsite observation: Observe the hardware running environment of devices.
  - Remote operation: Check the running status of the software and hardware of devices.

HUAWEI

Routine Maintenance: Checklist

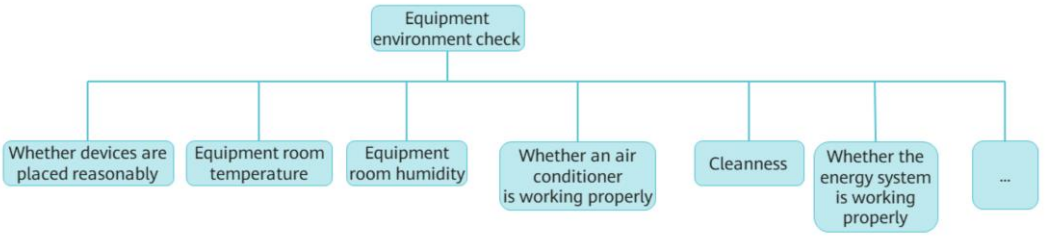
- Routine maintenance is a planned routine task. It is necessary to prepare a checklist for operations. For details about the checklists for different network devices, see related product documentation. Items to be checked during routine maintenance can also be defined by customers.



HUAWEI

Checklist: Equipment Environment Check (1)

- A proper running environment is the prerequisite for the proper running of a device.
- In practice, however, the device environment is not checked immediately once a fault occurs. Compared with other factors, the device environment is more stable and is less prone to faults.



HUAWEI

Checklist: Equipment Environment Check (2)

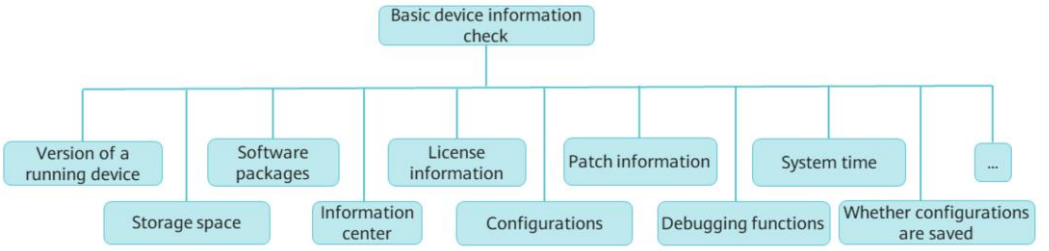
Check Item ping	Method/Tool	Evaluation Criteria and Description
Device location	Observation	The equipment should be placed in a well-ventilated and dry environment at a flat location so that it can be securely installed. No sundries are left around the equipment.
Equipment room temperature	Observation/Temperature meter	Generally, the long-term operating temperature of an equipment room is required to be 0° C to 45° C. Short-term operating temperature: –5° C to +55° C.
Humidity of an equipment room	Observation/Hygrometer	Generally, the relative humidity of the long-term operating environment of the equipment room should be between 5% RH and 85% RH, without condensation. The short-term relative humidity should be between 0% RH and 95% RH, without condensation.
Whether the air conditioner in the equipment room is running properly	Observation/Air conditioner	The air conditioner is running stably and continuously to keep the temperature and humidity in the equipment room within a specified range.
Cleanness	Observation	All items are clean and tidy without obvious dust. Clean or replace the dust filter in time to ensure proper ventilation and heat dissipation of cabinet doors and fan trays.
Heat dissipation	Observation	When a device is working properly, ensure that fan trays are running properly (except when the fans are being cleaned). If the fan trays are shut down, the device temperature will increase and boards may be damaged. Do not place any sundries at the air vents of a chassis. In addition, clean the dust filters of the fan trays periodically.
Cable layout	Observation	Power cables and service cables are routed separately. Power cables are routed neatly and orderly. Service cables are routed neatly and orderly. Cable labels are clear, correct, and standard.
Whether the grounding mode and grounding resistance meet the requirements	Observation	Generally, the working ground, protection ground, and lightning protection ground of the equipment room should be set separately. Joint grounding can be used due to the limited space in an equipment room. Especially for outdoor devices, grounding is very important. If the devices are not grounded, they may be damaged by lightning strikes.
Whether the power supply system is running properly	Observation/Voltage meter	The power supply system must be running stably. The rated DC voltage ranges from –48 V to –60 V. The rated AC voltage ranges from 100 V to 240 V.

HUAWEI

- A proper running environment is the prerequisite for the proper running of a device.
- Temperature and humidity easily affect the proper running of devices. Standard equipment rooms should be equipped with thermometers and hygrometers, and check and record of the temperature and humidity should be performed on a daily basis.
- The cleanness and neatness of the equipment room also affect the proper running of the equipment.
    - Cleanness affects heat dissipation.
    - Tidiness refers to the proper layout of devices and cables. Devices must be installed and cables must be routed according to installation and deployment requirements. However, during network operation, temporary adjustments, such as temporary jumper tests, are often made. After such activities are taken for a period of time, the equipment room becomes disordered. The purpose of checking the equipment environment is to find out and rectify these problems in time.
- For nonstandard equipment rooms, checking the equipment environment more carefully. For example, check the cleanness and heat dissipation of equipment rooms on floors.
- The preceding check items may vary according to devices. For details, see the product documentation of each type of device.

Checklist: Basic Device Information (1)

- Basic device information check involves checking the software version, license, and storage space.



HUAWEI

Checklist: Basic Device Information (2)

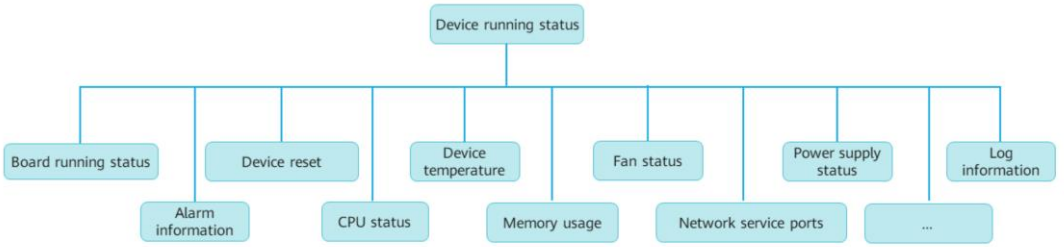
Check Item	Check Method	Evaluation Criteria
Version of a device	display version	The PCB version and software version of the board meet the requirements.
Software packages	display startup	Check whether the following system file names are correct: names of the software packages for the existing and next startup; name of the backup software package; names of the configuration files for the existing and next startup, license files, and patch file.
License information	display license display license state	Check whether the name, version, and configuration items of the GTL license file meet the requirements and determine whether the GTL license file needs to be upgraded. The value of **Master board license state** must be **Normal**. If the value of **Master board license state** is **Demo** or **Trial**, the license is valid.
Patch information	display patch-information	The patch file must meet actual requirements. It is recommended that you load the latest patch file matching the product version released by Huawei. The patches must have taken effect. That is, the total number of patches is the same as the number of running patches.
System time	display clock	The system time must be consistent with the time used on a network management server (the time difference must be less than or equal to 5 minutes), which facilitates troubleshooting.
Space of the flash memory, SD cards, and CF cards	dir flash dir slave#cfcard	Files in the flash memory, SD cards, and CF card must be useful. Otherwise, run the **delete** or **unreserved** command in the user view to delete the unwanted files.
Information Center	display info-center	The **Information Center** value must be **enabled**.
Configurations	display current-configuration	Check whether the device configuration is correct by viewing the currently effective configuration parameters.
Debug functions	display debugging	All debugging functions are disabled when the device is running properly.
Whether the configurations are saved	compare configuration	The existing configurations on a device are the same as those contained in the configuration file for a next startup.
...	...	...

HUAWEI

- Software version running on a device:
  - The running software version of a device should be confirmed in project implementation. In normal cases, the version information does not change. Pay attention to any change in version information. This situation is usually caused by nonstandard management.
  - If a device is newly added, the software version may be different from the existing software version. Some devices may be upgraded or downgraded due to other reasons. Especially on a large-scale network, the same type of device may run different versions. In this case, verify that different versions can meet the same network function requirements.
- Startup information:
  - Multiple software packages of different versions or configuration files may be stored on a device. In this case, changing startup information may cause great risks to the proper running of the network. Once the device is restarted (for example, if power supply is faulty), the running of the entire network may be adversely affected.
- License information:
  - License rules vary according to devices. The licenses of some devices have validity periods.

Checklist: Device Running Status (1)

- When checking the running status of a device, pay attention to the hardware running status, such as board status, power supply status, fan status, temperature, CPU usage, and memory usage. Alarm indicators on devices can help you find the abnormal status onsite.
- Check the running status of components such as cards, power modules, and fan trays according to instructions provided by manufacturers. If necessary, contact the manufacturers for instructions. If hardware is faulty, contact vendors who may replace hardware or provide support for troubleshooting.



Checklist: Device Running Status Check (2)

Check Item	Check Method	Evaluation Criteria
Board running status	display device	Check whether the in-position status and working status of a board are normal. The **Online** value is **Present**. The **Power** value is **PowerOn**. The **Register** value is **Registered**. The value of **Alarm** is **Normal**.
Device reset	display reset-reason display reboot-info	Check the reset information (including the reset time and cause) to ensure that no abnormal reset occurs.
Device temperature	display temperature display environment	The current temperature of each module must be between the upper and lower thresholds.
Fan status	display fan	If the **Present** value is **YES**, the status is normal.
Power supply status	display power	If the **State** value is **Supply**, the power supply is normal.
FTP network service port	display ftp-server	The FTP network service ports that are not used must be disabled.
Alarm information	display alarm all	No alarm is generated. If an alarm is generated, record the alarm. If a major or critical alarm is generated, analyze and handle the alarm immediately.
CPU status	display cpu-usage	The CPU usage of each module is normal. Pay special attention to CPU usage that exceeds 80%.
Memory usage	display memory-usage	Memory usage must be normal. Pay special attention to the **Memory Using Percentage** value that exceeds 60%.
Log information	display logbuffer display trapbuffer	No exception information exists.
Master/Slave status of main control boards	display switchover state	If both the master and slave main control boards are equipped, the status information about both the boards must be correct. After a master/slave main control board switchover is complete and the device is working properly, this command is expected to display **realtime or routine backup** for the master main control board.
...	...	...

HUAWEI

Checklist: Device Interface Check

- Network devices exchange data packets through interfaces. Interface information is very important. Abnormal interface status adversely affects network functions.
- If a large number of error packets are generated on an interface and the number of error packets increases continuously within a short period of time, a link (as well as physical interfaces) may be faulty.

Check Item	Check Method	Evaluation Criteria
Error packets on an interface	display interface	When services are running, check whether there are error packets on the interface, including CRC error packets.
Interface negotiation mode	display interface	The negotiation mode of the interface must be correct. The negotiation modes of the two interfaces on both ends must be the same. The half-duplex mode is not allowed.
Interface configuration	display current-configuration interface	Interface configuration items, such as the duplex mode, negotiation mode, rate, and loopback configuration, must be correct.
Interface status	display interface brief	The up/down status of an interface meets planning requirements. Check whether the traffic sent or received by the interface is too heavy. (The long-term traffic volume is greater than 70% of the interface capacity.)
PoE power supply	display port power-state interface *interface-type interface-number*	The PoE power supply status of an interface must be normal. The **Delivering-power** value must be **Delivering-power** for an interface whose **Port power ON/OFF** value is **ON**.
...	...	...

HUAWEI

Checklist: Service Running Status Check

- Service running status refers to the running status of network protocols.

Check Item	Check Method	Evaluation Criteria
MAC address table information	display mac-address	The MAC address table information must be correct.
VLAN information	display vlan	Basic information about all VLANs must be correct.
Routing table information	display ip routing-table	Check the default routes or others specific routes, information of which is used for remote fault locating. If the devices at the same layer of a network run the same routing protocol, the number of routes on each device should be close. (The number of routes may vary according to the number of static routes.)
OSPF neighbor status IS-IS neighbor status BGP peer status	display ospf peer display isis peer display bgp peer	OSPF neighbor status: The **State** value is **Full** or **2-Way**. IS-IS neighbor status: The **State** value is **Up**. BGP peer status: The **State** value is **Established**.
VRRP status	display vrrp display vrrp statistics	The VRRP status of devices in a VRRP group cannot be **Master** at the same time.
MSTP status	display stp brief	The **STP Status** values of the designated port and root port are **FORWARDING**.
...	...	...

HUAWEI

Routine Maintenance: Software and Configuration Backup

- The purpose of backup is to restore network functions in extreme cases. Backup is to transfer files to a backup server. There are many backup methods. Generally, a device functions as an FTP or TFTP client and transfers files to a server through the CLI.
- You are advised to back up configuration files every week. In addition, back up the configuration file before the configuration on a device is changed.
- Both software and configuration (including license files) need to be backed up. The purpose of backup is to restore network functions in extreme cases.
  - Service restoration will be delayed without a backup configuration file in case of a hardware fault-induced restart failure or replacement by a device of the same model.
  - The software package also needs to be backed up. The software package of a specific product and a specific version is backed up only once. You can also obtain the corresponding software package from the vendor's official website and store it on a local computer.
  - A license file is a special file that is configured for a specific product. If the license file is missing due to an incident (for example, by mistake), a new license file must be applied for through the vendor's process. In this process, some proof materials (such as the contract number and device SN) must be provided. Consequently, the application period is long. If a backup license file is available, the license file can be quickly restored on the device.

HUAWEI

Contents

1. Routine Maintenance
2. **Information Collection Tools**
   - Information Center
   - Packet Information Obtaining
   - LLDP
   - Traffic Statistics Collection

HUAWEI

Information Center Overview

- The information center is an information hub on a device. Log, trap, and debugging information generated by the device are sent to the information center. The information center manages and controls all information and flexibly outputs the information.
- The information center can be configured to classify or filter the information by type and severity and to output the information in different directions (such as the console, user terminal, and log host). In this way, users or network administrators can collect device information, which is used to monitor the device running status and locate faults.

Information Type	Description
Log information	Logs record information about user operations, system faults, and system security. • User logs: record user operations and system running information. • Security logs: record information about account management, protocols, attack defense, and attack defense status. • Diagnostic logs: record information that helps locate faults.
Trap information	Traps are notifications generated once a system detects faults. Traps record system status information, for example, faults. This type of information is different from log information. The major characteristic of trap information is that administrators must be notified and alerted immediately, and trap information is sensitive to time.
Debugging information	Debugging information is output by a device and used to monitor the internal running of the device. The device can generate debugging information only after the debugging of a specific module is enabled.

HUAWEI

Information Severity

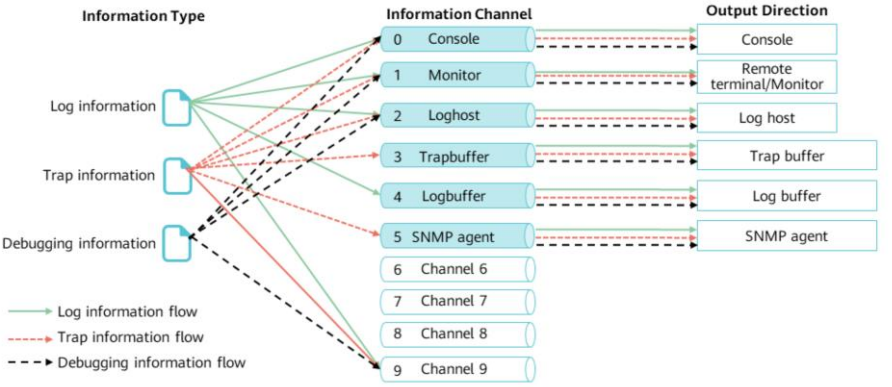
- When a large amount of information is generated on a device, it is difficult to tell which information can be ignored and which information indicates a fault to be dealt with. Information is classified into different levels, which helps users efficiently take measures or shield unnecessary information.
- Information is classified into eight levels based on its severity and urgency. A lower value indicates a higher severity level.

Value	Severity	Description
0	Emergency	A fatal exception occurs on a device, and the device cannot recover. You must restart the device. For example, a program exception causes the device to restart and memory usage is incorrect.
1	Alert	A major exception occurs on a device and measures must be taken immediately. For example, the memory usage of a device reaches a specified upper limit.
2	Critical	The device is abnormal. You need to take measures to handle a fault or analyze causes. For example, the memory usage of a device is lower than a specified lower threshold or the device is unreachable after BFD detection is performed.
3	Error	Incorrect operations or abnormal device processes occur, which do not affect subsequent services. However, you need to pay attention to and analyze causes. For example, incorrect commands, incorrect passwords, or incorrect protocol packets are detected.
4	Warning	A device running exception may cause a service fault. For example, a routing process is disabled, a packet loss event is detected by BFD or incorrect protocol packets are detected.
5	Notification	Key operation information about the proper running of the device. For example, the interface is shut down, neighbor discovery is performed, and the protocol state machine changes normally.
6	Informational	General operation information about the proper running of the device. For example, a user runs a display command.
7	Debugging	Informational only, and no action is required.

HUAWEI

Information Output

- The information generated by a device can be output to a remote terminal, console, log buffer, log file, or SNMP agent. To facilitate the output control of information in different directions, the information center defines 10 information channels. The information is output separately in each direction.



HUAWEI

- You can configure information output rules as needed to control the output of various types and levels of information along information channels in different output directions.
- A remote terminal is used to log in to a device through a VTY interface to receive logs, traps, and debugging information, facilitating remote maintenance.

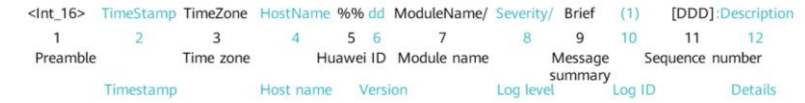
Information Filtering

- The information center can filter information to flexibly control the output of information. When a device is running properly, each module reports information during service processing. To filter out unnecessary information about service modules or information of specified levels, you can configure the information filtering function.
- The information center uses the information filtering table to filter information along the channels. The information filtering table is used to filter the information output in each direction based on the Information Severity, level, and source.
- The information filtering table provides the following information:
  - Sequence number of an information module
  - Status of the log output function
  - Level of log information to be output
  - Trap output status
  - Level of trap information to be output
  - Debugging information output status
  - Level of debugging information to be output

HUAWEI

Information Output Format

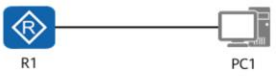
- Log output format



- Trap output format

HUAWEI

Application Scenarios of the Information Center

Outputting logs to a log host	Outputting traps to an NMS	Outputting debugging information to the console
		
The information center outputs logs of a specified level to a log host. Maintenance personnel can obtain the running status of the device by querying logs.	The information center outputs traps to an NMS. The NMS monitors the running status of a device based on the received traps.	The information center sends debugging information to the console. Maintenance personnel can obtain the running status of a device by checking the debugging information.

HUAWEI

Information Center Commands (1)

1. Enable the information center function.

   [HUAWEI] **info-center enable**

   By default, the information center is enabled.

2. Specify the name of an information channel with a specified ID.

   [HUAWEI] **info-center channel** *channel-number* **name** *channel-name*

3. Enable a device to filter logs or traps.

   [HUAWEI] **info-center filter-id** { *id* | **bymodule-alias** *modname alias* } [**bytime** *interval* | **bynumber** *number*]

   bymodule-alias: outputs log or trap information of a module with a specified name.

4. Enable the device to send log information to the log buffer.

   [HUAWEI] **info-center logbuffer**

   By default, a device is enabled to send logs to the log buffer.

HUAWEI

- Enable a device to send information to a log host.
  - [HUAWEI] info-center loghost ip-address { source-ip source-ip-address } | transport { udp | tcp ssl-policy policy-name }]

Information Center Commands (2)

1. Configure an information channel used to output information.

   [HUAWEI] **info-center** { **console** | **logbuffer** | **logfile** | **monitor** | **snmp** | **trapbuffer** } **channel** { *channel-number* | *channel-name* }

2. Enable the terminal to display information sent by the information center.

   [HUAWEI] **terminal monitor**

   By default, the console is enabled to display information, and a user terminal is disabled from displaying information.

3. Enable the terminal to display debugging information.

   [HUAWEI] **terminal debugging**

   By default, a terminal is disabled from displaying debugging information.

4. Enable the terminal to display log information.

   [HUAWEI] **terminal logging**

   By default, a terminal is enabled to display log information.

HUAWEI

Information Center Commands (3)

1. Query information recorded in the log buffer.

   [HUAWEI] **display logbuffer** [**size** *size* | **slot** *slot-id* | **module** *module-name* | **security** | **level** { *severity* | *level* }]

2. Query information in a log file.

   [HUAWEI] **display logfile** *file-name* [*offset* | **hex**]

3. Query information recorded in the trap buffer of the information center.

   [HUAWEI] **display trapbuffer**

4. Query debugging information that can be sent by a device.

   [HUAWEI] **display debugging**

5. Query the output direction configuration of the information center.

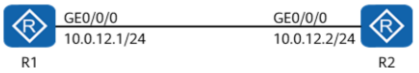
   [HUAWEI] **display info-center**

HUAWEI

Example for Configuring the Information Center

Enable OSPF on GE 0/0/0 of R1 and R2. Log in to R1 through the console and check the following information on R1:

- Trap information
- Log information
- Debugging information




```
<R1> terminal monitor     //By default, the console is enabled to display information.
<R1> terminal logging     //By default, a terminal is enabled to display log information.
<R1> terminal debugging   //By default, the terminal is disabled from displaying debugging information.
<R1> terminal trapping    //By default, the terminal is enabled to display trap information.
<R1> system-view
[R1] info-center enable   //By default, the information center function is enabled.
```


HUAWEI

Information Channel

```
<R1> display channel
 channel number: 0,  channel name: console
MODU_ID  NAME     ENABLE  LOG_LEVEL      ENABLE  TRAP_LEVEL  ENABLE  DEBUG_LEVEL
ffff0000 default  Y       warning        Y       debugging   Y       debugging
 channel number: 2,  channel name: loghost
MODU_ID  NAME     ENABLE  LOG_LEVEL      ENABLE  TRAP_LEVEL  ENABLE  DEBUG_LEVEL
ffff0000 default  Y       informational  Y       debugging   N       debugging
 channel number: 3,  channel name: trapbuffer
MODU_ID  NAME     ENABLE  LOG_LEVEL      ENABLE  TRAP_LEVEL  ENABLE  DEBUG_LEVEL
ffff0000 default  N       informational  Y       debugging   N       debugging
 channel number: 4,  channel name: logbuffer
MODU_ID  NAME     ENABLE  LOG_LEVEL      ENABLE  TRAP_LEVEL  ENABLE  DEBUG_LEVEL
ffff0000 default  Y       warning        N       debugging   N       debugging
...      ...      ...     ...
```

Information is classified into eight levels based on its severity and urgency. The more serious the information is, the lower the severity level is.
By default:

- Warning is level 4, which indicates an abnormal point in device running, which may cause service faults and requires attention. For example, a routing process is disabled, a packet loss event is detected by BFD or incorrect protocol packets are detected.
- Debugging is level 7, which indicates general information about the proper running of a device. No action is required for such information.

R1 provides four channels: console, log host, trap buffer, and log buffer.

HUAWEI

Traps

Query information recorded in the trap buffer of the information center.

```
<R1> display trapbuffer
Trapping buffer configuration and contents: enabled
Allowed max buffer size: 1024
Actual buffer size: 256
Channel number: 3, Channel name: trapbuffer
Dropped messages: 0
Overwritten messages: 0
Current messages: 1
#YY-MM-DD 08:38:51-08:00 R1 LLDP/4/ADDCHGTRAP:OID: [OID] Local management add
ress is changed. (LocManIPAddr=[IPADDR])
```

By default, traps are output through information channel 3.

HUAWEI

Logs

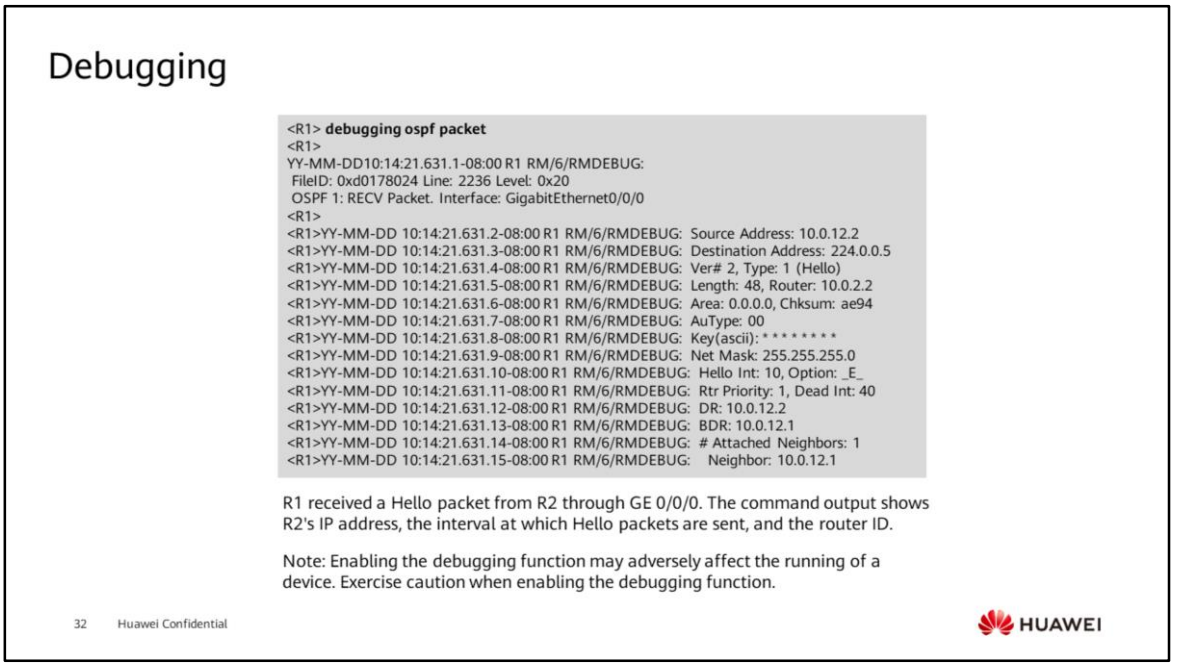
```
<R1> display log cli all
  ------------------------------------------------------------------------------
  No.   UserName                  Domain          IP-Address
    35                            --              Serial
  Time: YY-MM-DD 09:34:35-08:00
  Cmd:  quit
  ------------------------------------------------------------------------------
  No.   UserName                  Domain          IP-Address
    34                            --              Serial
  Time: YY-MM-DD 09:34:33-08:00
  Cmd:  ip address 10.0.12.1 24
  ------------------------------------------------------------------------------
  No.   UserName                  Domain          IP-Address
    33                            --              Serial
  Time: YY-MM-DD 09:34:29-08:00
  Cmd:  interface gi 0/0/0
  ------------------------------------------------------------------------------
  No.   UserName                  Domain          IP-Address
    32                            --              Serial
  Time: YY-MM-DD 09:34:26-08:00
  Cmd:  system-view
```

Query all commands entered by users. As shown in the figure, the IP address of GE 0/0/0 on R1 has been configured.

```
<R1>
YY-MM-DD 10:09:57-08:00 R1 %%01OSPF/4/NBR_CHANGE_E(l)[0]:Neighbor
changes event: neighbor status changed.
(ProcessId=256,NeighborAddress=10.0.12.2, NeighborEvent=HelloReceived,
NeighborPreviousState=Down, NeighborCurrentState=Init)
<R1>
YY-MM-DD 10:09:57-08:00 R1 %%01OSPF/4/NBR_CHANGE_E(l)[1]:Neighbor
changes event: neighbor status changed.
(ProcessId=256,NeighborAddress=10.0.12.2, NeighborEvent=2WayReceived,
NeighborPreviousState=Init, NeighborCurrentState=2Way)
...
<R1>
YY-MM-DD 10:09:57-08:00 R1 %%01OSPF/4/NBR_CHANGE_E(l)[5]:Neighbor
changes event: neighbor status changed.
(ProcessId=256,NeighborAddress=10.0.12.2, NeighborEvent=LoadingDone,
NeighborPreviousState=Loading, NeighborCurrentState=Full)
```

The OSPF neighbor status change of R1 is automatically displayed in the log format on the console port.

HUAWEI



- To facilitate display, the debugging information displayed on this page is adjusted.
- The content of the Hello packet sent by R1 through GE 0/0/0 is as follows:

  <R1>

  YY-MM-DD10:14:21.751.1-08:00 R1 RM/6/RMDEBUG:

  FileID: 0xd0178025 Line: 559 Level: 0x20

  OSPF 1: SEND Packet. Interface: GigabitEthernet0/0/0

  <R1>YY-MM-DD 10:14:21.751.2-08:00 R1 RM/6/RMDEBUG: Source Address: 10.0.12.1

  <R1>YY-MM-DD 10:14:21.751.3-08:00 R1 RM/6/RMDEBUG: Destination Address: 224.0.0.5

  <R1>YY-MM-DD 10:14:21.751.4-08:00 R1 RM/6/RMDEBUG: Ver# 2, Type: 1 (Hello)

  <R1>YY-MM-DD 10:14:21.751.5-08:00 R1 RM/6/RMDEBUG: Length: 48, Router: 10.0.12.1

  <R1>YY-MM-DD10:14:21.751.6-08:00 R1 RM/6/RMDEBUG: Area: 0.0.0.0, Chksum: ae94

  <R1>YY-MM-DD 10:14:21.751.7-08:00 R1 RM/6/RMDEBUG: AuType: 00

  <R1>YY-MM-DD10:14:21.751.8-08:00 R1 RM/6/RMDEBUG: Key(ascii): * * * * * * * *

Contents

1. Routine Maintenance
2. **Information Collection Tools**
   - Information Center
   - Packet Information Obtaining
   - LLDP
   - Traffic Statistics Collection

HUAWEI

Packet Information Obtaining

When the service traffic of a device becomes abnormal, for example, the traffic status is inconsistent with the traffic model, you can use the packet information obtaining function for analysis. In this way, invalid packets can be processed in time to ensure proper transmission of network data.

```
<R1> system-view
[R1] capture-packet interface gigabitethernet 0/0/0 destination terminal
Info: Captured packets will be shown on terminal.
[R1]ping 10.0.12.2
[R1]
  Packet: 1
---------------------------------------------------------------
ff ff ff ff ff ff 00 e0 fc 9f 4b 1d 08 06 00 01
08 00 06 04 00 01 00 e0 fc 9f 4b 1d 0a 00 0c 02
00 00 00 00 00 00 0a 00 0c 02 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00
---------------------------------------------------------------

  Packet: 2
---------------------------------------------------------------
ff ff ff ff ff ff 00 e0 fc 9f 4b 1d 08 06 00 01
08 00 06 04 00 01 00 e0 fc 9f 4b 1d 0a 00 0c 02
00 00 00 00 00 00 0a 00 0c 02 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00
---------------------------------------------------------------
```

ARP request packets sent by R2
Destination MAC address: ffff-ffff-ffff
Source MAC address: 00e0-fc9f-4b1d
Type: 0806, indicating ARP

HUAWEI

- Only one packet information obtaining instance can run at a time. That is, if a previous process is not complete, a next process cannot be started.
- The rate of packets whose information is to be obtained is limited. If burst traffic exceeds the rate limit configured for obtained packet information, packet loss may occur.
- The capture-packet command obtains header information in the service packets that match the configured rules and sends the obtained information to the terminal for display or saves the obtained information on a local device.
    - capture-packet interface interface-type interface-number [acl acl-number] destination { terminal | file file-name } * [car cir car-value | time-out time | packet-num number | packet-len { length | total-packet }] *
        - terminal: sends the obtained information to the terminal for display.
        - file file-name: saves the obtained information in a specified file.

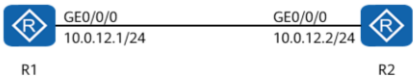
Contents

1. Routine Maintenance
2. **Information Collection Tools**
   - Information Center
   - Packet Information Obtaining
   - LLDP
   - Traffic Statistics Collection

HUAWEI

LLDP Application Example

- Link Layer Discovery Protocol (LLDP) is a link layer topology discovery protocol defined in IEEE 802.1ab. It can accurately locate the interfaces on devices and the interfaces connected to other devices, and display information about paths between clients, switches, routers, application servers, and network servers.
- In actual networking, you can use LLDP to obtain physical connection information of devices.



Enable LLDP on R1 and R2.

```
<R1> system-view
[R1] lldp enable
```

```
<R2> system-view
[R2] lldp enable
```

```
<R1> display lldp neighbor
GigabitEthernet0/0/0 has 1 neighbors:

Neighbor index          1
Chassis type            macAddress
Chassis ID              00e0-fc9f-4b1d
Port ID type            interfaceName
Port ID                 GigabitEthernet0/0/0
Port description        HUAWEI, AR Series, GigabitEthernet0/0/0 Interface
System name             R2
System description      Huawei AR2220 Huawei Versatile Routing Platform
                        Software VRP (R) software,Version 5.130 (AR2220
                        V200R003C00) Copyright (C) 2011-2012 Huawei
                        Technologies Co., Ltd

Management address type ipV4
Management address      10.0.12.2
Expired time            109s
```

The preceding command output shows that GE 0/0/0 on R1 is connected to GE 0/0/0 on R2. R2 is an AR2220, and GE 0/0/0's IP address is 10.0.12.2.

HUAWEI

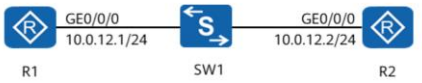
Contents

1. Routine Maintenance
2. **Information Collection Tools**
   - Information Center
   - Packet Information Obtaining
   - LLDP
   - Traffic Statistics Collection

HUAWEI

Traffic Statistics Collection

- Traffic statistics collection helps you learn about traffic passing and discarding after a traffic policy is applied. You can analyze and determine whether a traffic policy is properly applied, and locate faults.



1. Configure an ACL rule.
[R2] **acl 2000**
[R2-acl-basic-2000] **rule permit source 10.0.12.1 0**
[R2-acl-basic-2000] **quit**
2. Configure a traffic classifier.
[R2] **traffic classifier c1**
[R2-classifier-c1] **if-match acl 2000**
[R2-classifier-c1] **quit**
3. Configure a traffic behavior.
[R2] **traffic behavior b1**
[R2-behavior-b1] **statistic enable**
[R2-behavior-b1] **quit**

4. Create a traffic policy.
[R2] **traffic policy p1**
[R2-trafficpolicy-p1] **classifier c1 behavior b1**
[R2-trafficpolicy-p1] **quit**
5. Apply the traffic policy to an interface.
[R2] **interface GigabitEthernet 0/0/0**
[R2-GigabitEthernet0/0/0] **traffic-policy p1 inbound**
[R2-GigabitEthernet0/0/0] **quit**

```
[R2]display traffic policy statistics interface GigabitEthernet 0/0/0 inbound
 Interface: GigabitEthernet0/0/0
 Traffic policy inbound: test
 Rule number: 1
 Current status: OK!
Item                  Sum(Packets/Bytes)        Rate(pps/bps)
-------------------------------------------------------------
Matched               0/0                       0/0
Passed                0/0                       0/0
Dropped               0/0                       0/0
Filter                0/0                       0/0
CAR                   0/0                       0/0
Queue Matched         0/0                       0/0
```

The preceding command output shows that R2 does not receive ICMP messages. Based on the preceding information, you can check whether a physical link is faulty or whether the VLAN configuration of SW1 is incorrect.

HUAWEI

Quiz

1. (Multiple-answer question) Which of the following statements about network maintenance functions are correct?( )
   A. Routine maintenance is a type of preventive work.
   B. Routine maintenance helps you obtain the network baseline, which lays a solid foundation for troubleshooting.
   C. Routine maintenance poses high technical requirements for operators, but does not pose high requirements for operation standardization.
   D. Network maintenance is not only a technical issue, but also a management issue.

HUAWEI

1. ABD

Quiz

1. (True or false) packet information obtained by a tool can be displayed only on the CLI of a device and cannot be saved in a file.( )
   A. True
   B. False
2. (True or false) R1 and R2 are directly connected. If the interface IP addresses of R1 and R2 are on different network segments, the host names of the two devices cannot be obtained using LLDP.( )
   A. True
   B. False

HUAWEI

1. B
2. B

Summary

- Network maintenance includes routine maintenance and troubleshooting. This course describes the routine maintenance and its check items, such as equipment environment check, basic equipment information check, and equipment running status check. During routine maintenance, you must periodically back up device configurations and software packages. By doing so, you can use the backup data to restore network functions in extreme situations.
- Troubleshooting is network maintenance driven by fault events. A large amount of information needs to be collected during troubleshooting. A device provides the information center function. The information center provides log, trap, and debugging information and 10 information channels. You can configure rules for outputting information to control the output of various types and levels of information through different information channels.
- This course also introduces tools for obtaining information, such as Packet Information Obtaining, traffic statistics collection, and LLDP. The use of these tools helps you efficiently rectify faults.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Network Troubleshooting

HUAWEI

Foreword

- Digital transformation of medium- and large-sized enterprises is implemented using multiple technologies, such as cloud computing, big data, artificial intelligence (AI), and Internet of Things (IoT). These technologies are all supported by data communications networks. A stable data communications network requires fully prepared network design, construction, and maintenance.
- An enterprise data communications network accommodates various types of devices that are connected by multiple types of physical links. In addition, to accurately forward data packets, the devices run multiple protocols. Network devices, cables, and protocols may encounter faults. How to quickly rectify faults is a basic skill of senior network engineers.
- This course describes common network faults, how to troubleshoot them in an effort to help network engineers build capabilities of troubleshooting faults in various scenarios.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Understand the troubleshooting methods.
  - Analyze loop faults.
  - Analyze failures to establish neighbor relationships of routing protocols.
  - Write a troubleshooting guide.

Contents

1. **Troubleshooting Data Communication Network Faults**
   - Overview of Network Faults
   - Structured Troubleshooting Process
   - Core Ideas and Methods of Network Troubleshooting
2. Troubleshooting Common Network Faults

HUAWEI

What Is a Network Fault?

- A network fault refers to the phenomenon that a network loses a specific function and adversely affects services due to some reasons.
- From the perspective of users, any phenomenon that adversely affects services can be defined as a fault.
- The common fault symptoms and categories are as follows:

Symptom / Category	Alarm	Loop	Service Forwarding Failure	Service Interruption	Transient Service Interruption	Packet Loss	Protocol Anomaly	Protocol Flapping	Route Anomaly
Hardware	√			√		√			
Configuration		√	√				√		√
Network		√	√	√	√	√	√	√	√
Performance	√				√	√		√	√
Software							√		√
Interconnection		√	√				√		
Others	√		√	√	√	√			

HUAWEI

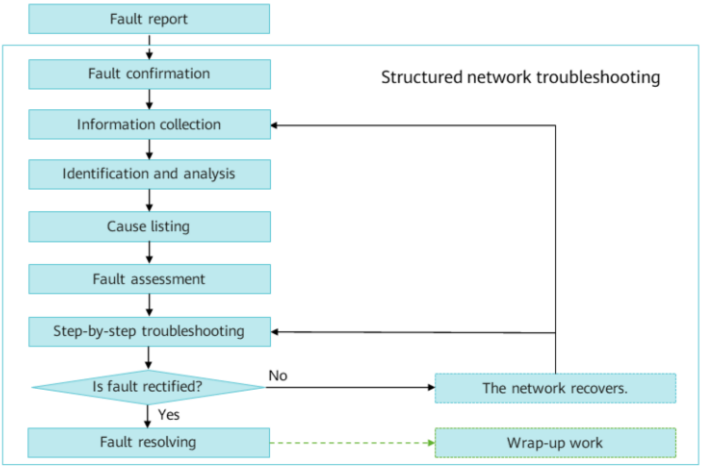
- Mapping between the preceding fault symptoms and categories varies according to scenarios.

Contents

1. **Troubleshooting Data Communication Network Faults**
   - Overview of Network Faults
   - Troubleshooting Process
   - Core Ideas and Methods of Network Troubleshooting
2. Troubleshooting Common Network Faults

HUAWEI

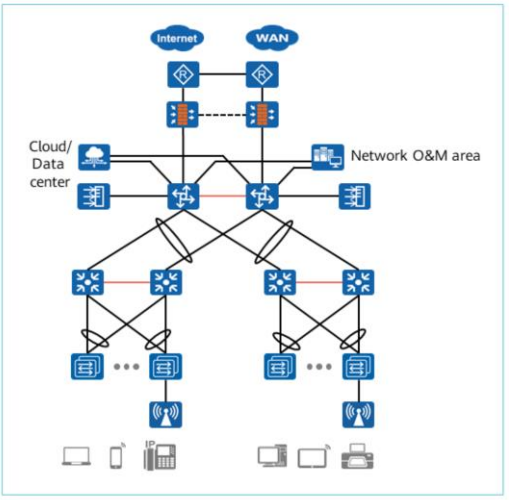
Structured Network Troubleshooting Process



HUAWEI

- If an unstructured network troubleshooting is carried out, steps are performed repeatedly, leading to low efficiency even though a solution to the fault is found.
- In a complex network environment, a new fault may be caused due to an unstructured network fault rectification process, making network fault rectification more difficult.

Fault Report



- An enterprise has multiple departments, such as finance, human resource, logistics, marketing, and R&D departments. These departments need to communicate. To properly guarantee network operations, enterprises may take the following measures:
  - Large- and medium-sized enterprises set up network maintenance departments to build professional network teams.
  - To reduce expenses, small-sized enterprises do not set up an independent network maintenance department. Instead, they entrust their networks to professional network maintenance companies.
  - Contact device manufacturers for after-sales service.
- Generally, the person who first percepts network faults is from a department related to services, rather than a network maintenance engineer. Network engineers often receive calls for help, such as "the computer suddenly cannot access the Internet", "the web page cannot be displayed normally", and "the game is stuck".

HUAWEI

Fault Report | Fault Confirmation | Information Collection | Identification and Analysis | Cause Listing | Fault Assessment | Step-by-Step Troubleshooting | Fault Resolving | Wrap-up Work

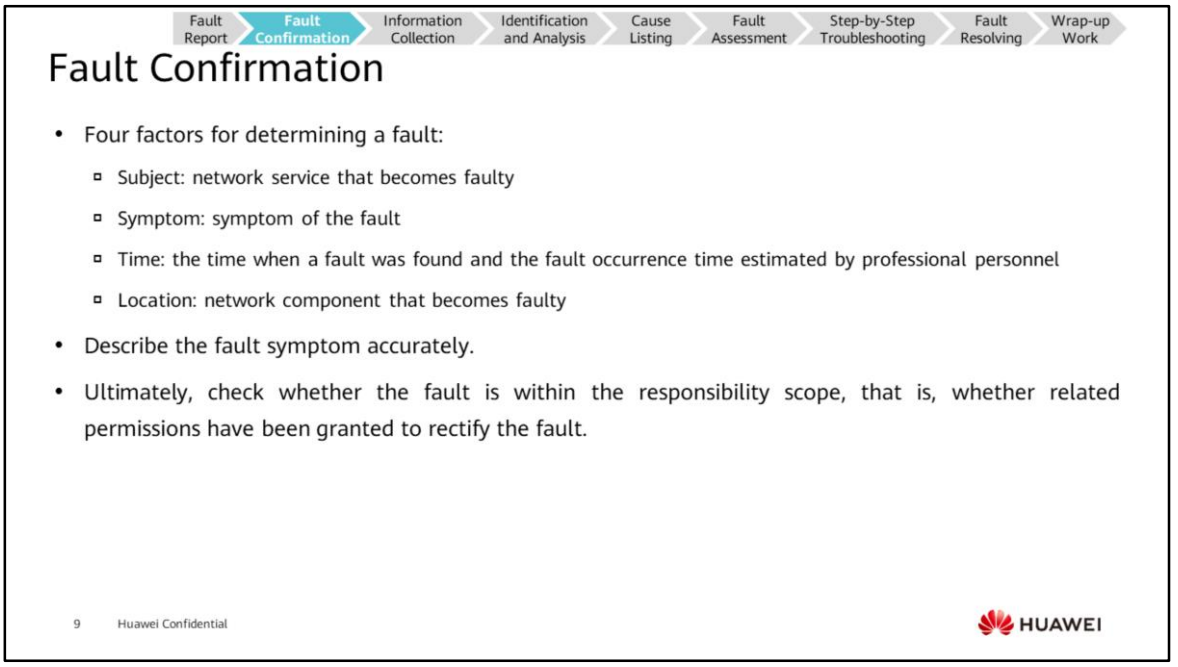
Fault Report: Proactive Communication and Confirmation

- Ask a user for the preceding information through a phone and record the information in a troubleshooting report.

Fault reporter	Name, department, position, work content, computer location (floor, room, wireless or wired access), and website that the computer attempts to access.
Fault frequency	Check whether the fault occurs suddenly, occasionally, or frequently.
User operation	Operations performed by a user on a terminal before and after a fault occurs. For example, the IP address and DNS parameters are changed, desktop firewall software is installed, and security control software is installed.

HUAWEI

- Why do we need to know the positions and work content of users?
  - In an enterprise environment, network access permissions to be granted vary according to positions. Even users of the same position may have only the permission to use network services related to their work content.



- Why do we need to confirm a fault?
  - The user description may be ambiguous, and the reported fault may not be the actual faulty point. In this situation, experienced engineers have to confirm the fault.

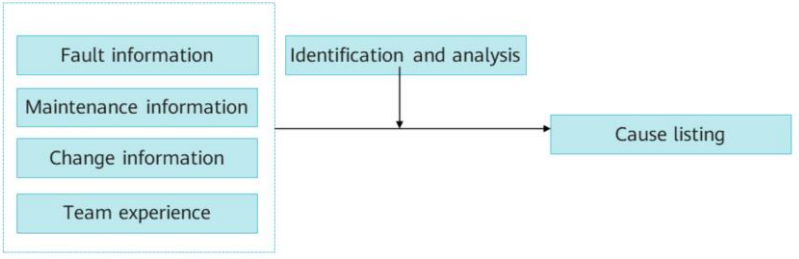
Information Collection

- Which information needs to be collected?
  - In the information collection phase, fault-related information, such as documents and network changes, is collected.
- How to collect the information:
  - Run commands on an involved device. Use information collection tools, such as the packet information obtaining tool and network management software.
- Obtaining permissions:
  - In a network environment that poses high requirements on information security, information collection must be authorized. Sometimes, a written authorization file must be signed.
- Risk assessment in the information collection phase:
  - Some information collection operations, such as running a debug command on a router or switch, may cause high CPU usage. In worse cases, a device may even stop responding to instructions, causing more faults. When collecting information, you must evaluate risks, balance the risks of introducing new faults and the urgency of rectifying existing faults, and notify users of the risks. Then, users determine whether to collect information in case of high risks.

HUAWEI

Identification and Analysis

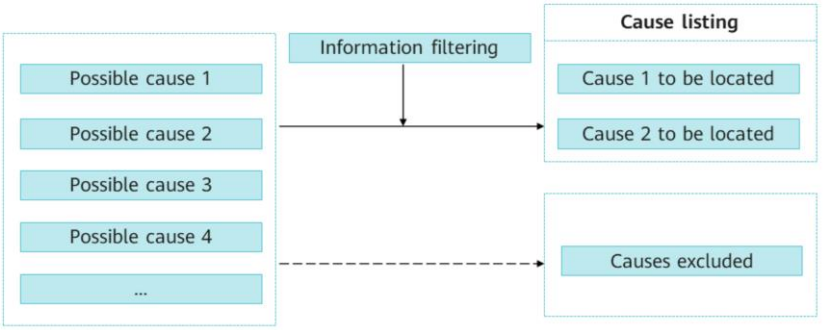
- In the identification and analysis phase, the collected information is analyzed and sorted.
  - By summarizing the fault, maintenance, and version-specific change information and using team (or personal) experience, you can obtain the list of possible causes for network faults.



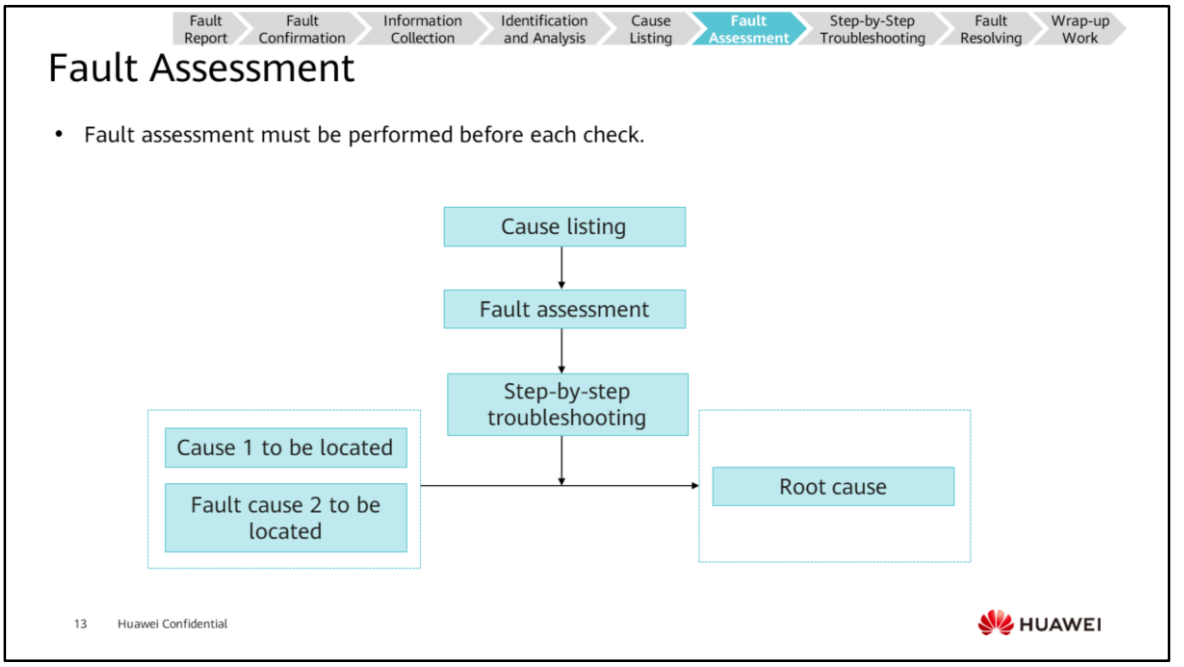
HUAWEI

Cause Listing

- In the cause listing phase, you must list all possible fault causes, sort out the most likely causes, and exclude the least possible causes to narrow down the troubleshooting scope.



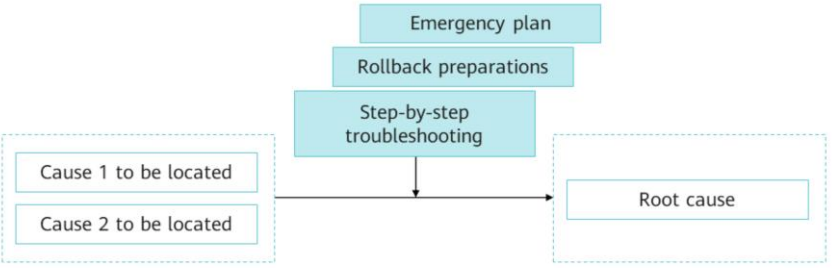
HUAWEI



- A temporary network environment may need to be built for fault evaluation.
  - If a complex network fault cannot be rectified within a short period of time after being evaluated and a user wants to immediately restore network availability, you advise the user to temporarily skip the faulty node and build an alternative network environment.
  - When building a temporary network environment, fully consider the urgency of solving problems and the risk of bypassing certain security restrictions. Fully communicate with users and implement the environment only after obtaining permissions.

Step-by-Step Troubleshooting

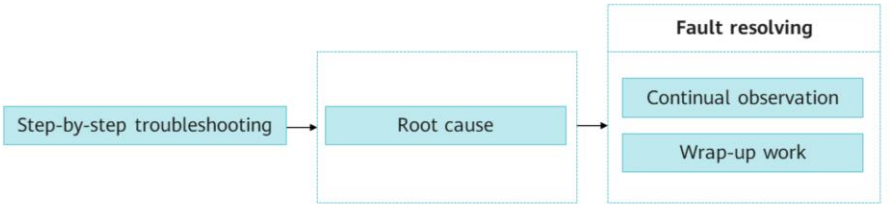
- In the phase of step-by-step troubleshooting, the conflict between the urgency of solving problems and the risk of introducing new faults must be balanced. Therefore, users must be clearly informed of the risks that may be induced the process. Perform the check only after being authorized.
- In some cases, network changes may be involved in the verification process. In this case, a complete emergency plan and rollback preparations must be made.



HUAWEI

Fault Resolving

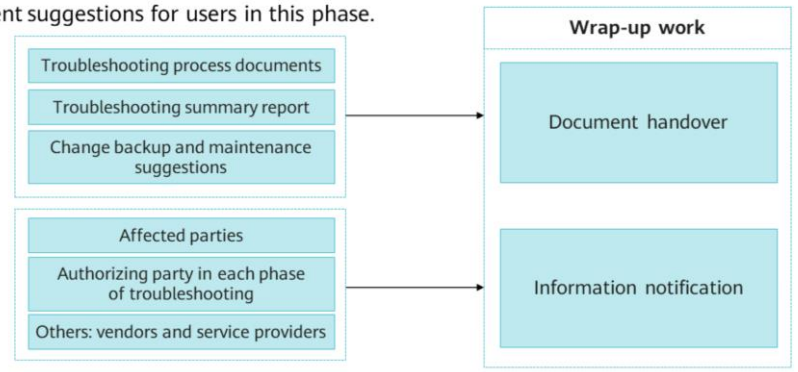
- After the root cause is found and the fault is rectified, the troubleshooting is complete.
- In a complex network environment, you have to observe the network for a period of time after the fault symptom disappears. On the one hand, you can confirm that the fault reported by the user has been rectified. On the other hand, you can confirm that no new fault is introduced during the troubleshooting process.



HUAWEI

Wrap-up Work

- Wrap-up work involves arranging related documents and sending notifications. Back up all changed configurations or software in the previous network troubleshooting process, and sort out and hand over troubleshooting documents. To prevent the same fault from occurring again, provide improvement suggestions for users in this phase.



HUAWEI

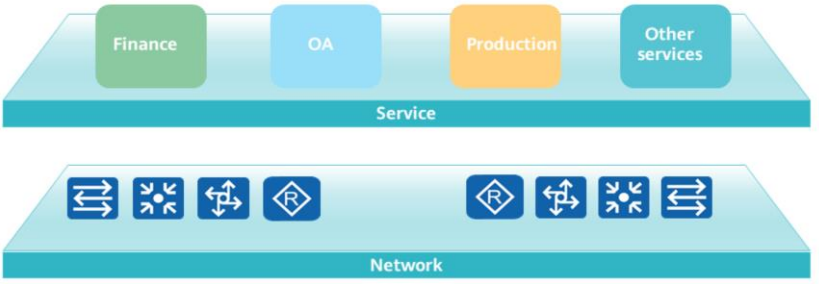
Contents

1. **Troubleshooting Data Communication Network Faults**
   - Overview of Network Faults
   - Troubleshooting Process
   - Core Ideas and Methods of Network Troubleshooting
2. Troubleshooting Common Network Faults

HUAWEI

Service Traffic Path-Centric Troubleshooting Ideals

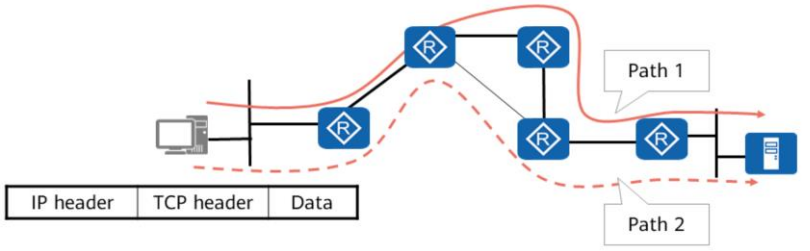
- The path along which service traffic passes is usually designed in the network planning phase. You merely need to know the round-trip path of service traffic adversely affected by a network fault, trace the path, and rectify the fault step by step.



HUAWEI

Determining a Service Traffic Path: Network Layer

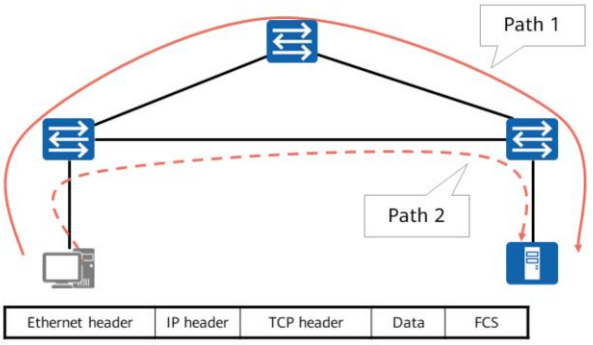
- Multiple paths may exist during packet forwarding. Therefore, you need to determine the path over which service traffic is transmitted based on the packet forwarding process.



HUAWEI

Determining a Service Traffic Path: Data Link Layer

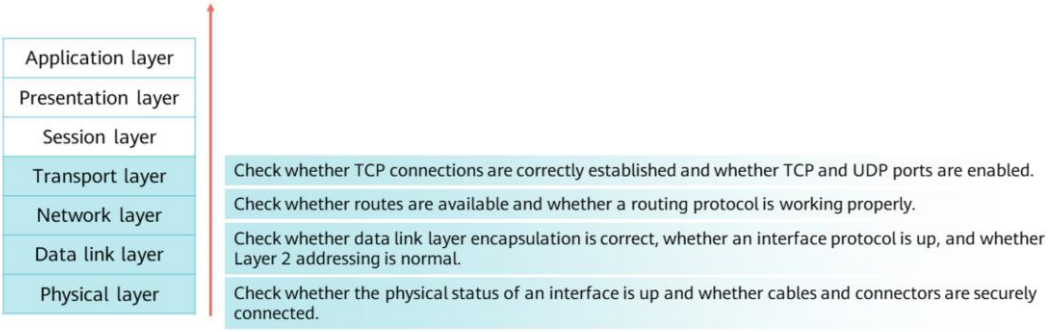
- Check how data frames of service traffic are forwarded by switches on a Layer 2 network.



HUAWEI

Layered Troubleshooting Approach

- The layered troubleshooting approach is simple, because all working models follow a simple rule: the upper-layer structure of any model can work properly as long as the lower-layer structure is working properly.



HUAWEI

Configuration Comparison Approach

- Compare configurations, software versions, and hardware models in normal and faulty states to find differences.
- Network troubleshooting personnel with less experience will use this method more frequently in practice.

```
#
 sysname r1
#
isis 1
 network-entity 49.0001.1000.0000.0001.00
#
interface Serial4/0/0
 link-protocol ppp
 ip address 10.0.12.1 255.255.255.0
 isis enable 1
 isis timer hello 30
#
interface LoopBack0
 ip address 10.0.1.1 255.255.255.255
 isis enable 1
#
```

Compare them

```
#
 sysname r1
#
isis 1
 network-entity 49.0001.1000.0000.0001.00
#
interface Serial4/0/0
 link-protocol ppp
 ip address 10.0.12.1 255.255.255.0
 isis enable 1
 isis timer hello 120
#
interface LoopBack0
 ip address 10.0.1.1 255.255.255.255
 isis enable 1
#
```


HUAWEI

Block-based Troubleshooting Approach

- The configuration files of Huawei network devices, such as switches and routers, are edited in a clear structure.
- If a fault occurs, you can narrow down the fault locating scope by classifying the fault into one or several categories:
  - Management (router name, password, service, and log)
  - Ports (address, encapsulation, cost, and authentication)
  - Routing protocols (static route, RIP, OSPF, BGP, and route import)
  - Policies (routing policy, policy-based routing, and security configuration)
  - Access (console port login, Telnet login, dial-up)
  - Applications (DNS, DHCP, and VPN configuration)

HUAWEI

Block-based Troubleshooting Approach: Example

- After the **display ip routing-table** command is run, only direct routes are displayed. What are possible causes?

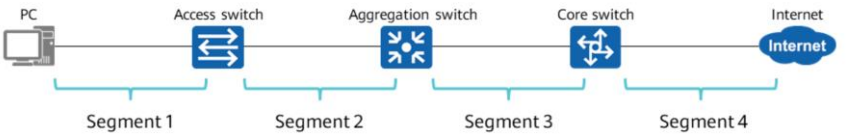
```
<R2>display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: Public
         Destinations : 16        Routes : 16

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

        10.0.12.1/32  Direct  0    0           D   10.0.12.1       Serial4/0/0
        10.0.12.2/32  Direct  0    0           D   127.0.0.1       Serial4/0/0
       10.0.12.255/32 Direct  0    0           D   127.0.0.1       Serial4/0/0
        10.0.23.0/24  Direct  0    0           D   10.0.23.2       GigabitEthernet0/0/0
        10.0.23.2/32  Direct  0    0           D   127.0.0.1       GigabitEthernet0/0/0
      10.0.23.255/32  Direct  0    0           D   127.0.0.1       GigabitEthernet0/0/0
```

- The fault is related to the following blocks: routing protocols, policies, and ports. If no routing protocol is configured or a routing protocol is incorrectly configured, the routing table may be empty. If an ACL is incorrectly configured, route update may be adversely affected. If the IP address, mask, or authentication configuration of an interface is incorrect, the routing table may be incorrect.

HUAWEI

Segment-based Troubleshooting Approach

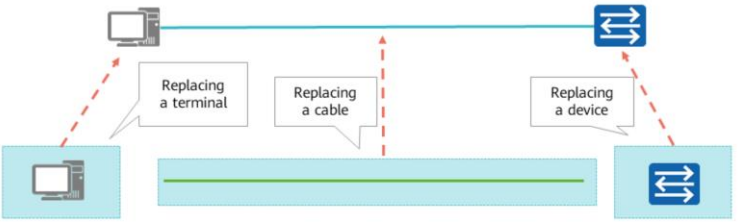
- Since data packets may pass through multiple routers and physical links, each segment may encounter a fault. In this situation, the segment-based approach applies.



HUAWEI

Replacement Approach

- The replacement approach is one of the most common methods for checking hardware problems.
- If a network cable may be faulty, replace it with another one in good condition. If an interface module may fail, replace it with another interface module that is working properly.



Requirements for Network Maintenance and Management Personnel

- Have an in-depth understanding of protocol requirements.
- Be able to guide a customer to describe the fault symptom and related information in detail.
- Fully understand the networks managed and maintained.
- Record troubleshooting documents and summarize troubleshooting experience.
- Be familiar with network troubleshooting approaches and combine them flexibly.

HUAWEI

Quiz

1. (Multiple-answer question) In the wrap-up work of a structured network troubleshooting process, which of the following parties should be notified of information? ( )
   A. Related parties affected by the fault
   B. Authorizing parties in each phase of troubleshooting
   C. Manufacturer and service providers
   D. Other irrelevant personnel who are interested in the root cause
2. (True or false) On a large-scale network, the comparison approach is the most effective method for troubleshooting faults. ( )
   A. True
   B. False

HUAWEI

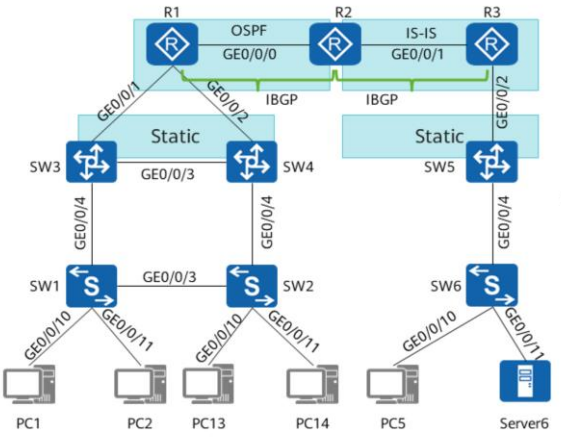
1. ABC
2. False

Contents

1. Troubleshooting Data Communication Network Faults
2. **Troubleshooting Common Network Faults**
   - LAN Faults
   - Route Faults
   - Service Faults

HUAWEI

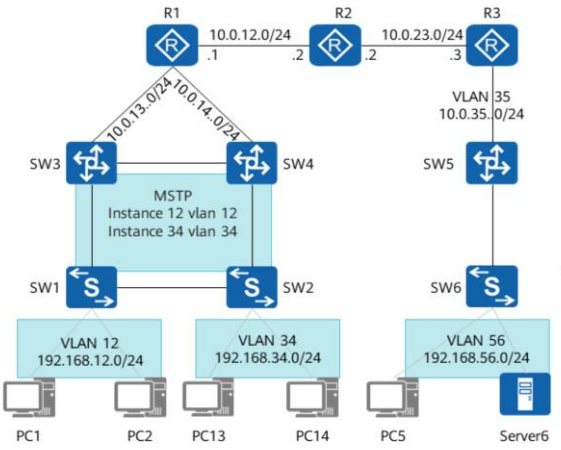
Troubleshooting Common Network Faults: Topology (1)



- During network maintenance, network engineers may encounter various network faults, such as login, route, and IP service faults. The left figure shows a part of the network architecture, which is used as an example to describe how to troubleshoot common network faults.
- Routing protocol overview:
  - OSPF: runs between R1 and R2. OSPF is enabled on all interfaces of R1, and GE 0/0/0 belongs to area 0.
  - IS-IS: runs between R2 and R3.
  - BGP: R1 and R3 establish IBGP peer relationships with R2 and function as clients of R2, namely, the RR.
  - Static route: SW3, SW4, and SW5 use static routes to connect to routers.

HUAWEI

Troubleshooting Common Network Faults: Topology (2)

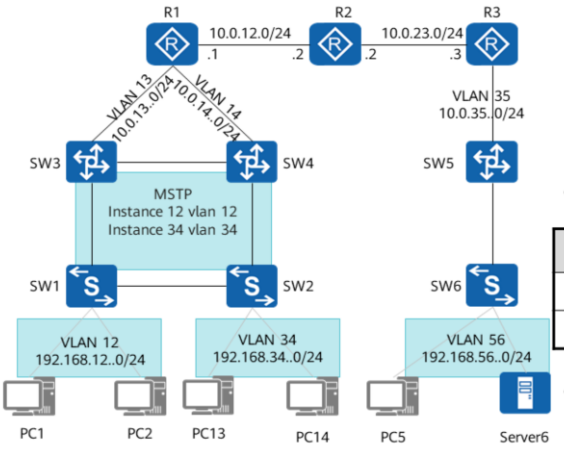


- IP address planning:
  - Loopback0: are 10.0.1.1/32, 10.0.2.2/32, and 10.0.3.3/32 on R1, R2, and R3, respectively.
  - Interconnection interfaces: The network segment is shown in the preceding figure, and the decimal number of the right-most octet in an IP address is a device ID.
  - Terminals: The network segment is shown in the figure. A gateway has the largest IP address on a network segment. PC5 obtains an IP address using DHCP, and R3 functions as a DHCP server. Other terminals use static IP addresses, with the decimal number of the right-most octet in an IP address indicating a device ID.
- Route advertisement planning:
  - On R2, configure OSPF and IS-IS to import routes from each other, so that Loopback0 addresses on R1, R2, and R3 are reachable.
  - R1 imports static routes destined for 192.168.12.0/24 and 192.168.34.0/24 into the BGP routing table.
  - R3 imports the static route destined for 192.168.56.0/24 into the BGP routing table.

HUAWEI

- R3 and SW5 are connected through Layer 3 sub-interfaces.

Troubleshooting Common Network Faults: Topology (3)



- MSTP planning:

Instance ID	VLAN ID	Root Bridge	Backup Bridge
Instance 12	VLAN 12	SW3	SW4
Instance 34	VLAN 34	SW4	SW3

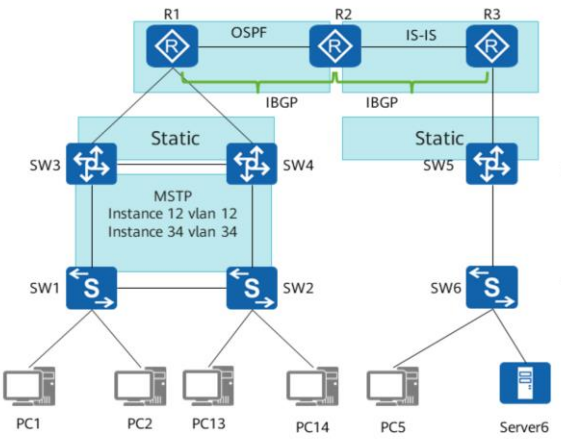
- VRRP planning:

Network Segment	Master Gateway	Backup Gateway	VRID	Virtual IP
192.168.12.0/24	SW3	SW4	1	192.168.12.254
192.168.34.0/24	SW4	SW3	2	192.168.34.254

- The Telnet username and password for login are Huawei and Huawei@123, respectively.

HUAWEI

Common Network Troubleshooting: Symptom

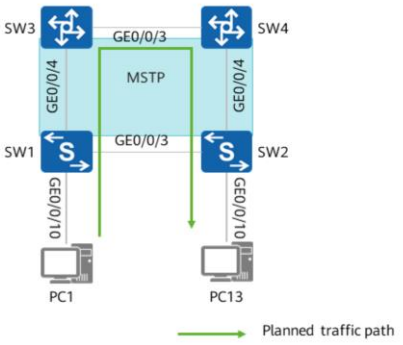


- The following symptoms are found:
  - PC1 and PC13 cannot communicate.
  - Server 6 provides the FTP service, but PC1 cannot use this service.
  - PC5 cannot communicate with any host.
- There are multiple possible causes. The preceding approaches are used to demonstrate how to troubleshoot the three faults.
- Assume that the symptoms have been confirmed. Skip the following steps in the subsequent troubleshooting: fault report, fault confirmation, information collection, and wrap-up work.

HUAWEI

PC1 and PC13 Cannot Communicate (1)

Simplified topology:



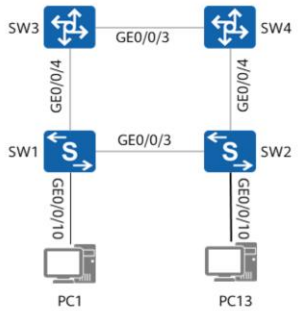
- The figure on the left shows the planned path for PC1-to-PC13 traffic. Use the layered, segment-based, and forwarding path-centric approaches to analyze the faults. PC1 and PC13 fail to communicate due to the following causes:
  - Physical link fault
  - Incorrectly configured IP address
  - VLAN configuration error
  - Loop
  - VRRP fault

HUAWEI

- This section describes common troubleshooting methods and tools, providing guidance for network maintenance personnel. The processing sequence in actual scenarios can be different from that in the example.

PC1 and PC13 Cannot Communicate (2)

Physical link fault



Incorrectly configured IP address

VLAN configuration error

Loop

VRRP fault

- On PC1 and PC13, choose **Control Panel** > **Network and Internet** > **Network Connection** > **Ethernet Cable** to check Ethernet cables and ensure that the physical cable connections to the PCs are correct. (The preceding path varies according to an operating system and version.)
- Check the physical interface status on each involved switch (SW1, for example). If the physical status of an interface is not up, use another interface to connect the switch to a PC.

```
<SW1>display interface brief | include up
```

Interface	PHY	Protocol	InUti	OutUti	inErrors	outErrors
GigabitEthernet0/0/3	**up**	up	0%	0%	0	0
GigabitEthernet0/0/4	**up**	up	0%	0%	0	0
GigabitEthernet0/0/10	**up**	up	0%	0%	0	0
GigabitEthernet0/0/11	**up**	up	0%	0%	0	0

- Ping 192.168.34.13 from PC1 and check whether the number of packets sent and received by the interface increases. If so, the physical link is working properly. If not, use another interface or replace the network cable.

```
<SW1>display interface GigabitEthernet 0/0/10
Description:
Switch Port, PVID :  12, TPID : 8100(Hex), The Maximum Frame Length is 9216
IP Sending Frames' Format is PKTFMT_ETHNT_2, Hardware address is 4c1f-cc69-6d7b
Hardware address is 4c1f-cc69-6d7b
    Last 300 seconds input rate 0 bytes/sec, 0 packets/sec
    Last 300 seconds output rate 0 bytes/sec, 0 packets/sec
    Input: 430 bytes, 6 packets
    Output: 197137 bytes, 1659 packets
```


HUAWEI

- This section uses the Windows 10 OS as an example to describe how to check the physical connection status of a PC.
- InUti: input bandwidth utilization
- OutUti: output bandwidth utilization

PC1 and PC13 Cannot Communicate (3)

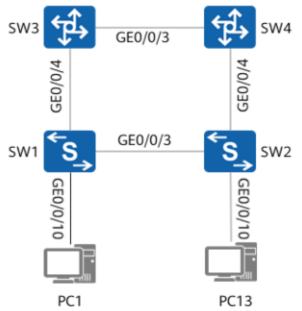
Physical link fault

Incorrectly configured IP address

VLAN configuration error

Loop

VRRP fault



- Check that PC1's physical IP address is set to 192.168.12.1 and the gateway address is set to 192.168.12.254.
- Check the IP addresses on SW3 and SW4 and ensure that the IP addresses are correctly configured. (A VLANIF interface without an IP address assigned will not go up and cannot implement Layer 3 forwarding.)

<SW3>display ip interface brief

Interface	IP Address/Mask	Physical	Protocol
MEth0/0/1	unassigned	down	down
NULL0	unassigned	up	up(s)
Vlanif1	unassigned	up	down
Vlanif12	**192.168.12.3/24**	**up**	**up**
Vlanif34	**192.168.34.3/24**	**up**	**up**

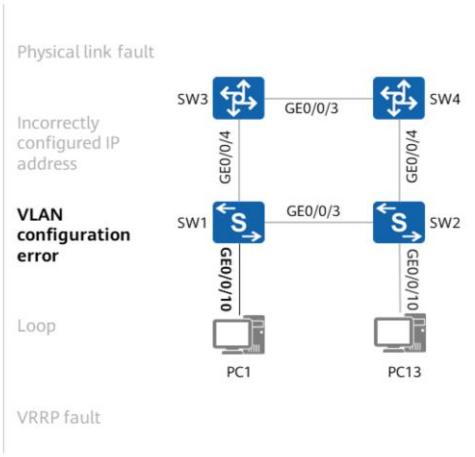
<SW4>display ip interface brief

Interface	IP Address/Mask	Physical	Protocol
MEth0/0/1	unassigned	down	down
NULL0	unassigned	up	up(s)
Vlanif1	unassigned	up	down
Vlanif12	**192.168.12.4/24**	**up**	**up**
Vlanif34	**192.168.34.4/24**	**up**	**up**

- The preceding information indicates that the IP addresses of the interfaces on SW3 and SW4 have been correctly configured.

HUAWEI

PC1 and PC13 Cannot Communicate (4)



- Query the switch port and VLAN configuration.

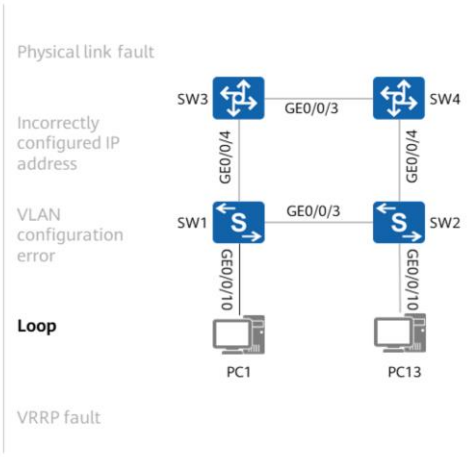
```
[SW1]display vlan
The total number of vlans is : 3
--------------------------------------------------------------------------------
U: Up;         D: Down;         TG: Tagged;         UT: Untagged;
MP: Vlan-mapping;               ST: Vlan-stacking;
#: ProtocolTransparent-vlan;    *: Management-vlan;
--------------------------------------------------------------------------------
VID  Type    Ports
--------------------------------------------------------------------------------
12   common  TG:GE0/0/11(U)      TG:GE0/0/10(U)
34   common  TG:GE0/0/11(U)      TG:GE0/0/10(U)
VID  Status  Property      MAC-LRN Statistics Description
--------------------------------------------------------------------------------
12   enable  default       enable  disable    VLAN 0012
34   enable  default       enable  disable    VLAN 0034
```

- According to the preceding information, VLAN configurations of GE 0/0/10 and GE 0/0/11 on SW1 are incorrect. Correct VLAN configuration errors. Repeat the preceding step to verify the configurations on the other three switches.
- After the VLAN is correctly configured on the switches, check whether PC1 can communicate with other IP addresses on the same network segment. For example, run the **ping 192.168.12.13** command on PC1. The command output shows that packet loss occurs and the delay is long.

HUAWEI

- GE 0/0/10 belongs to VLAN 12 and VLAN 34 and works in tagged mode, indicating that the interface is configured as a trunk interface and the PVID is not 12.

PC1 and PC13 Cannot Communicate (5)



- Check the MSTP status on each switch. All ports on SW4 are in the Forwarding state.

```
<SW4>display stp brief
 MSTID        Port                  Role  STP          State
    0         GigabitEthernet0/0/2  DESI  FORWARDING   NONE
    0         GigabitEthernet0/0/3  DESI  FORWARDING   NONE
    0         GigabitEthernet0/0/4  DESI  FORWARDING   NONE
   12         GigabitEthernet0/0/2  DESI  FORWARDING   NONE
   12         GigabitEthernet0/0/3  DESI  FORWARDING   NONE
   12         GigabitEthernet0/0/4  DESI  FORWARDING   NONE
   34         GigabitEthernet0/0/2  DESI  FORWARDING   NONE
   34         GigabitEthernet0/0/3  DESI  FORWARDING   NONE
   34         GigabitEthernet0/0/4  DESI  FORWARDING   NONE
```

- MSTP faults may be caused by an incorrect domain name setting, incorrect binding between instances and VLANs, or incorrect binding between ports and VLANs. Check the MSTP configuration on SW4.

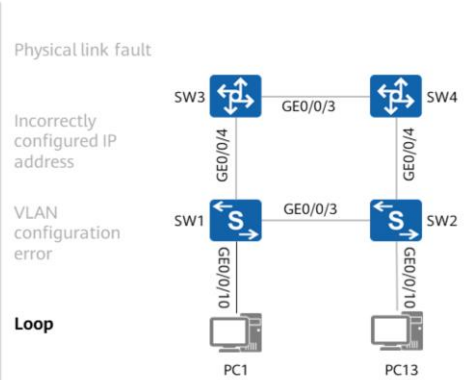
```
<SW4>display current-configuration | begin region-configuration
stp region-configuration
 region-name TEST //The correct domain name is test, not TEST.
 instance 12 vlan 12
 instance 34 vlan 34
 active region-configuration
#
```

- Correct the domain name on SW4. Ping 192.168.12.13 from PC1. As a result, packet loss occurs now and then.

HUAWEI

- A Layer 2 loop causes the following failures:
  - An attempt to remotely log in to a device fails.
  - An interface receives a large number of broadcast packets, which can be viewed in the **display interface** command output.
  - An attempt to log in to a device through the serial port is time consuming.
  - CPU usage exceeds 70%.
  - High packet loss occurs when a ping command is used.
  - The indicator of the VLAN interface with the loop occurring frequently blinks.
  - A PC receives a large number of broadcast packets.
  - A loop alarm is generated if loop detection is configured on a switch.

PC1 and PC13 Cannot Communicate (6)



- Check whether packet loss is caused by the loop. Check whether MAC address flapping detection is enabled on each involved switch and whether MAC address flapping is detected.

```
[SW3]display mac-address flapping
Mac-address Flapping Configurations :
------------------------------------------------------
 Flapping detection                 : Enable
 Aging  time(sec)                   : 300
 Quit-vlan Recover time(min)        : 10
 Exclude vlan-list                  : -
------------------------------------------------------
<SW3>display mac-address flapping record
Info: The mac-address flapping record does not exist.
```

- After observing for a while, find that the fault occurs during working hours. When the fault occurs, check the MAC address table, and find that the MAC address table is unstable. Then check STP statistics.

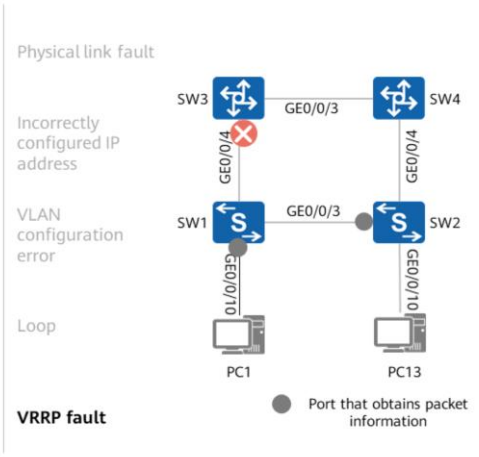
```
<SW3>display stp tc-bpdu statistics
------------------------- STP TC/TCN information ---------------
MSTID        Port                TC(Send/Receive)   TCN(Send/Receive)
  12    GigabitEthernet0/0/1          13/56              -/-
  12    GigabitEthernet0/0/3          22/18              -/-
  12    GigabitEthernet0/0/4          29/66              -/-
```

- During working hours, a switch port frequently alternates between up and down, and sends a large number of TC BPDUs. In this case, configure the switch port connected to the PC as an edge port.

HUAWEI

- After receiving STP TC BPDUs, the STP-enabled switch clears the MAC address table and re-learns MAC addresses. During this period, data forwarding is interrupted for a short period, causing packet loss.

PC1 and PC13 Cannot Communicate (7)



- Shut down SW3's GE 0/0/4. PC1 cannot ping PC13, causing a large number of packets to be discarded in a short period. After SW3 is restarted, the following alarm is generated on SW3:

```
<SW3>
ARP/4/ARP_DUPLICATE_IPADDR(l)[0]:Received an
ARP packet with a duplicate IP address from the interface.
(IpAddress=192.168.12.254, InterfaceName=Vlanif12,
MacAddress=0000-5e00-0101)
```

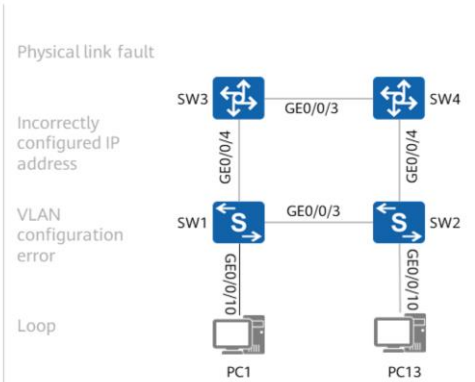
- The IP address is a virtual IP address. VRRP may be defective. Obtain packet information on SW1's GE 0/0/10 and SW2's GE 0/0/3. The addresses marked bold are source MAC and IP addresses. SW2 fails to obtain data packet information.

```
[SW1]capture-packet interface GigabitEthernet 0/0/10 destination
terminal
Packet: 6
----------------------------------------------------------------
00 00 5e 00 01 03 54 89 98 1f 5a 8d 81 00 00 0c
08 00 45 00 00 3c 39 aa 40 00 80 01 11 b8 c0 a8
0c 01 c0 a8 22 0d 08 00 f9 d1 82 3b 0a 71 08 09
0a 0b 0c 0d 0e 0f 10 11 12 13 14 15 16 17 18 19
1a 1b 1c 1d 1e 1f 20 21 22 23 24 25 26 27
----------------------------------------------------------------
```


HUAWEI

- The **capture-packet** command obtains information in service packets that match a configured rule. The obtained information is saved in a local file.
  - **capture-packet** { **interface** *interface-type interface-number* | **acl** *acl-number* } * [**vlan** *vlan-id* | **cvlan** *cvlan-id*] * **destination terminal** [**car** *cir car-value* | **time-out** *time-out-value* | **packet-num** *number* | *packet-len length*] *
  - Information in packets on the management interface cannot be obtained.
  - This command can only obtain information received by an interface, not information sent by an interface.

PC1 and PC13 Cannot Communicate (8)



VRRP fault

- As shown in the preceding slide, packets are dropped between SW1 and SW2, and the destination MAC address is **00 00 5e 00 01 03. The VRRP ID is 3, which should have been 1 as planned.** Check the VRRP status and configuration of SW3.

```
<SW3>display vrrp
        Vlanif12 | Virtual Router 3
          State : Master
          Virtual IP         : 192.168.12.254
          Master IP          : 192.168.12.3
          PriorityRun        : 100
          PriorityConfig     : 100
          MasterPriority     : 100
          Virtual MAC        : 0000-5e00-0103
<SW3>display current-configuration interface Vlanif 12
       #
       interface Vlanif12
        ip address 192.168.12.3 255.255.255.0
        vrrp vrid 3 virtual-ip 192.168.12.254
```

- According to the preceding analysis, a VRRP dual-master fault occurs because the VRID is incorrectly set. As a result, packet loss occurs during a VRRP switchover. Correct the configuration of SW3.
- Carry out a reliability test again. Find that no packet loss occurs during the switchover. Then, the fault is rectified.

HUAWEI

- The VRRP group numbers on SW3 and SW4 are different. After the VRRP group on SW3 detects a downlink fault, the VRRP status on SW4 does not change. The VRRP status on SW4 remains in the Master state. In this situation, sending gratuitous ARP messages is not triggered for an ARP entry update on the terminal.
- The destination MAC address of data frames sent from PC1 to a gateway is still 00 00 5e 00 01 03.
- After the link between SW1 and SW3 is disconnected, SW1 cannot forward packets to SW2, because SW1 does not have the MAC address entry of 00 00 5e 00 01 03.

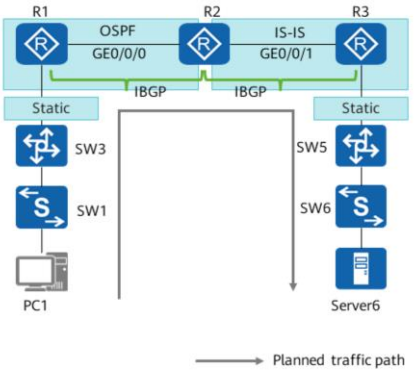
Contents

1. Troubleshooting Data Communication Network Faults
2. **Troubleshooting Common Network Faults**
   - LAN Faults
   - Route Faults
   - Service Faults

HUAWEI

PC1 Cannot Use the FTP Service (1)

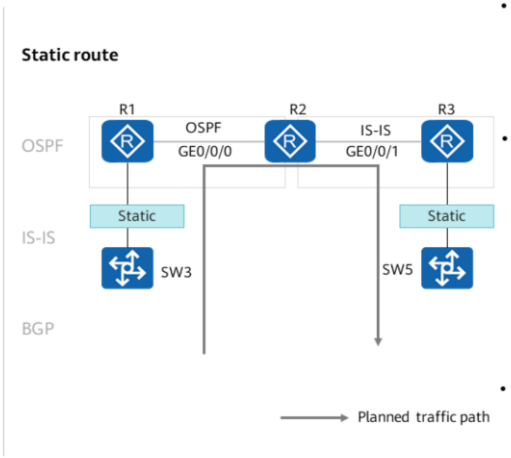
Simplified topology:



- In the preceding example, measures are taken to ensure no fault occurs between PC1 and SW3 and between server 6 and SW5.
- As shown in the figure on the left, possible causes for an FTP failure on PC1 are as follows:
  - Physical link fault (done)
  - Route faults
    - Static route
    - OSPF
    - BGP
    - IS-IS
  - Traffic control fault
  - Server fault (done)
- This section describes how to troubleshoot route.

HUAWEI

PC1 Cannot Use the FTP Service (2)



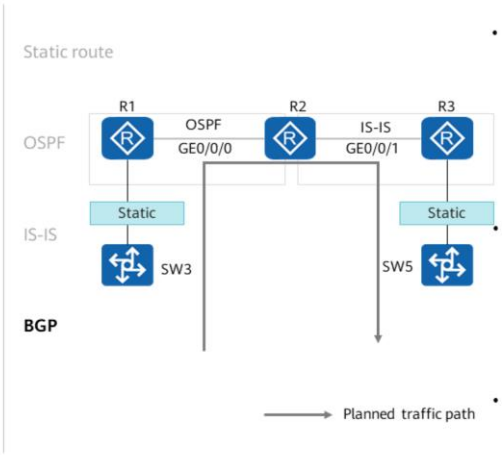
- Data packets are forwarded hop by hop. All routing devices along a path must have routes to the destination. First, check whether routes destined for server 6 exist on all devices through which data packets sent from PC1 to server 6 pass.
- Check the static route configuration on SW3.
  - When configuring a static route, specify only the outbound interface name for a P2P interface. For a broadcast interface, also specify a next-hop IP address.

```
<SW3>display ip routing-table protocol static
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
      0.0.0.0/0     Static  60   0           RD  10.0.13.1       Vlanif13
```

- The command output indicates that the static route has been correctly configured on SW3.

HUAWEI

PC1 Cannot Use the FTP Service (3)



- Check the routing table on R1.

```
<R1>display  ip routing-table 192.168.56.0
```

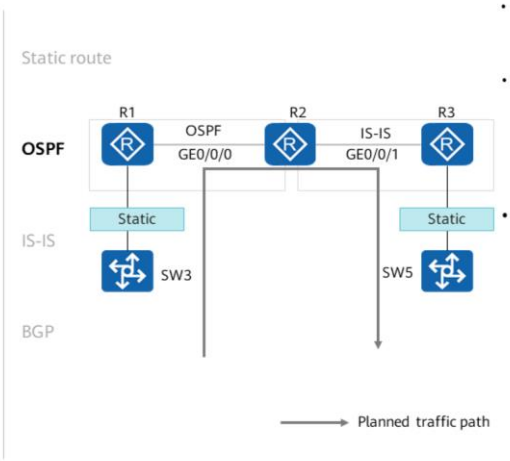
- The command output shows that R1 does not have a route to 192.168.56.0. Check whether a BGP peer relationship is properly established between R1 and R2.

```
<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 : 0
  Peer        AS  MsgRcvd MsgSent OutQ Up/Down   State  PrefRcv
  10.0.2.2   100       0       0    0 0:00:05    Idle        0
```

- The preceding command output shows that the BGP peer relationship fails to be established. Possible causes are as follows:
  - The loopback0 interface of the remote device is unreachable.
  - The AS number of the local or remote device is incorrect.
  - The **peer ebgp-max-hop** command used to allow the establishment of an indirect EBGP peer relationship is not run.
  - Router IDs on both ends are the same.
- According to the command output, the numbers of sent and received BGP packets are 0, indicating that the loopback0 interface on the remote device may be unreachable.

HUAWEI

PC1 Cannot Use the FTP Service (4)



- On R1, check the route destined for the BGP peer.

```
<R1>display ip routing-table 10.0.2.2
```

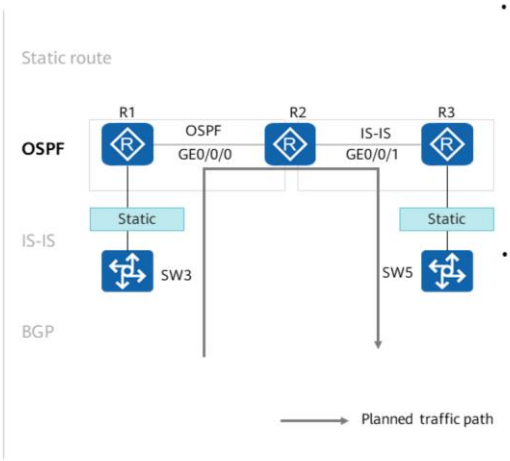
- The command output shows that R1 does not have a route to 10.0.2.2. Check whether an OSPF neighbor relationship is properly established between R1 and R2.

```
<R1>display ospf peer
                OSPF Process 1 with Router ID 10.0.1.1
```

- The preceding command output shows that the OSPF neighbor relationship fails to be established. Possible causes are as follows:
  - Router IDs on both ends are the same.
  - Area IDs do not match on both ends.
  - Network masks do not match on both ends.
  - MTUs do not match on both ends.
  - On an MA network, DR priorities of all devices are set to 0.
  - Authentication passwords do not match on both ends.
  - An interface is configured as a silent interface.
  - Time parameters do not match on both ends.

HUAWEI

PC1 Cannot Use the FTP Service (5)



- Check OSPF error information on R1.

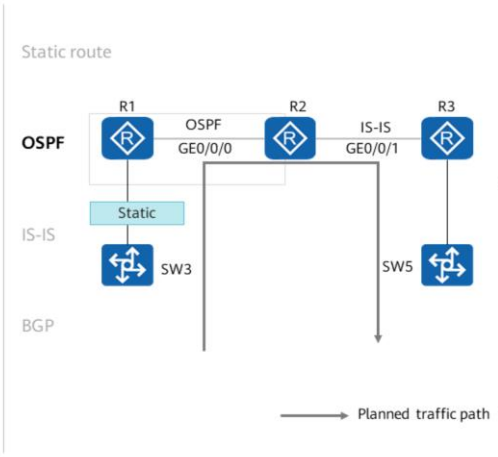
```
[R1]display ospf error
General packet errors:
 0    : IP: received my own packet      0    : Bad packet
 0    : Bad version                     0    : Bad checksum
 0    : Bad area id                     0    : Drop on unnumbered interface
 0    : Bad virtual link                0    : Bad authentication type
 0    : Bad authentication key          0    : Packet too small
 0    : Packet size > ip length         0    : Transmit error
 0    : Interface down                  0    : Unknown neighbor
 0    : Bad net segment                 0    : Extern option mismatch
 133  : Router id confusion
```

- The preceding command output shows that a router ID conflict may cause a failure to establish an OSPF neighbor relationship. To accurately locate the fault, enable OSPF debugging on R1.

```
<R1>terminal debugging
Info: Current terminal debugging is on.
<R1>debugging ospf packet interface GigabitEthernet 0/0/0
```


HUAWEI

PC1 Cannot Use the FTP Service (6)



- Check the configuration of the local OSPF protocol.

```
<R1>display ospf interface GigabitEthernet 0/0/0 verbose
            OSPF Process 1 with Router ID 10.0.1.1
 Interface: 10.0.12.1 (GigabitEthernet0/0/0)
 Cost: 1        State: DR       Type: Broadcast    MTU: 1500
 Designated Router: 10.0.12.1
 Backup Designated Router: 0.0.0.0
 Timers: Hello 10 , Dead 40 , Poll  120 , Retransmit 5 , Transmit Delay 1
  IO Statistics
        Type       Input     Output
       Hello          36         36
```

- Check OSPF error information in the debugging information on R1. Some information is as follows:

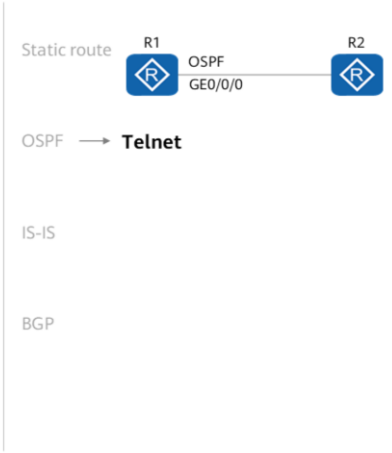
```
: Source Address: 10.0.12.2
: Destination Address: 224.0.0.5
: Ver# 2, Type: 1 (Hello)
: Length: 44, Router: 10.0.1.1
: Area: 0.0.0.0, Chksum: db9c
: AuType: 00
: Key(ascii): ********
: Net Mask: 255.255.255.0
: Hello Int: 10, Option: _E_
: Rtr Priority: 1, Dead Int: 40
```

- After comparison, it is found that the interval at which Hello packets are sent, mask, and authentication information on one end matches those on the other end, and only the router ID conflict occurs.

HUAWEI

- The debugging information on R1 shows that the OSPF router ID carried in the Hello packets sent from 10.0.12.2 is the same as the OSPF router ID on R1.

PC1 Cannot Use the FTP Service (7)



- Log in to R2 and change the OSPF router ID. However, the attempt to use Telnet to log in to R2 fails.
- Common causes of Telnet login failures are as follows:
  - A route is unavailable, and a TCP connection cannot be established between the client and server.
  - Telnet is disabled on the server.
  - The number of users logging in to a device reaches a specified upper limit.
  - An ACL is bound to a VTY user interface.
  - An access protocol configured in the VTY user interface view is incorrect. If the **protocol inbound ssh** command is used, the attempt to use Telnet for login fails.
- Log in to R2 through the console port and check whether Telnet is enabled.

```
[R2]display telnet server status
 TELNET IPV4 server                :Enable
 TELNET IPV6 server                :Enable
 TELNET server port                :23
```

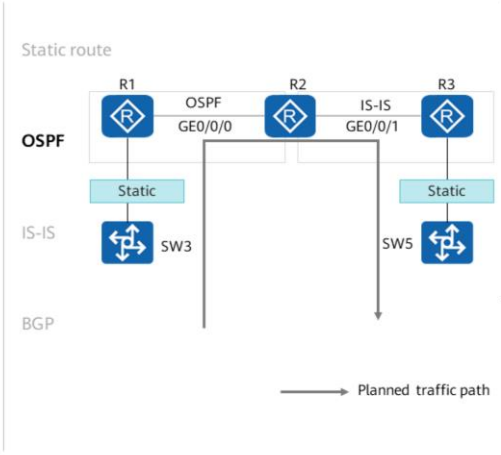
- Check whether Telnet is allowed in the VTY view.

```
[R2-ui-vty0-4]display this
user-interface vty 0 4
 authentication-mode aaa
 protocol inbound ssh
```

- Modify the configuration of R2 to support Telnet in the VTY user interface view. The test result shows that the attempt to log in to R2 is successful.

HUAWEI

PC1 Cannot Use the FTP Service (8)



- Change the OSPF router ID on R2, restart the OSPF process to make the router ID take effect, and check the OSPF neighbor relationship status.

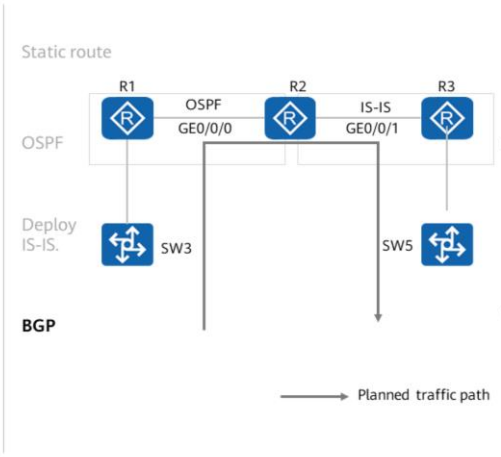
```
<R2>display ospf peer
                OSPF Process 1 with Router ID 10.0.12.2
                        Neighbors
 Area 0.0.0.0 interface 10.0.12.2(GigabitEthernet0/0/0)'s neighbors
 Router ID: 10.0.1.1          Address: 10.0.12.1
   State: Full  Mode:Nbr is  Slave  Priority: 1
   DR: 10.0.12.2  BDR: 10.0.12.1  MTU: 0
   Dead timer due in 35  sec
   Retrans timer interval: 5
   Neighbor is up for 00:09:17
   Authentication Sequence: [ 0 ]
```

- Check whether the route to 10.0.2.2 exists on R1.

```
<R1>display ip routing-table 10.0.2.2
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
      10.0.2.2/32   OSPF    10   1           D   10.0.12.2       GE0/0/0
```


HUAWEI

PC1 Cannot Use the FTP Service (9)



- Check the BGP peer relationship status on R1.

```
<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      AS  MsgRcvd  MsgSent Up/Down      State
 10.0.2.2   100     25       26    0:19:22  Established
```

- Check whether the BGP routing table of R1 contains a route destined for 192.168.56.0/24.

```
<R1>display bgp routing-table
 BGP Local router ID is 10.0.1.1
 Total Number of Routes: 2
      Network          NextHop  MED  LocPrf  Path/Ogn
 *>192.168.12.0/24     0.0.0.0   0      0        ?
```

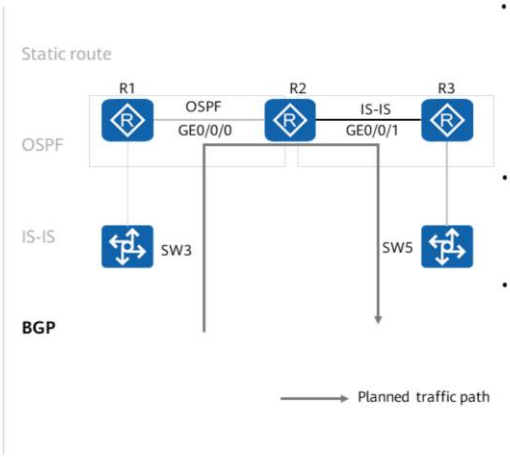
- R1 still does not have an available route. As R1 should have get the route from R3, check whether the route is imported into the BGP routing table on R3.

```
<R3>display bgp routing-table
 BGP Local router ID is 10.0.3.3
 Total Number of Routes: 1
      Network          NextHop  MED  LocPrf  Path/Ogn
 *> 192.168.56.0       0.0.0.0   0      0        ?
```


HUAWEI

- On R3, the command output shows that the route to 192.168.56.0/24 has been imported into the BGP routing table.

PC1 Cannot Use the FTP Service (10)



- Check the BGP peer status on R3.

```
<R3>display bgp peer
 BGP local router ID : 10.0.3.3
 Local AS number : 100
 Total number of peers : 1    Peers in established state : 0
  Peer        AS  MsgRcvd  MsgSent  OutQ  Up/Down   State  PrefRcv
  10.0.2.2   100        0        0     0  0:00:05    Idle        0
```

- The BGP peer relationship is not established between R3 and R2. Check whether a route destined for 10.0.2.2/32 exists on R3.

```
<R3>display ip routing-table 10.0.2.2
```

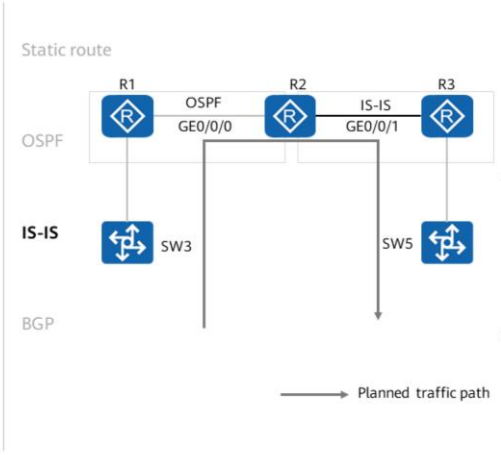
- R3 does not have the route to 10.0.2.2/32. As IS-IS runs between R2 and R3, check whether an IS-IS neighbor relationship is properly established between R3 and R2.

```
<R3>display isis peer
                  Peer information for ISIS(1)
  SystemId      Interface     CircuitId          State Type PRI
  0100.0000.2002 GE0/0/1      0100.0000.2002.01  Up    L2   64
```


HUAWEI

- Possible causes for the failure to establish an IS-IS neighbor relationship are as follows:
  - Area IDs do not match on both ends. (The inconsistency adversely affects only level-1 neighbor relationships.)
  - IS-IS levels do not match on both ends. (Note that on Huawei devices if the system level differs from the interface circuit level, the system level takes effect.)
  - Interface authentication settings do not match on both ends.
  - System ID lengths do not match or system ID conflict occurs.
  - The IP addresses are on different network segments. (Source check is enabled for IS-IS on a broadcast network, and can be disabled.)

PC1 Cannot Use the FTP Service (11)



- The IS-IS neighbor relationship is properly established, but R3 cannot obtain the route to 10.0.2.2/32. Possible causes are as follows:
  - IS-IS is not enabled on an interface.
  - Cost styles do not match on both ends.
  - A routing policy is configured on the device.
  - Network types do not match on both ends.
- Check information about an IS-IS interface on R2.

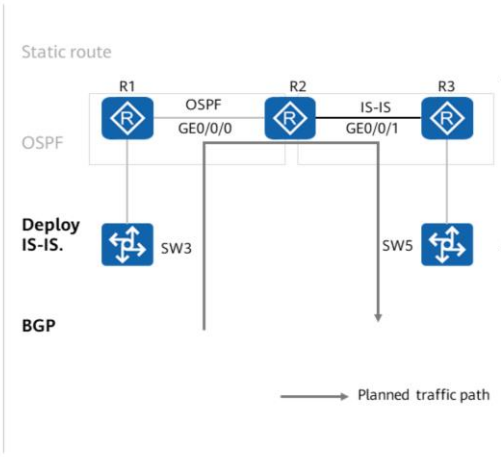
```
<R2>display isis interface
                Interface information for ISIS(1)

 Interface  Id  IPV4.State  MTU  Type    DIS
 GE0/0/1    1       Up     1497 L1/L2  No/Yes
  Loop0     1       Up     1500 L1/L2    --
```

- Check the IS-IS configuration on R2.

```
isis 1
 is-level level-2
 cost-style wide
 network-entity 49.0001.0100.0000.2002.00
 import-route ospf 1
```


HUAWEI

PC1 Cannot Use the FTP Service (12)



- Check the IS-IS configuration on R3.

```
isis 1
 is-level level-2
 network-entity 49.0001.0100.0000.3003.00
 #
```

- The following information shows that the cost style of R2 does not match that of R3. Change the cost style of R3 to wide. Then, check whether R3 has a route to 10.0.2.2/32.

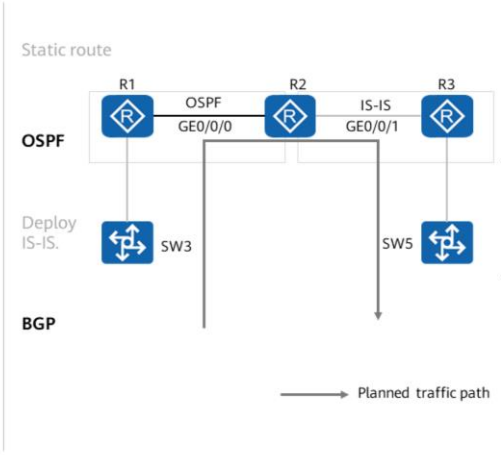
```
<R3>display ip routing-table 10.0.2.2
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
10.0.2.2/32         ISIS-L2 15   10        D     10.0.23.2       GE0/0/1
```

- Check whether the BGP peer relationship on R3 is restored and check the BGP routing table of R3.

```
<R3>display bgp peer
 BGP local router ID : 10.0.23.3
  Peer        AS  MsgRcvd  MsgSent  Up/Down      State
  10.0.2.2   100        8        7  0:04:42  Established
<R3>display bgp routing-table
Total Number of Routes: 2
     Network            NextHop        MED        LocPrf     Path/Ogn
*>I  192.168.12.0       10.0.1.1       0          100        ?
*>   192.168.56.0       0.0.0.0        0          0          ?
```


HUAWEI

PC1 Cannot Use the FTP Service (13)



- R3 has correctly advertised routes and learned the route to 192.168.12.0/24. Check whether R1 has a route to 192.168.56.0/24.

```
<R1>display bgp routing-table
BGP Local router ID is 10.0.1.1
     Network         NextHop      MED    LocPrf    Path/Ogn
*>  192.168.12.0     0.0.0.0      0        0         ?
I  192.168.56.0      10.0.3.3     0        100       ?
```

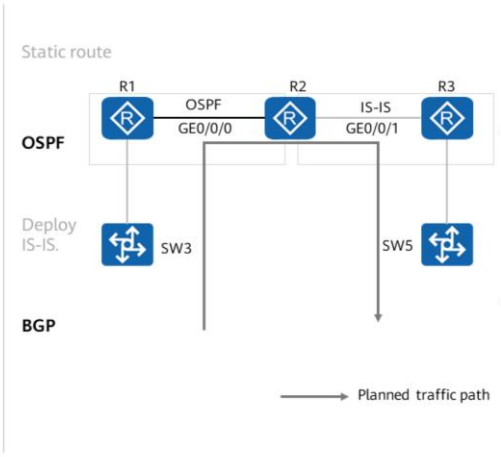
- R1 has received the BGP route from R3, but the route is unavailable. The possible cause is that the next hop is unreachable. On R1, check whether there is a route to 10.0.3.3/32.

```
<R1>display ip routing-table 10.0.3.3
```

- The command output shows that the routing table of R1 does not contain the route to 10.0.3.3/32. This route should have been imported by R2 from IS-IS into the OSPF routing table. Possible causes are as follows:
  - A routing policy is configured on R1.
  - R2 does not import IS-IS routes into the OSPF routing table.
  - Type 5 LSAs are filtered out in the outbound direction of R2's interface.

HUAWEI

PC1 Cannot Use the FTP Service (14)



- Check the LSDB of R1.

```
<R1>display ospf lsdb
  OSPF Process 1 with Router ID 10.0.1.1
        Area: 0.0.0.0
  Type      LinkStateID     AdvRouter     Age  Len  Sequence   Metric
  Router    10.0.2.2        10.0.2.2      5    48   80000003   1
  Router    10.0.1.1        10.0.1.1      3    48   8000000D   1
  Network   10.0.12.1       10.0.1.1      3    32   80000002   0
```

- R1 does not have Type 5 LSAs. Check whether R1 imports IS-IS routes into the OSPF routing table.

```
<R2>display current-configuration configuration ospf
#
ospf 1 router-id 10.0.2.2
 area 0.0.0.0
  network 10.0.2.2 0.0.0.0
  network 10.0.12.2 0.0.0.0
#
```

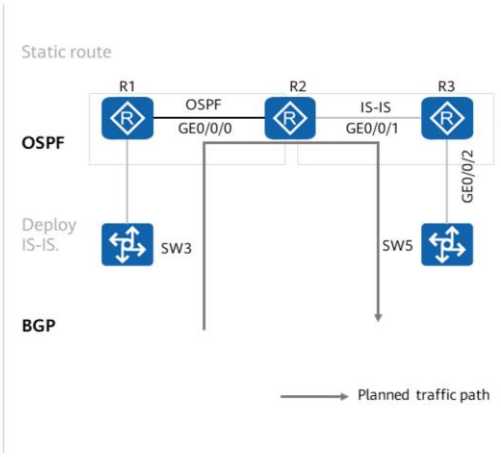
- Modify the configuration of R2, import IS-IS routes into the OSPF routing table, and then check the routing table of R1.

```
<R1>display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
      10.0.3.3/32   O_ASE   150  1           D   10.0.12.2       GE0/0/0
  192.168.56.0/24   IBGP    255  0           RD  10.0.3.3        GE0/0/0
```


HUAWEI

- After the configuration of R2 is modified, the route to 10.0.3.3/32 is displayed on R1.

PC1 Cannot Use the FTP Service (15)



- Run the **traceroute 192.168.56.6** command on PC1.

```
traceroute to 192.168.56.6, 8
 1  192.168.12.3   63 ms  46 ms  47 ms
 2  10.0.13.1   78 ms  63 ms  62 ms
 3  10.0.12.2   94 ms  63 ms  78 ms
 4  10.0.23.3   94 ms  62 ms  63 ms
 5  * * *
```

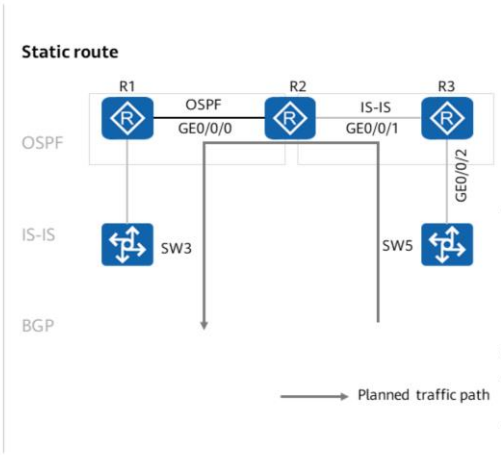
- R3 does not respond to the received data packets. Enable traffic statistics collection on R3's GE 0/0/2 and check whether R3 sends the data packets through GE0/0/2.

```
[R3] acl 3000
[R3-acl-adv-3000]rule 5 permit ip source 192.168.12.1 0 destination
192.168.56.6 0
[R3-acl-adv-3000]quit
[R3]traffic classifier trafficSta
[R3-classifier-trafficSta]if-match acl 3000
[R3-classifier-trafficSta]quit
[R3]traffic behavior trafficSta
[R3-behavior-trafficSta]statistic enable
[R3-behavior-trafficSta]quit
[R3]traffic policy trafficSta
[R3-trafficpolicy-trafficSta]classifier trafficSta behavior trafficSta
[R3-trafficpolicy-trafficSta]quit
[R3]interface GigabitEthernet0/0/2.35
[R3-GigabitEthernet0/0/2.35]traffic-policy trafficSta outbound
```


HUAWEI

- After R1 learns the route, PC1 still cannot access the FTP service provided by server 6. In this case, run the traceroute command to check connectivity between R1 and server 6.
- Based on traffic statistics, the analysis is as follows:
  - Check whether the traffic reaches the inbound interface of the device and determine whether packet loss occurs on the upstream device.
  - Check whether the traffic is forwarded to the outbound of the device and determine whether packet loss occurs on the device.
  - Check whether Layer 2 and Layer 3 information about traffic on the inbound interface of the device is correct and determine whether the upstream device forwards and encapsulates packets properly.
  - Check whether the Layer 2 and Layer 3 information about the outbound interface is correct and determine whether the device forwards and encapsulates packets properly.
  - Check whether transient traffic flapping occurs due to MAC address flapping, route changes, or IP address conflicts.
- Procedure for configuring traffic statistics collection:
  - Configure an ACL rule to match traffic to be collected.
  - Configure a traffic classifier.
  - Configure a traffic behavior and configure traffic statistics collection in the traffic behavior.
  - Configure a traffic policy; bind the traffic classifier and behavior to the traffic policy; apply the traffic policy to the inbound direction of the switch to collect statistics on packets of different users.

PC1 Cannot Use the FTP Service (16)



- Check traffic statistics on R3. No packet loss occurs on R3.

```
<R3> display traffic policy statistics interface GigabitEthernet0/0/2.35
outbound
Interface:GigabitEthernet0/0/2.35
 Traffic policy inbound: trafficSta
 Rule number: 1
 Current status: OK!
Item                  Sum(Packets/Bytes)
---------------------------------------
Matched                 50/400
  Passed                50/400
  Dropped               0/0
```

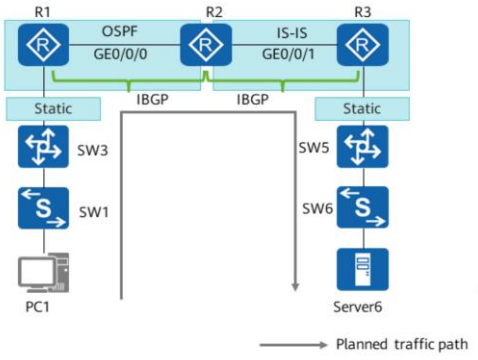
- Check the routing table of SW5 in the direction where server 6 sends data packets.

```
[SW5]display ip routing-table
 Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
      10.0.35.0/24   Direct  0    0           D   10.0.35.5       Vlanif35
    192.168.56.0/24  Direct  0    0           D   192.168.56.5    Vlanif56
```

- SW5 does not have a static route to PC1. Configure a static route on SW5.
- Check whether PC1 and server 6 can communicate and use the FTP service properly.
- After the preceding operations are complete, the troubleshooting is complete.

HUAWEI

PC1 Cannot Use the FTP Service Summary



- The following faulty points are involved in the troubleshooting:
  - Incorrect static route configuration
  - Failure to establish an OSPF neighbor relationship
  - Incorrect IS-IS route calculation
  - Failure to establish a BGP peer relationship
  - Incorrect BGP route selection
  - Telnet login failure
- The following tools are used:
  - packet information obtaining tool
  - Traffic statistics collection tool
  - Tracert

HUAWEI

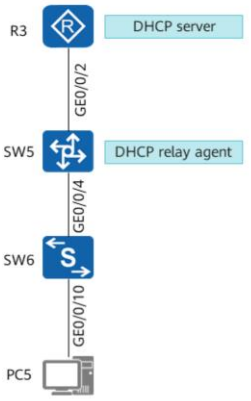
Contents

1. Troubleshooting Data Communication Network Faults
2. **Troubleshooting Common Network Faults**
   - LAN Faults
   - Route Faults
   - Service Faults

HUAWEI

PC5 Cannot Communicate with Any Host (1)

- Simplified topology:



- PC5 cannot communicate with any host. The possible cause is that the physical link to PC5 is abnormal, or PC5's IP address is incorrect.
- Check the IP address of PC5. PC5 fails to obtain an IP address.

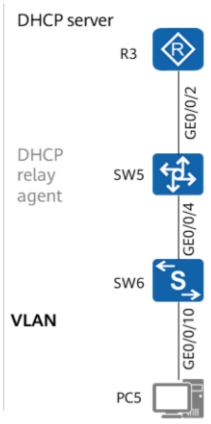
```
PC>ipconfig
IPv4 address        0.0.0.0
Subnet mask         0.0.0.0
Gateway             0.0.0.0
Physical address    54-89-98-39-22-B7
DNS server          0.0.0.0
```

- PC5 obtains an IP address using DHCP. The common causes of DHCP faults are as follows:
  - The link between the client and server becomes faulty.
  - The DHCP function is disabled on a device.
  - The DHCP address allocation mode is not selected on a VLANIF interface.
  - No IP address is available in an address pool.

HUAWEI

- If the client and server are on different network segments and a relay agent is deployed between them
  - The link between the DHCP relay agent and server becomes faulty.
  - The DHCP function is not enabled globally on a device. As a result, the DHCP function does not take effect.
  - No DHCP server is specified on the DHCP relay agent.
  - The DHCP relay agent and server are unreachable.

PC5 Cannot Communicate with Any Host (2)



- Check whether the physical link to PC5 is normal.
- Check whether a Layer 2 loop occurs on SW6 and check the VLAN configuration.

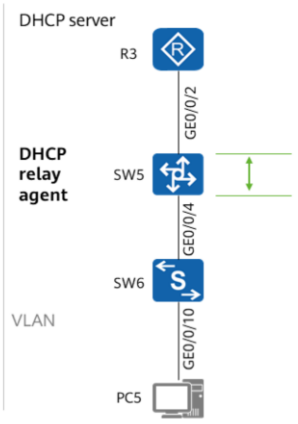
```
<SW6>display stp brief
 MSTID        Port                      Role  STPState
    0         GigabitEthernet0/0/4      ROOT  FORWARDING
    0         GigabitEthernet0/0/10     DESI  FORWARDING
    0         GigabitEthernet0/0/11     DESI  FORWARDING

<SW6>display vlan
The total number of vlans is : 2
--------------------------------------------------------------------------------
56   common  TG:GE0/0/4(U) GE0/0/10(U) UT: GE0/0/11(U)
```

- The preceding information shows that the physical status of GE 0/0/10 is normal, but GE 0/0/10 is configured as a trunk interface. Modify the configuration on SW6 to configure GE 0/0/10 as an access interface and set the VLAN ID to 56.
- After the modification, check whether PC5 can properly obtain an IP address.

```
PC>ipconfig
IPv4 address                  0.0.0.0
Subnet mask                   0.0.0.0
Gateway                       0.0.0.0
Physical address              54-89-98-39-22-B7
DNS server                    0.0.0.0
```


HUAWEI

PC5 Cannot Communicate with Any Host (3)



- SW5 is a DHCP relay agent. Query the global configuration of SW5.

```
<SW5>display current-configuration
dhcp enable
#
interface Vlanif35
 ip address 10.0.35.5 255.255.255.0
 dhcp select relay
 dhcp relay server-ip 10.0.35.3
#
```

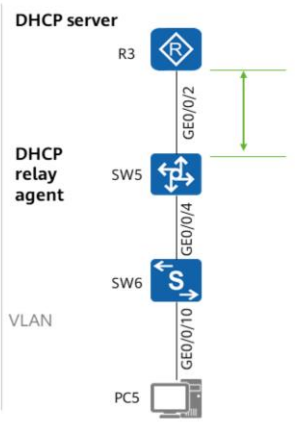
- The DHCP service has been enabled on SW5 and the DHCP relay agent has been configured. However, after the data packets sent by PC5 pass through SW6, SW6 adds the tag with VLAN ID 56 to the packets before forwarding them. As a result, the DHCP relay interface configured on SW5 is incorrect.
- Modify the configuration of SW5:

```
[SW5]interface Vlanif 56
[SW5-Vlanif56]dhcp select relay
[SW5-Vlanif56]dhcp relay server-ip 10.0.35.3
```

- After the modification, PC5 still cannot obtain an IP address.

HUAWEI

PC5 Cannot Communicate with Any Host (4)



- Query the DHCP relay status of SW5.

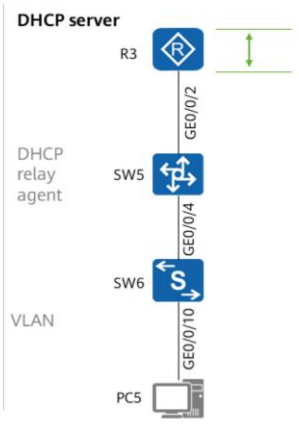
```
[SW5]display dhcp relay statistics
 The statistics of DHCP RELAY:
    DHCP packets received from clients  : 11
    DHCP DISCOVER packets received      : 11
    DHCP REQUEST packets received       : 0
    DHCP RELEASE packets received       : 0
    DHCP INFORM packets received        : 0
    DHCP DECLINE packets received       : 0
    DHCP packets sent to clients        : 0
    DHCP packets received from servers  : 0
    DHCP packets sent to servers        : 11
```

- SW5 has sent packets to the DHCP server, but does not receive any responses.
- Check the DHCP server configuration on R3.

```
<R3>display  current-configuration
dhcp enable
#
ip pool test
 gateway-list 192.168.56.254
 network 192.168.56.0 mask 255.255.255.0
 excluded-ip-address  192.168.56.6
 dns-list 192.168.1.1
#
interface GigabitEthernet0/0/2
dhcp select global
```


HUAWEI

PC5 Cannot Communicate with Any Host (5)



- R3 and SW5 are connected through sub-interfaces. Therefore, enable the DHCP server service on a sub-interface, instead of the physical interface GE0/0/2.

```
[R3]interface GigabitEthernet 0/0/2.35
[R3-GigabitEthernet0/0/2.35]dhcp select global
```

- After the preceding configurations are complete on R3, check the status of the DHCP server on R3.

```
<R3>display dhcp server statistics
 DHCP Server Statistics:
 Client Request                  2
  Dhcp Discover                  1
  Dhcp Request                   1
 Server Reply                    2
  Dhcp Offer                     1
  Dhcp Ack                       1
 Bad Messages                    0
```

- Check the IP address on PC5.

```
PC>ipconfig
IPv4 address        192.168.56.253
Subnet mask         255.255.255.0
Gateway             192.168.56.254
Physical address    54-89-98-39-22-B7
DNS server          192.168.1.1
```

- PC5 has obtained an IP address and can use it to communicate with all hosts. The troubleshooting is complete.

HUAWEI

Quiz

1. (Multiple-answer question) Which of the following causes are possible for a failure to establish an OSPF neighbor relationship? ( )
   A. Router ID conflict
   B. Area ID inconsistency
   C. Interface mask inconsistency
   D. Process ID inconsistency
2. (True or false) If the level of an interface on an IS-IS router is different from the global router level, the level of the interface takes effect. ( )
   A. True
   B. False
3. (Multiple-answer question) Which of the following faults may occur in case of a Layer 2 loop? ( )
   A. An attempt to remotely log in to a device fails.
   B. An interface receives a large number of broadcast packets, which can be viewed in the display interface command output.
   C. An attempt to log in to a device through the serial port is time consuming.
   D. CPU usage exceeds 70%.

HUAWEI

1. ABC
2. B
3. ABCD

Summary

- The structured troubleshooting process involves fault report, fault confirmation, information collection, identification and analysis, cause listing, assessment, step-by-step troubleshooting, fault resolving, and wrap-up work.
- The purpose of troubleshooting is to restore the proper service running status. First, determine a service traffic path before troubleshooting. The layered, comparison, block-based, segment-based, and replacement approaches are used.
- On a LAN, the commonly used methods are to replace hardware to rectify link or device faults and to use STP to rectify LAN loops.
- Network-layer faults are mainly caused by unavailable routes. This course describes the causes and troubleshooting procedures for the failures to establish OSPF and IS-IS neighbor relationships and BGP peer relationships.
- Troubleshooting personnel must have abundant knowledge and be skilled in using multiple troubleshooting approaches. Troubleshooting experience summary also matters.

HUAWEI

Thank you.

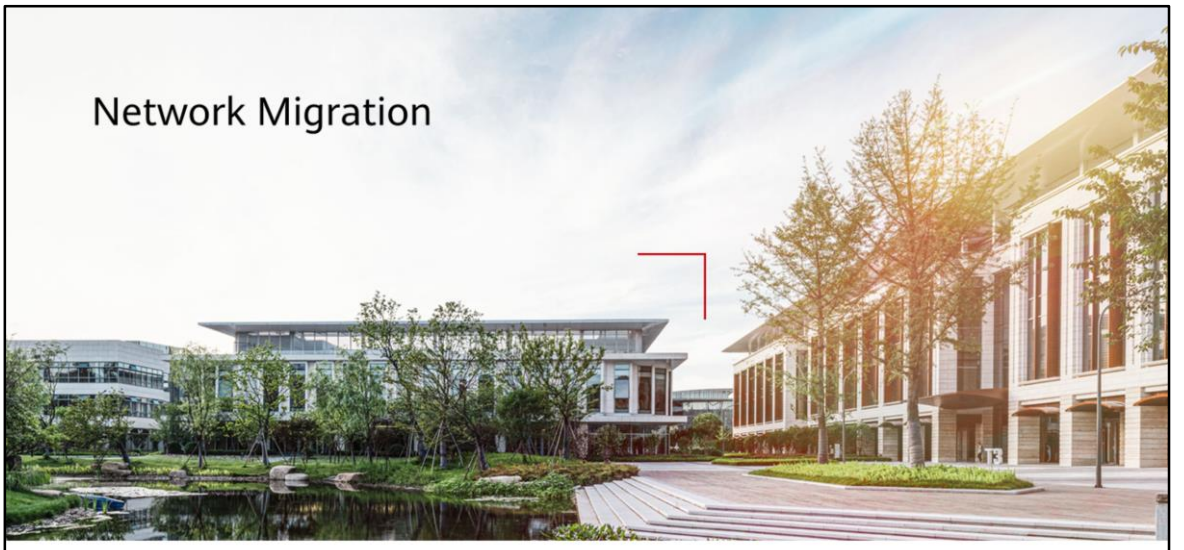
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Network Migration



HUAWEI

Foreword

- With the development of enterprise services, enterprise networks need to be reconstructed and optimized to meet service requirements. For example, IPv4 networks need to be upgraded to IPv6 networks, Intermediate System-to-Intermediate System (IS-IS) levels need to be modified, and the virtual private network (VPN) architecture needs to be modified. Regardless of hardware capacity expansion, software upgrade, and configuration change, enterprises formulate strict operation processes and risk mitigation measures based on service security level requirements for operations that affect services on the live network (for example, service interruption), and define these operations as migration projects.
- This course introduces the migration process, operation specifications, and risk control measures, and describes how to efficiently and smoothly complete network migration.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Clarify the operation procedure and specifications of the migration.
  - Describe common migration scenarios.

HUAWEI

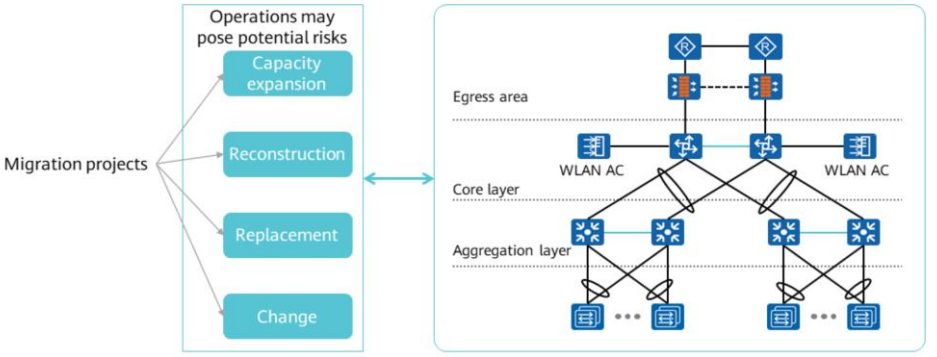
Contents

1. **Basic Concepts of Migration**
2. Migration Process
3. Migration Cases

HUAWEI

Basic Concepts

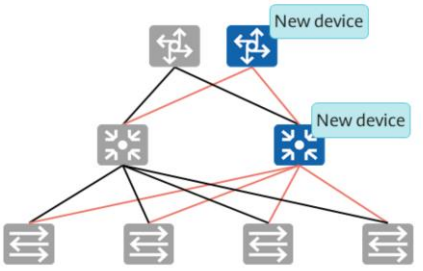
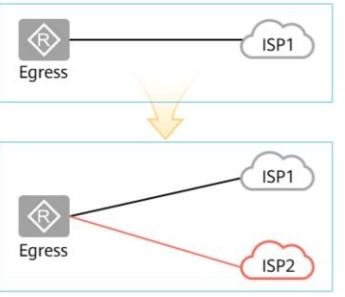
- Migration project: If the technical migration performed on the network affects the services running on the live network, strictly comply with the preset operation process and risk control measures during the implementation of the technical migration project. Generally, this type of project is defined as a migration project.



HUAWEI

Common Migration Scenarios (1)

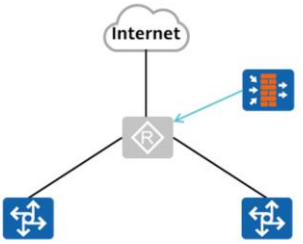
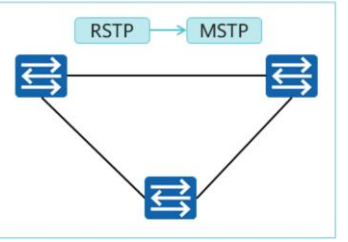
- As service traffic increases, network devices and links need to be added to expand the network capacity.

Network capacity expansion	Network reconstruction
	 The network architecture, including the physical architecture and logical architecture, needs to be adjusted.

HUAWEI

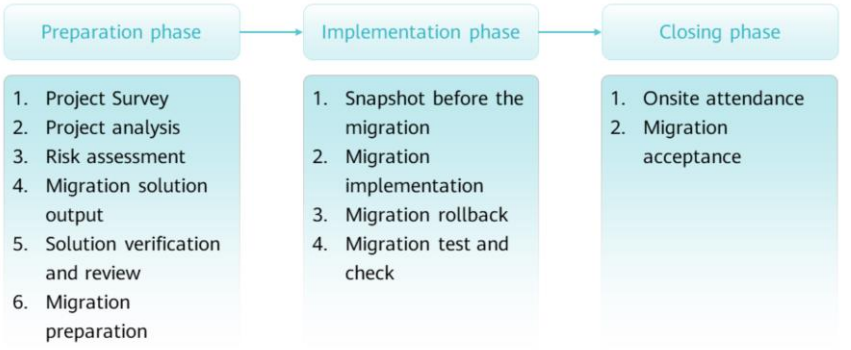
Common Migration Scenarios (2)

- Replace old devices with new devices, devices from other vendors, or devices of other types.

Device replacement	Configuration change
	 If you change the device configuration without changing the physical topology, the running services may be affected.

HUAWEI

Migration Process



HUAWEI

Contents

1. Basic Concepts of Migration
2. **Migration Process**
3. Migration Cases

HUAWEI

Project Investigation

- Project survey
- Project analysis
- Risk assessment
- Migration solution output
- Solution verification and review

- Before the migration, communicate with the customer network information owner, frontline maintenance engineers, ISP technical contact person, and device vendor representatives, and collect customer network information.

Static information collection and analysis	Dynamic information analysis
Service model analysis	Hardware environment survey on the live network

HUAWEI

Information Collection and Analysis

- Project survey
- Project analysis
- Risk assessment
- Migration solution output
- Solution verification and review

- Static and dynamic information on the live network is used to analyze the network status and compare the network status before and after the migration to determine whether the service volume is normal before and after the migration.

Static information collection and analysis

Detailed topology information	Device type	License
Device configuration	Device version	Interface type

Dynamic information collection and analysis

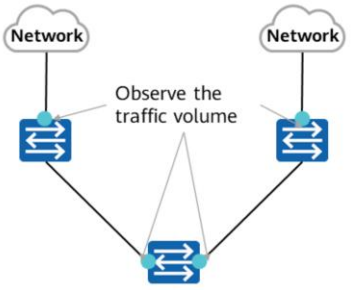
Network traffic	Bandwidth information	Protocol status
Protocol entry	Latency and jitter	Packet loss rate

HUAWEI

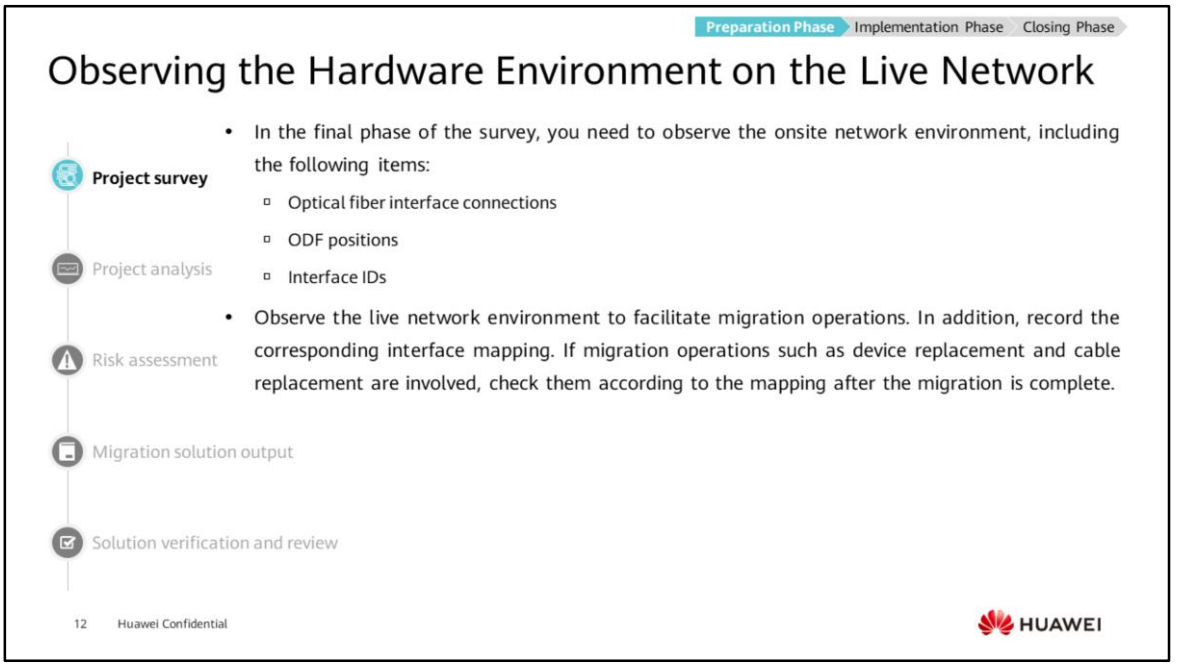
Traffic Model Analysis

- Project survey
- Project analysis
- Risk assessment
- Migration solution output
- Solution verification and review

- In the project survey phase, observe the customer's service traffic direction and volume, including the traffic direction change and link traffic volume, which can be used for comparison before and after the migration.



HUAWEI



- The optical distribution frame (ODF) is mainly used on backbone networks, metropolitan area networks (MANs), and optical fiber and cable networks. It connects, terminates, distributes, splits, and schedules backbone optical cables.

Project Analysis

- Project survey
- **Project analysis**
- Risk assessment
- Migration solution output
- Solution verification and review

- After the project survey is complete, analyze the customer's new requirements on the network after the migration, such as the bandwidth, network KPIs, and new service bearer capability.
- In addition, clarify the migration change requirements in the migration solution in this phase.
- Output the Customer Requirement Analysis Form.

HUAWEI

Risk Assessment

- Project survey
- Project analysis
- **Risk assessment**
- Migration solution output
- Solution verification and review

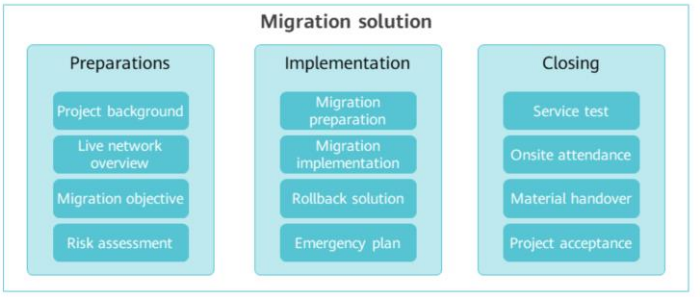
- Analyze and evaluate migration risks based on the survey result, requirement analysis result, and cutover solution framework, formulate countermeasures for projects with potential risks in advance, and confirm the owners of the countermeasures for corresponding risk items.
- The technical personnel involved in the risk assessment need to participate in the discussion, and the specific technical personnel needs to be specified for each risk.

HUAWEI

Migration Solution Output

Project survey
Project analysis
Risk assessment
Migration solution output
Solution verification and review

- Formulate the migration solution based on the survey result, project analysis result, and risk evaluation of technical personnel.
- In the migration solution, the detailed steps and procedures of the preparation, implementation, and closing phases must be specified.



HUAWEI

Implementation Solution

- Project survey
- Project analysis
- Risk assessment
- **Migration solution output**
- Solution verification and review

- To implement efficient and standard migration, plan each step in advance, including the specific action, specific command lines, and time required for performing each step.
- In the migration implementation solution, prepare the corresponding migration operation confirmation record table. The table records the operation time, confirmation result after operations are performed, and exception information.

Migration solution

Migration Procedure

Step	Operation	Time
Power off old devices.	Power off old devices and remove cables.	0:00-0:30
Deploy devices on the rack.	Power on new devices and connect cables.	0:30-1:00
Configure new devices.	Import the pre-configured script to the new device.	1:00-1:20
Check the network status.	Check network entries.	1:20-2:00
Test services.	Coordinate Party A's personnel to test services and applications.	2:00-2:30

HUAWEI

Rollback Solution

- Project survey
- Project analysis
- Risk assessment
- **Migration solution output**
- Solution verification and review

- Rollback solution: If the migration fails (for example, service test fails after the migration or the migration is not complete after the migration time), roll back the network to the status before the migration. In this case, perform operations according to the rollback solution to restore the network.
- A detailed rollback solution must be included in the migration solution. Similar to migration operations, each step in the rollback solution must be specified. In this way, rollback can be performed in an orderly manner when the migration fails.

Rollback Procedure

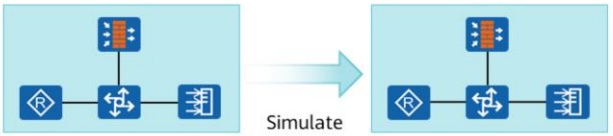
Step	Operation	Time
Remove new devices.	Power off new devices and remove cables.	3:00-3:20
Place the old devices on the rack.	Power on the old devices and connect cables.	3:20-3:40
Test services.	Coordinate Party A's personnel to test services and applications.	3:50-4:00

HUAWEI

Solution Verification and Review

- Project survey
- Project analysis
- Risk assessment
- Migration solution output
- **Solution verification and review**

- For large-scale and important migration projects, an environment must be set up in the lab for verification in advance. The previous risk points must be tested and the feasibility of the entire solution must be confirmed. This type of verification in the lab environment is called the pilot office test.
- In addition to FOA solution verification, a technical review meeting needs to be held with the customer and implementation party to verify the technologies in the migration solution so as to understand actual requirements and difficulties of both parties, and solve problems face to face.



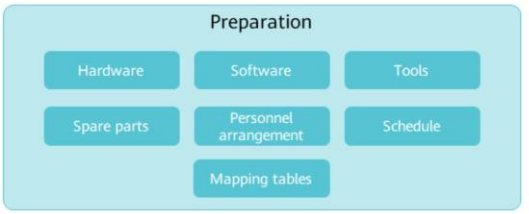
Customer network　　Pilot office test

HUAWEI

Migration Preparation

- Project analysis
- Risk assessment
- Migration solution output
- Solution verification and review
- **Migration preparation**

- Migration preparation is a key procedure before migration and is the basis for successful migration.
- The following items should be included: environment preparation (hardware, software, tools, and spare parts), personnel preparation (party A, party B, and supervisor), and procedure preparation (execution time arrangement). All aspects should be considered to ensure migration success.



HUAWEI

Software and Hardware Preparation

- Project analysis
- Risk assessment
- Migration solution output
- Solution verification and review
- **Migration preparation**

Hardware

Device	Power-on test
Board	Check whether the tested board is normal.
Cable	Connectivity test

Software

Versions	Check the device version and interconnection version mapping.
License	Check whether the authorized license is available.
Scripts	Overall script, segment configuration script, and rollback script

HUAWEI

Tool and Spare Part Preparation

- Project analysis
- Risk assessment
- Migration solution output
- Solution verification and review
- **Migration preparation**

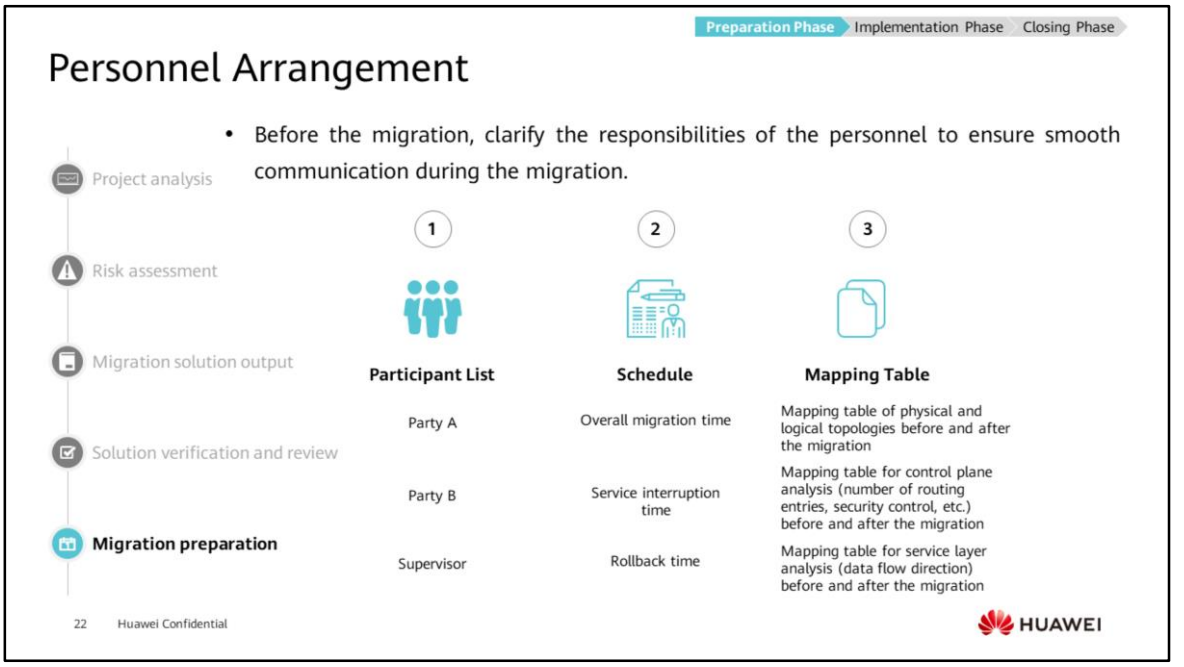
Tools

Transportation tools	Vehicles, landing tools, and installation tools
Test tools	Network cable tester, network tester, and optical power meter
Debugging tools	PC terminal and debugging software CRT
Monitoring tools	Service monitoring tool, traffic monitoring platform, and network analysis tool

Spare parts

- Important boards and device spare parts
- Backup of software of old and new versions
- Backup of old and new scripts

HUAWEI

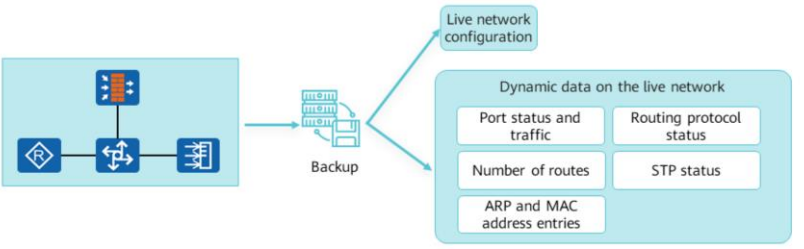


- Time arrangement preparation:
  - Negotiate the time arrangement with the customer and obtain customer's approval.
  - Make an overall time schedule.
  - Specify actions to be performed in each time segment.
  - In the migration phase, time arrangement should be accurate to minutes.
  - Reserve some time for major operations to avoid engineering accidents due to timeout.
  - Do not perform migration in peak hours (such as holidays and off-duty time).

Snapshot Before the Migration

- **Snapshot before the migration**
- Migration implementation
- Migration rollback
- Migration test and check

- To facilitate migration rollback and service comparison after the migration, you need to take snapshots of the live network configuration and data before the migration.
  - Back up live network configuration.
  - Collect dynamic data on the live network, including the port status, traffic, routing protocol status, number of routes, STP status, and ARP/MAC address entries of each interface.
  - Test services before the migration to ensure that the services involved in the migration are normal.

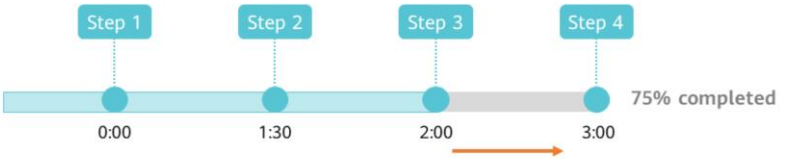


HUAWEI

Migration Implementation

- Snapshot before the migration
- **Migration implementation**
- Migration rollback
- Migration test

- The onsite migration personnel must strictly follow the prepared migration procedure. The implementation procedure cannot be changed temporarily unless there are special reasons.
- Record the time, actions, and results of each step.

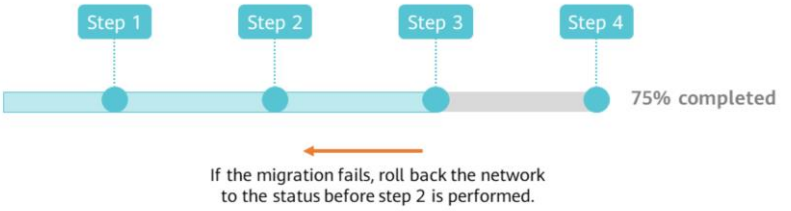


HUAWEI

Migration Rollback

- Snapshot before the migration
- Migration implementation
- **Migration rollback**
- Migration test and check

- If the migration is not complete before the specified time, perform the rollback step by step according to the prepared rollback scheme to restore the network.
- You can negotiate with the customer to determine whether to perform partial or full rollback based on the actual situation.

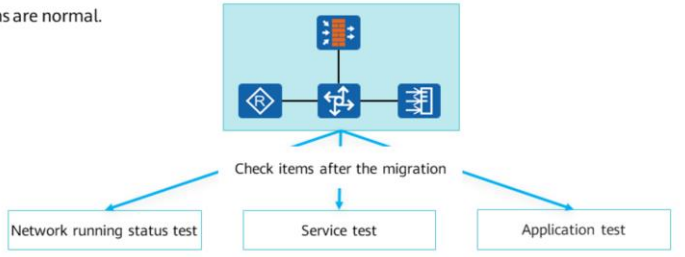


HUAWEI

Migration test

- Snapshot before the migration
- Migration implementation
- Migration rollback
- **Migration test and check**

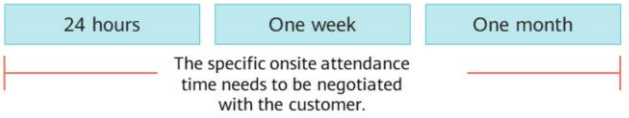
- After the migration is complete, you need to test the network and services from the following aspects to check whether the migration is successful:
  - Network running status test: Collect dynamic data on the live network and compare the data with that before the migration.
  - Service test: Perform the ping or tracert operation or use third-party software on the client side to check whether the network connectivity, delay, and jitter meet service requirements.
  - Customer application test: Test customer applications carried by the network and check whether applications are normal.



HUAWEI

Onsite Attendance

- After the migration is complete and the customer's application service test is passed, the network enters a special observation period. During this period, engineers usually stay at the customer's site and observe the network running status to prevent unexpected faults.



HUAWEI

Migration Acceptance

- After the migration is complete and no exception occurs, we need to provide maintenance training for the customer and hand over documents to the customer.

Transfer-to-maintenance training

If the migration project involves new devices and new configurations, organize training for the customer after the migration is complete.

Material handover

To facilitate subsequent maintenance, transfer the documents involved in the migration to the customer after the migration is complete.

HUAWEI

Contents

1. Basic Concepts of Migration
2. Migration Process
3. **Migration Cases**

HUAWEI

Migration Cases: Precautions

- The key point is to master the migration process and method based on actual migration cases, in addition to device configuration planning required by the migration.
- Discuss migration scenarios based on the migration process and output the migration document based on the Project Migration Solution Template. The document includes the following contents:
  - Project survey and analysis: Analyze the live network topology and services, and identify the devices whose dynamic and static information needs to be collected.
  - Risk evaluation: Evaluate the risks that may occur during the migration and how to avoid the risks.
  - Migration solution: Specify the commands to be executed in each step and the time point of each step.
  - Test and check items: Determine the contents to be checked and tested after the migration.

HUAWEI

Class Activities

Background 15 minutes	• The teacher introduces the scenario of the migration case, inputs the topology diagram and service planning, and conducts group discussion, make a summary, and output the migration cutover process, and cutover-related documents).
Grouping communication 5 minutes	• Each group discusses and selects a team leader. The team leader is responsible for the final summary and report.
Discussion 30 minutes	• Each group discusses the migration procedure, designs the contents to be configured, and outputs the migration document.
Summary 20 minutes	• Each group reports the discussion results and submits related documents.

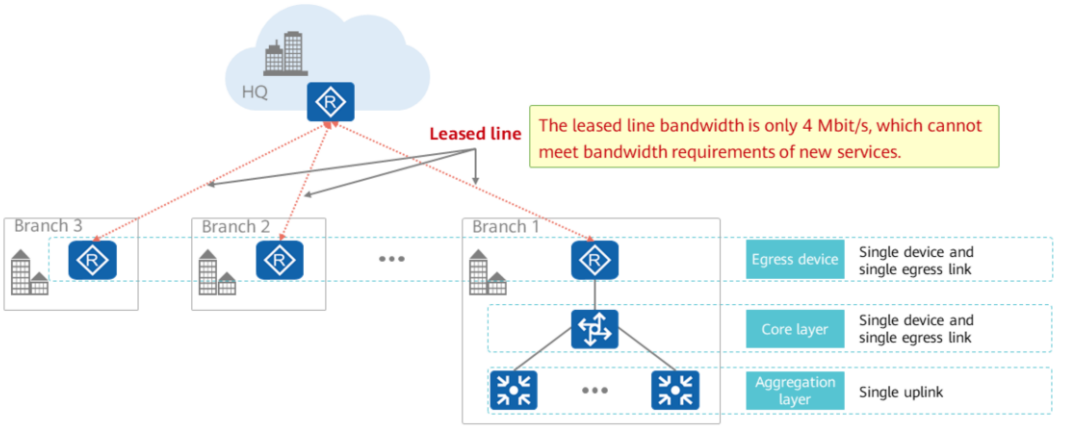
HUAWEI

Introduction to the Early Stage of the Reconstruction Project

- A company has multiple branches that are connected to the headquarters network through leased lines. The following problems exist:
  - The leased line rate is low and cannot meet bandwidth-intensive requirements of new services.
  - A single egress device and a single uplink (leased line) are deployed at the egress of a branch, which may cause a single point of failure (SPOF).
  - Increasing the leased line rate causes a high cost.
  - The standalone core switch is deployed, which may cause an SPOF.
  - The performance of core switches and egress routers cannot provide good support for increasing end users in branches.
- To solve the problems, the company decides to reconstruct the branch network to improve the reliability of the branch network and increase the bandwidth for communication with the headquarters.

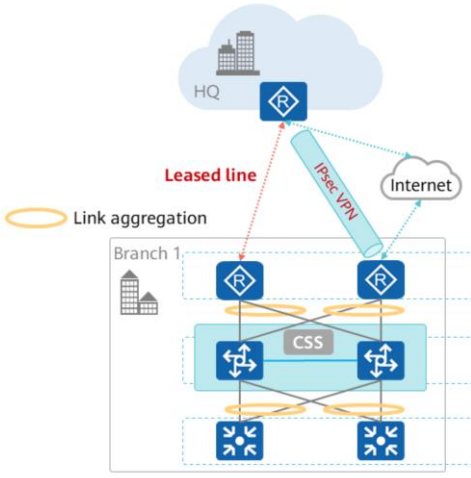
HUAWEI

Original Topology



HUAWEI

Target Topology

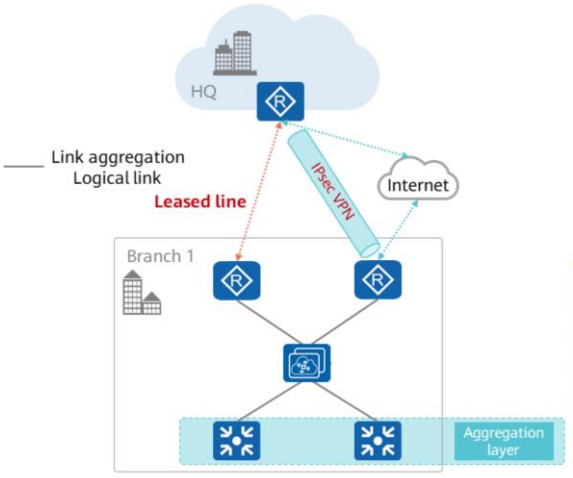


Device redundancy and link redundancy are used to improve reliability. Internet connections are added at the egress, and IPsec VPNs are established with the headquarters to carry services that demand low delay and high bandwidth.

Egress device	Add an Internet connection and establish an IPsec VPN to carry some services.
Core layer	The CSS composed of core switches establishes inter-chassis link aggregation with uplink and downlink devices.
Aggregation layer	Dual uplinks set up a link aggregation group with the CSS composed of core switches.

HUAWEI

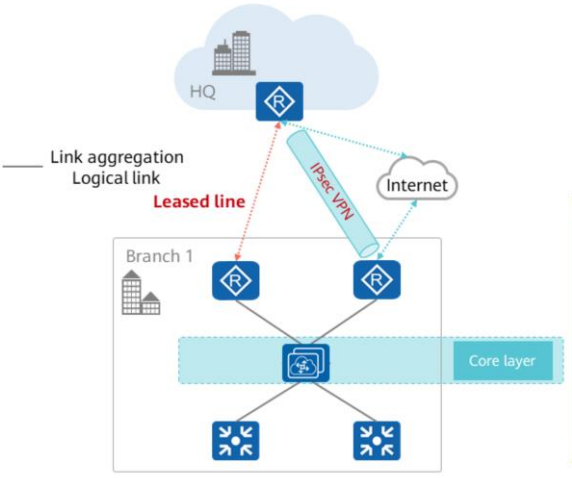
Target Logical Topology (1)



The aggregation layer uses dual uplinks to connect to the CSS composed of core switches.
The gateway is configured on the VLANIF interface of the core switch. The CSS composed of core switches and link aggregation technology are deployed to improve reliability, so VRRP does not need to be deployed.

HUAWEI

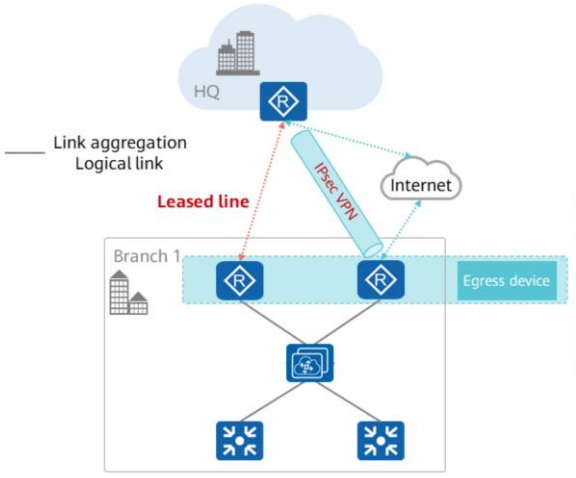
Target Logical Topology (2)



The core device serves as the gateway of intranet terminals. Static routes are configured to direct traffic of different services to the leased line and IPsec VPN egress devices, and floating routes are configured for route backup.
NQA is configured to detect the availability of the leased line and IPsec VPN. Static routes are associated with NQA so that traffic can be switched to the backup path when a link fault occurs.

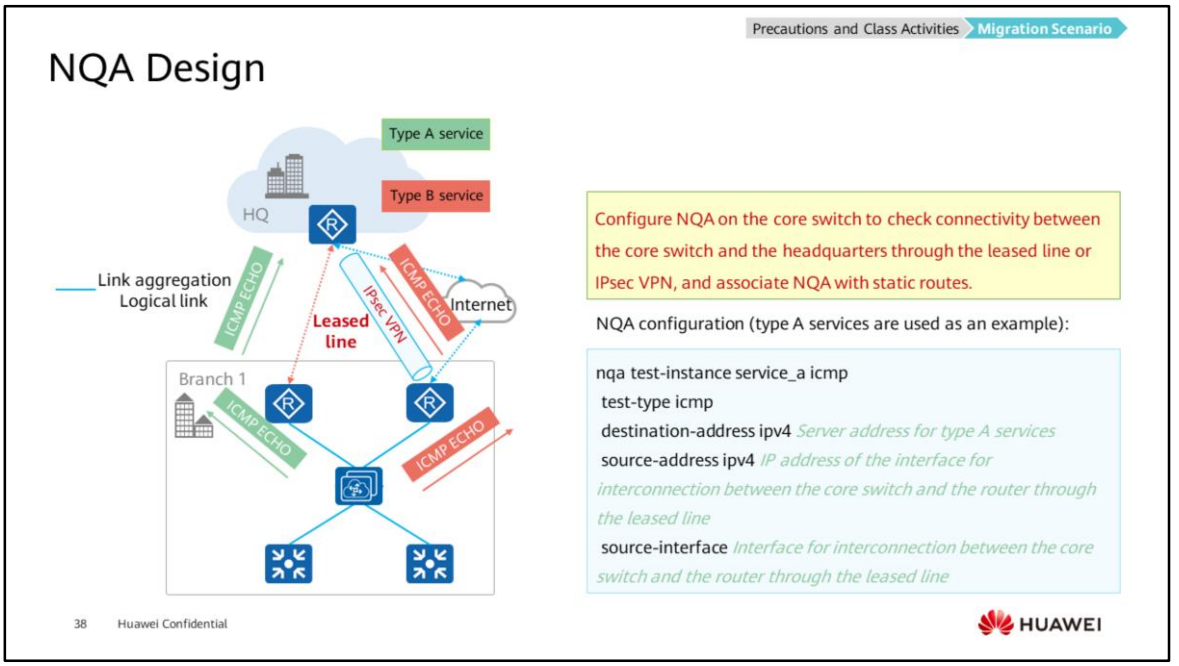
HUAWEI

Target Logical Topology (3)



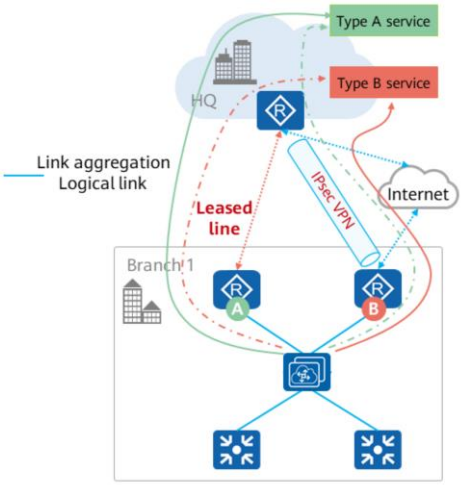
The two egress devices are connected to the leased line and Internet to carry different services. Traffic of intranet terminals' access to different services is forwarded based on routes of the core switch. If the routes change, the traffic forwarding path changes accordingly.

HUAWEI



- Type A service: service demanding low latency and bandwidth. These services are carried over leased lines.
- Type B service: service that has low requirements on the latency but occupies much bandwidth. These services are carried over IPsec VPNs.
- Static return routes are manually specified for the headquarters, and NQA is used to switch services to the standby path upon faults. This case focuses on the branch network and does not involve the headquarters network.

Route Backup Design



Active path of type A service
Active path of type B service
Standby path of type A service
Standby path of type B service

On the core switches, the routes to type A and type B services are destined for different egress devices and carried by different links. Floating routes are configured so that the two egress devices to back up each other.

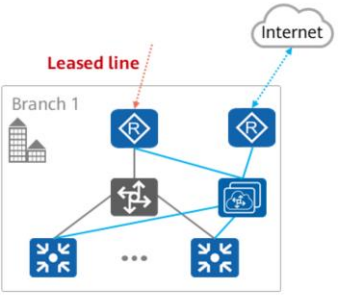
Association between static routes and NQA on core switches:

ip route-static *IP address of type A service mask A* track *nqa service_a icmp*
ip route-static *IP address of type A service mask B*

ip route-static *IP address of type B service mask B* track *nqa service_b icmp*
ip route-static *IP address of type B service mask A*

HUAWEI

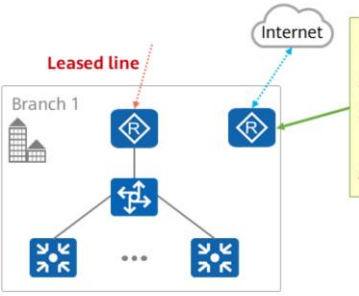
Migration Solution (1)



- Perform the following steps to migrate the original network to the target network (overview):
  - Deploy a new egress router and connect it to the Internet. Establish an IPsec VPN and configure routes.
  - Deploy two new core switches, establish the CSS and routes, and configure inter-chassis aggregation between the core switches and egress routers. In this case, the core switches are not connected to the aggregation switch.
  - Disconnect the original core switches from uplink and downlink devices, and connect the aggregation switch to the new core switches.

HUAWEI

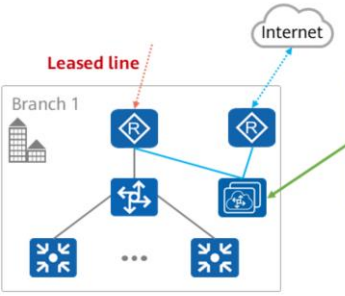
Migration Solution (2)



1. Connect the egress router to the Internet.
2. Deploy IPsec VPN with the headquarters.
3. Configure routes to network segments of type A and type B services of the headquarters.
4. Configure a return route to the intranet.

HUAWEI

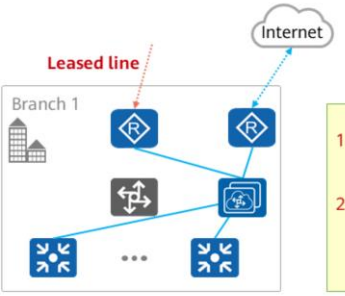
Migration Solution (3)



1. Deploy core switches and configure the CSS.
2. Configure routes to network segments of type A and type B services and complete the NQA configuration.

HUAWEI

Migration Solution (4)



1. Connect the original aggregation switch to the new core switch in the CSS and disconnect the original core switches.
2. Modify the configuration of the original egress router and change the next hop of the return route to the intranet to the interconnection address of the new core switch in the CSS.

HUAWEI

Migration Document Output

- Output the ***Project Migration Solution Template*** based on the preceding migration project information.

Project Migration Solution Template

HUAWEI

Quiz

1. (Short-answer question) What operations can be performed locally to verify risks before the migration?
2. (Short-answer question) Which of the following items must be backed up for the customer network before the migration?

HUAWEI

1. We can set up a local pilot office and simulate the customer's network to verify the feasibility of the entire migration solution.
2. The configuration of the live network needs to be backed up. To verify the network status before and after the migration, collect dynamic data of the live network, including the port status, traffic, status of each routing protocol, number of routes, STP status, and ARP/MAC address entries of each port.

Summary

- The migration changes the network and may affect the customer's services. Therefore, the migration must be based on the customer's requirements to control risks in real time.
- Before the migration, perform survey and verification, and prepare each step in advance. After the migration, verify services. The migration can be successful and efficient only when the migration process is strictly followed.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Enterprise Network Introduction

HUAWEI

Foreword

- Nowadays, data has become a new production factor and important asset of enterprises. Enterprise networks, as the infrastructure, promote the efficient transmission of various data and accelerate the digital transformation of thousands of industries.
- Enterprise networks are classified into campus networks, wide area networks (WANs), and data center networks (DCNs). The three types of networks are used to meet different service requirements and have different network architectures.
- This course introduces the general enterprise network architecture, enterprise networks in typical industries, and Huawei enterprise network solutions for specific scenarios.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Understand the overall architecture of enterprise networks.
  - Understand the typical enterprise network architectures of different industries.
  - Understand the relationship between Huawei datacom certification and enterprise networks.
  - Understand Huawei enterprise network solutions.

HUAWEI

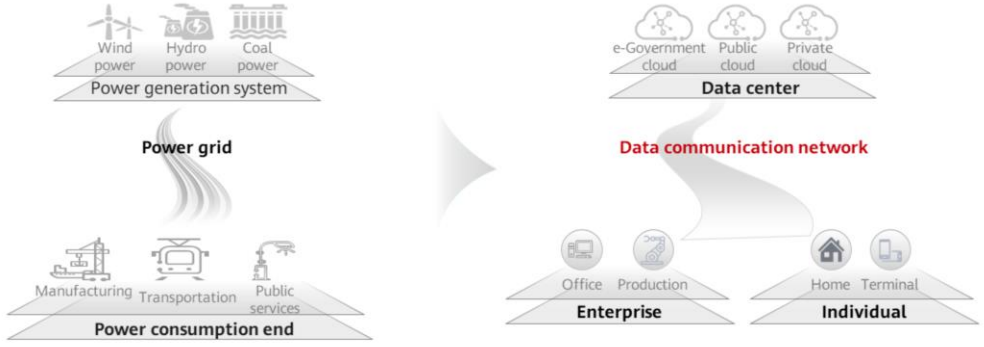
Contents

1. **Enterprise Network Overview**
2. Trends and Challenges of Enterprise Networks
3. Huawei Enterprise Network Solutions

HUAWEI

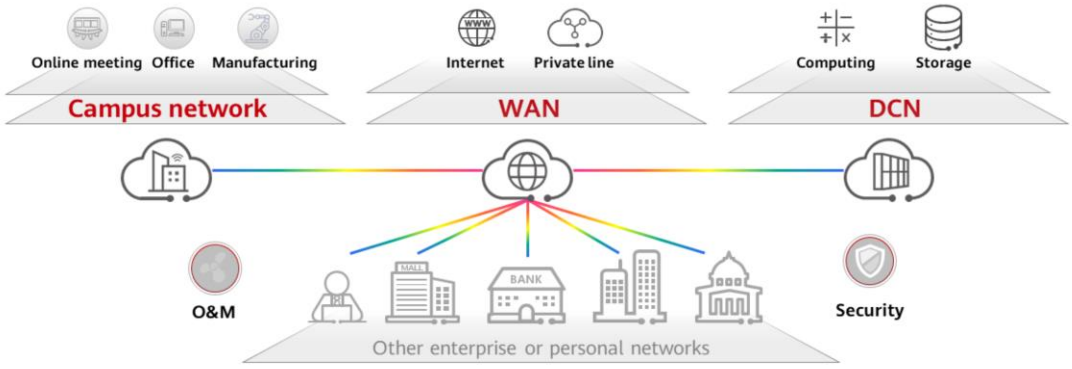
“Power Grid” in the Digital World : Data Communication Network

- The symbol of industrialization is "electricity", transmitting electricity through the power grid.
- The symbol of digitalization is "cloudification", transmitting computing power through the cloud network.



HUAWEI

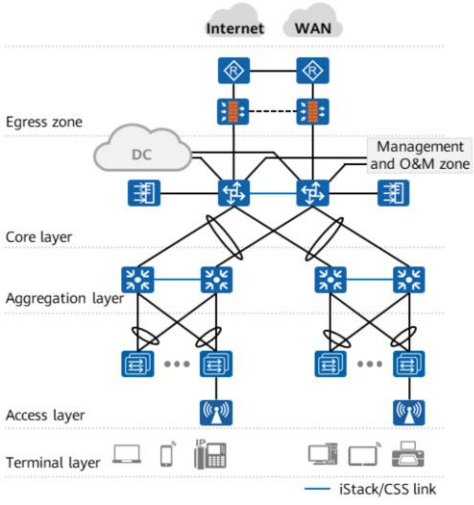
Enterprise Networks in the Digital Society



After more than 30 years of development, networks have laid a solid foundation for the intelligent society. With the acceleration of enterprise digitalization, enterprise networks need to carry more and more key services, playing an important role in the digitalization process.

HUAWEI

Enterprise Campus Network Overview

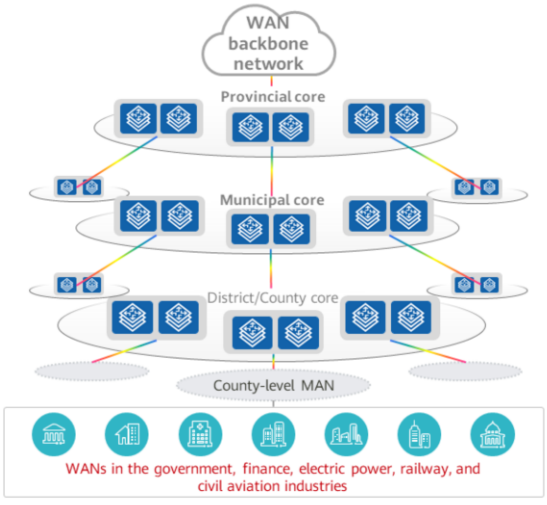


A campus network is used to implement terminal access in a small range and interworking within an enterprise. Campus networks can be classified into small, midsize, and large ones based on the number of terminal users or NEs. A typical large campus network consists of the following modules:

- **Egress zone**: the border between the campus internal network and external network. It is used for internal users to access the public network and for external users to remotely access the internal network.
- **Core layer**: the core of campus data switching that connects all components of the campus network, such as the DC, management and O&M zone, and campus egress.
- **Aggregation layer**: forwards horizontal traffic between users and vertical traffic to the core layer. The aggregation layer can function as the switching core of a department or area and can connect more terminals to the network.
- **Access layer**: provides various access modes for users and is the first network layer to which terminals connect.
- **Terminal layer**: involves various terminals that access the campus network, such as PCs, printers, IP phones, mobile phones, and cameras.
- **DC**: deploys servers and application systems to provide data and application services for internal and external users of an enterprise.
- **Management and O&M zone**: deploys the network management system (such as the NMS and authentication server) and O&M system.

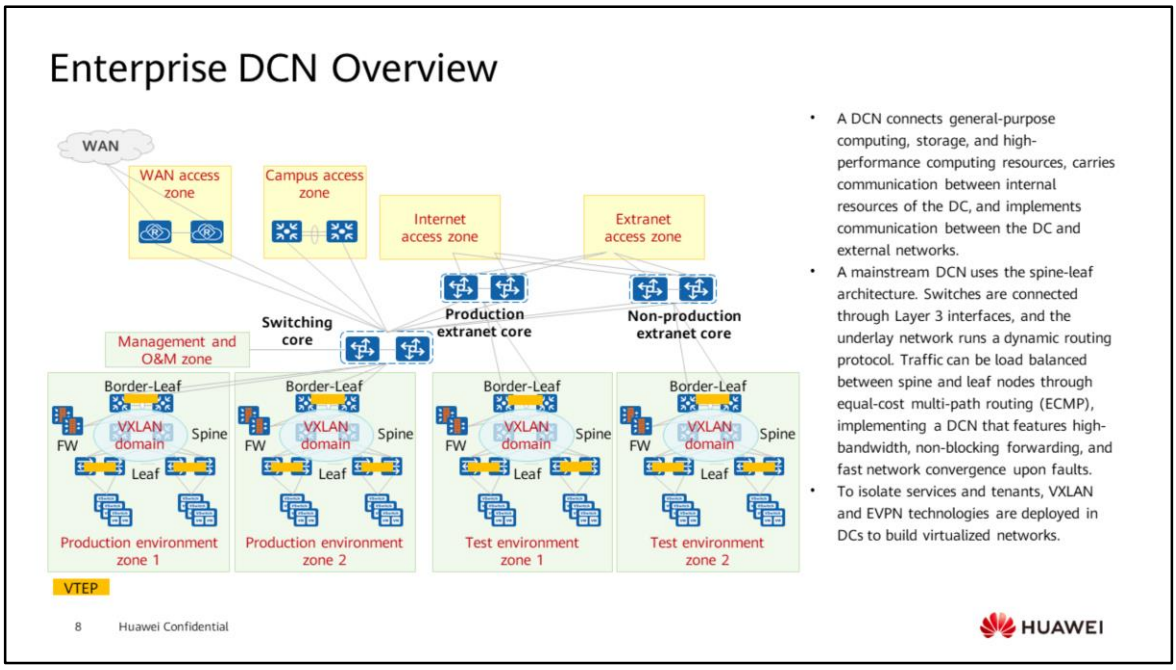
HUAWEI

Enterprise WAN Overview

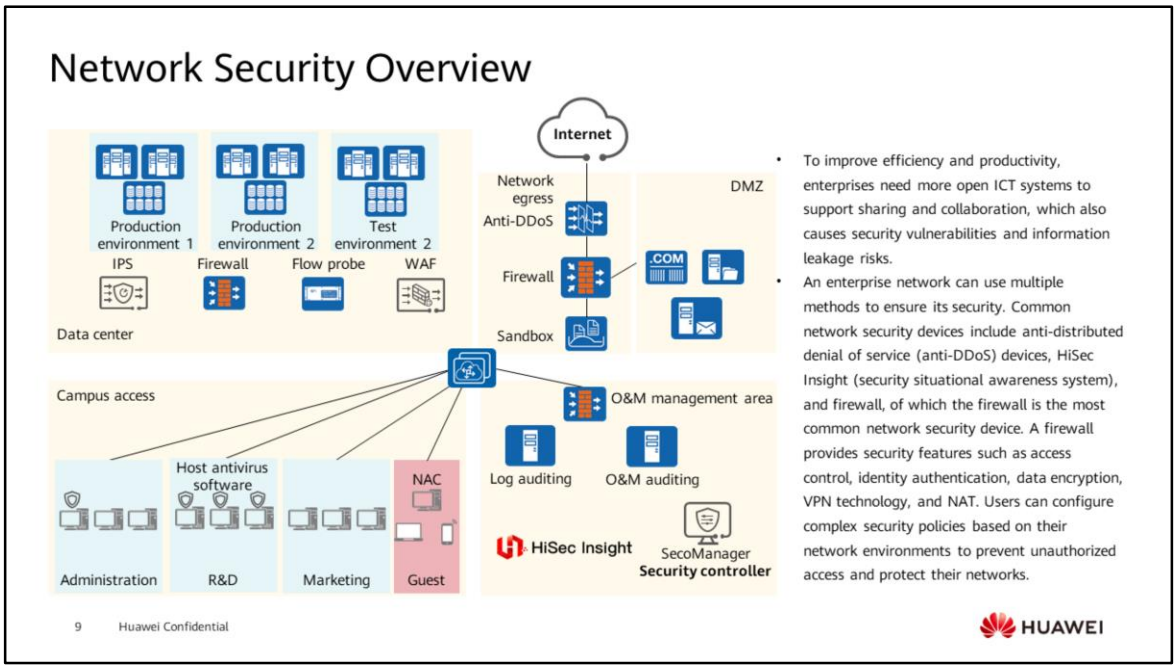


- An enterprise WAN connects the enterprise headquarters, branches, and data center for interconnection.
- By users, enterprise WANs can be classified into self-built for internal use and self-built for external use.
  - **Self-built, for internal use**: To meet the interconnection requirements of the headquarters, branches, and data centers in a large span, various industries such as large enterprise, government, finance, and electric power need to build their own WANs.
  - **Self-built, for external use**: Carriers build WANs to provide WAN connection services for enterprises, including the Internet, MPLS private lines, and transmission private lines.
- Generally, a WAN uses a hierarchical structure. For example, a WAN can be divided into the county-level MAN, municipal MAN, provincial core network, and backbone network. MANs at different levels provide access for networks in the current area and connect to the upper-level MAN or backbone network through the core network.

HUAWEI

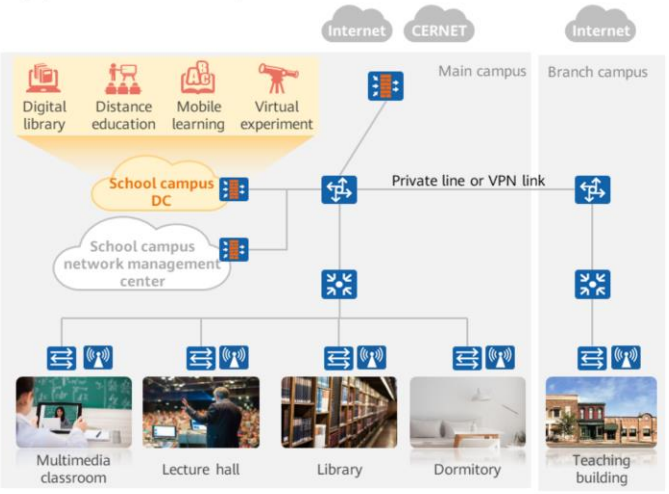


- DCNs do not have a fixed zone division mode. Different industries and enterprises have different area division modes.
- The zone division in the slide uses the financial industry as an example:
  - WAN access zone: connects to the WAN built by the enterprise.
  - Campus access zone: accesses traffic from enterprise campus users.
  - Extranet access zone: connects to external networks, such as networks of other enterprises and partners.
  - Internet access zone: accesses user traffic from the Internet.
  - Production environment zone: runs services.
  - Test environment zone: provides services for external systems informally. Before a service is brought online to the production environment zone, the service is developed, tested, and verified in the test environment zone.



- Different network security solutions can be used to cope with different network security problems.
  - Deploying firewalls in the server area of a data center: implements security isolation, access control, and intrusion prevention between servers and VMs in the data center.
  - Deploying Web Application Firewalls (WAFs) in the server area of a data center: protects website servers and prevents website text and images from being tampered with.
  - Deploying anti-DDoS abnormal traffic detecting and cleaning devices: mitigates various DDoS attacks from the Internet.
  - Deploying egress firewalls: performs security protection, such as NAT, application protocol identification and control, security isolation, and access control, to prevent unauthorized access.
  - Implementing O&M auditing in the management area: centrally manages and controls accounts, authentication, authorization, and audit of various IT resources such as core service systems, hosts, databases, and network devices to comply with related laws and standards and implement unified access management and O&M audit of core resources.

Typical Enterprise Network: Education Industry

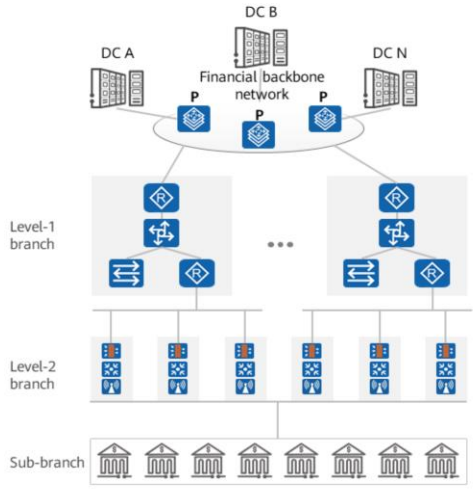


- Take the campus network of a university in China as an example.
- The campus network is a computer network that provides teaching, scientific research, and comprehensive information services for teachers and students (including their family members and visitors).
- The higher education campus network refers to the campus network of higher education institutions.
- The higher education campus network is generally divided into dormitories, living areas, teaching areas, and public areas. It provides wired and wireless network access services to help the campus enter the digital era and improve talent cultivation and innovation capabilities.

HUAWEI

- China Education and Research Network (CERNET): a nationwide education and research computer network that is funded by the Chinese government and directly managed by the Chinese Ministry of Education. It is constructed and operated by Tsinghua University and the other leading Chinese universities.

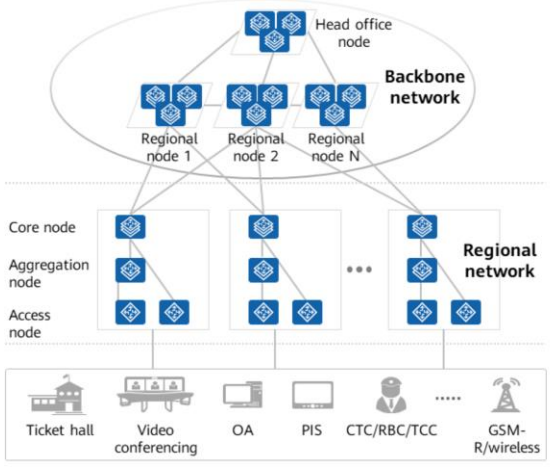
Typical Enterprise Network: Financial Industry



- To meet their own business development and operation management requirements, financial enterprises generally build their own backbone networks across the country to support the development of enterprise informatization and core services.
- The DC and its branches are connected through the backbone network. Network services are mainly centralized in the DC. Users are widely distributed inside and outside the banks. Branches access the DC of the head office through the backbone network.

HUAWEI

Typical Enterprise Network: Transportation Industry

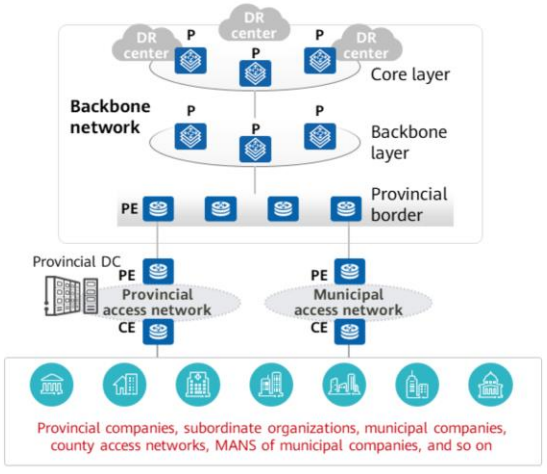


- The railway system in China is used as an example.
- In the railway system, the network provides unified transmission services for signaling, dispatching, ticketing, integrated services, and GSM-R services.
- The railway data communications network uses a two-layer architecture:
  - A backbone network consisting of multiple high-performance core routers located at regional nodes and railway bureaus.
  - A regional network set by each railway bureau. The regional network is connected to the backbone network through the backbone network's regional node of the local railway bureau to implement communication between railway bureaus.
- As an independent autonomous system, the backbone network forwards traffic between the head office and railway bureaus and between railway bureaus.
- Each railway bureau is configured with a backbone network access node and multiple backbone network forwarding nodes. The head office node functions as the service aggregation node of the entire railway system and does not forward traffic.

HUAWEI

- Generally, the regional nodes of the backbone network are deployed in the equipment rooms of each railway bureau.
- GSM-R: Global System for Mobile Communications - Railway
- CTC: Centralized Traffic Control
- RBC: Radio Block Center
- PIS: Passenger Information System
- TCC: Track Communications Center

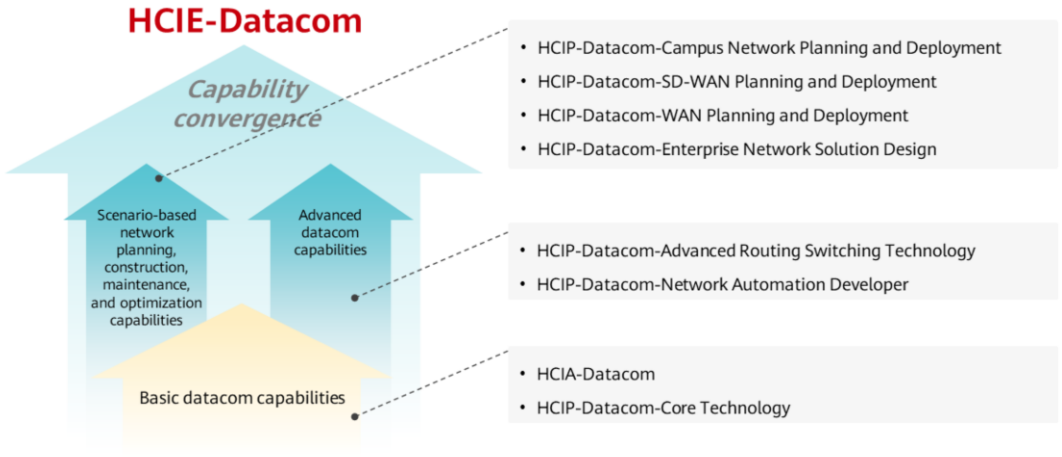
Typical Enterprise Network: Electric Power Industry



- The electric power industry in China is as an example.
- The electric power industry network is divided into two independent networks based on the carried services: dispatching network and integrated data network.
- The dispatching network, as one of the core networks for electric power informatization, mainly carries power-related service control systems.
- The integrated data network carries applications such as OA, production management, lightning weather information, and remote monitoring of unmanned substations. It is physically independent and parallel with the power dispatching network.
- The integrated data network adopts a multi-level architecture, including the national backbone network, provincial integrated data network, and municipal integrated data network. The national backbone network is divided into the core layer (consisting of multiple core nodes deployed in the country) and the backbone layer (consisting of backbone nodes distributed in provincial companies in the country).
- The data networks of provincial companies and branches are connected to the provincial border nodes of the backbone network to implement communication between provincial networks.

HUAWEI

Huawei Datacom Certification and Enterprise Network



HUAWEI

Contents

1. Enterprise Network Overview
2. **Trends and Challenges of Enterprise Networks**
3. Huawei Enterprise Network Solutions

HUAWEI

Industry Digital Transformation, Improving Production Efficiency and Customer Satisfaction

Digital OA

All-wireless access, OA anytime and anywhere, one-click conference scheduling using an app, and automatic adjustment of lighting and temperature

Source: According to *Worldwide Digital Transformation 2019 Predictions* released by IDC, "at least 55% of organizations will be digitally determined by 2020."

Digital education

Change from stuffing the duck to immersion, high-quality teaching resources on demand, and real-world teaching

Source: Gartner's *2019 CIO Agenda: Higher Education Industry Insights* reveals that 86% of higher education CIOs would regard IT as a key factor in enabling education business transformation.

Digital manufacturing

Always-online mass terminals/sensors, real-time production data collection, and automatic and precise control

Source: IDC's 2018 manufacturing industry survey results shows that 84.9% manufacturing enterprises are undergoing digital transformation, driving business model innovations and reshaping the business ecosystem.

HUAWEI

In the Digital Era, How Campus Networks Support Digital Transformation Across Industries?

Ubiquitous connectivity

- Access anytime, anywhere
- High-quality service transport

On-demand services

- Quick service deployment and adjustment
- Rapid rollout of value-added applications

Perceptible experience

- Accurate measurement and evaluation of user experience
- AI-powered automatic adjustment of resource allocation

Easy O&M

- Service-based configuration delivery
- Quick and accurate fault locating

HUAWEI

Service Requirements and Challenges of Large- and Medium-Sized Campus Networks

Converged transport

Requirements:

Access terminals and services are diversified, and a converged transport network is required.

Challenges:

- Wi-Fi and IoT services are independently planned, deployed, and managed. The overall network construction cost is high.
- The workload of network management and O&M is heavy.

User experience awareness

Requirements:

Network O&M needs to be automated and intelligent to perceive user experience anytime and anywhere.

Challenges:

- Service faults cannot be detected in a timely manner.
- After a fault occurs, the fault cause is determined based on the O&M experience of professional personnel, and the fault cannot be quickly located.
- The network cannot be automatically optimized.

Network automation

Requirements:

Networks need to be automated to cope with the complexity of deployment and policies caused by the surge of applications and services.

Challenges:

- The workload is repetitive and heavy, and manual configuration is complex.
- New services need to be configured on devices one by one, which is time-consuming and costly.
- The workload of network policy deployment and adjustment is heavy.

HUAWEI

Service Requirements and Challenges of Small- and Medium-Sized Campus Networks

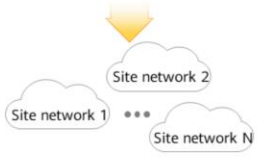
Plug-and-play network devices, improving deployment efficiency

Unified management and centralized configuration

- The network configurations of multiple sites are centrally delivered, reducing onsite configuration and commissioning workload and improving deployment efficiency.
- The network is plug-and-play and able to be expanded on demand, requiring low costs for upgrades.

Centralized O&M on the cloud, simplifying O&M of multiple sites

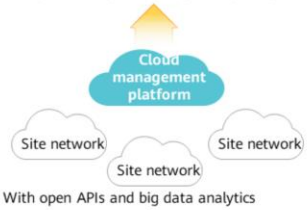
Centralized management of multiple branches on the cloud and remote automatic O&M



Scattered campus branch networks are centrally managed on the cloud through the Internet, and automation tools for troubleshooting, monitoring, and other management operations are integrated, implementing remote automated O&M.

Open APIs, accelerating integration of business applications

Open and big data analytics capability



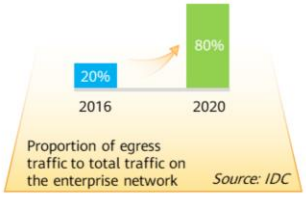
With open APIs and big data analytics capabilities, multiple management systems can be combined and interconnected to achieve unified network management. More value-added applications are available, leading enterprises into digital transformation.

HUAWEI

Development of Enterprise SD-WAN Technologies

Migration of enterprise services to the cloud

Enterprise services are migrated to the cloud, and enterprise egress traffic increases sharply.



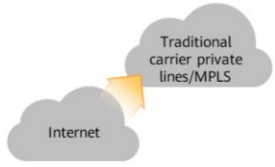
Emergence of Software defined networking (SDN)

Unified management and centralized configuration

SDN introduces a new role, that is, the centralized controller, so that it can not only understand the network requirements proposed by the administrator, but also fully manage and configure the physical network, implementing automatic deployment and fast provisioning of applications on the network.

High-speed development of the Internet

Rapidly narrowing gap between the quality of the Internet and traditional private lines



The coverage and network performance of the 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.

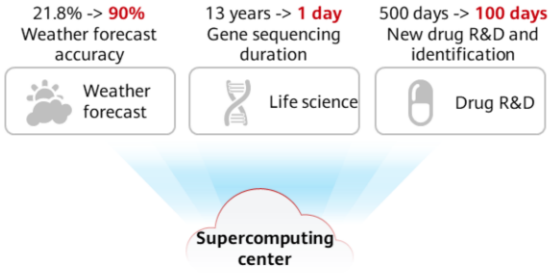
HUAWEI

Service Requirements and Challenges of Multi-Campus Network Interconnection

Branch interconnection cost	Service provisioning period	Service experience	Management and O&M	Unified management
Requirements: The cost of WAN interconnection between branches needs to be reduced. **Challenges:** Carriers' transmission private lines or MPLS VPN private lines are of high quality but expensive.	**Requirements:** The network deployment efficiency needs to be improved and the service provisioning period needs to be shortened. **Challenges:** • It takes a long time to provision traditional private lines. • Service provisioning is mainly performed by network engineers on site, which has high skill requirements and low efficiency.	**Requirements:** The application identification capability needs to be enhanced to ensure key service experience. **Challenges:** • Different enterprise applications have different requirements on link quality. • Traditional private lines cannot detect the application status and cannot guarantee key services at any time.	**Requirements:** Visualized O&M is required to simplify management and O&M. **Challenges:** • CLI-based O&M is inefficient and lacks visualized O&M methods. • If an enterprise has a large number of branches, onsite O&M will increase costs.	**Requirements:** LAN-side and WAN-side services need to be managed in a unified manner. **Challenges:** • Unified configuration management and O&M cannot be implemented.

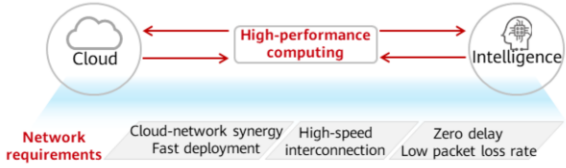
HUAWEI

Cloud + Intelligence, Accelerating HPC

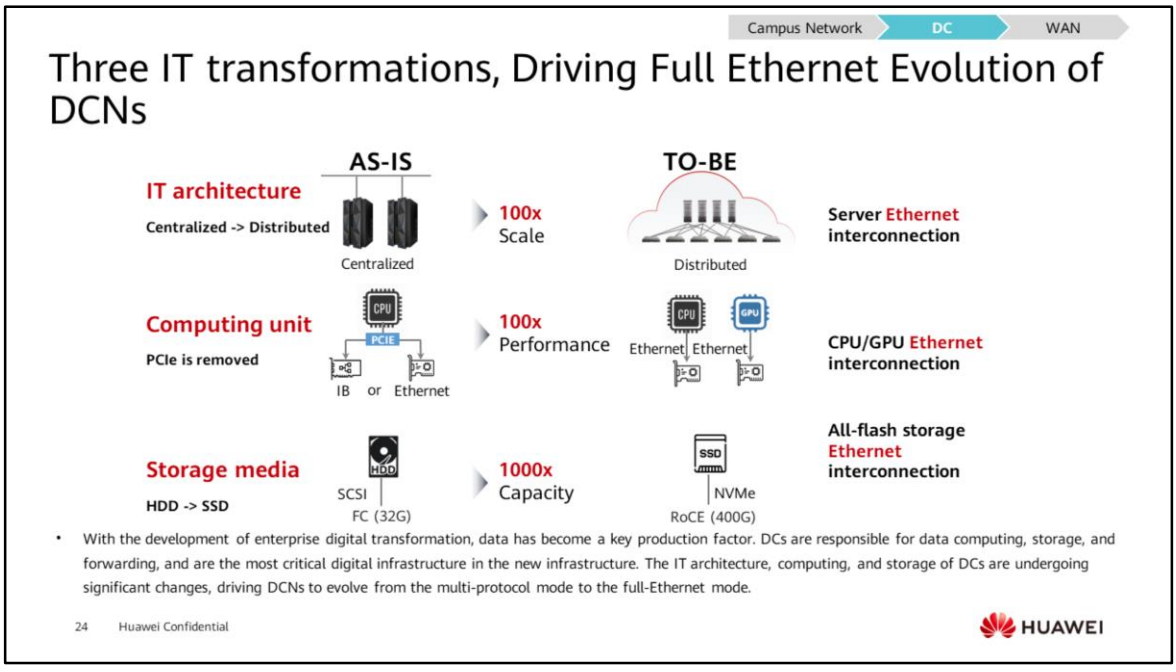


- Supercomputing — the core of high-performance computing (HPC) — provides the computing power needed to drive development of the digital economy and plays a key role in numerous scientific and technological fields.

- Speed is still the primary factor in supercomputing realizing faster and more accurate computing and simulation. This is where cloud and intelligence come in. In addition, the supercomputing of cloudification and artificial intelligence is gradually integrated, which poses higher requirements on the DCN that carries supercomputing services.



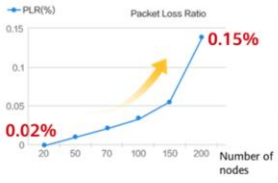
HUAWEI



- InfiniBand (IB): an input/output (I/O) switching technology that uses a central InfiniBand switch to establish a single link between remote storage devices, networks, and servers as well as to direct traffic. It has a compact structure, greatly improving system performance, reliability, and effectiveness, and relieving data traffic congestion between hardware devices.
- Remote Direct Memory Access over Converged Ethernet (RoCE): a network protocol that allows remote direct memory access (RDMA) over the Ethernet. There are two RoCE versions: RoCEv1 and RoCEv2. RoCEv1 is a data link layer protocol that allows any two hosts in the same Ethernet broadcast domain to communicate with each other. RoCEv2 is a network layer protocol and its packets can be routed.

Three Challenges Facing DCN Full-Ethernet Evolution

Zero packet loss required for high-performance computing



The packet loss rate increases exponentially with the increase of network nodes on traditional Ethernet.

Zero packet loss required for dual-active storage

0.2–0.3% packet loss rate

The delay increases in intra-city (long distance) transmission, and it is difficult to perform flow control across DCNs on traditional Ethernet.

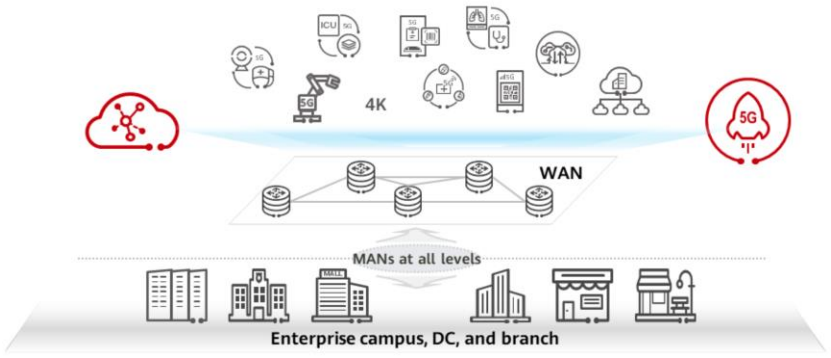
More complex for large-scale network O&M

1000 nodes, **million-level** configuration

Traditional Ethernet lacks effective O&M methods, and the network is too complex to be handled manually.

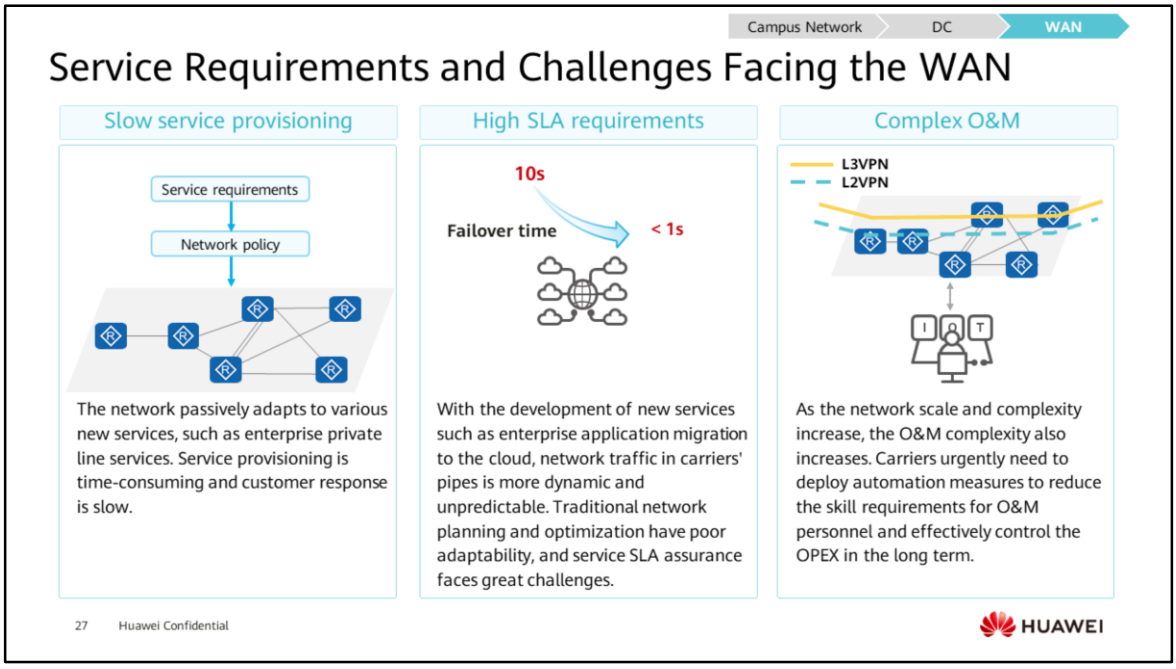
HUAWEI

5G + Cloud, Increasingly Complex WAN Transport Services



- With the advent of the 5G and cloud era, various innovative services such as VR/AR, live streaming, and autonomous driving are emerging. The traffic on the entire network increases explosively, and the dynamic complexity of services also increases the complexity of the entire network.

HUAWEI



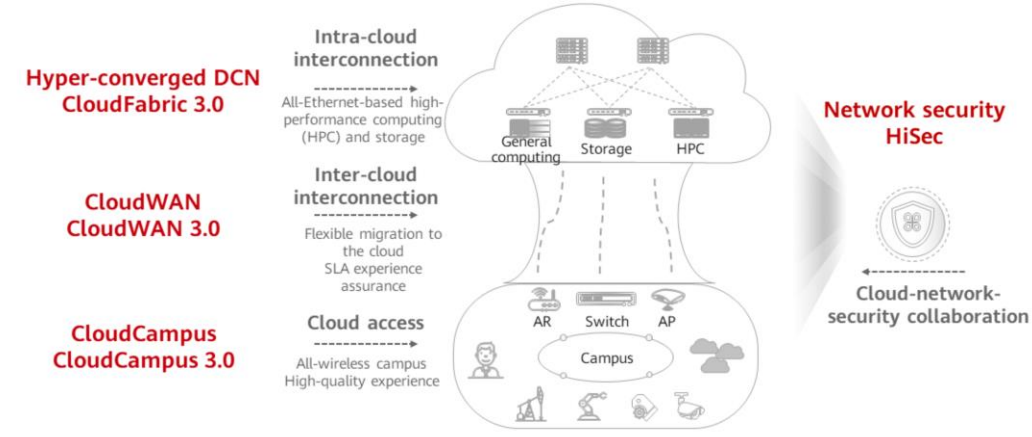
- Service Level Agreement (SLA): an agreement between a service provider and a user. The SLA defines the service type and quality provided by the service provider for the user, and the commitment to the performance and reliability of the user assurance service. For example, the SLA ensures that the service reliability is higher than 99.99%, the fault response time is within 30 minutes.
- Operating Expense (OPEX): the sum of the maintenance cost, marketing expense, labor cost, and depreciation expense during the enterprise operations.

Contents

1. Enterprise Network Overview
2. Trends and Challenges of Enterprise Networks
3. **Huawei Enterprise Network Solutions**

HUAWEI

Huawei Intelligent Cloud-Network Solution



HUAWEI

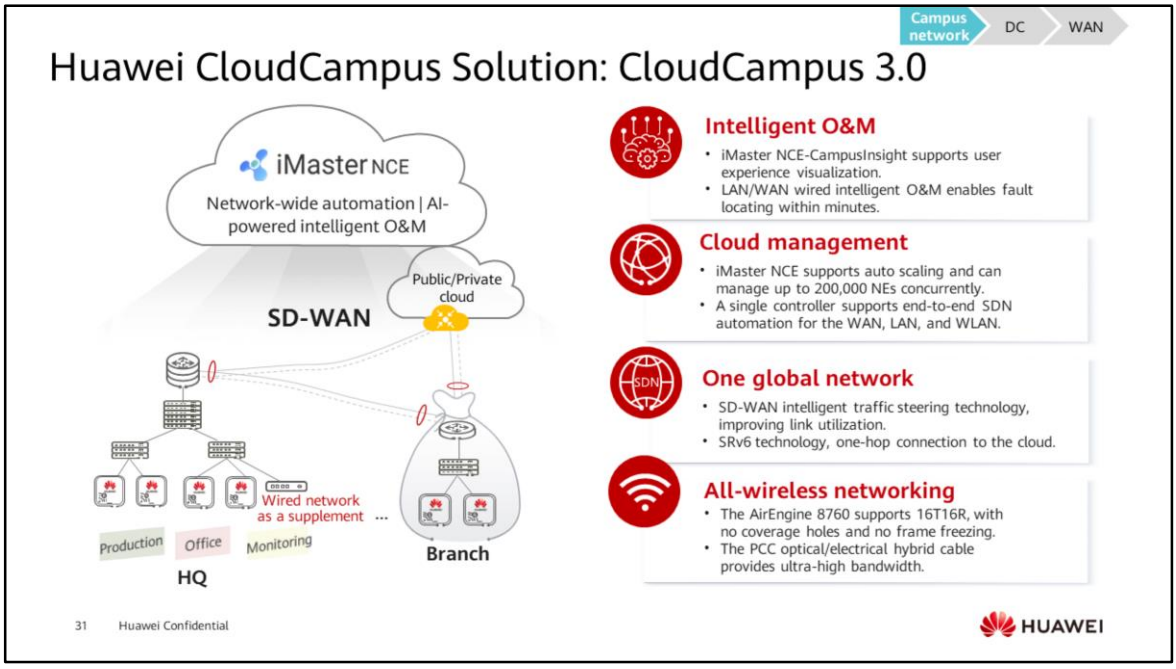
Four Engines + Platform for Manager, Controller, and Analyzer Convergence

iMaster NCE

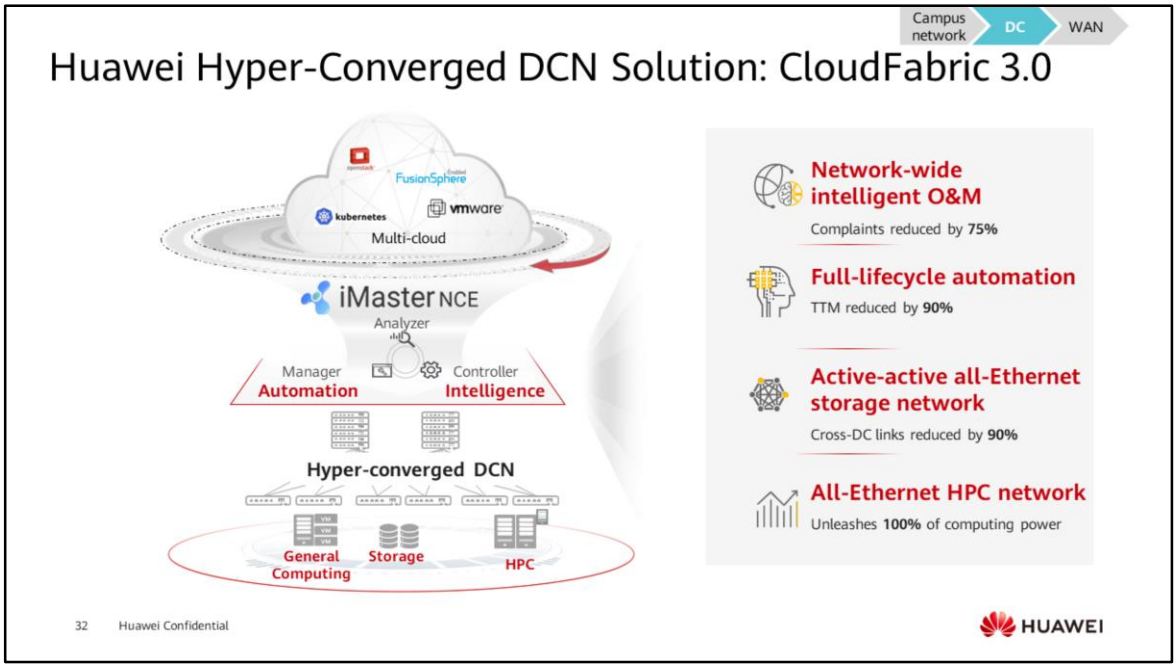
Management, control, and analysis convergence platform

AirEngine	CloudEngine	NetEngine	HiSecEngine
WLAN	Switch	Router	Network security

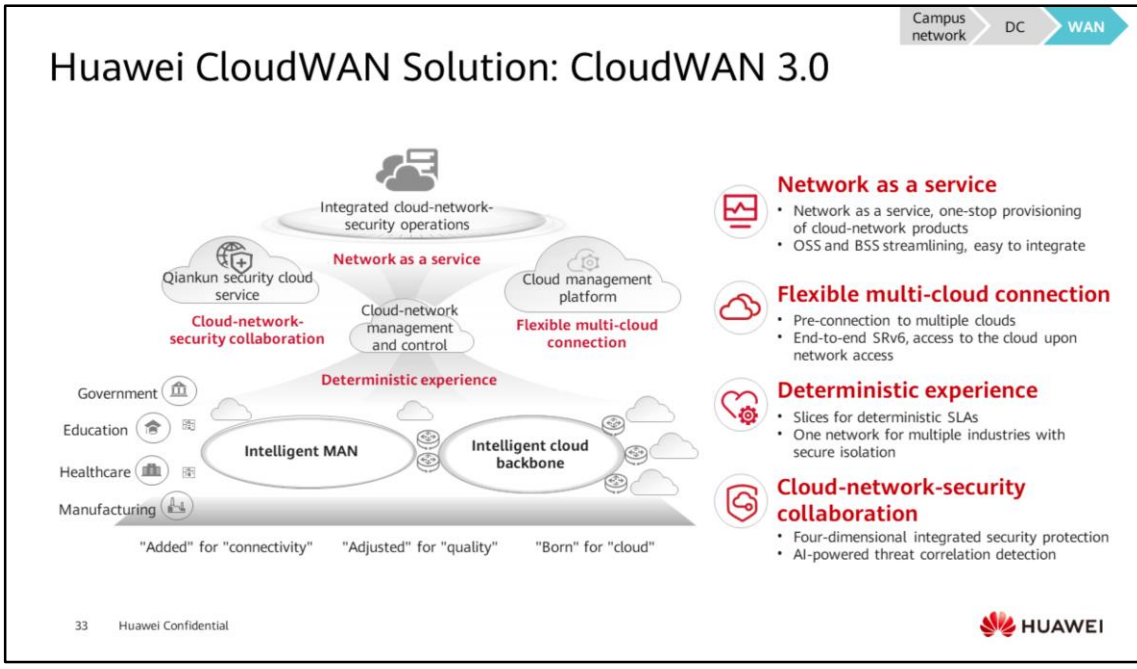
HUAWEI



- In the campus network scenario, iMaster NCE-Campus is used as the iMaster NCE controller.

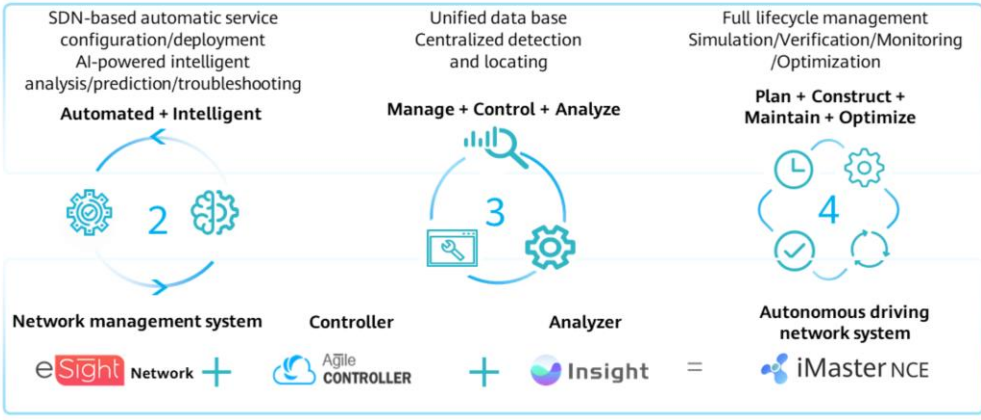


- In the data center scenario, iMaster NCE-Fabric is used as the iMaster NCE controller.



- In this scenario, iMaster NCE-IP is used as the iMaster NCE controller.

What Is iMaster NCE?



iMaster NCE is a network automation platform that integrates management, control, analysis, and AI capabilities.

HUAWEI

All-New iMaster NCE

iMaster NCE

DC	iMaster NCE-Fabric
Enterprise campus	iMaster NCE-Campus
SD-WAN	iMaster NCE-WAN
IP WAN	iMaster NCE-IP
WAN transmission	iMaster NCE-T

HUAWEI

Quiz

1. (Multiple-answer question) Which of the following controllers have the SD-WAN feature? ( )
   A. iMaster NCE-Campus
   B. iMaster NCE-WAN
   C. iMaster NCE-IP
   D. iMaster NCE-SD-WAN

HUAWEI

1. AB

Summary

- This course briefly introduces the overall enterprise network architecture and the campus networks, DCNs, and WANs in different scenarios.
- With the continuous development of services and new technologies, enterprise networks are facing great challenges. Therefore, Huawei provides corresponding solutions for networks in different scenarios. These contents will be described in more details in subsequent courses.

HUAWEI

Thank you.

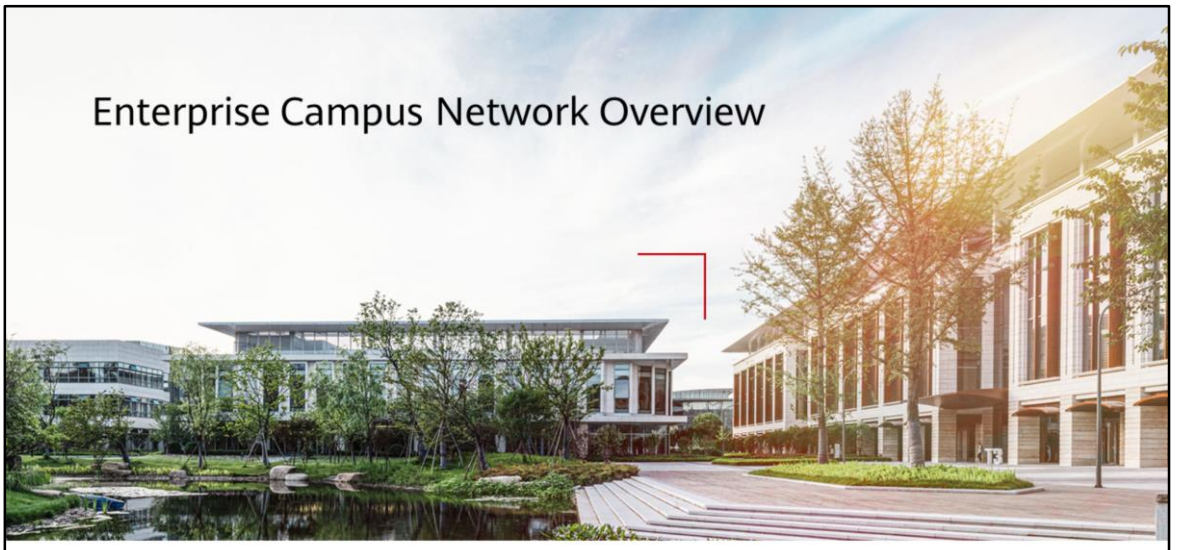
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Enterprise Campus Network Overview



HUAWEI

Foreword

- Campuses are everywhere in our cities. Some examples of campuses include factories, government buildings and facilities, shopping malls, office buildings, school campuses, and parks. According to statistics, 90% of urban residents work and live in campuses, 80% of gross domestic product (GDP) is created in campuses, and each person stays in campuses for 18 hours every day. Campus networks, as the infrastructure for campuses to connect to the digital world, are an indispensable part of campus construction and play an increasingly important role in daily working, R&D, production, and operation management.
- This course describes campus networks, common architectures and technologies of campus networks, and typical applications of these technologies. This information helps readers better understand the campus network design.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe what a campus network is like.
  - Describe the typical architecture of a campus network.
  - Describe the trends and challenges of campus networks.
  - Describe Huawei CloudCampus Solution.
  - Understand common technologies used on campus networks.
  - Understand typical applications of campus network technologies.

HUAWEI

Contents

1. **Introduction to Campus Networks**
2. Campus Network Challenges and Huawei CloudCampus Solution
3. Typical Campus Network Technologies
4. Typical Applications of Campus Network Technologies

HUAWEI

Campuses Are Everywhere

90%+

of city residents work and live in campuses.

5 h & 22 h

spent in using smart terminals & staying in the campus every day

80%+

of GDP is created in campuses.

90%+

of innovations are made in campuses.

HUAWEI

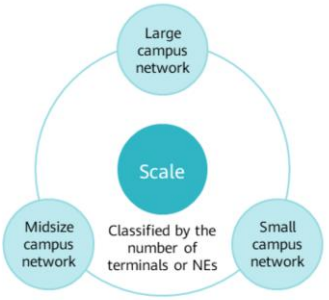
Overview of a Campus Network

- A campus network generally refers to the internal network of an enterprise or organization, which is connected to the wide area network (WAN) and data center network.
- A campus network is built to ensure that key enterprise services are running more efficiently.
- Campus networks can be classified into large- and medium-sized campus networks and small- and medium-sized campus networks by scale.
- Some enterprises have branches dispersed in different geographical locations. Each branch network can be considered as a single campus network.

HUAWEI

Campus Network Classification (1)

- Number of terminals > 2000
- Number of NEs > 100

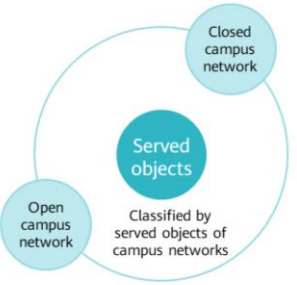


- Number of terminals: 200–2000
- Number of NEs: 25–100

- Number of terminals < 200
- Number of NEs < 25

Networks of different scales have different requirements and pain points.

- Users: internal personnel only
- Security requirements: network access control and external threat defense



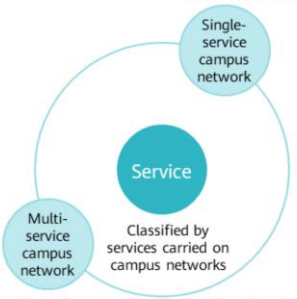
- Users: including external personnel, such as the public
- Security requirements: access control, identity identification, behavior control, security protection, etc.

A running campus network usually has both closed and open subnets.

HUAWEI

Campus Network Classification (2)

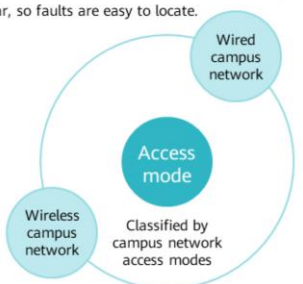
- Single services
- Simple network architecture



- The network needs to carry a large number of services, and the network scale is large. Different services need to be isolated and guaranteed.
- The campus network architecture is becoming complex and virtualized.

The complexity of the campus network architecture depends on the complexity of services carried on the campus network.

- Each device accessing a network must be connected to the preset network port through a network cable.
- The architecture is structured and hierarchical, and the logic is clear, so faults are easy to locate.



- The network is based on the 802.11 protocol (Wi-Fi) and is also called WLAN.
- AP deployment and installation affect the coverage effect. Interference and conflicts exist, so the network needs to be optimized periodically.
- Faults are difficult to locate.

Currently, most campus networks are a mix of wired and wireless networks.

HUAWEI

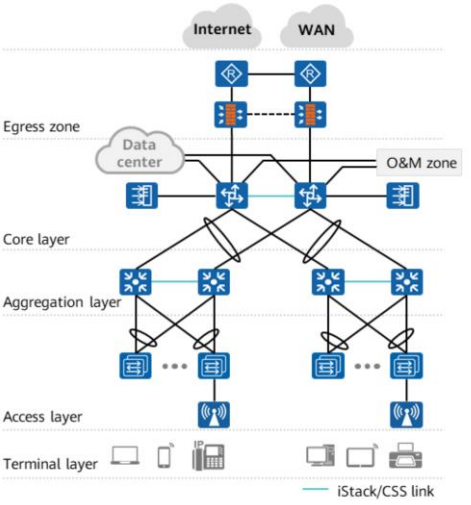
Campus Network Classification (3)

To meet requirements of different industries, the campus network architecture is designed based on the characteristics of the industries that campus networks serve, ultimately building industry-specific campus network solutions.

Enterprise campus network	School campus network	Government campus network	Business campus network
• It refers to the Ethernet-based enterprise office network. • The enterprise campus network focuses on network reliability and advancement, continuously improves employees' office experience, and ensures the efficiency and quality of operation and production.	• School campus networks are classified into primary/secondary education and higher education campus networks. • Higher education campus networks are complex and usually have teaching and research networks, student networks, and operational dormitory networks. • There are high requirements on network manageability and security, and specific requirements on network advancement.	• It usually refers to the internal network of a government agency. • High security is required. Generally, the internal network and external network are isolated to ensure high security of confidential information.	• This type of campus networks apply to shopping malls, supermarkets, hotels, and parks. • Such networks mainly serve consumers. In addition, they include subnets for internal office work. • Such networks provide Internet access services and help build business intelligence (BI) systems for a better user experience, lower O&M cost, higher efficiency, and value transfer.

HUAWEI

Typical Physical Architecture of a Campus Network



- **Egress zone**: serves as the border between the campus internal network and the external network. Through this egress zone, internal users can access the public network and external users (including customers, partners, branch users, and remote users) can access the internal network. Firewalls can be deployed at the egress zone to ensure the security of the internal network.
- **Core layer**: serves as the core of data switching on the campus network. It connects various parts of the campus network, such as the data center, O&M zone, and egress zone.
- **Aggregation layer**: forwards not only horizontal traffic between users, but also vertical traffic to the core layer. It can also function as the switching core for a department or zone and further extend the quantity of access terminals.
- **Access layer**: provides various access modes for users and is the first layer of the network to which a terminal accesses.
- **Terminal layer**: has terminals deployed to connect to the campus network. Terminals include computers, printers, IP phones, mobile phones, and cameras.
- **Data center**: has servers and application systems deployed to provide data and application services for internal and external users of the enterprise.
- **O&M zone**: manages network servers such as the NMS and authentication server.

HUAWEI

Contents

1. Introduction to Campus Networks
2. **Campus Network Challenges and Huawei CloudCampus Solution**
3. Typical Campus Network Technologies
4. Typical Applications of Campus Network Technologies

HUAWEI

Industry Digital Transformation Improves Efficiency and Customer Satisfaction

Digital workspace

- All-wireless access
- Anytime, anywhere workstyles
- One-click conference reservation via an app
- Auto-adjustable lighting and temperature

"At least **55%** of the organizations will be firm advocates of digitalization by 2020."
Source: IDC FutureScape: Worldwide Digital Transformation 2019 Predictions

Digital education

- Shift from spoon-feeding education to immersive education
- On-demand access to high-quality teaching
- Practical-scene teaching

"**86%** of higher education CIOs will regard IT as a key factor in enabling education business transformation."
Source: 2019 Higher Education Industry Insights, Gartner

Digital manufacturing

- Always-on mass terminals/sensors
- Real-time collection of production data
- Automated and precise control

"**84.9%** of manufacturing enterprises are going digital, driving business model innovations and reshaping the business ecosystem."
Source: IDC's 2018 manufacturing industry survey

HUAWEI

- CIO: Chief Information Officer

In the Digital Era, How Should Campus Networks Support Digital Transformation Across Industries?

Ubiquitous connectivity

- Access anytime, anywhere
- High-quality service support

On-demand services

- Quick service deployment and adjustment
- Rapid rollout of value-added applications

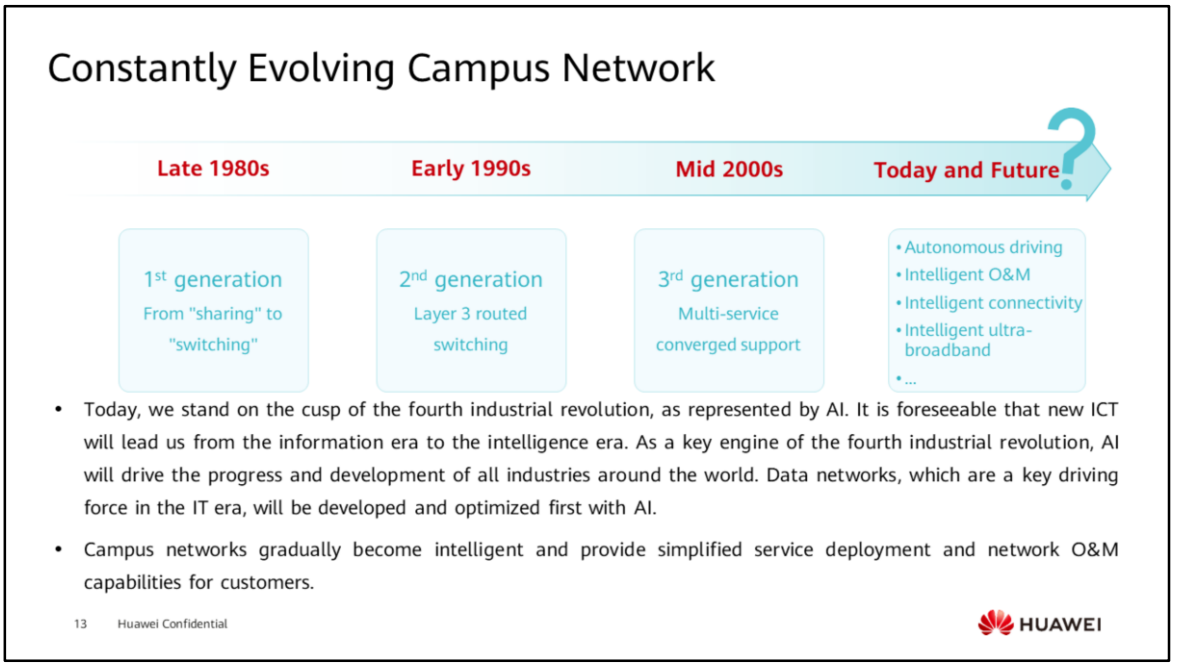
Perceptible and visible experience

- Precise measurement and evaluation of user experience
- AI-powered automatic adjustment of resource allocation

Efficient and intelligent O&M

- Configuration delivery based on services
- Fast and accurate fault locating

HUAWEI



- 1st-generation campus network:
  - In 1980, IEEE released the IEEE 802.3 standard, signaling the birth of Ethernet technology. By using twisted pair connections, Ethernet was more cost-effective and easier to implement than previous networking technologies. Consequently, Ethernet quickly became the mainstream technology for campus networks. During the early days, campus networks used hubs as access devices. A hub was a shared-medium device that worked at the physical layer. It was limited in the number of users it could support for concurrent access. If many users connected to a hub simultaneously, network performance degraded severely due to the expanded collision domain. In the late 1980s, Ethernet switches emerged. Due to their more advantageous working scheme than hubs, Ethernet switches therefore quickly replaced hubs to become the standard components of campus networks.

Service Requirements and Challenges of Large- and Medium-Sized Campus Networks

Converged transport

Requirements:

Access terminals and services are diversified, and a converged transport network is required.

Challenges:

- Wi-Fi and IoT services are independently planned, deployed, and managed. The overall network construction cost is high.
- The workload of network management and O&M is heavy.

User experience awareness

Requirements:

Network O&M needs to be automated and intelligent to perceive user experience anytime and anywhere.

Challenges:

- Service faults cannot be detected in a timely manner.
- After a fault occurs, the fault cause is determined based on the O&M experience of professional personnel, and the fault cannot be quickly located.
- The network cannot be automatically optimized.

Network automation

Requirements:

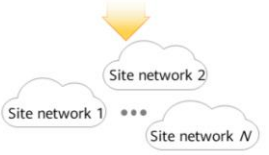
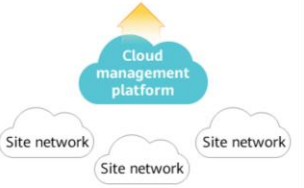
Networks need to be automated to cope with the complexity of deployment and policies caused by the surge of applications and services.

Challenges:

- The workload is repetitive and heavy, and manual configuration is complex.
- New services need to be configured on devices one by one, which is time-consuming and costly.
- The workload of network policy deployment and adjustment is heavy.

HUAWEI

Service Requirements and Challenges of Small- and Medium-Sized Campus Networks

Plug-and-play network devices, improving deployment efficiency	Centralized O&M on the cloud, simplifying O&M of multiple sites	Open APIs, accelerating integration of business applications
Unified management and centralized configuration • The network configurations of multiple sites are centrally delivered, reducing onsite configuration and commissioning workload and improving deployment efficiency. • The network is plug-and-play and able to be expanded on demand, requiring low costs for upgrades.	Centralized management of multiple branches on the cloud and remote automatic O&M  Scattered campus branch networks are centrally managed on the cloud through the Internet, and automation tools for troubleshooting, monitoring, and other management operations are integrated, implementing remote automated O&M.	Open and big data analytics capability  With open APIs and big data analytics capabilities, multiple management systems can be combined and interconnected to achieve unified network management. More value-added applications are available, leading enterprises into digital transformation.

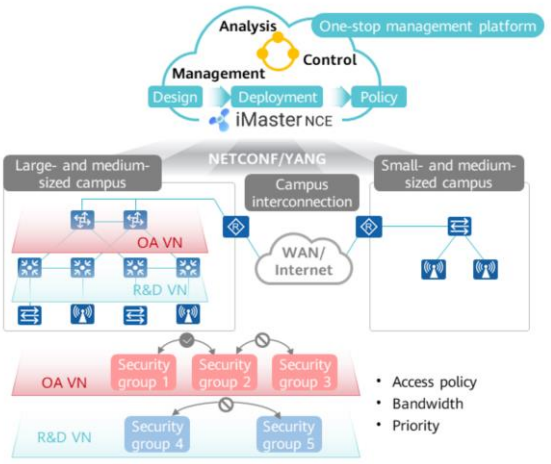
HUAWEI

Service Requirements and Challenges of Multi-Campus Network Interconnection

Branch interconnection cost	Service provisioning period	Service experience	Management and O&M	Unified management
Requirements: The cost of WAN interconnection between branches needs to be reduced. **Challenges:** Carriers' transmission private lines or MPLS VPN private lines are of high quality but expensive.	**Requirements:** The network deployment efficiency needs to be improved and the service provisioning period needs to be shortened. **Challenges:** • It takes a long time to provision traditional private lines. • Service provisioning is mainly performed by network engineers on site, which has high skill requirements and low efficiency.	**Requirements:** The application identification capability needs to be enhanced to ensure key service experience. **Challenges:** • Different enterprise applications have different requirements on link quality. • Traditional private lines cannot detect the application status and cannot guarantee key services at any time.	**Requirements:** Visualized O&M is required to simplify management and O&M. **Challenges:** • CLI-based O&M is inefficient and lacks visualized O&M methods. • If an enterprise has a large number of branches, onsite O&M will increase costs.	**Requirements:** LAN-side and WAN-side services need to be managed in a unified manner. **Challenges:** • Unified configuration management and O&M cannot be implemented.

HUAWEI

CloudCampus: One-Stop Autonomous Driving Solution for Campus Networks



Automated network design, accurately aligning with service intents

- The one-stop management platform can accurately convert service intents into configuration commands.

Automated network deployment, provisioning services in minutes

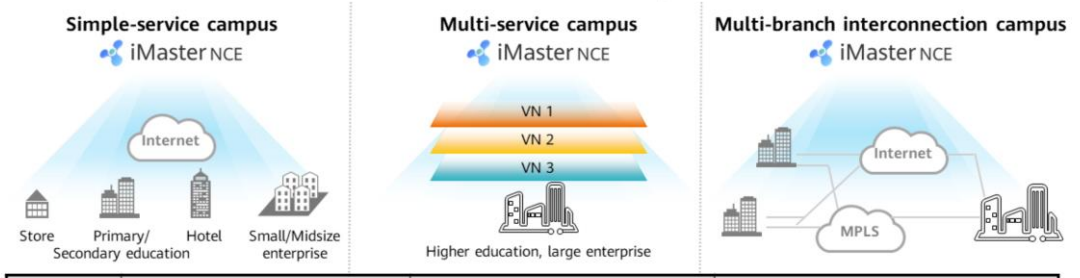
- Plug-and-play devices and automatic deployment of physical networks.
- Network resource pooling, multi-purpose network, and automatic service provisioning.
- Fine-grained control using intelligent policies (covering access policies, bandwidth, priorities, etc.) based on virtual networks, security groups, and applications.

Automated intelligent O&M, improving network-wide performance

- **Real-time visualized experience**: Visualized network experience for each user in each area at each moment.
- **Precise fault analysis**: Proactively identifying typical network issues and providing remediation suggestions.
- **Intelligent network optimization**: Predictive optimization of wireless networks based on historical data.

HUAWEI

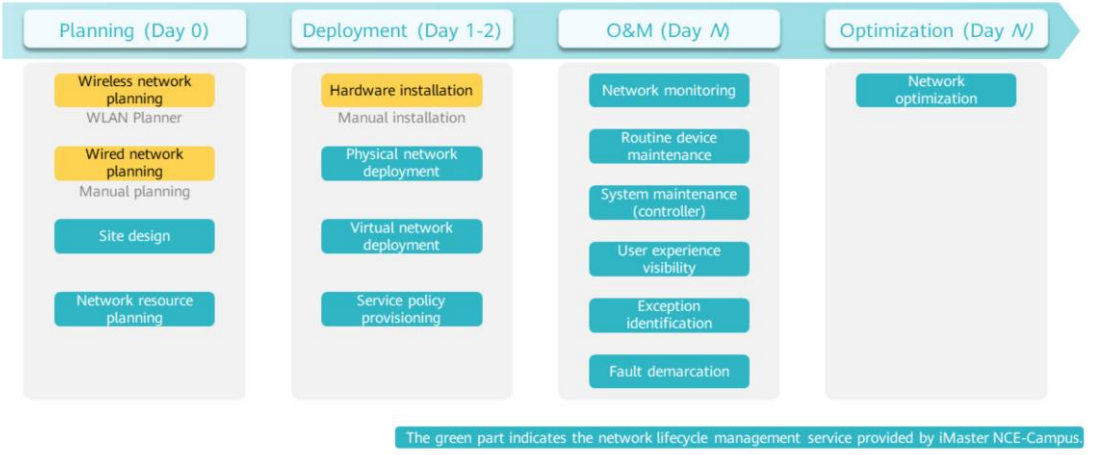
Full Scenarios: Full Coverage from Simple-Service Campuses to Multi-Branch Interconnection Campuses



	Simple-service Campus	Multi-service Campus	Multi-branch Interconnection Campus
Network characteristics	Small scale, simple services Large numbers of sites, with similar models	Large scale, complex services, and coexistence of multiple services	Multiple branch sites, which need to communicate with each other through hybrid WAN links
Typical scenarios	Multi-branch and small/midsize enterprise campuses, such as hotels and primary/secondary schools	Higher education institutions, governments, large enterprise campuses, etc.	Large enterprises, financial branches, etc.

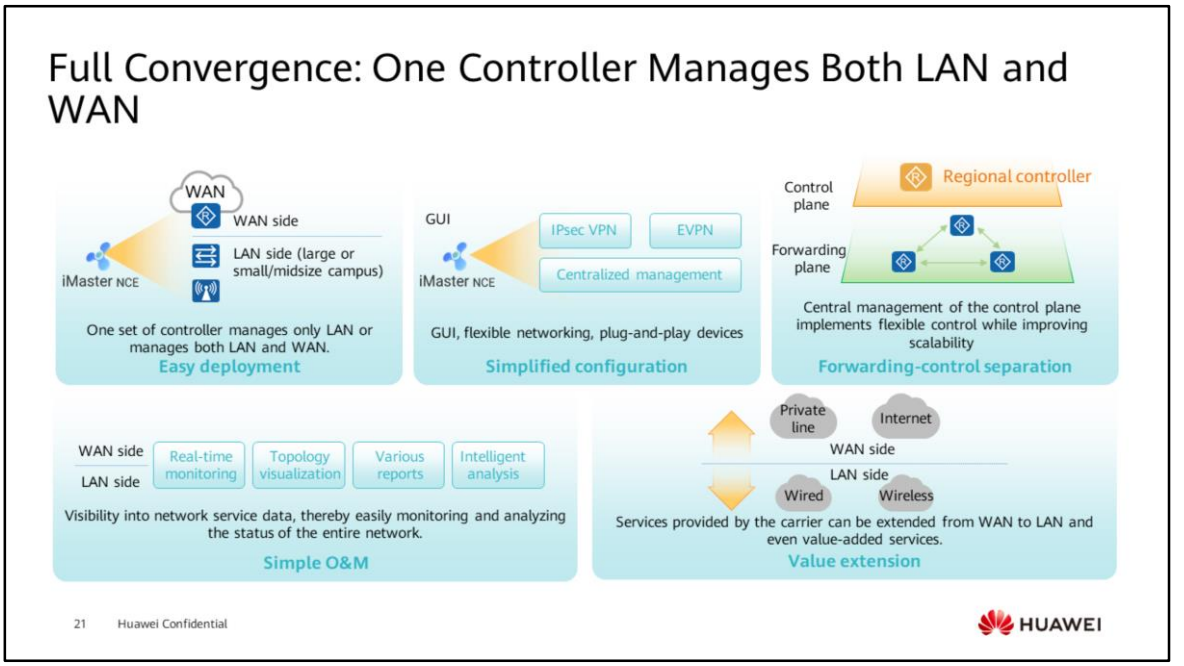
HUAWEI

Full Lifecycle: Planning, Deployment, O&M, and Optimization



The green part indicates the network lifecycle management service provided by iMaster NCE-Campus.

HUAWEI



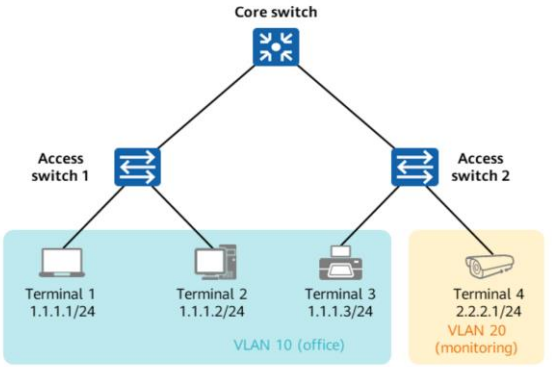
- Huawei CloudCampus Solution integrates service configuration and management models across LAN and WAN. It achieves LAN-WAN convergence by not only configuring and managing LAN services, but also managing WAN interconnection services.

Contents

1. Introduction to Campus Networks
2. Campus Network Challenges and Huawei CloudCampus Solution
3. **Typical Campus Network Technologies**
4. Typical Applications of Campus Network Technologies

HUAWEI

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 specify different broadcast domains. Devices in a VLAN can directly communicate with each other, whereas devices in different VLANs cannot communicate directly, but require a Layer 3 device for mutual communication.
- Typically, a VLAN is a logical subnet.
- Members of a VLAN are assigned based on interfaces of a switch. After an interface of a switch is added to a VLAN, the device connected to the interface is also added to this VLAN.

HUAWEI

- In addition to interface-based VLAN assignment, you can also use the following methods to assign VLANs:
  - MAC address-based assignment: assigns VLANs based on the source MAC addresses of frames. This mode applies to small-scale networks where physical locations of user terminals frequently change but their network adapters seldom change.
  - IP subnet-based assignment: assigns VLANs based on the source IP addresses of frames. This mode applies to scenarios where there are high requirements for mobility and simplified management and low requirements for security.
  - Protocol-based assignment: assigns VLANs based on the protocol (suite) types and encapsulation formats of frames. This mode applies to networks running multiple protocols.
  - Policy-based assignment: assigns VLANs based on a specified policy, which means VLANs are assigned based on combinations of interfaces, MAC addresses, and IP addresses. This mode applies to complex networks.

Voice VLAN

- Multiple types of flows are usually transmitted on a network simultaneously. Packet loss and delay seriously affect voice communication quality. Users are more sensitive to the quality of the voice service than to the quality of data and video services. Therefore, voice data flows must be preferentially transmitted when there is limited bandwidth available.
- A voice VLAN is used to exclusively transmit voice data.
- After a voice VLAN is configured on a switch, the switch can identify voice data flows, transmit the voice data flows in the voice VLAN, and provide QoS guarantee for the voice data flows.

Typical networking of voice VLAN technology

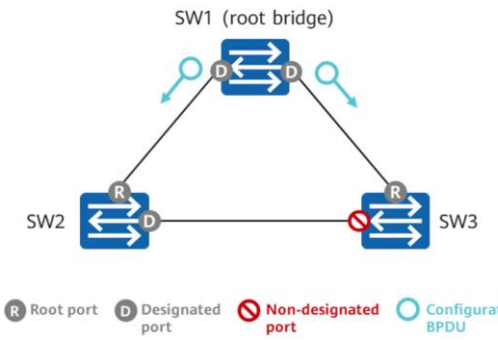
You can use a single interface of a switch to transmit both voice and data services.

An IP phone is connected to a switch independently.

HUAWEI

- A device configured with voice VLAN can identify voice flows in either of the following modes:
  - MAC address-based identification: A device identifies voice flows based on the source MAC addresses of received data packets. If a source MAC address of a packet matches the organizationally unique identifier (OUI) of a voice device, the packet is considered a voice packet. OUIs must be preconfigured and are used in scenarios where IP phones send untagged voice packets.
  - VLAN-based identification: Configuring OUIs for a large number of IP phones is time-consuming. In this case, you can configure a switch to identify voice packets based on VLAN IDs. If the VLAN ID of a received packet matches the configured voice VLAN ID, the packet is considered a voice packet. This simplifies configuration when a large number of IP phones are connected to the switch. However, the IP phones must be able to obtain voice VLAN information from the switch.

Spanning Tree Protocol (STP)



- On a switching network with physical loops, switches running STP automatically generate a loop-free working topology, which is also called an STP tree. Each node of an STP tree is a specific switch, and each branch is a specific link.
- 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 proactively sends configuration BPDUs during initialization. After the network topology becomes stable, only the root bridge periodically sends configuration BPDUs. Other switches send their own configuration BPDUs only after receiving configuration BPDUs from upstream devices.

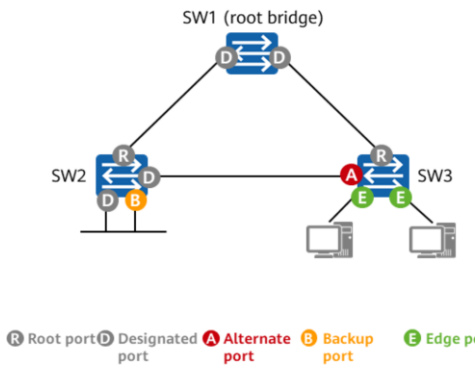
PID	PVI	BPDU Type	Flag	Root ID	RPC	Bridge ID	Port ID	Message Age	Max Age	Hello Time	Forward Delay

Format of a configuration BPDU

HUAWEI

- How is an STP tree generated?
  - STP compares four parameters: root bridge ID, root path cost (RPC), bridge ID (BID), and port ID (PID). A smaller value indicates a higher priority. All these parameters are BPDU fields.
    - Root bridge election: The device with the smallest root bridge ID is elected as the root bridge.
    - Root port election: A device compares the RPC, peer BID, peer PID, and local PID of its ports in sequence. The port with the smallest value is elected as the root port.
    - Designated port election: A device compares the RPC, local BID, and local PID of its ports in sequence. The port with the smallest value is elected as the root port.
    - After the root port and designated port are determined, all the non-root ports and non-designated ports on the switch will be blocked.

Rapid Spanning Tree Protocol (RSTP)

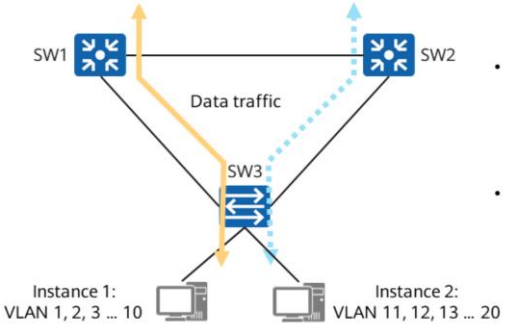


R Root port D Designated port A Alternate port B Backup port E Edge port

- As LANs are increasing in scale, the problem of slow STP topology convergence is becoming more prominent. To address this problem, the IEEE released the 802.1w standard in 2001 that defined the RSTP protocol. RSTP is an enhanced version of STP and implements rapid convergence of the switching network.
- RSTP makes the following improvements over STP:
  - Defines more port roles, making the spanning tree protocol easier to understand and configure.
  - Eliminates the listening state, reducing the number of port states from 5 to 3.
  - Uses the Flags field of STP configuration BPDUs to define port roles.
  - Processes configuration BPDUs differently from STP. For example, it allows non-root bridges to proactively generate and send configuration BPDUs, and reduces the BPDU timeout interval to three times that of the Hello Time.
  - Provides rapid convergence mechanisms, such as the Proposal/Agreement (PA) mechanism and edge port mechanism.
  - Provides protection mechanisms to ensure the stability of the switching network.

HUAWEI

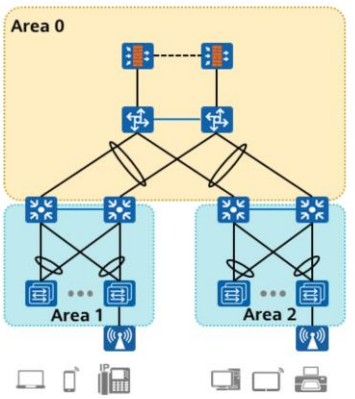
Multiple Spanning Tree Protocol (MSTP)



- Defined in IEEE 802.1s and compatible with STP and RSTP, MSTP implements fast convergence while providing multiple redundant paths for forwarding data, effectively load balancing traffic for VLANs.
- An MSTP network is composed of one or more MST regions, and each MST region contains one or more multiple spanning tree instances (MSTIs). An MSTI is a tree network that consists of switches running STP, RSTP, or MSTP.
- The switches in the same MST region share the following characteristics:
  - Have MSTP enabled.
  - Have the same region name.
  - Have the same VLAN-to-MSTI mappings.
  - Have the same MSTP revision level.

HUAWEI

OSPFv2

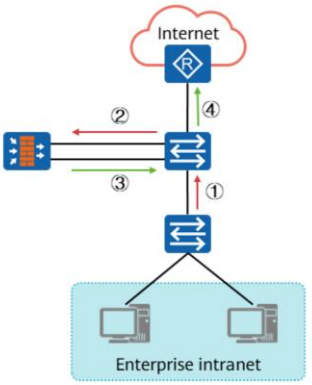


On a large-scale enterprise network, the aggregation layer may consist of Layer 3 devices, such as routers and switches. In such scenarios, static routes are not flexible because the configuration is complex and they cannot quickly respond to topology changes. Therefore, a dynamic routing protocol — typically OSPF in IGP — can be deployed on the enterprise intranet.

- The aggregation devices are divided into different non-backbone OSPF areas based on the network structure.
- Core devices, network egress devices, and uplink interfaces of aggregation devices belong to Area 0. The downlink interfaces of aggregation devices and the access devices are added to Area X.
- Redundancy links and the OSPF triggered update mechanism are used to implement backup of intranet paths.
- Aggregation devices function as area border routers (ABRs) to execute route filtering policies.

HUAWEI

Policy-Based Routing (PBR)



PBR can be configured using the modular QoS command line interface (MQC) or an ACL-based simplified traffic policy.

In the conventional routing and forwarding process, devices search their IP routing tables for routes based on packets' destination addresses and then forward the packets accordingly. However, more and more users require packet routing based on user-defined policies.

Traffic classifier

Configure a traffic classifier for matching interested data flows.

Matching criteria: VLAN ID, source or destination MAC address, Ethernet protocol type, DSCP priority, IP precedence, inbound or outbound interface, ACL rule

Traffic behavior

Redirects interested packets.

You can set the next-hop IP address or outbound interface for redirection.

Traffic policy

Traffic classifier > Traffic behavior

Apply the traffic policy

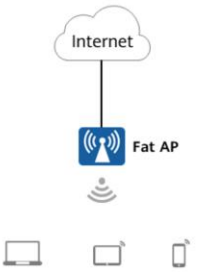
- Apply the traffic policy to the inbound direction of an interface.
- Apply the traffic policy to the incoming packets that belong to the VLAN and match the criteria defined in the traffic classifier.
- Apply the traffic policy globally or to a card.

HUAWEI

- PBR is used in agile campus service orchestration, multi-egress, and anti-DDoS off-path deployment scenarios.
- MQC: Modular QoS Command Line Interface

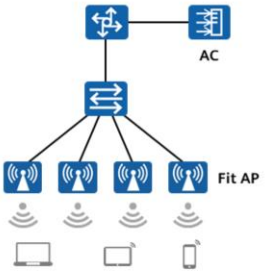
WLAN Network Architectures (1)

Fat AP



- **Networking characteristics**: A Fat AP works independently and needs to be configured separately. It provides only simple functions and is cost-effective.
- **Application scenarios**: families or mini-stores

AC + Fit AP

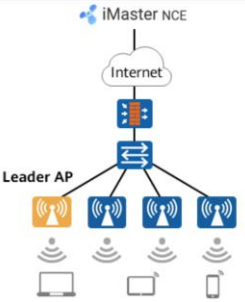


- **Networking characteristics**: A Fit AP provides a wide range of functions and is used with an AC. Fit APs are managed and configured by an AC in a unified manner, posing high requirements on skills of maintenance personnel.
- **Application scenarios**: medium- and large-sized enterprises

HUAWEI

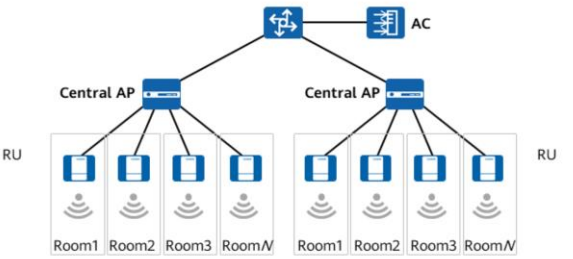
WLAN Network Architectures (2)

Leader AP



- **Networking characteristics**: A leader AP can work independently or manage a small number of common APs to implement basic roaming functions. A leader AP has a low price and low requirements for maintenance skills.
- **Application scenarios**: small and micro enterprises

Agile distributed AP

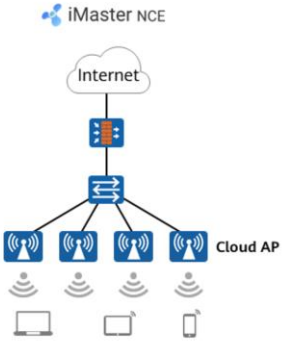


- **Networking characteristics**: The agile distributed architecture divides APs into central APs and remote units (RUs). A central AP can manage multiple RUs, and this architecture provides good coverage and reduces costs. RUs can be used in the Fat AP, AC + Fit AP, and cloud management architectures.
- **Application scenarios**: scenarios where rooms are densely distributed

HUAWEI

WLAN Network Architectures (3)

Cloud management mode

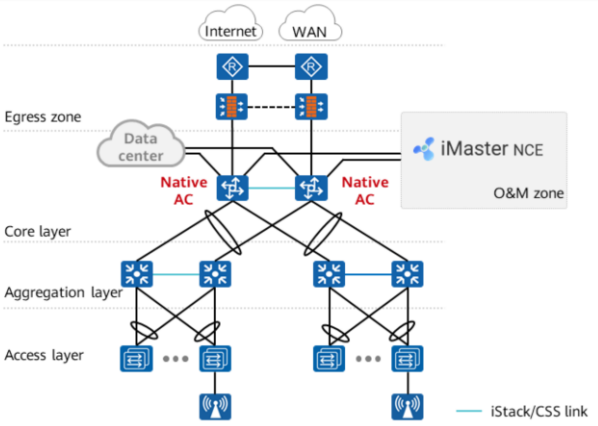


- **Networking characteristics**: APs are centrally managed and configured on iMaster NCE-Campus. They provide rich functions and support plug-and-play.
- **Application scenarios**: small- and medium-sized enterprises

HUAWEI

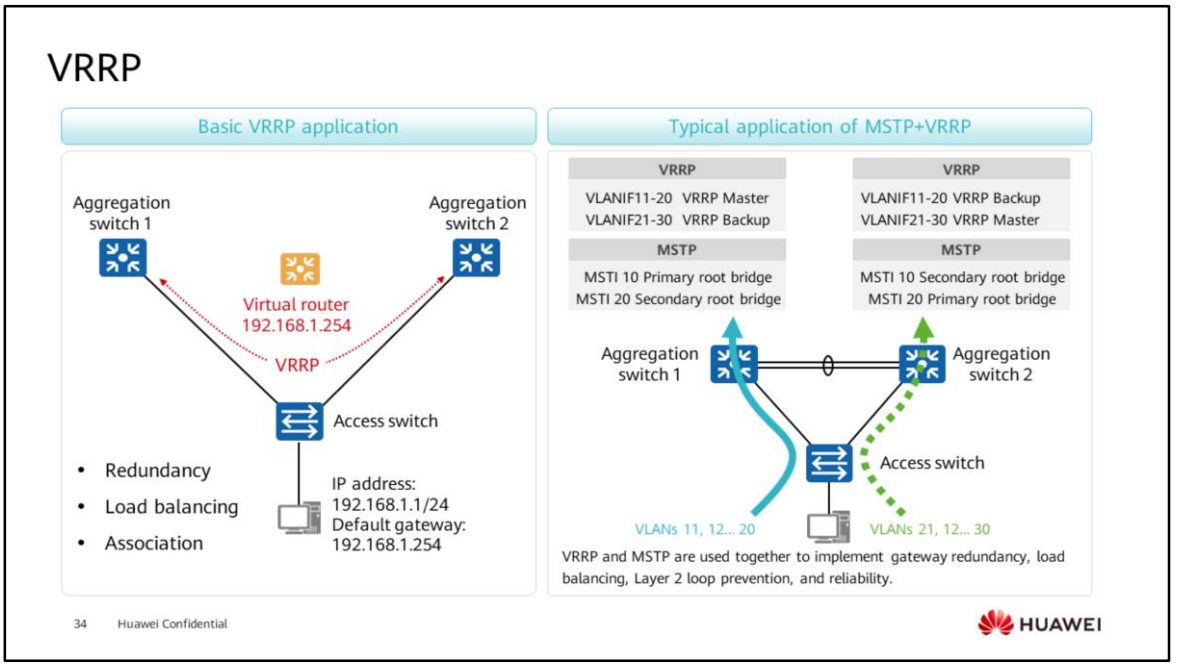
WLAN Network Architectures (4)

Native AC



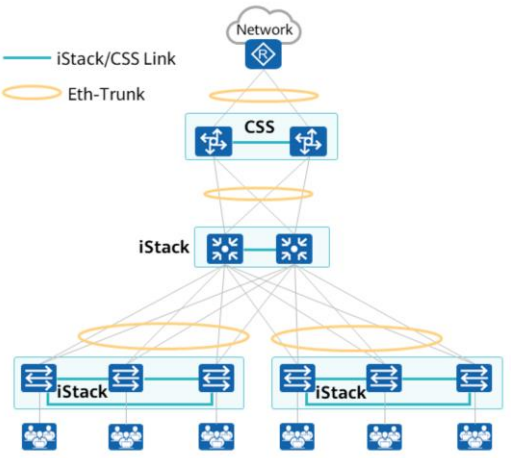
- **Networking characteristics**: This architecture uses the native AC function of switches to manage APs on a network, implementing wired and wireless network convergence. This architecture also leverages big data and AI technologies to implement simplified, smart, and secure campus networks.
- **Application scenarios**: large- and medium-sized enterprises

HUAWEI



- 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, which then forwards the packets to other network segments, enabling hosts to communicate with external networks. If the gateway fails, hosts using this gateway address as the next hop of their default route cannot communicate with external networks.
- The Virtual Router Redundancy Protocol (VRRP) virtualizes several routing devices into a virtual router and uses the IP address of the virtual router as the default gateway address for the communication between users and external networks. If a gateway fails, VRRP selects another 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 (SPOF) occurs, services are transmitted through the backup link. This reduces the possibility of network faults and ensures non-stop transmission of services.
    - Load balancing: VRRP enables multiple available routers to share the load, reducing the traffic burden on the master.
    - Association: VRRP can monitor faults on uplinks. When the uplink interface or uplink is faulty, the priority of the original master decreases, and an optimal backup becomes the master, ensuring proper traffic forwarding. Association between VRRP and BFD speeds up the active/standby switchover. To speed up the active/standby switchover in the VRRP group, configure a BFD session between the master and backup and associate the BFD session with the VRRP group. This is because BFD can fast detect faults. When the link between the master and backup becomes Down, the backup immediately switches to the master and takes over traffic.

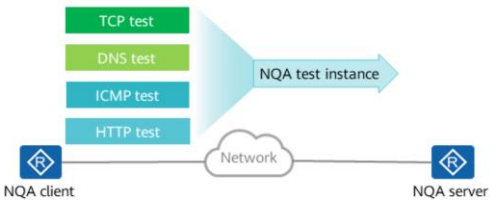
Link Aggregation, iStack, and CSS



- Ethernet link aggregation, also known as Eth-Trunk, bundles multiple physical links into a logical link to increase link bandwidth, without having to upgrade hardware.
- 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), also known as a cluster, combines two clustering-capable switches into a single logical switch. CSS is applicable to Huawei modular switches.
- Link aggregation can be used with iStack/CSS to implement link-level and device-level reliability and increase network bandwidth.

HUAWEI

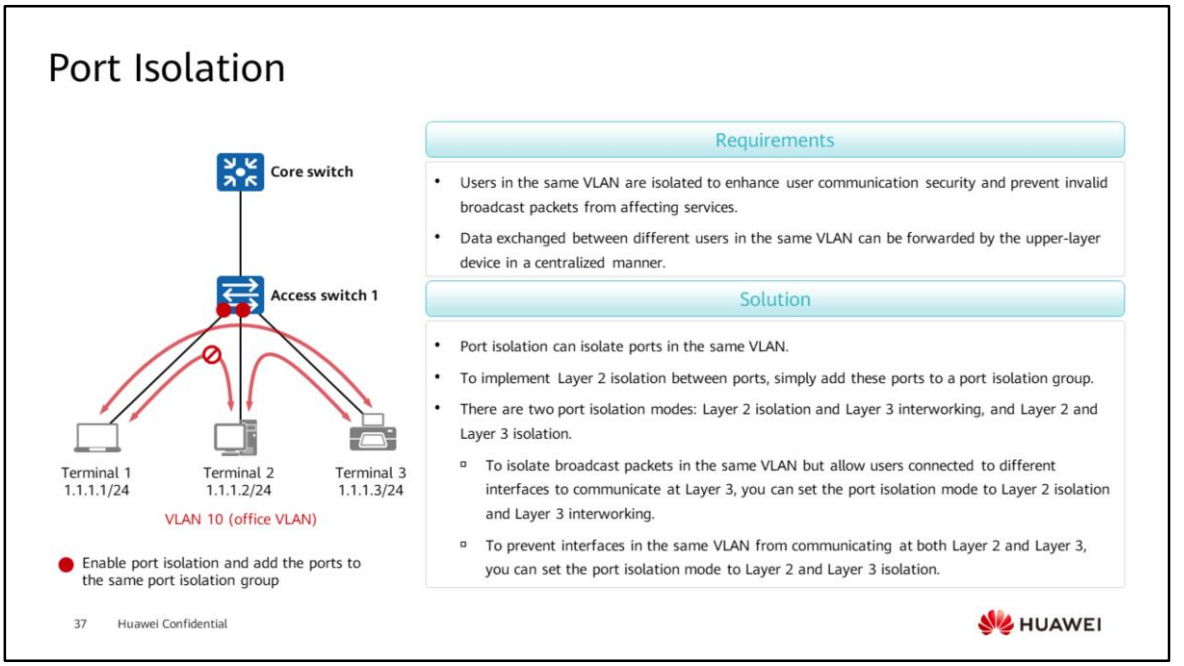
Network Quality Analysis (NQA)



- To visualize the quality of network services and allow users to check whether the quality of network services meets requirements, the following measures must be taken:
  - Enable the device to provide network service quality information.
  - Deploy probe devices to monitor network service quality.
- The preceding measures require devices to provide statistical parameters such as the delay, jitter, and packet loss ratio and require dedicated probe devices. These requirements increase investments on devices.
- NQA can precisely test the network operating status and output statistics without using dedicated probe devices, effectively reducing costs.
- NQA measures network performance and collects statistics on the delay, jitter, and packet loss ratio in real time.

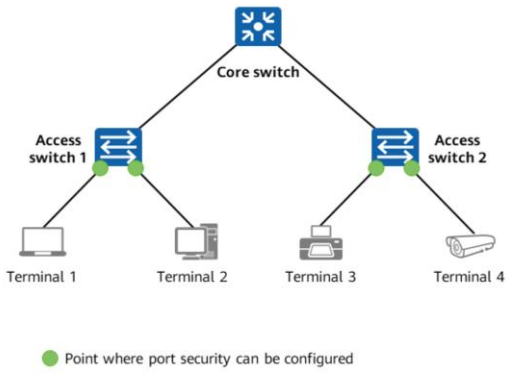
HUAWEI

- NQA measures the performance of various protocols running on networks, which helps users collect statistics about network operation indexes in real time.



- Port isolation provides more secure and flexible networking solutions.
- In some scenarios, data exchanged between terminals in a VLAN needs to be forwarded by an upper-layer device instead of an access switch. This ensures that traffic is forwarded by the upper-layer device instead of the access switch, so that traffic management and control policies can be deployed on the upper-layer device. This mode is called centralized forwarding mode. In this mode, port isolation is configured for all downstream Layer 2 devices. Therefore, proxy ARP should be configured on the gateway. In most cases, intra-VLAN proxy ARP is used.

Ethernet Port Security



Requirements

- An enterprise requires that each access switch interface connected to terminals allow only one PC to access the network so as to limit the number of MAC addresses.
- If an employee attempts to connect a small switch or hub to an interface to increase the number of Internet access interfaces, this behavior should be detected and prohibited.
- Some enterprises may require that a switch forward only data frames sent by terminals with trusted MAC addresses to the upper-layer network. Employees are not allowed to change their locations, that is, change access interfaces of the switch.

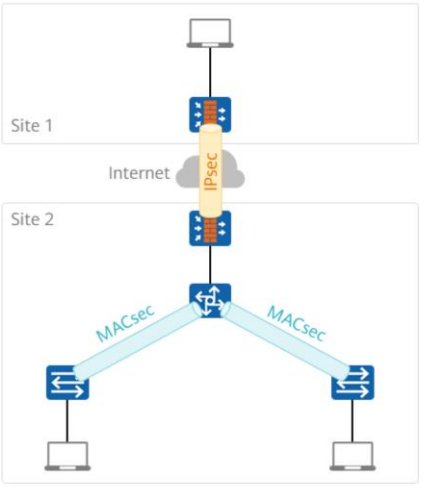
Solution

Port security converts the dynamic MAC addresses learned on an interface into secure MAC addresses (including dynamic and static secure MAC addresses, and sticky MAC addresses). This function prevents unauthorized users from using this interface to communicate with the switch, thereby enhancing device security.

HUAWEI

- Configuring port security on an interface of a switch can limit the number of MAC addresses learned by the interface. Then punishment measures can be taken when a violation occurs.
- The interface configured with port security can convert the learned MAC addresses into secure MAC addresses, preventing devices with other MAC addresses from accessing the network through the interface.

Media Access Control Security (MACsec)



Background

Most data is transmitted in plaintext on LAN links, failing to meet security requirements in scenarios demanding high security.

MACsec overview

Media Access Control Security (MACsec), in compliance with the IEEE 802.1AE standard, defines a method for secure data communication based on the Ethernet. It provides hop-by-hop encryption to secure data transmission.

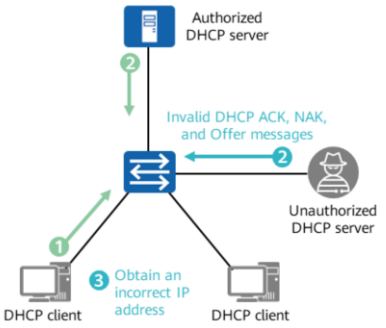
- Data integrity check
- User data encryption
- Data source authenticity verification
- Replay protection

Typical application scenarios

- MACsec is deployed between switches to ensure data security. For example, MACsec is deployed between access switches and uplink aggregation or core switches.
- When transmission devices exist between switches, MACsec can be deployed to ensure data security.

HUAWEI

DHCP Snooping

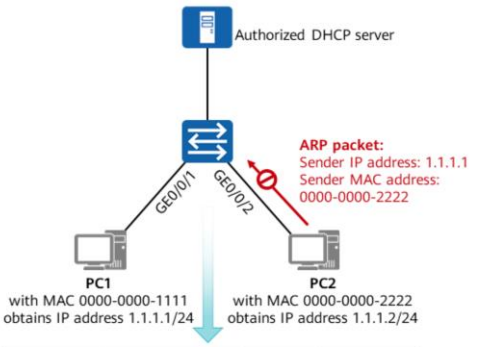


- Some attacks are launched against DHCP on the network, including bogus DHCP server attacks, DHCP server DoS attacks, and bogus DHCP message attacks.
- DHCP snooping ensures that DHCP clients obtain IP addresses from authorized DHCP servers and records mappings between IP addresses and MAC addresses of DHCP clients, preventing DHCP attacks on the network.

HUAWEI

- The DHCP snooping-enabled device forwards DHCP Request messages of users (DHCP clients) to an authorized DHCP server through the trusted interface, and then generates DHCP snooping binding entries based on the DHCP ACK messages received from the DHCP server. When receiving DHCP messages from users through the DHCP snooping-enabled interfaces, the device checks the messages against the binding table, thereby preventing attacks initiated by unauthorized users. In addition, DHCP snooping supports multiple security features, such as limiting the rate for sending DHCP messages.

Dynamic ARP Inspection (DAI)



IP Address	MAC Address	VLAN ID	Interface
1.1.1.1	0000-0000-1111	10	GE0/0/1
1.1.1.2	0000-0000-2222	10	GE0/0/2

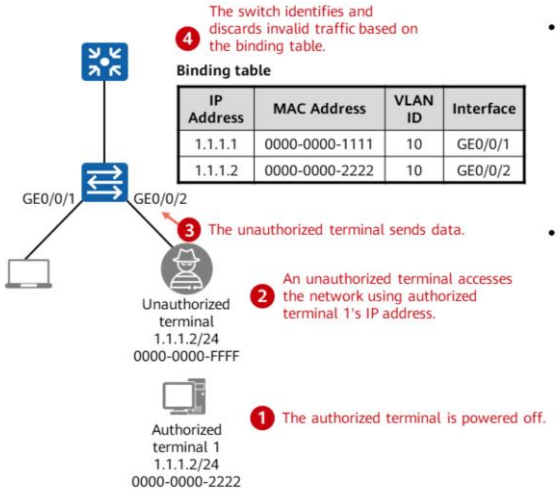
DHCP snooping binding table

- ARP attacks are frequently launched on networks. A man-in-the-middle (MITM) attack is a common ARP spoofing attack.
- To defend against MITM attacks, configure DAI on a switch.
- DAI defends against MITM attacks using a DHCP snooping binding table. When a switch receives an ARP packet, it compares the source IP address, source MAC address, VLAN ID, and interface number of the ARP packet with those in DHCP snooping binding entries.
  - If the ARP packet matches a binding entry, the switch considers the ARP packet valid and allows the packet to pass through.
  - If the ARP packet does not match any binding entry, the switch considers the ARP packet invalid and discards the packet.

HUAWEI

- An MITM attacker establishes independent connections with two parties that intend to communicate and relays messages between them. The two parties consider that they are directly communicating with each other over a private connection, but the entire conversation is in fact controlled by the attacker. In an MITM attack, the attacker can intercept all packets exchanged between the two parties and insert new ones.
- To defend against MITM attacks, configure DAI on a switch.
  - DAI defends against MITM attacks using a DHCP snooping binding table. When the switch receives an ARP packet, it compares the source IP address, source MAC address, VLAN ID, and interface number of the ARP packet with those in DHCP snooping binding entries. If the ARP packet matches a binding entry, the switch considers the ARP packet valid and allows the packet to pass through. If the ARP packet does not match any binding entry, the switch considers the ARP packet invalid and discards the packet.
  - DAI is available only when DHCP snooping is configured. A DHCP snooping-enabled switch automatically generates DHCP snooping binding entries when DHCP users go online. If a user is configured with a static IP address, you need to manually configure a static binding entry for the user.
  - When an attacker connected to the DAI-enabled switch sends bogus ARP packets, the switch detects the attack based on the binding entries and discards the bogus ARP packets. If the packet discarding alarm function is also enabled on the DAI-enabled switch, the switch will generate an alarm when the number of ARP packets discarded due to failure to match any binding entry exceeds the alarm threshold.

IP Source Guard (IPSG)

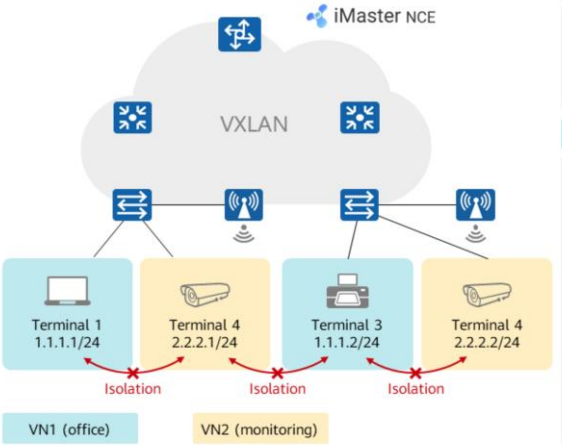


- IPSG implements source IP address filtering based on Layer 2 interfaces. IPSG prevents malicious hosts from simulating authorized hosts using forged IP addresses. In addition, IPSG prevents unauthorized hosts from accessing or attacking networks using forged IP addresses.
- IPSG checks IP packets on Layer 2 interfaces against a binding table that contains the bindings of source IP addresses, source MAC addresses, VLAN IDs, and inbound interfaces. Only packets matching the binding table are forwarded, and other packets are discarded.

HUAWEI

- As networks continue to increase in scale, a growing number of attackers are forging source IP addresses to initiate network attacks (IP address spoofing attacks). Some attackers forge IP addresses of authorized users to obtain network access rights and access networks. As a result, authorized users are unable to access networks or sensitive information may be intercepted.
- IPSG provides a mechanism to effectively defend against IP address spoofing attacks.
    - The binding table can be a static binding table or a dynamic DHCP snooping binding table.
    - IPSG only checks the IP packets from hosts. It does not check non-IP packets such as ARP packets.

Using VRF for VN Isolation



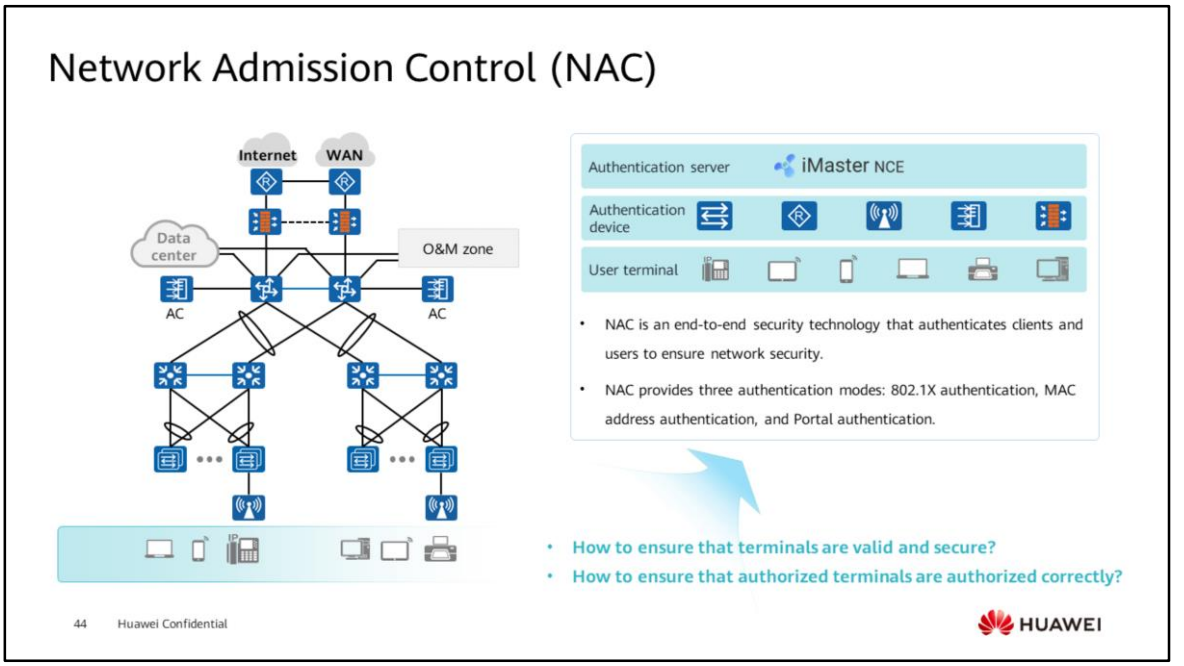
Requirements

- Multiple logical networks, such as the office network and monitoring network, can be built on a simple physical campus network.
- The office network is completely isolated from the monitoring network.

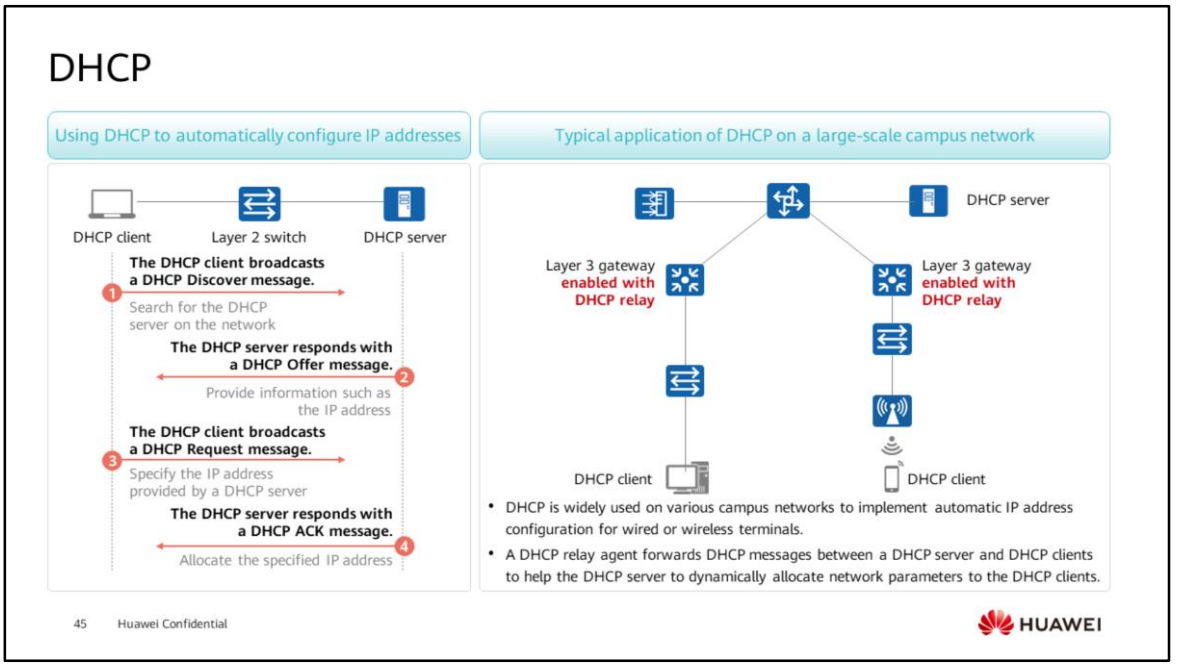
Solution

- Deploying ACLs: ACLs can be used to isolate traffic between VLANs. However, ACL configuration is complex and ACLs do not apply to all scenarios.
- Build independent physical networks: Separate physical networks can be built for office and monitoring purposes, so services can be isolated. However, this increases network construction costs.
- Virtual Routing and Forwarding (VRF): It is also called a VPN instance, which is similar to a virtual device. A VPN instance uses a routing table independent of the root device to completely isolate the VPN instance from the root device, without increasing hardware costs.

HUAWEI

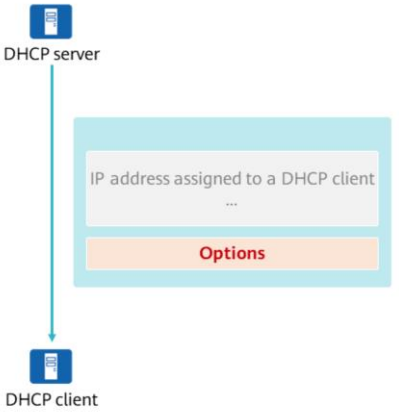


- In addition to using NAC to authenticate access users and control their rights, a campus network also needs to authenticate and control rights of administrators (also called login users) who can log in to devices through FTP, HTTP, SSH, Telnet, or console ports.



- DHCP dynamically configures and uniformly manages IP addresses of hosts. 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.
    - Instead of statically specifying an IP address for a host, DHCP enables a host to obtain an IP address dynamically.
    - DHCP can allocate other configuration parameters, such as the startup configuration file of a client, so that the client can obtain all the required configuration information by using only one message.
    - DHCP supports dynamic and static IP address allocation. A network administrator can select different address allocation modes for hosts as required.
        - 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 or the number of idle IP addresses is less than the total number of hosts that do not require permanent connections.
        - Static allocation: DHCP allocates fixed IP addresses to clients. Compared with manual IP address configuration, DHCP static allocation prevents manual configuration errors and enables unified maintenance and management.
- DHCP has the following benefits:
    - Reduced client configuration and maintenance cost.
    - Centralized management of limited IP addresses.

DHCP Options



Options field in DHCP messages

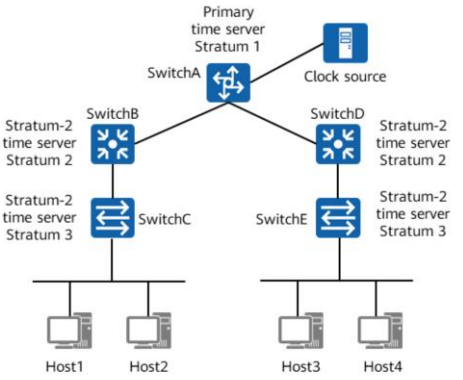
- The Options field stores control information and parameters allocated to a DHCP client.
- A DHCP message can carry multiple options.
- DHCP predefines a large number of options for various purposes, for example, setting the gateway address and DNS server address.

Typical application scenarios

- The Options field can be used to set the IP address of the NTP server.
- The Options field can be vendor-defined. For example, the CloudCampus solution uses Option 148 to implement plug-and-play of network devices. The information is used to notify the DHCP client of controller information so that the device can obtain the controller address and port number.
- Option 43 is used to register an AP with an AC that resides on a different network segment from the AP in WLAN scenarios.

HUAWEI

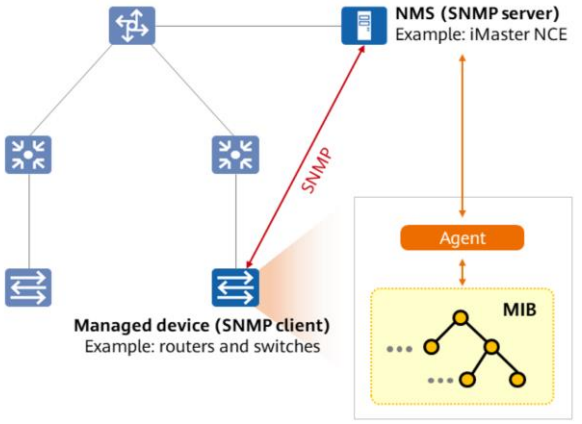
Network Time Protocol (NTP)



- NTP is an application layer protocol in the TCP/IP suite that synchronizes time between time servers and clients. NTP is implemented based on IP and UDP. NTP packets are transmitted over UDP using port 123.
- Key concepts in the NTP network architecture:
  - **Synchronization subnet:** consists of the primary time server, stratum-2 time servers, PC clients, and transmission paths connecting them.
  - **Primary time server:** directly synchronizes its clock with a standard reference clock through a cable or radio. Typically, the standard reference clock is either a radio clock or the Global Positioning System (GPS).
  - **Stratum-2 time server:** synchronizes its clock with either the primary time server or other stratum-2 time servers on the network. Stratum-2 time servers use NTP to transmit time information to other hosts in a LAN.
  - **Stratum:** is a hierarchical standard for clock synchronization. It represents the precision of a clock. The value of a stratum ranges from 1 to 16. A smaller value indicates higher precision. The value **1** indicates the highest clock precision, and the value **16** indicates that the clock is not synchronized.

HUAWEI

SNMP



Overview of SNMP

- SNMP is a network management protocol widely used on TCP/IP networks.
- SNMP provides a method for managing devices through a central computer that runs network management software — known as a network management station (NMS).
- Using in-band management, SNMP achieves efficient and batch network device management. In addition, SNMP enables unified management of different types of network devices from different vendors.

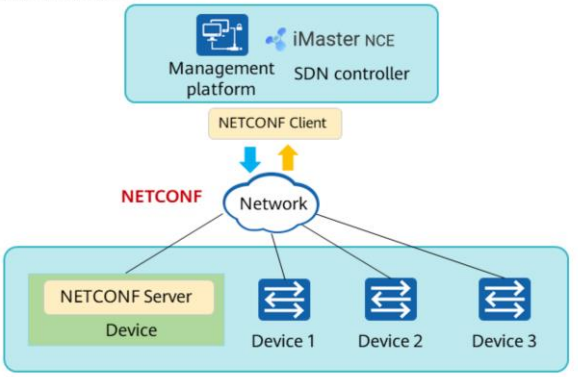
HUAWEI

Network Configuration Protocol (NETCONF)

- NETCONF provides a set of mechanisms for managing network devices. To be specific, users can use NETCONF to add, modify, and delete configurations of network devices, as well as obtain configurations and status of network devices.

NETCONF has three objects:

- NETCONF client
- NETCONF server
- NETCONF message



HUAWEI

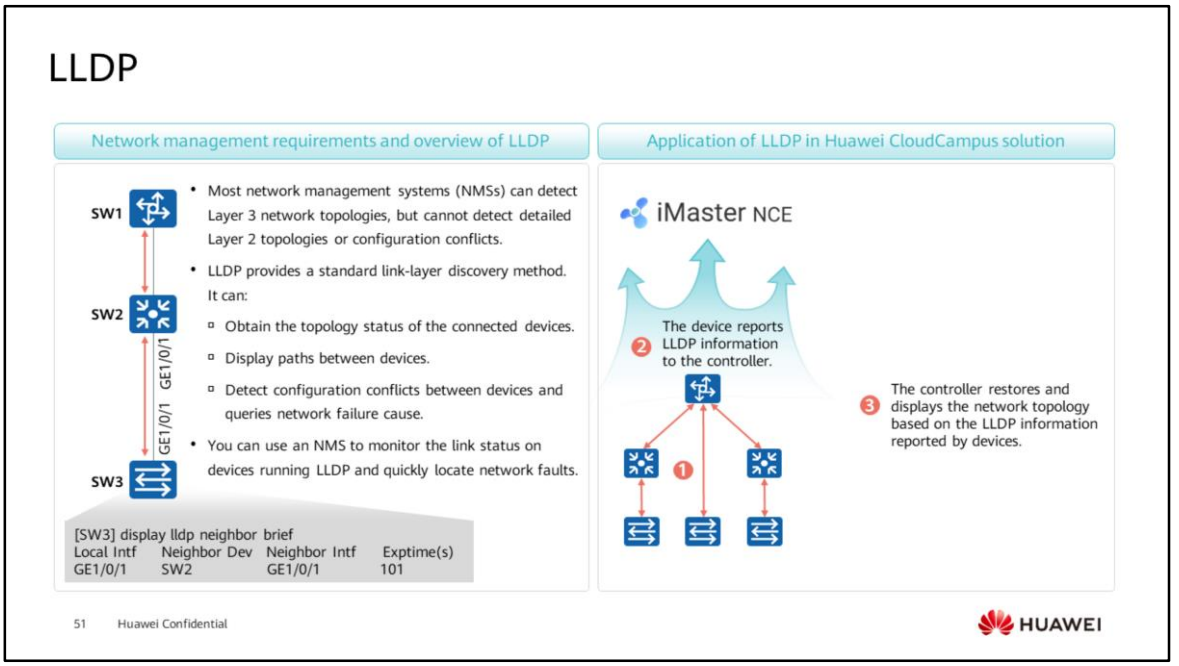
- The plug-and-play of a device is implemented by establishing a NETCONF session between the device and the controller, so that the controller can deliver configurations to the device.

Basic Operations of NETCONF

- NETCONF defines a series of operations:

Scenario	Operation	Function
Query data	<get-config>	Queries configuration data.
	<get>	Queries the current configuration and status data of the device.
Edit data	<edit-config>	Creates, modifies, or deletes configuration data.
Backup/Restoration	<copy-config>	Exports configuration data, or replaces one configuration datastore with another configuration datastore.
	<delete-config>	Deletes the configuration datastore and clears the startup configuration.
Locking/Unlocking	<lock>	Exclusively occupies the permission to modify the configuration datastore.
	<unlock>	Releases the exclusive permission to modify the configuration datastore.
Transactional operations	<commit>	Copies the candidate configuration to the device's running configuration.
	<cancel-commit>	Cancels an ongoing confirmed commit.
	<discard-changes>	Discards uncommitted changes to the candidate configuration.
	<validate>	Checks whether the syntax and semantics of the specified configuration data are correct.
Session operations	<close-session>	Normally ends the NETCONF session.
	<kill-session>	Forcibly terminates other NETCONF sessions. You must have the administrator permission to perform this operation.

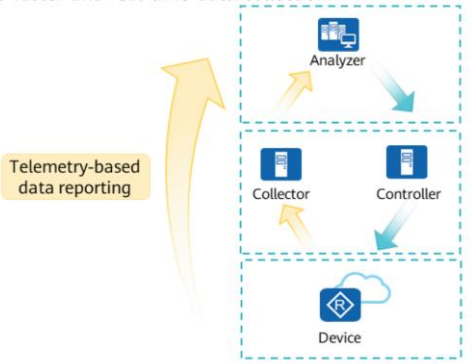
HUAWEI



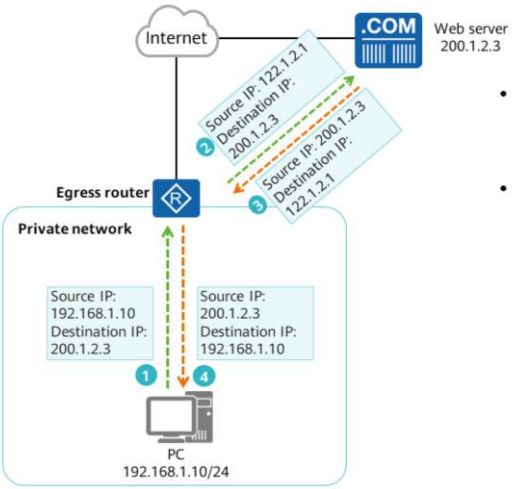
- LLDP is a standard Layer 2 topology discovery protocol defined in IEEE 802.1ab. LLDP collects local device information including the management IP address, device ID, and port ID and advertises the information to neighboring devices. Neighboring devices save the received information in their management information bases (MIBs). The NMS can query required information in MIBs to determine link status.
- As networks grow in scale, network devices are increasing in diversity with complex configurations, posing higher requirements on network management capabilities. Most NMSs can detect Layer 3 network topologies, but they cannot detect detailed Layer 2 topologies or configuration conflicts. To address the network management problems, a standard protocol is required to exchange Layer 2 information between network devices.
- LLDP provides a standard link-layer discovery method. By obtaining Layer 2 information from devices, LLDP allows users to detect the topology of neighboring devices, and display paths between clients, switches, routers, application servers, and network servers. The information also helps 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.

Telemetry

- Telemetry, also called network telemetry, is a technology that remotely collects data from physical or virtual devices at a **high** speed. Devices periodically send information such as interface traffic statistics, CPU usage, and memory usage to collectors in push mode. Compared with the conventional pull mode (request-response interaction), the push mode provides faster and real-time data collection.



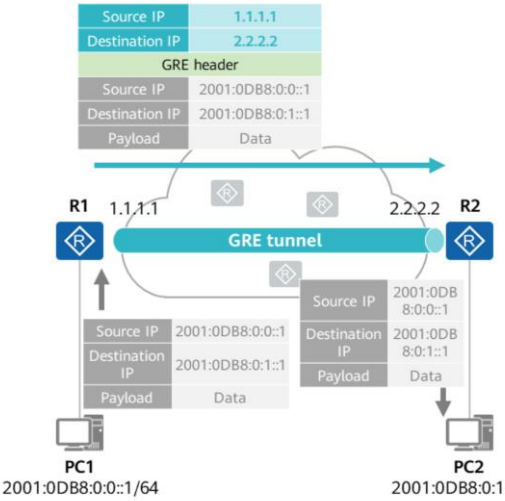
Network Address Translation (NAT)



- NAT technology translates IP addresses in IP data packets. It is widely used on live networks and is usually deployed on network egress devices, such as routers or firewalls.
- Typical application scenarios:
  - When users with private IP addresses attempt to access the Internet, NAT deployed on the egress device of the campus network can translate the source IP addresses of data packets sent from the intranet to the Internet into specific public IP addresses.
  - The campus network provides some services for Internet users. When Internet users access these services using public IP addresses, NAT can translate the destination IP addresses of data packets from the Internet.

HUAWEI

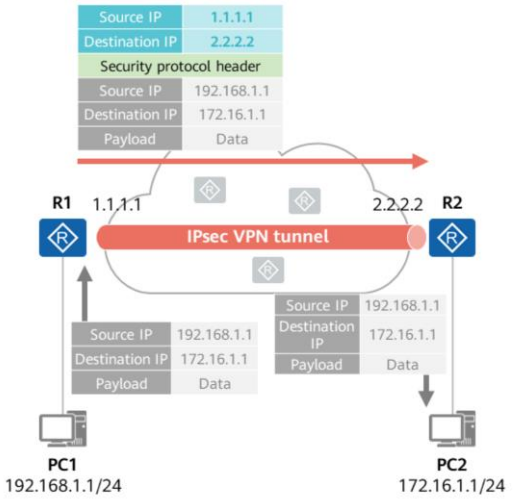
Generic Routing Encapsulation (GRE)



- GRE is a protocol that encapsulates data packets of some network layer protocols, such as IPX, IPv6, and IPv4. The encapsulated data packets can then be transmitted over a network using a different network layer protocol, such as IPv4 and IPv6.
- GRE is a Layer 3 tunneling technology that transparently transmits packets over GRE tunnels. It solves the transmission problems on the network that uses different network-layer protocols.
  - GRE is easy to implement and increases only a few loads on devices at both ends of a tunnel.
  - GRE sets up tunnels over an IPv4 network to connect networks running different protocols, reusing the original network architecture and reducing costs.
  - GRE connects non-contiguous subnets and sets up VPNs to ensure secure connections between the enterprise headquarters and branches.
  - GRE provides only simple password authentication and does not provide the encryption function.

HUAWEI

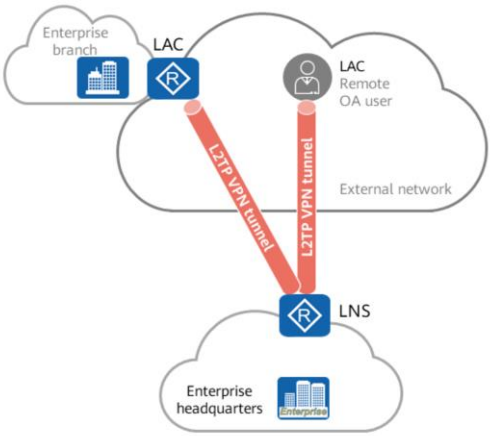
IPsec VPN



- Internet Protocol Security (IPsec) is a set of open network security protocols defined by the Internet Engineering Task Force (IETF) to secure data transmission and reduce the risk of information leakage.
- Through encryption and authentication, IPsec ensures secure service data transmission over the Internet from the following dimensions:
  - Data source authentication
  - Data encryption
  - Data integrity verification
  - Anti-replay

HUAWEI

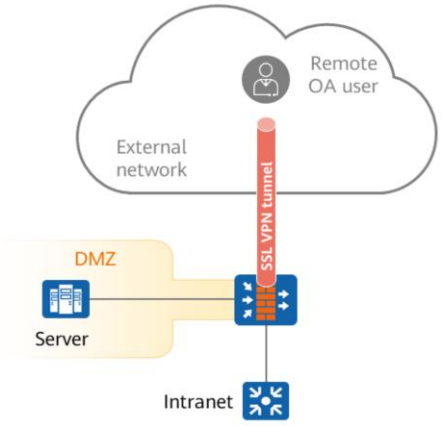
L2TP VPN



- Layer 2 Tunneling Protocol (L2TP) extends the application of the Point-to-Point Protocol (PPP). It is a VPN tunneling protocol that allows traveling employees or enterprise branches to remotely access intranet resources.
- The L2TP network architecture includes an L2TP access concentrator (LAC) and an L2TP network server (LNS).
  - An LAC provides PPP and L2TP processing capabilities, and establishes an L2TP tunnel with the LNS. LACs can be different devices in different networking environments, for example, an LAC can be a gateway or a terminal. An LAC can initiate requests for establishing multiple L2TP tunnels to isolate data flows.
  - An LNS is the peer of an LAC. That is, an L2TP tunnel is established between the LAC and the LNS. The LNS is located at the border between the private network of the enterprise headquarters and public network, and is usually the gateway of the enterprise headquarters.

HUAWEI

SSL VPN



- Secure Shell (SSL) VPN is an SSL-based remote access VPN technology. It allows mobile users (referred to as remote users in SSL VPN) to securely and conveniently access enterprise intranets and intranet resources, improving work efficiency.
- SSL VPN uses the browser/server (B/S) architecture. Remote users can directly use a web browser to access intranet resources securely and quickly, without the need to install extra client software. SSL VPN provides the following functions:
  - Fine-grained permission control based on the type of intranet resources accessed by remote users.
  - Flexible identity authentication.
  - Host check policies that check whether the operating systems, ports, processes, and antivirus software of remote users' devices meet security requirements. It also provides anti-nested remote desktop connection and anti-snapshot functions to eliminate security risks from remote users' devices.

HUAWEI

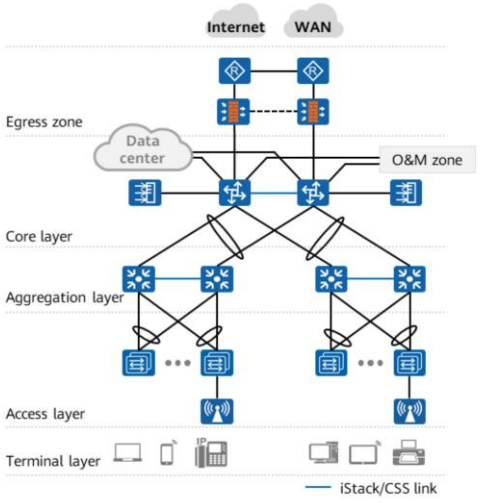
- B/S: Browser/Server

Contents

1. Introduction to Campus Networks
2. Campus Network Challenges and Huawei CloudCampus Solution
3. Typical Campus Network Technologies
4. **Typical Applications of Campus Network Technologies**

HUAWEI

Case 1 - Traditional Campus Network



Network requirements

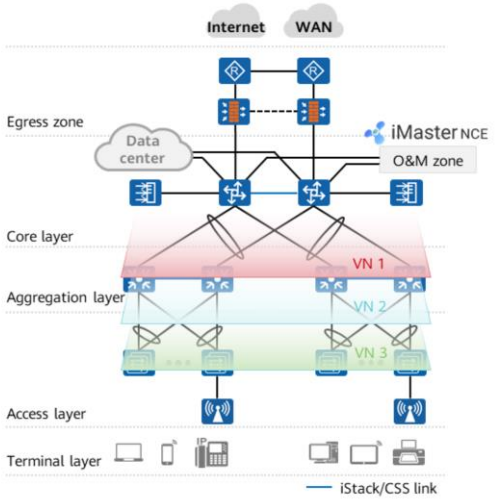
- There are about 3000 wireless and wired terminals in total, so the campus network needs to provide both wireless and wired access services.
- Network segments are divided based on departments and services. The gateways are configured on aggregation switches, and terminal addresses are automatically allocated.
- Network devices are centrally managed and monitored.
- Unified policy control is implemented for access users, and the intranet is protected from attacks from external networks.
- Network reliability must be ensured.

Solution

- Access, aggregation, and core switches are stacked or set up a CSS, and Eth-Trunks are configured on uplinks to ensure reliability.
- The AC + Fit AP wireless networking mode is used, and hot standby is configured between ACs to ensure wireless network reliability. VLANs are configured on access switches to differentiate users on different service network segments, and IP addresses are allocated to these users by the corresponding aggregation switches through DHCP. Single-area OSPF is deployed on aggregation devices and upper-layer devices.
- Access authentication is deployed for all users on the network, and corresponding policies are configured using ACLs.
- Firewalls are deployed in the egress zone and security zones are configured to control traffic.
- SNMP is enabled on all devices on the network, and the network management software is used to manage and maintain network devices in a unified manner.

HUAWEI

Case 2 - Virtualized Campus Network



Network requirements

- Reliability and security requirements similar to those of traditional campus networks.
- Plug-and-play and centralized management of network devices, and flexible provisioning of service configurations and policies.
- Isolation of multiple services on the campus network.
- Access control of terminals, and consistent network access rights and experience for users while they are moving within the campus so long as their identities remain unchanged.

Solution

- iMaster NCE-Campus is used to manage network-wide devices. Protocols such as NETCONF and Telemetry are used to implement unified management and maintenance of network devices as well as quick onboarding of devices.
- VXLAN, BGP EVPN, and VRF technologies are used to divide VNs based on service requirements. iMaster NCE-Campus implements simplified deployment of the underlay and overlay networks as well as VN isolation.
- iMaster NCE-Campus is used to authenticate users on the entire network. Free mobility is deployed to define different user groups or resources on the network as different security groups and implement policy control based on security groups.

HUAWEI

Case 3 - Small- and Medium-Sized Cloud Managed Campus Network

Network requirements

- There are a large number of stores, most of which use wireless networks and have simple network topologies. In addition, onsite personnel do not have network O&M skills.
- There are many types of wireless terminals, including common wireless terminals and IoT terminals such as electronic shelf labels (ESLs).
- The stores want to push product discounts and promotion activity advertisements to customers who are connecting to the Wi-Fi network in the stores.

Solution

- Deployment by scanning barcodes achieves minute-level AP onboarding and quick provisioning of wireless network services. The cloud management platform provides one-stop management of the entire lifecycle (from planning and construction to maintenance and optimization) of all store networks.
- Deploy IoT APs and use the IoT slots built in the IoT APs to implement co-site deployment of IoT and Wi-Fi, unified planning, and shared network for data backhaul. The ESL management system connects to and interacts with the supermarket management and ERP systems to dynamically display real-time price changes and issue out-of-stock warnings.
- Customers access the network through Portal authentication, and can view customized advertisements displayed on the authentication page.

HUAWEI

- ERP: Enterprise Resource Planning

Quiz

1. (Multiple-Answer Question) Which of the following technologies can be used to improve network reliability? ( )
   A. STP
   B. Eth-Trunk
   C. iStack
   D. CSS
2. (True or False) On an MSTP network, devices in the same MST region must be configured with the same VLAN-to-MSTI mappings. ( )
   A. True
   B. False

HUAWEI

1. BCD
2. A

Summary

- This course provides the definition of campus networks, common technologies used in campus networks, and application scenarios of these technologies.
- In subsequent courses, we will discuss how to design networks based on customer requirements, which requires a comprehensive and in-depth understanding of campus network technologies. Due to limited space, this document cannot list every technology used on campus networks. To better understand subsequent courses, you can read related documents for more campus network technologies and a deeper understanding of the key technologies discussed in this course.

HUAWEI

Thank you.

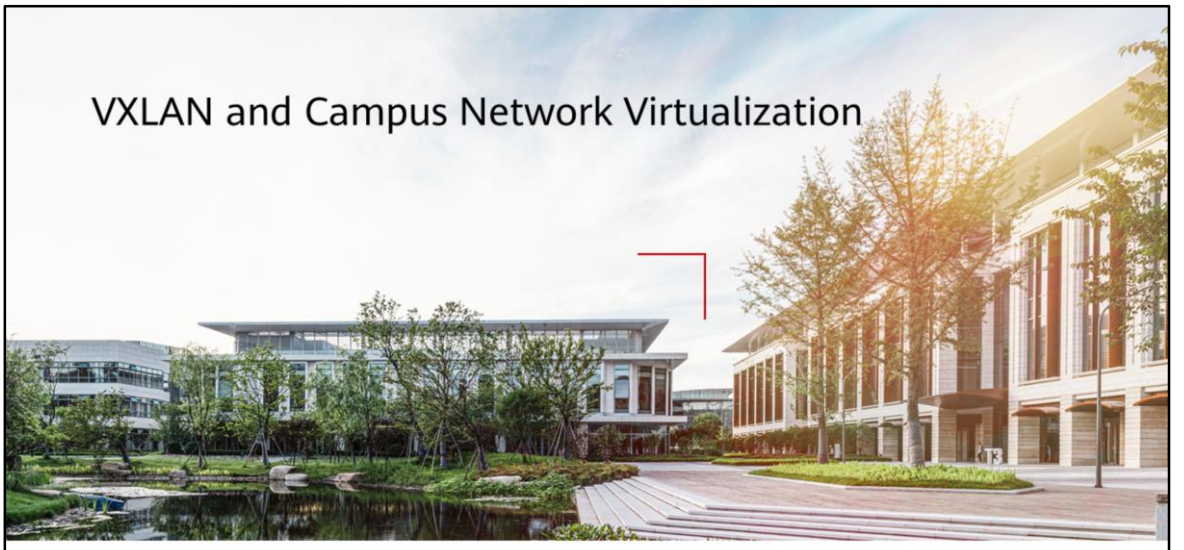
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

VXLAN and Campus Network Virtualization



HUAWEI

Foreword

- Virtual eXtensible Local Area Network (VXLAN), defined in RFC 7348, is a Network Virtualization over L3 (NVO3) technology that uses MAC-in-User Datagram Protocol (MAC-in-UDP) encapsulation.
- VXLAN has been widely used on data center networks (DCNs). As campus networks have increasingly flexible service requirements and growing virtualization and network automation requirements, VXLAN is introduced to campus networks and works with the SDN controller to provide more benefits for customers.
- This course describes basic concepts and fundamentals of VXLAN and how to use VXLAN to build a multi-purpose campus network.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the network requirements of VM live migration in the data center (DC) scenario and how VXLAN meets these requirements.
  - Describe basic concepts of VXLAN.
  - Describe fundamentals of VXLAN.
  - Describe the application of VXLAN in campus network virtualization scenarios.
  - Describe the application of BGP EVPN in campus network virtualization scenarios and how BGP EVPN works with VXLAN.

HUAWEI

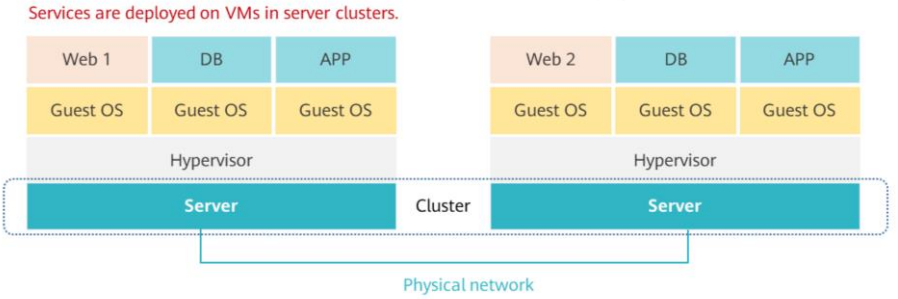
Contents

1. **Overview of VXLAN and Campus Network Virtualization**
2. Basic Concepts and Fundamentals of VXLAN
3. BGP EVPN
4. Campus Network Virtualization

HUAWEI

Technical Background: Virtualization Are Widely Deployed by Enterprises

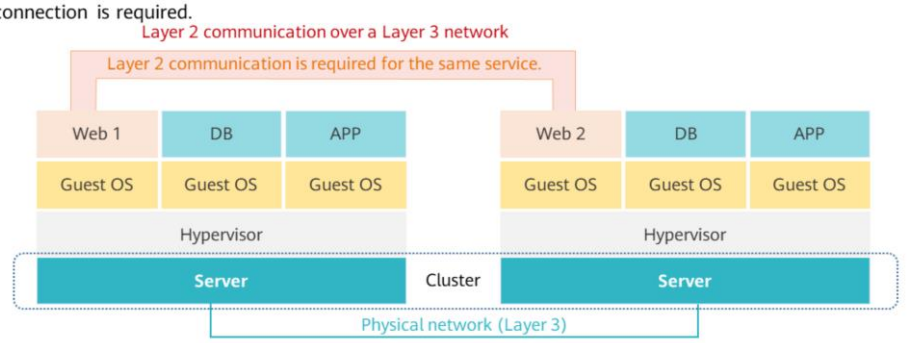
- Virtualization technologies reduce IT costs, improve service deployment flexibility, and reduce O&M costs. More and more enterprises choose to use cloud computing or virtualization technologies in their DCs or campus IT infrastructure.
- After an enterprise chooses the virtualization architecture, services are deployed on VMs in server clusters.



HUAWEI

New Network Requirement: Layer 2 Extension

- VMs in a virtualization or cloud computing cluster can be migrated flexibly. As a result, VMs running the same service (on the same network segment) may run on different servers, or the same VM (with the same IP address) may run on different servers (physical locations) at different times.
- Physical servers may be located in equipment rooms with long geographical distances. Therefore, Layer 3 interconnection is required.

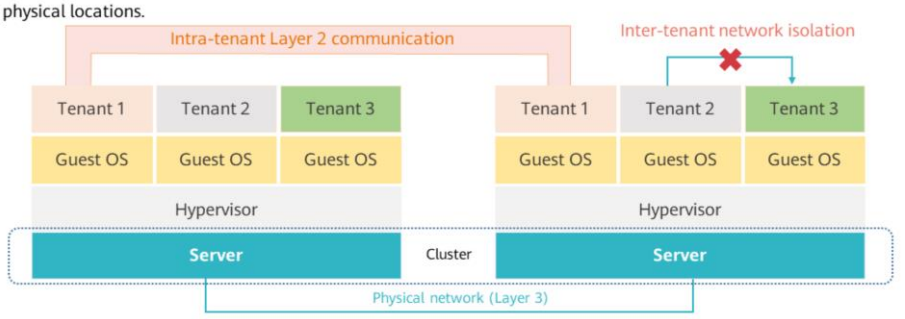


HUAWEI

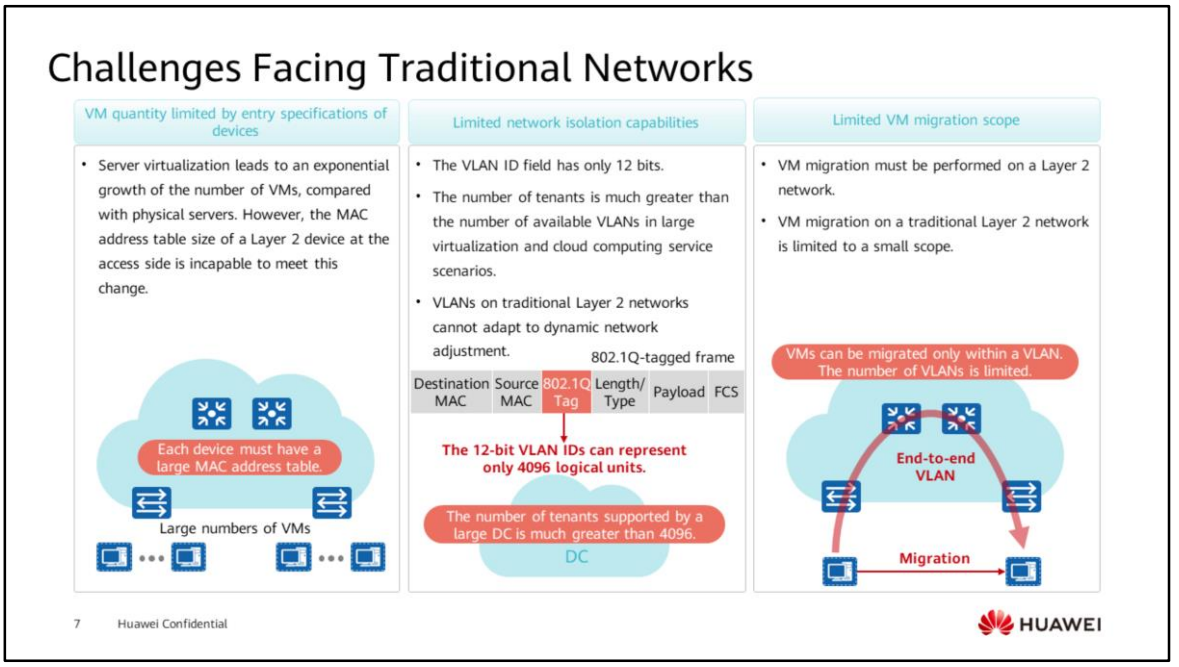
- After servers are virtualized, services are encapsulated in VMs. VMs can be live migrated to any host in a cluster. One of the features of live migration is that the network status does not change. As a result, the IP addresses of service VMs may be in different network locations. Therefore, a large Layer 2 network is required to solve this problem.

New Network Requirement: Multi-Tenant Isolation

- In cloud-based scenarios, multi-tenancy is supported, that is, different tenants share physical resources. This poses two requirements on the network: inter-tenant isolation and intra-tenant communication.
  - Inter-tenant isolation: Tenants may be configured with the same MAC address and IP address. Therefore, physical network isolation needs to be considered. In addition, a large number of potential tenants need to be isolated.
  - Intra-tenant communication: VMs on the same network segment of a tenant require Layer 2 communication, which is irrelevant to physical locations.



HUAWEI



- VXLAN is used to meet requirements of DCNs. On traditional enterprise campus networks, VXLAN is used to construct virtual networks instead of solving some urgent problems.

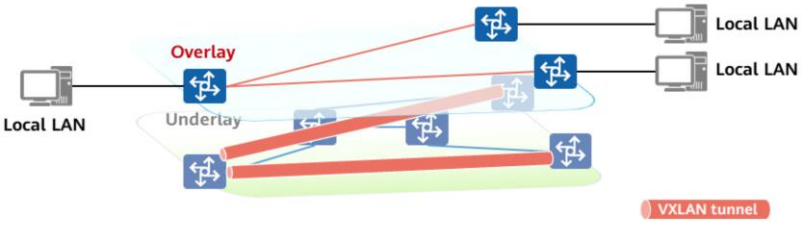
Contents

1. Overview of VXLAN and Campus Network Virtualization
2. **Basic Concepts and Fundamentals of VXLAN**
   - Basic Concepts
   - Fundamentals
   - Basic Configuration
3. BGP EVPN
4. Campus Network Virtualization

HUAWEI

Overview of VXLAN

- VXLAN is essentially a virtual private network (VPN) technology and can be used to build a Layer 2 virtual network over any networks with reachable routes. VXLAN uses VXLAN gateways to implement communication within a VXLAN network and communication between a VXLAN network and a non-VXLAN network.
- VXLAN utilizes MAC-in-UDP encapsulation to extend Layer 2 networks. It encapsulates Ethernet packets into IP packets for these Ethernet packets to be transmitted over routes, without considering the MAC addresses of VMs. In addition, the routed network has no limitation on the network structure and supports large-scale expansion. As such, VM migration is not constrained by the network architecture.

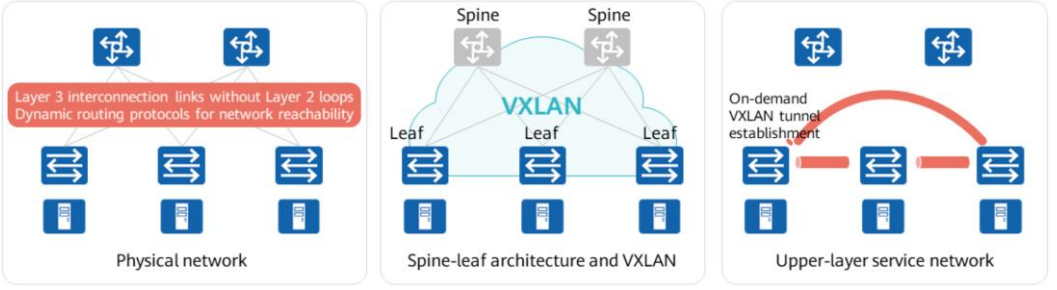


HUAWEI

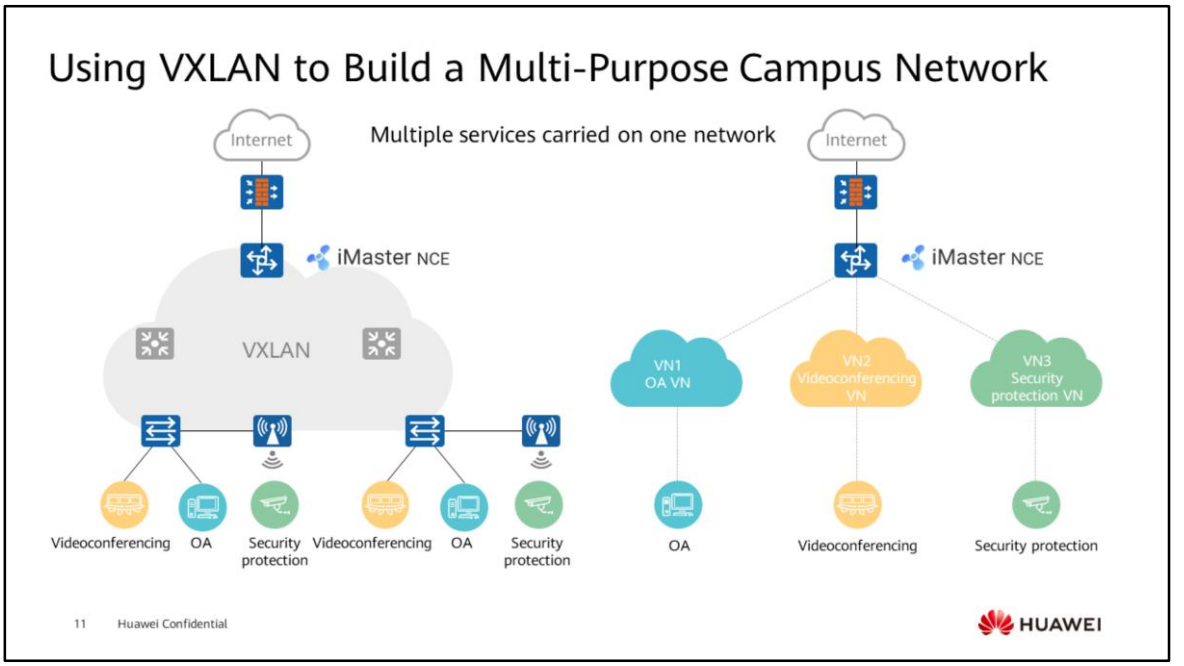
- VXLAN solves the following problems on traditional networks:
  - For VM quantity limited by entry specifications of devices.
    - VXLAN encapsulates original data packets sent from VMs in the same domain into UDP packets, with the IP and MAC addresses used on the physical network in outer headers. Devices on the VXLAN network are aware of only the encapsulated parameters but not the inner data.
    - Except VXLAN edge devices, other devices on the network do not need to identify MAC addresses of VMs. This reduces the burden of learning MAC addresses and improves device performance.
  - For limited network isolation capabilities.
    - VXLAN uses a VXLAN Network Identifier (VNI) field similar to the VLAN ID field to identify users. The VNI field has 24 bits and can identify up to 16 million VXLAN segments, effectively isolating a large number of tenants.
  - For limited VM migration scope.
    - VMs using IP addresses in the same network segment are in a Layer 2 domain logically, even if they are on different physical Layer 2 networks. VXLAN technology constructs a virtual large Layer 2 network over a Layer 3 network.
- Underlay network: a physical network, which serves as the basic layer of the upper-layer logical network.
- Overlay network: a logical network established on the underlay network using tunneling technologies.

Application of VXLAN in DCs

- VXLAN can be applied to a DCN that uses a spine-leaf two-layer physical architecture.
- Spine nodes forward traffic based on routes and are unaware of the VXLAN tunnel during traffic forwarding. Leaf nodes provide network access for device resources such as servers, and perform VXLAN encapsulation and decapsulation.
- All services in the DC are carried by the VXLAN network.



HUAWEI



- The virtualization technology is introduced to create multiple virtual networks (VNs) on a physical network on a campus network. Different VNs are used for different services, such as OA, videoconferencing, and security protection.

Why VXLAN Is Used to Implement Campus Network Virtualization?

Technology	Underlay Network	Overlay Network	Configuration and Management	Carrying User Group Information	Networking Requirements
VLAN	Layer 2	Layer 2	CLI or controller-based	None	Widely supported by devices.
MPLS VPN	Layer 3	Layer 2/Layer 3	Carrier-level technology requirements	None	End-to-end MPLS must be supported.
VXLAN	Layer 2/Layer 3 Built on any physical network	Layer 2/Layer 3	CLI or controller-based	Supported	Nodes at both ends of a tunnel support VXLAN.

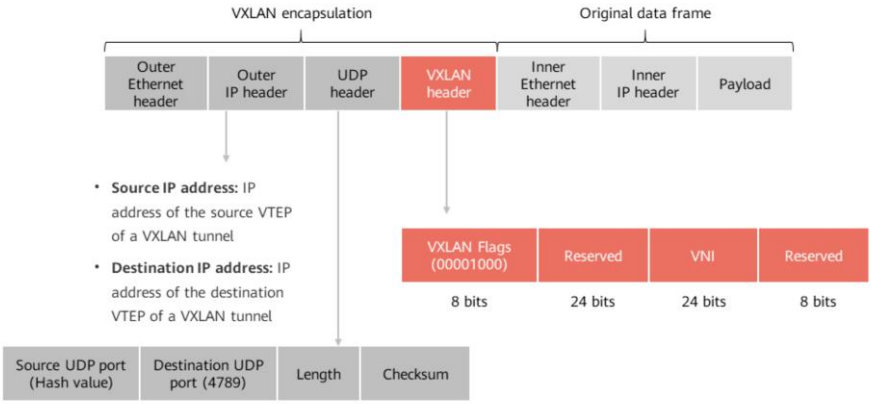
A VXLAN network can be built on any complex Layer 3 network.

Layer 3 virtual network supported

VXLAN works with the SDN controller to implement centralized deployment and automation. There are successful practices.

HUAWEI

VXLAN Packet Format

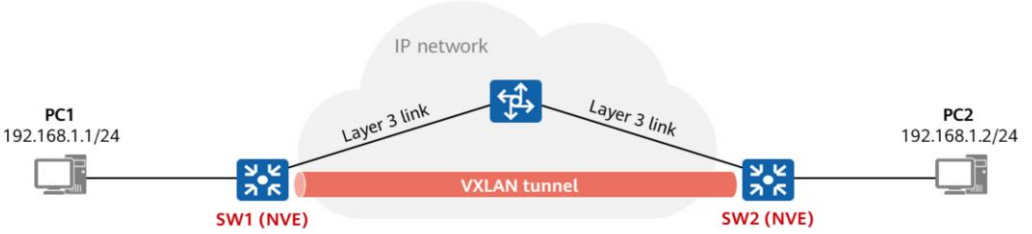


HUAWEI

- The preceding packet format is a standard VXLAN packet format. Huawei CloudEngine S series switches use customize reserved fields based on the standard one.

Basic Concepts of VXLAN: NVE

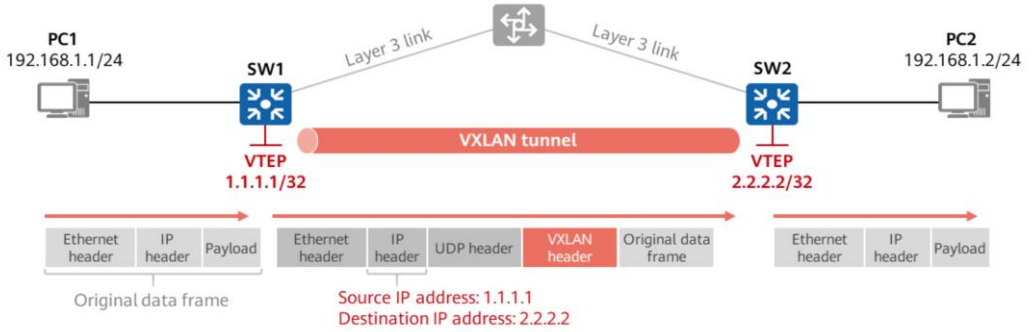
- Network Virtualization Edge (NVE):
  - A network entity that implements network virtualization functions. A physical or software switch can work as an NVE.
  - NVEs run VXLAN and construct a Layer 2 virtual network over a Layer 3 network. SW1 and SW2 in the figure are NVEs.



HUAWEI

Basic Concepts of VXLAN: VTEP

- VXLAN tunnel endpoint (VTEP):
  - A VTEP is located on an NVE and encapsulates and decapsulates VXLAN packets.
  - In the outer IP header of VXLAN packets, the source IP address is the IP address of the source VTEP, and the destination IP address is the IP address of the destination VTEP.

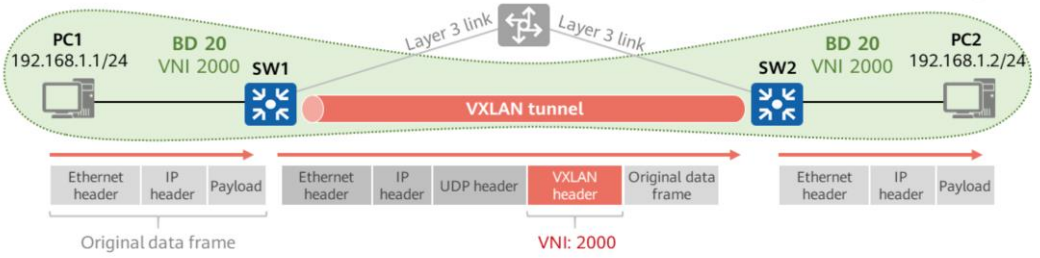


HUAWEI

- A pair of VTEP IP addresses identifies a VXLAN tunnel.
- The source VTEP encapsulates packets and sends the encapsulated packets to the destination VTEP through the VXLAN tunnel. After receiving the encapsulated packets, the destination VTEP decapsulates the packets.
- Generally, the IP address of a loopback interface on a device is used as the VTEP address.

Basic Concepts of VXLAN: VNI and BD

- VXLAN Network Identifier (VNI):
  - VNIs are similar to VLAN IDs and are used to differentiate VXLAN segments. VMs in different VXLAN segments cannot communicate with each other at Layer 2.
  - A tenant can have one or more VNIs. The VNI field has 24 bits.
- Bridge domain (BD):
  - VLANs are used to divide broadcast domains on a traditional network. Similarly, BDs are used to divide broadcast domains on a VXLAN network. A BD identifies a large Layer 2 broadcast domain on a VXLAN network.
  - VNIs are mapped to BDs in 1:1 mode. Terminals in the same BD can communicate with each other at Layer 2.

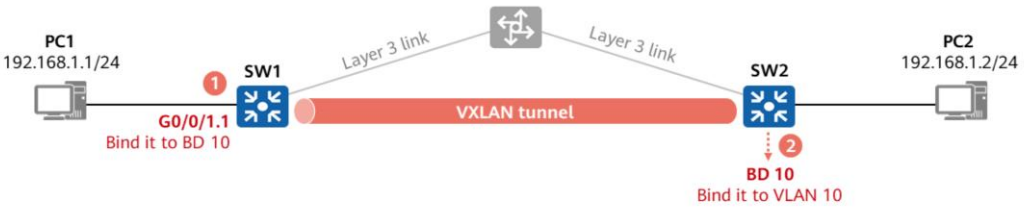


HUAWEI

NVE VTEP VNI and BD VAP Border and Edge Gateway

Basic Concepts of VXLAN: VAP

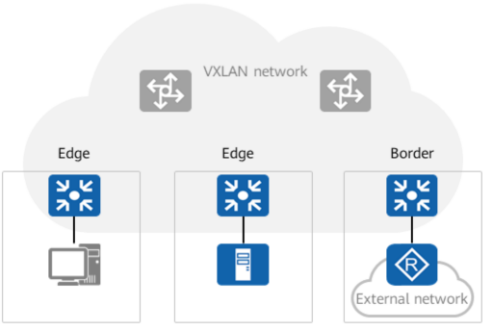
- Virtual access point (VAP):
  - A VAP implements VXLAN service access. A VAP can be configured in Layer 2 sub-interface mode or VLAN binding mode:
    - Layer 2 sub-interface: In this example, a Layer 2 sub-interface is created on SW1 and associated with BD 10, indicating that only specific traffic on the sub-interface is forwarded to BD 10.
    - VLAN binding: In this example, VLAN 10 is configured on SW2 and associated with BD 10, indicating that all traffic from VLAN 10 is forwarded to BD 10.



HUAWEI

- After traffic from a traditional network enters a VXLAN network, a Layer 2 sub-interfaces or VLAN is bound to a BD. A VXLAN VNI is specified in the BD to implement mapping from the traditional VLAN network to the VXLAN network.

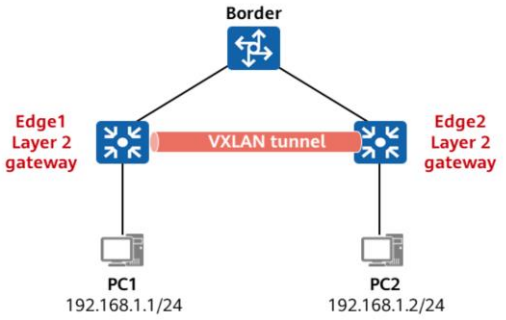
Basic Concepts of VXLAN: Border and Edge Nodes



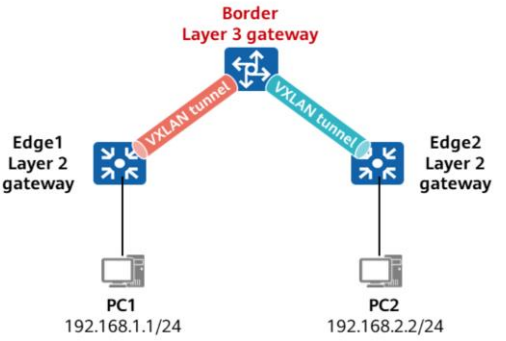
- Edge: an edge access device on a VXLAN network, through which traffic from a traditional network enters the VXLAN network.
- Border: a node for communication between a VXLAN network and an external network. It is used for external traffic to enter the VXLAN network or internal traffic to access the external network. It is usually connected to devices (such as routers and firewalls) that have Layer 3 forwarding capabilities.

HUAWEI

Basic Concepts of VXLAN: Layer 2 and Layer 3 VXLAN Gateways



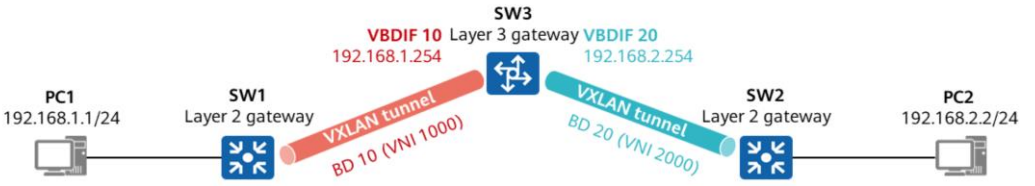
Layer 2 gateway: forwards traffic to a VXLAN network or is used for intra-subnet communication on the same VXLAN network.



Layer 3 gateway: is used for inter-subnet communication on a VXLAN network and allows access to an external network (non-VXLAN network).

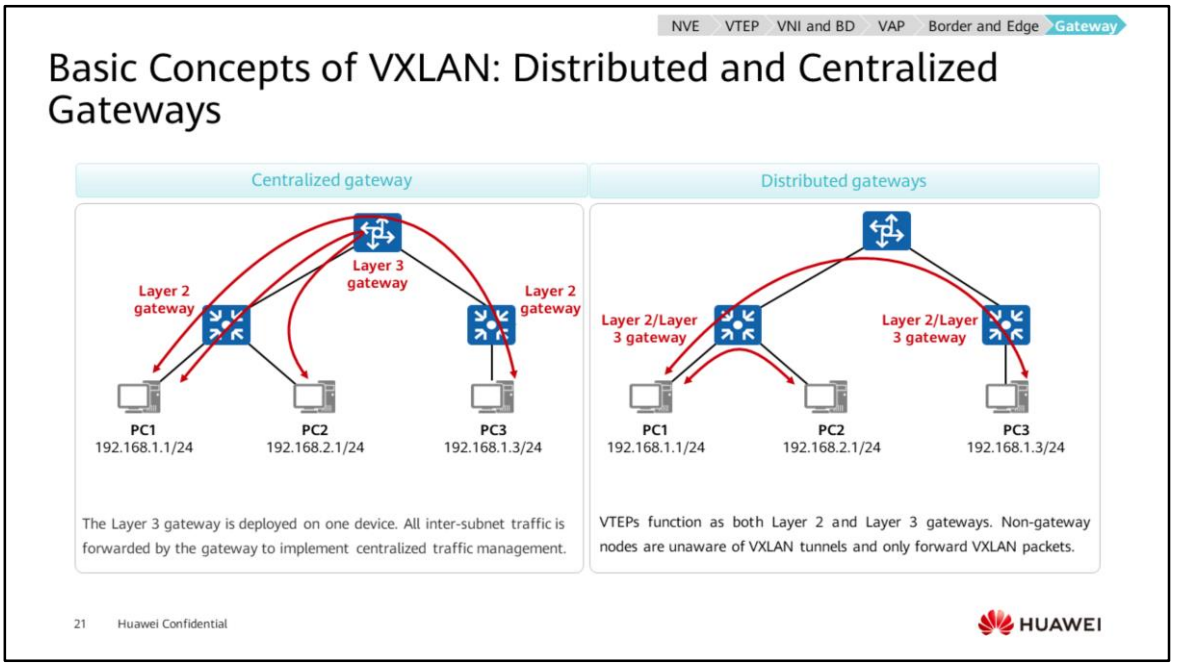
HUAWEI

Basic Concepts of VXLAN: VBDIF Interface



- VLANIF interfaces are used for communication between broadcast domains on a traditional network. Similarly, VBDIF interfaces are used for communication between BDs on a VXLAN network.
- A VBDIF interface is a Layer 3 logical interface created for a BD on a Layer 3 VXLAN gateway.
- IP addresses can be configured for VBDIF interfaces to implement communication between different VXLAN segments and between VXLAN and non-VXLAN networks, and to connect a Layer 2 network to a Layer 3 network.

HUAWEI



- Centralized gateway:
  - Advantage: Inter-subnet traffic is managed in a centralized manner, simplifying gateway deployment and management.
  - Disadvantage: The forwarding path is not optimal. The ARP entry specification is a bottleneck. Because a centralized Layer 3 gateway is deployed, ARP entries must be generated on the gateway for terminals whose traffic is forwarded through the gateway.
- Distributed gateways:
  - Advantage: VTEPs only need to learn ARP entries of terminals connected to them, which eliminates the ARP entry specification bottleneck in the centralized Layer 3 gateway scenario and improves the network scalability.
  - Disadvantage: Gateway deployment in this scenario is more complex than that in the centralized gateway scenario.

Contents

1. Overview of VXLAN and Campus Network Virtualization
2. **Basic Concepts and Fundamentals of VXLAN**
   - Basic Concepts
   - Fundamentals
   - Basic Configuration
3. BGP EVPN
4. Campus Network Virtualization

HUAWEI

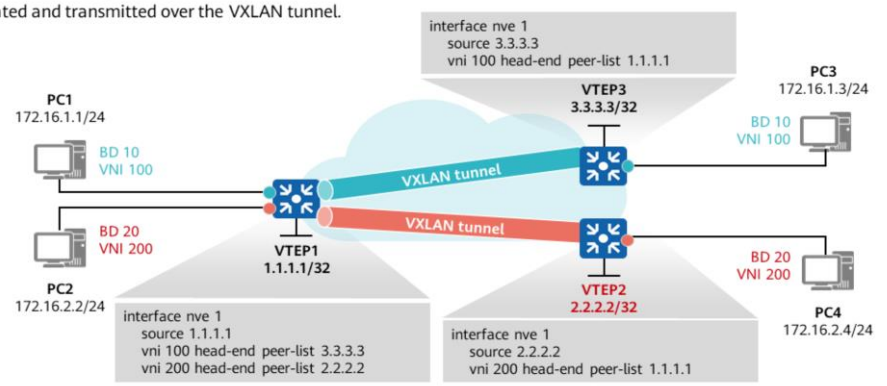
VXLAN Tunnel Establishment Modes

- A VXLAN tunnel is identified by a pair of VTEP IP addresses. Packets are encapsulated on VTEPs and then transmitted based on routes in the VXLAN tunnel. After a VXLAN tunnel is configured, it can be established successfully as long as the VTEPs at both ends of the tunnel have reachable routes to each other's IP address at Layer 3.
- VXLAN tunnels are classified into the following types based on the VXLAN tunnel creation mode:
  - Static tunnel: You need to manually configure local and remote VNIs, VTEP IP addresses, and ingress replication list.
  - Dynamic tunnel: VXLAN tunnels are dynamically established using BGP Ethernet Virtual Private Network (Ethernet VPN). When BGP EVPN is used to dynamically establish a VXLAN tunnel, the local and remote VTEPs first establish a BGP EVPN peer relationship before exchanging BGP EVPN routes to learn the VNIs and VTEP IP addresses from each other.

HUAWEI

Static VXLAN Tunnel

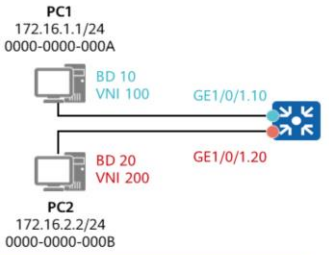
- A static VXLAN tunnel is not a stateful tunnel (for example, an IPsec VPN tunnel). It is similar to a GRE VPN tunnel for data transmission.
- The address in the ingress replication list is the tunnel destination address that is encapsulated when VXLAN packets are encapsulated and transmitted over the VXLAN tunnel.



HUAWEI

VXLAN MAC Address Entries

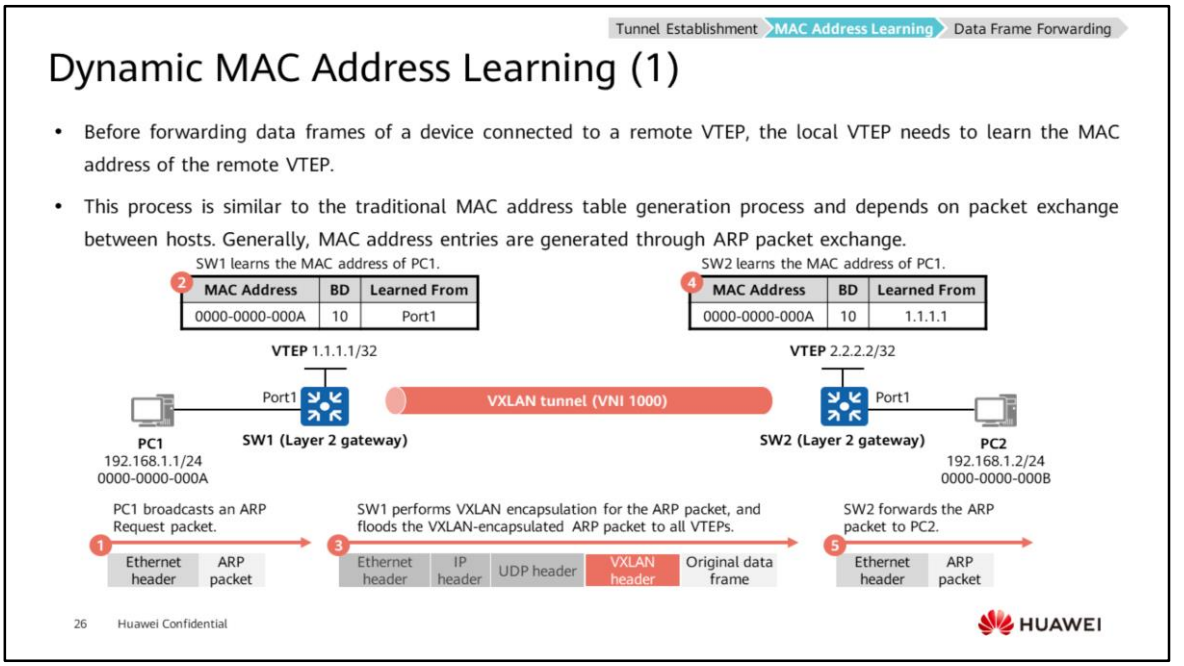
- VXLAN implements Layer 2 forwarding on the overlay network. Unicast data frames are forwarded based on MAC address entries.
- When a VTEP receives a data frame from the local device in a BD, the VTEP adds the source MAC address of the data frame to the MAC address table of the BD and sets the outbound interface to the interface that receives the data frame.
- This entry is used to guide the forwarding of data frames sent to a device connected to the VTEP.



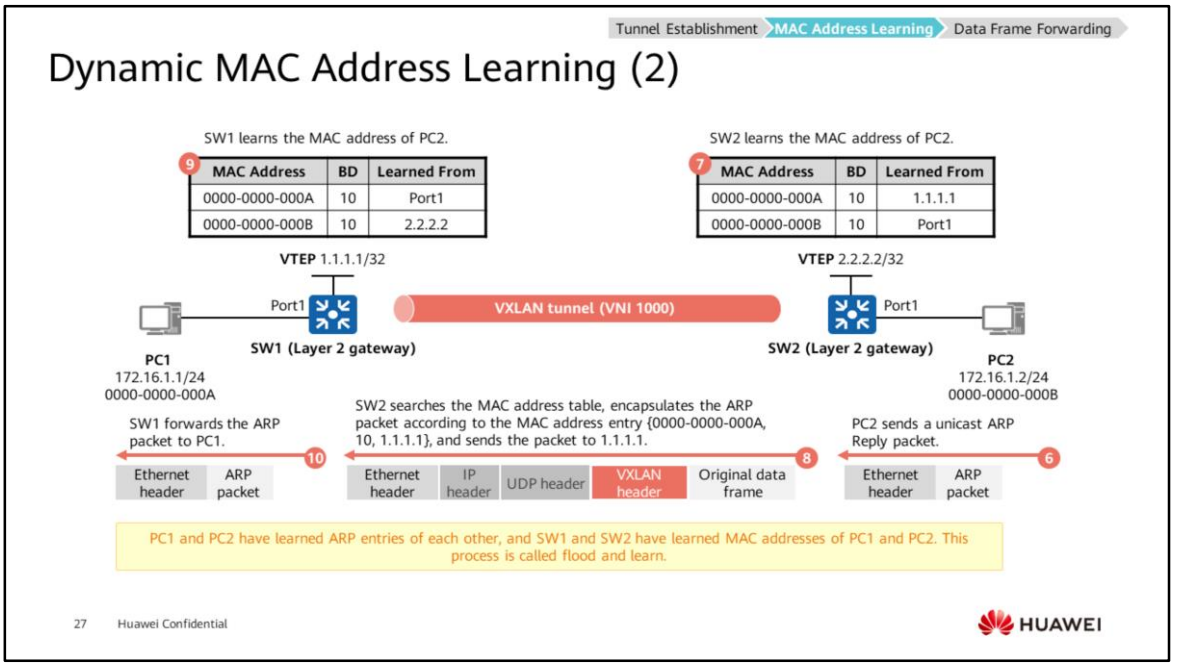
How are data frames forwarded to a device connected to a remote VTEP?

```
<S1>display mac-address bridge-domain 10
--------------------------------------------------------------------------------
MAC Address    VLAN/VSI/BD   Learned-From        Type
--------------------------------------------------------------------------------
0000-0000-000a -/-/10        GE1/0/1.10          dynamic

<S1>display mac-address bridge-domain 20
--------------------------------------------------------------------------------
MAC Address    VLAN/VSI/BD   Learned-From        Type
--------------------------------------------------------------------------------
0000-0000-000b -/-/20        GE1/0/1.20          dynamic
```


HUAWEI



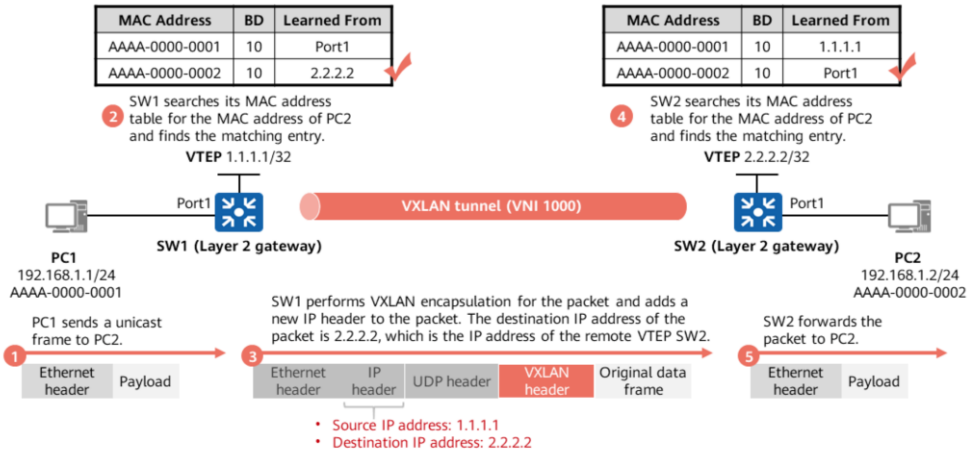
- Communication between PC1 and PC2 is as follows:
    1. To communicate with PC2, PC1 broadcasts an ARP Request packet to obtain the MAC address of PC2.
    2. After receiving the packet, SW1 determines the BD ID, destination VXLAN tunnel, and VNI of the traffic based on VAP information. In addition, SW1 learns the MAC address of PC1 and records the BD ID and the interface that receives the packet in the corresponding MAC address entry.
    3. SW1 performs VXLAN encapsulation for the ARP Request packet and forwards the encapsulated packet based on the ingress replication list.
    4. After receiving the VXLAN packet, SW2 decapsulates the packet to obtain the original data frame. In addition, SW2 learns the MAC address of PC1 and records the BD ID and the VTEP address of SW1 in the corresponding MAC address entry.
    5. SW2 floods the ARP packet in the local BD. PC2 then receives the packet and learns the ARP information of PC1.



6. PC2 sends a unicast ARP Reply packet.
7. SW2 has the MAC address entry of PC1; therefore, SW2 unicasts the packet and learns the source MAC address of PC2 in the MAC address entry.
8. SW2 encapsulates the ARP Reply packet with a VXLAN header and sends it to the remote VTEP at 1.1.1.1.
9. After SW1 receives the VXLAN packet, it decapsulates the packet and records the source MAC address of PC2 in the MAC address table. The outbound interface is the remote VTEP.
10. SW1 forwards the packet to PC1.

- By doing this, PC1 and PC2 learn ARP entries of each other, and SW1 and SW2 learn MAC addresses of each other.

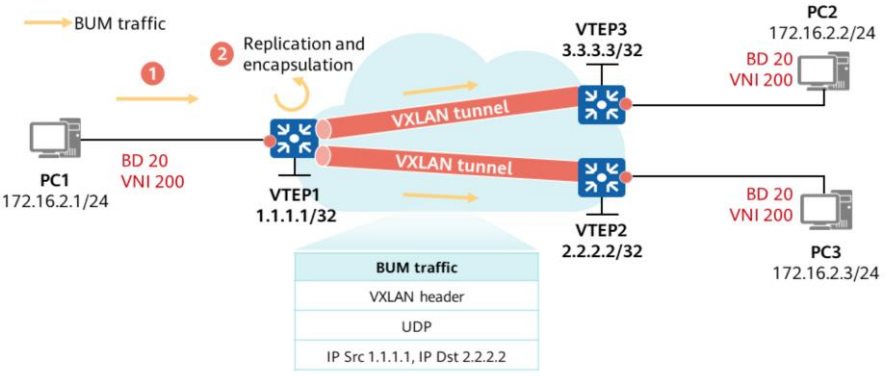
Intra-Subnet Forwarding of Unicast Packets with Known Destination Addresses



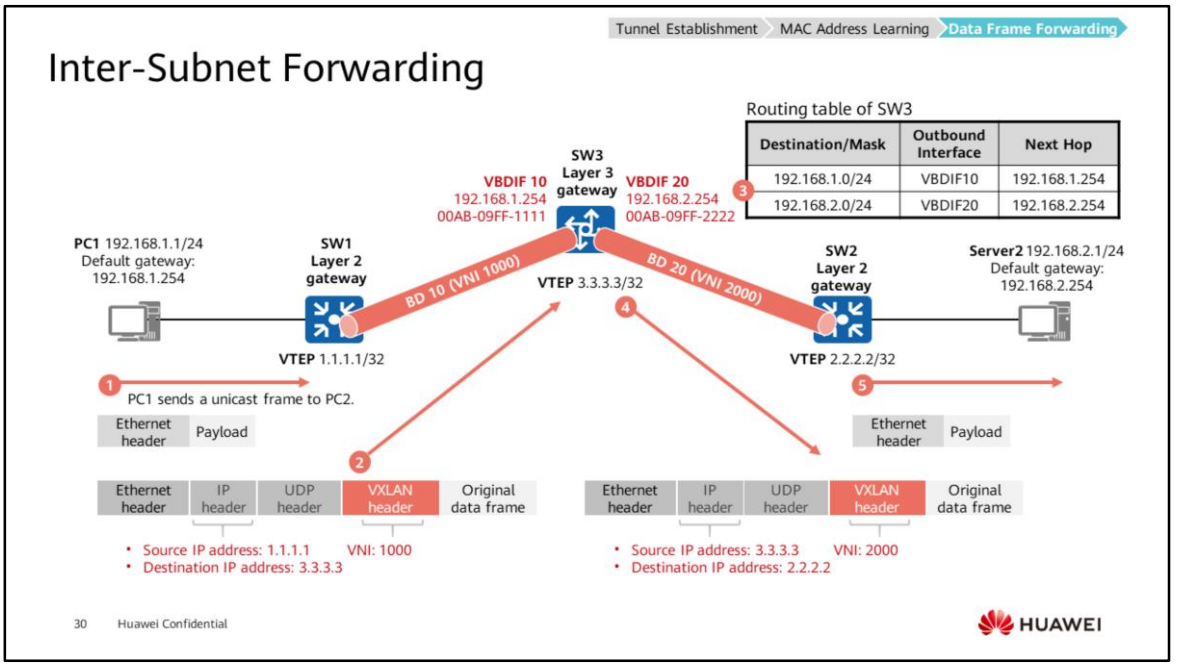
HUAWEI

BUM Traffic Forwarding

- When transmitting broadcast, unknown, and multicast (BUM) traffic, the VTEP replicates the traffic and sends multiple copies to the remote VTEP in the ingress replication list. In this way, BUM traffic is flooded on the overlay network.



HUAWEI



- PC1 wants to communicate with Server2. When finding that Server2 is on a different subnet, PC1 sends the packet to the gateway.
- PC1 sends a data packet to Server2. The destination MAC address of the data packet is 00AB-09FF-1111 (gateway MAC address). After receiving the data packet, SW1 searches the Layer 2 forwarding table and finds that the outbound interface is a remote VTEP (Layer 3 gateway). Therefore, SW1 adds a VXLAN header (VNI = 1000) to the data packet. Then the packet is sent to SW3.
- After receiving the packet, SW3 decapsulates the VXLAN packet and finds that the destination MAC address of the internal original data packet is 00AB-09FF-1111, which is the MAC address of its own interface VBDIF10. Then SW3 needs to search the Layer 3 forwarding table.
- SW3 searches the routing table and finds that the destination IP address 192.168.2.1 matches the direct route generated by VBDIF 20. SW3 then searches the ARP table for the destination MAC address of the packet and searches the MAC address table for the outbound interface of the packet. On SW3, the outbound interface for the MAC address corresponding to 192.168.2.1 is the remote VTEP at 2.2.2.2. SW3 encapsulates the packet into a VXLAN packet and sends it to SW2.
- After receiving the packet, SW2 decapsulates the VXLAN packet and finds that the destination MAC address is not the MAC address of any local interface. SW2 then searches the Layer 2 forwarding table and forwards the packet through the local interface based on the MAC address table.

Contents

1. Overview of VXLAN and Campus Network Virtualization
2. **Basic Concepts and Fundamentals of VXLAN**
   - Basic Concepts
   - Fundamentals
   - Basic Configuration
3. BGP EVPN
4. Campus Network Virtualization

HUAWEI

VXLAN Configuration (1)

1. Create a BD.

   [Huawei] **bridge-domain** *bd-id*

   Create a BD in the system view and enter the BD view. The value ranges from 1 to 16777215.

2. Create a VNI.

   [Huawei-bd100] **vxlan vni** *vni-id*

   Create a VNI in the BD view and associate the VNI with the BD. The value ranges from 1 to 16777215.

3. Create an NVE interface.

   [Huawei] **interface nve** *nve-number*

   Create an NVE interface and enter its view. The NVE interface number must be 1.

4. Configure an IP address for the source VTEP.

   [Huawei-Nve1] **source** *ip-address*

   Configure an IP address for the source VTEP. A loopback interface's address is recommended.

HUAWEI

VXLAN Configuration (2)

5. Configure an ingress replication list.

   [Huawei-Nve1] **vni** *vni-id* **head-end peer-list** *ip-address*&<1-10>

   Configure a VNI in the NVE view and specify the IP addresses of multiple remote VTEPs to create an ingress replication list. The local NVE replicates and forwards packets based on the list.

6. Configure an encapsulation mode on a sub-interface.

   [Huawei-GE1/0/1.1] **encapsulation** { **dot1q** [**vid** *low-vid* [**to** *high-vid*]] | **default** | **untag** | **qinq [vid** *id* **]}**

   Configure an encapsulation mode to determine the type of data packet that can pass through the Layer 2 sub-interface.

7. Bind a VLAN to the BD.

   [Huawei-bd] **l2 binding vlan** *vlan-id*

   This command is configured in the BD view. Before binding a global VLAN to a BD, ensure that the global VLAN has been created and interfaces have been added to the global VLAN.

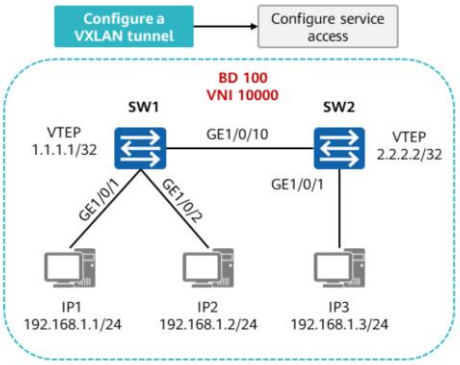
8. Configure a Layer 3 gateway.

   [Huawei] **interface vbdif** *bd-id*

   Create a VBDIF interface in the system view, enter the VBDIF interface view, and configure a gateway IP address in the VBDIF interface view.

HUAWEI

Configuration Example: Intra-Subnet Communication (1)



- Configure a VXLAN tunnel between the two switches to enable the three PCs on the same network segment to communicate with each other.

SW1 configuration:

```
[SW1] bridge-domain 100
[SW1-bd100] vxlan vni 10000
```

```
[SW1] interface Nve 1
[SW1-Nve1] source 1.1.1.1
[SW1-Nve1]vni 10000 head-end peer-list 2.2.2.2
```

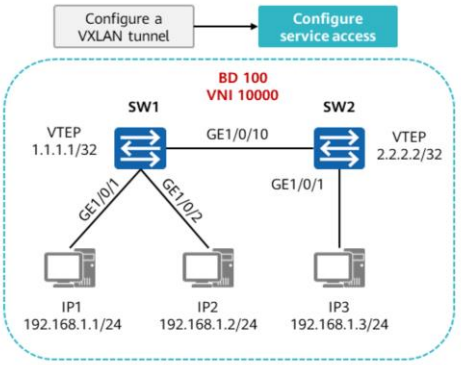
SW2 configuration:

```
[SW2] bridge-domain 100
[SW2-bd100] vxlan vni 10000
```

```
[SW2] interface Nve 1
[SW2-Nve1] source 2.2.2.2
[SW2-Nve1]vni 10000 head-end peer-list 1.1.1.1
```


HUAWEI

Configuration Example: Intra-Subnet Communication (2)



- Configure a VXLAN tunnel between the two switches to enable the three PCs on the same network segment to communicate with each other.

SW1 configuration:

```
[SW1]interface GigabitEthernet 1/0/1.1 mode l2
[SW1-GigabitEthernet1/0/1.1]encapsulation untag
[SW1-GigabitEthernet1/0/1.1]bridge-domain 100
[SW1]interface GigabitEthernet 1/0/2.1 mode l2
[SW1-GigabitEthernet1/0/2.1]encapsulation untag
[SW1-GigabitEthernet1/0/2.1]bridge-domain 100
```

SW2 configuration:

```
[SW2] bridge-domain 100
[SW1-bd100] l2 binding vlan 1
```

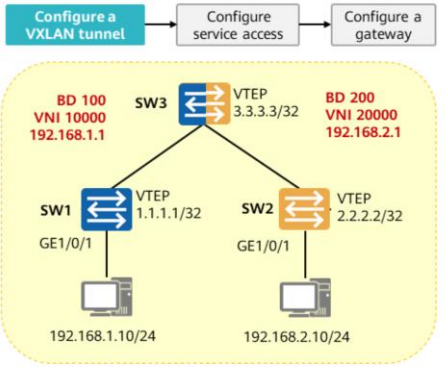
Question 1: What is the difference between access modes of SW2 and SW1?

Question 2: How is the gateway 192.168.1.254 configured?

HUAWEI

1. SW1 provides Layer 2 sub-interface access, and SW2 uses the VLAN binding mode.
2. Create a VBDIF interface as the gateway for terminals in the BD.

Configuration Example: Inter-subnet Communication (Centralized Gateway) (1)



- VXLAN tunnels are configured between the three switches to implement mutual access between two PCs on different network segments. The centralized VXLAN gateway is deployed on SW3.

SW1 configuration:

```
[SW1] bridge-domain 100
[SW1-bd100] vxlan vni 10000

[SW1] interface Nve 1
[SW1-Nve1] source 1.1.1.1
[SW1-Nve1]vni 10000 head-end peer-list 3.3.3.3
```

SW2 configuration:

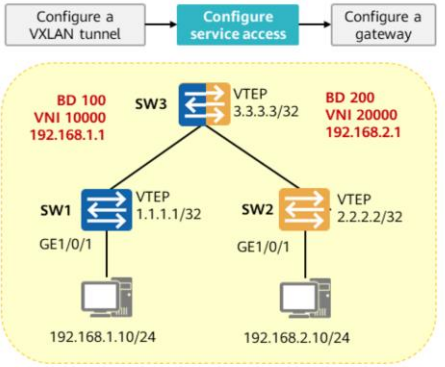
```
[SW2] bridge-domain 200
[SW2-bd100] vxlan vni 20000

[SW2] interface Nve 1
[SW2-Nve1] source 2.2.2.2
[SW2-Nve1] vni 20000 head-end peer-list 3.3.3.3
```

Question: How is SW3 configured?

HUAWEI

- On SW3, configure NVE interfaces to connect to SW1 and SW2, and create VBDIF 100 and VBDIF 200 as gateways of terminals in BD 100 and BD 200.

Configuration Example: Inter-subnet Communication (Centralized Gateway) (2)



- VXLAN tunnels are configured between the three switches to implement mutual access between two PCs on different network segments. The centralized VXLAN gateway is deployed on SW3.

SW1 configuration:

```
[SW1] interface GigabitEthernet 1/0/1.1 mode l2
[SW1-GigabitEthernet1/0/1.1] encapsulation untag
[SW1-GigabitEthernet1/0/1.1] bridge-domain 100
```

SW2 configuration:

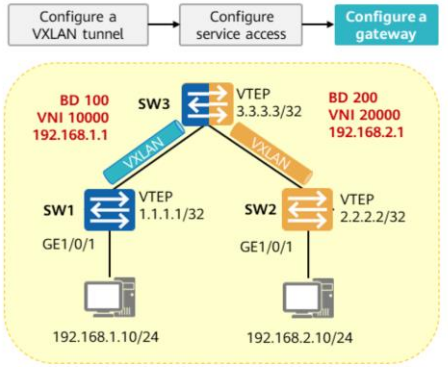
```
[SW1] interface GigabitEthernet 1/0/1.1 mode l2
[SW1-GigabitEthernet1/0/1.1] encapsulation untag
[SW1-GigabitEthernet1/0/1.1] bridge-domain 200
```

If traffic from 192.168.2.10 carries VLAN 20, how do we configure SW2?

HUAWEI

- Configure GE1/0/1 of SW2 as a trunk interface to allow packets from VLAN 20 to pass through (you can also configure the interface as an access interface and the default VLAN as VLAN 20), and bind VLAN 20 to BD 200.

Configuration Example: Inter-subnet Communication (Centralized Gateway) (3)



- VXLAN tunnels are configured between the three switches to implement mutual access between two PCs on different network segments. The centralized VXLAN gateway is deployed on SW3.

SW3 configuration:

```
[SW3] bridge-domain 100
[SW3-bd100] vxlan vni 10000
```

```
[SW3] bridge-domain 200
[SW3-bd200] vxlan vni 20000
```

```
[SW3] interface Nve 1
[SW3-Nve1] source 3.3.3.3
[SW3-Nve1]vni 10000 head-end peer-list 1.1.1.1
[SW3-Nve1]vni 20000 head-end peer-list 2.2.2.2
```

```
[SW3] interface Vbdif100
[SW3-Vbdif100] ip address 192.168.1.1 24
```

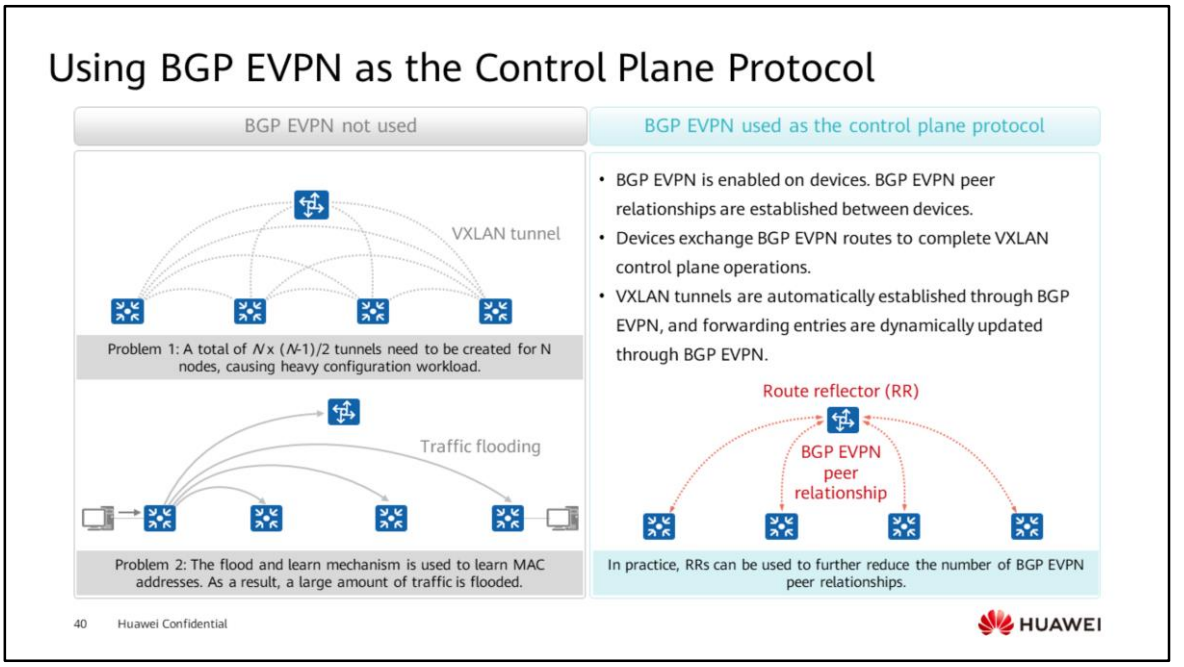
```
[SW3] interface Vbdif200
[SW3-Vbdif200] ip address 192.168.2.1 24
```


HUAWEI

Contents

1. Overview of VXLAN and Campus Network Virtualization
2. Basic Concepts and Fundamentals of VXLAN
3. **BGP EVPN**
   - Basic Concepts
   - BGP EVPN Routes
   - BGP EVPN Features
4. Campus Network Virtualization

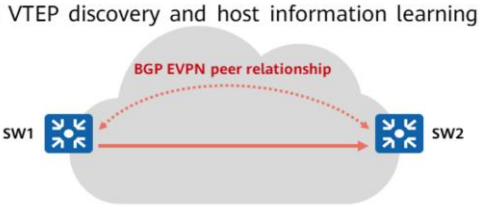
HUAWEI



- The static VXLAN solution does not have a control plane. VTEP discovery and learning of host information (including IP addresses, MAC addresses, VNIs, and gateway VTEP IP addresses) are performed through traffic flooding on the data plane. As a result, there is a lot of flooded traffic on VXLAN networks. To address this problem, VXLAN uses EVPN as the control plane protocol. EVPN allows VTEPs to exchange BGP EVPN routes to implement automatic VTEP discovery and host information advertisement, preventing unnecessary traffic flooding.
- Problems in configuring VXLAN in static mode:
    - If N devices need to establish VXLAN tunnels, you need to manually configure the ingress replication list a maximum of N(N-1)/2 times.
    - A static VXLAN tunnel only has the data forwarding plane.
    - Remote MAC addresses can be learned only through broadcast ARP packets.

Overview of BGP EVPN

- EVPN extends BGP to define several types of BGP EVPN routes (the MP_REACH_NLRI attribute defines several new NLRIs, which are called EVPN NLRIs).
- The BGP EVPN routes can be used to transmit VTEP addresses and host information. EVPN is applied to VXLAN networks to move VTEP discovery and host information learning from the data plane to the control plane.



- **Type 2 routes (MAC/IP routes):** are used to advertise host MAC addresses, ARP entries, and IP routes.
- **Type 3 routes (inclusive multicast routes):** are used to transmit Layer 2 VNI (L2VNI) and VTEP IP address information, implement automatic VTEP discovery, dynamic VXLAN tunnel establishment, and BUM packet forwarding.
- **Type 5 routes (IP prefix routes):** are used to advertise host MAC addresses, ARP entries, IP routes, and external network routes.

HUAWEI

- In a network virtualization overlay (NVO) scenario, BGP EVPN is used together with VXLAN as the control plane protocol for VXLAN.

EVPN NLRI

- EVPN NLRI is carried in the path attribute MP_REACH_NLRI. The address family identifier (AFI) is 25, indicating L2VPN; the sub-address family identifier (SAFI) is 70.

Path Attribute - MP_REACH_NLRI	
Flags: Optional, Non-transitive	
Type Code: MP_REACH_NLRI (14)	
Length	
Address family identifier (AFI): **Layer-2 VPN (25)**	
Subsequent address family identifier (SAFI): **EVPN (70)**	
Next hop network address (4 Bytes)	
Route Type (1 octet)	EVPN NLRI
Length (1 octet)	
Route Type specific (variable)	

HUAWEI

Extended Community

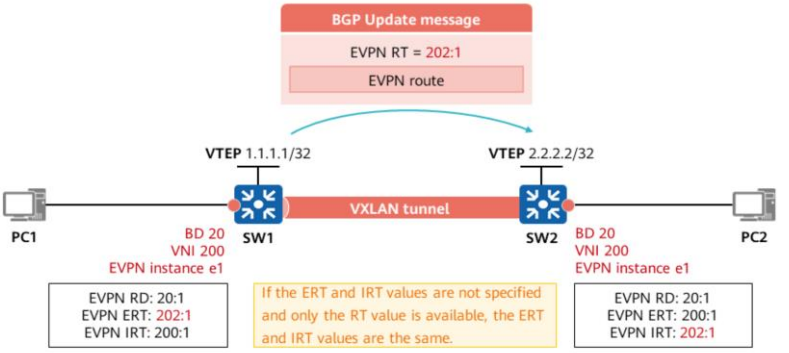
- BGP EVPN is similar to MPLS VPN. To control the sending and receiving of routes, BGP EVPN uses the EVPN instance, which is the same as the traditional IP VPN instance. An EVPN instance has RD and RT values. When routes are transmitted, the extended community attribute is used to carry the RT value of the EVPN instance.
- In addition to the RT value, BGP EVPN adds some new sub-types to the extended community attribute: MAC Mobility and EVPN Router's MAC Extended Community.

HUAWEI

- For details about the RD and RT, see the *HCIP-Advanced Routing Switching - 08 MPLS VPN Principles and Configuration*.

EVPN VPN Instance

- After an EVPN instance is bound to a BD, MAC address entries of the BD are transmitted through BGP EVPN routes, carrying the ERT of the EVPN instance bound to the BD. After receiving the EVPN routes, the remote end compares the ERT carried in the EVPN routes with the IRT of the local EVPN instance and adds the EVPN routes to the routing table of the EVPN instance. The remote end parses the EVPN routing table to obtain MAC address entries and adds them to the MAC address table of the BD bound to the EVPN instance.



HUAWEI

Contents

1. Overview of VXLAN and Campus Network Virtualization
2. Basic Concepts and Fundamentals of VXLAN
3. **BGP EVPN**
   - Basic Concepts
   - BGP EVPN Routes
   - BGP EVPN Features
4. Campus Network Virtualization

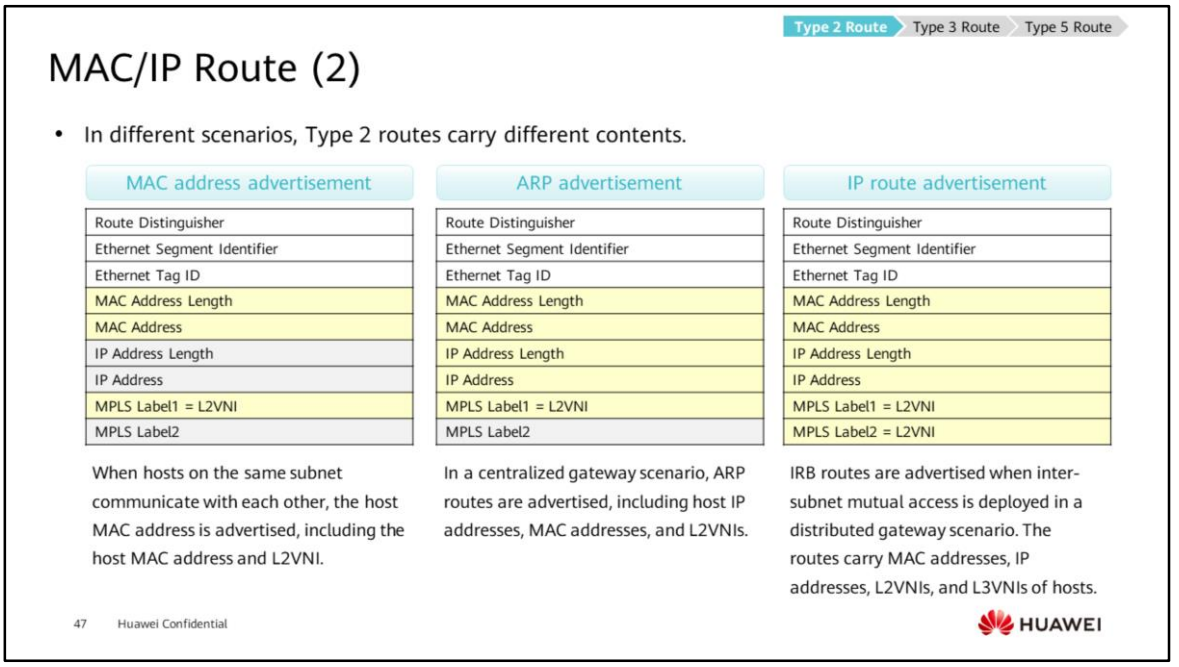
HUAWEI

MAC/IP Route (1)

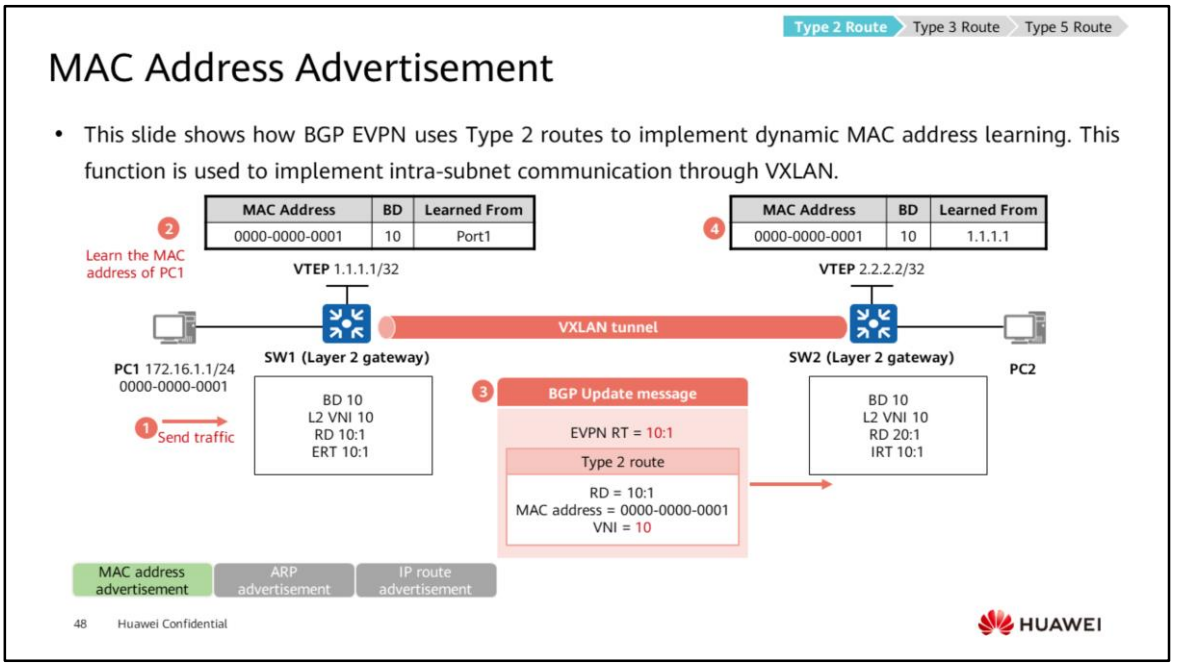
- Type 2 routes (MAC/IP routes): are used to advertise host MAC addresses, ARP entries, and IP routes.

Format of a MAC/IP route	Field description
Route Distinguisher (8 bytes)	Route distinguisher (RD) of an EVPN instance.
Ethernet Segment Identifier (10 bytes)	Unique ID for defining the connection between local and remote devices.
Ethernet Tag ID (4 bytes)	VLAN ID configured on the device.
MAC Address Length (1 byte)	Length of the host MAC address carried in the route.
MAC Address (6 bytes)	Host MAC address carried in the route.
IP Address Length (1 byte)	Mask length of the host IP address carried in the route
IP Address (0, 4, or 16 bytes)	Host IP address carried in the route.
MPLS Label1 (3 bytes)	L2VNI carried in the route.
MPLS Label2 (0 or 3 bytes)	L3VNI carried in the route.

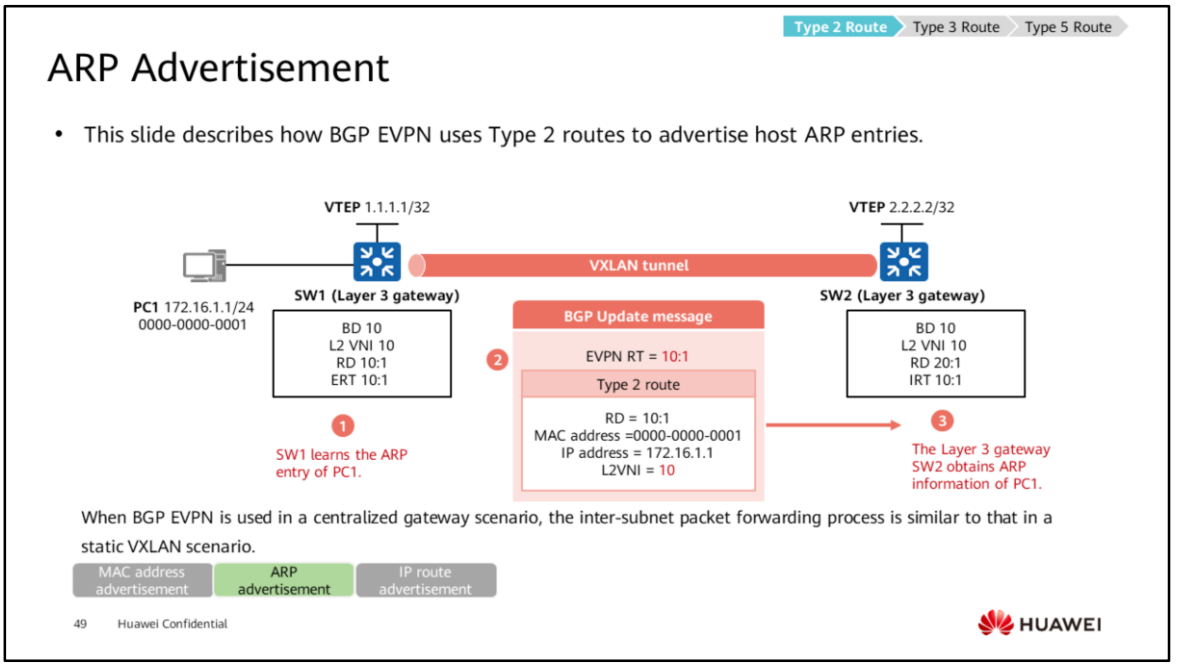
HUAWEI



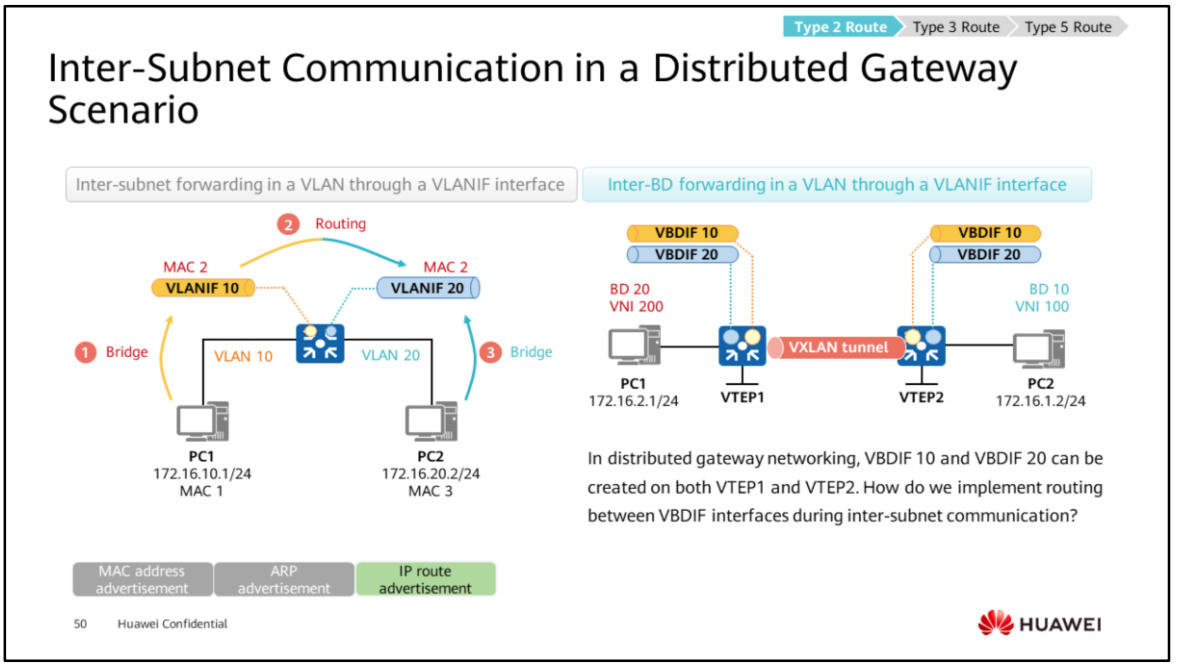
- The first three fields (RD, Ethernet Segment Identifier, and Ethernet Tag ID) of a Type 2 route are the same in different scenarios, and only the last six fields are different.



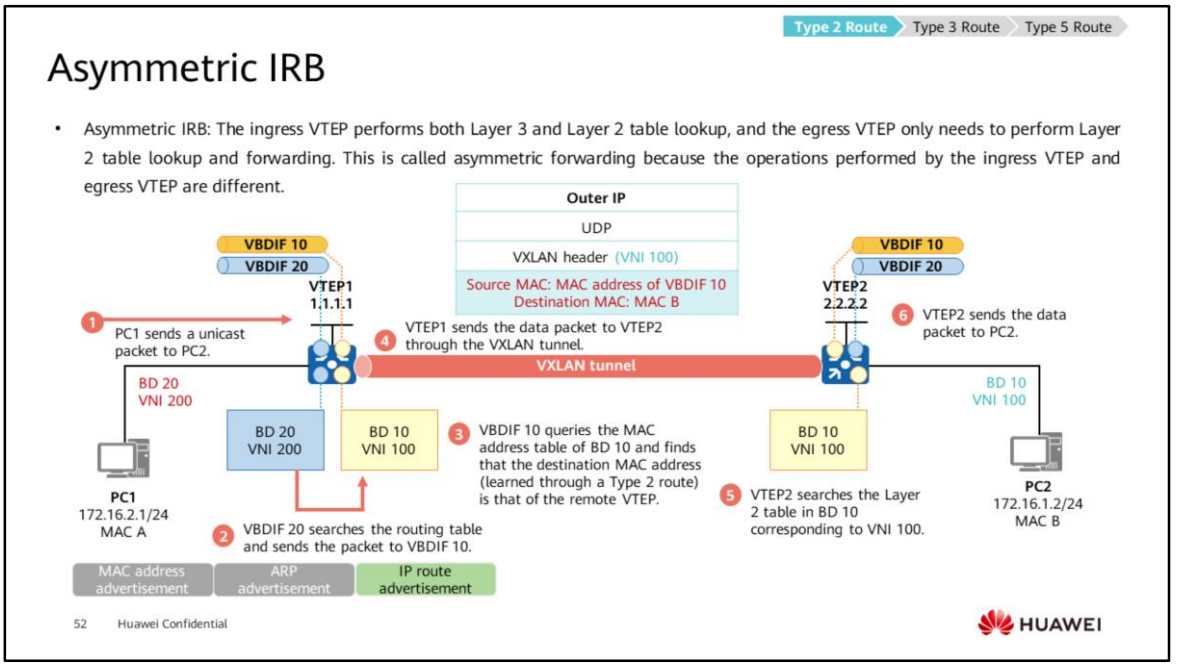
- Intra-subnet host MAC address advertisement:
    1. PC1 generates data traffic and sends the traffic to SW1.
    2. SW1 obtains the MAC address of PC1 and creates an entry in the MAC address table to record the MAC address, BD ID, and inbound interface.
    3. SW1 generates a BGP EVPN route based on this entry and sends the route to SW2. The route carries the RT value of the local EVPN instance and Type 2 route (MAC route). In the MAC route, the MAC address of PC1 is stored in the MAC Address field and the L2VNI is stored in the MPLS Label1 field.
    4. After receiving the BGP EVPN route from SW1, SW2 checks the RT (similar to the RT concept in MPLS VPN) carried in the route. If the RT is the same as the import RT of the local EVPN instance, SW2 accepts the route. Otherwise, SW2 discards the route. After accepting the route, SW2 obtains the MAC address of PC1 and the mapping between the BD ID and the VTEP IP address (next hop network address in MP_REACH_NLRI) of SW1, and generates the MAC address entry of PC1 in the local MAC address table. Based on the next hop, the outbound interface of the MAC address entry recurses to the VXLAN tunnel destined for SW1.



- A MAC/IP route can carry both the MAC and IP addresses of a host, and therefore can be used to advertise ARP entries between VTEPs. The MAC Address and MAC Address Length fields identify the MAC address of the host, whereas the IP Address and IP Address Length fields identify the IP address of the host. This type of MAC/IP route is called the ARP route. ARP advertisement applies to the following scenarios:
  - ARP broadcast suppression. After a Layer 3 gateway learns the ARP entries of a host, it generates host information that contains the host IP and MAC addresses, Layer 2 VNI, and gateway's VTEP IP address. The Layer 3 gateway then transmits an ARP route carrying the host information to a Layer 2 gateway. When the Layer 2 gateway receives an ARP request, it checks whether it has the host information corresponding to the destination IP address of the packet. If such host information exists, the Layer 2 gateway replaces the broadcast MAC address in the ARP request with the destination unicast MAC address and unicasts the packet. This implementation suppresses ARP broadcast packets.
  - VM migration in distributed gateway scenarios. After a VM migrates from one gateway to another, the new gateway learns the ARP entry of the VM (after the VM sends gratuitous ARP packets) and generates host information that contains the host IP and MAC addresses, Layer 2 VNI, and gateway's VTEP IP address. The new gateway then transmits an ARP route carrying the host information to the original gateway. After the original gateway receives the ARP route, it detects a VM location change and triggers ARP probe. If ARP probe fails, the original gateway withdraws the ARP and host routes of the VM.
- ARP advertisement is mainly used in the centralized VXLAN gateway+BGP EVPN scenario. In BGP EVPN, ARP or IRB advertisement to peers is mutually exclusive. Only one of these routes can be configured to advertise. Generally, ARP advertisement is selected in the centralized VXLAN gateway+BGP EVPN scenario, in the distributed VXLAN gateway+BGP EVPN scenario, IRB routes are advertised.



- In distributed gateway networking, VTEPs function as both Layer 2 and Layer 3 gateways. In this networking, inter-subnet communication is implemented in multiple modes. According to the processing mode of ingress VTEPs, inter-subnet communication can be implemented through asymmetric and symmetric Integrated Routing and Bridging (IRB).
- Inter-subnet forwarding in a VLAN through a VLANIF interface:
    - Based on the local IP address, local mask, and peer IP address, PC1 finds that the destination device PC2 is on a different network segment from itself. Therefore, PC1 determines Layer 3 communication and sends the traffic destined for PC2 to the gateway. In the data packet sent by PC1, the source MAC address is MAC1 and the destination MAC address is MAC2.
    - After receiving the packet sent from PC1 to PC2, the switch decapsulates the packet and finds that the destination MAC address is the MAC address of VLANIF 10. Therefore, the switch considers that the packet is sent to itself and sends the packet to the routing module for further processing.
    - The routing module parses the packet and finds that the destination IP address is 192.168.20.2, which is not the IP address of the local interface. Therefore, the routing module needs to forward the packet at Layer 3. When the routing module searches the routing table, it matches the direct route generated by VLANIF 20 against the packet.

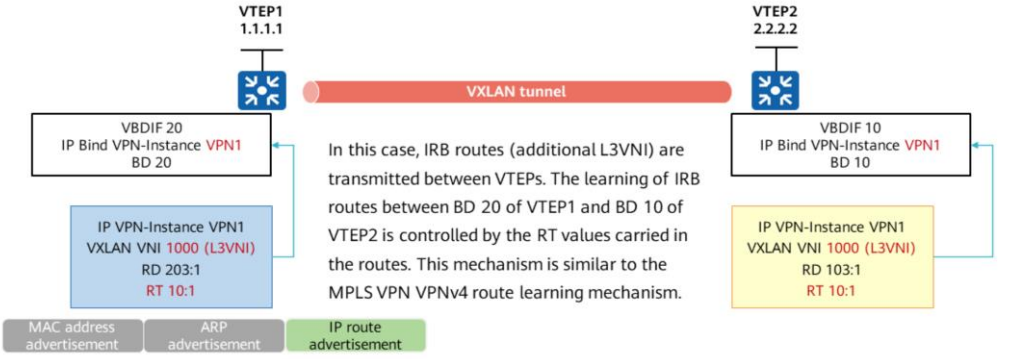


- During asymmetric IRB, VTEPs do not transmit host IP routes between each other. That is, VTEP1 and VTEP2 do not transmit the 32-bit host route (generated through an ARP entry) of the connected PC. Therefore, VTEP1 searches the routing table in step 2, and matches the packet against the direct route generated by VBDIF 10.
- In step 5, VTEP2 decapsulates the VXLAN packet and finds that the destination MAC address is not the MAC address of the local VBDIF interface corresponding to the BD. Therefore, VTEP2 searches the Layer 2 forwarding table for the MAC address entry of the corresponding BD based on the VNI carried in the packet and then forwards the packet at Layer 2.

Type 2 Route | Type 3 Route | Type 5 Route

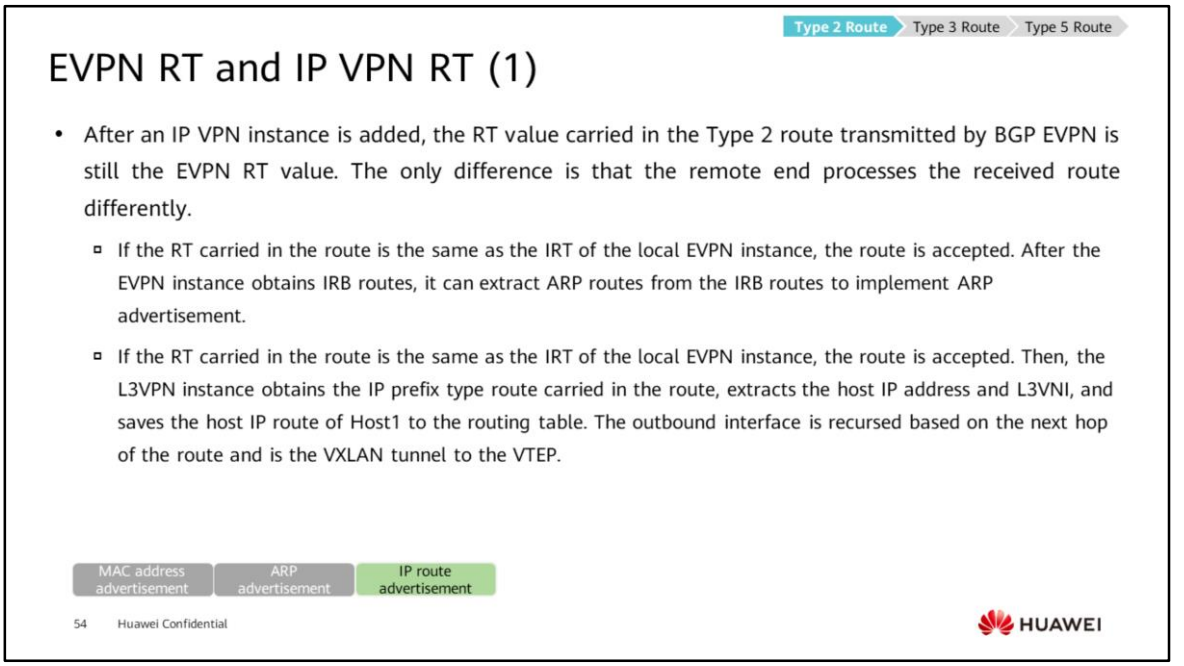
Symmetric IRB

- Symmetric IRB: Both the ingress and egress VTEPs perform Layer 3 table lookup and forwarding.
- Compared with asymmetric IRB, symmetric IRB adds an IP VPN instance and its bound L3VNI. In asymmetric IRB mode, the VNI in the VXLAN header of packets transmitted between VTEPs is the L2VNI. The VBDIF interface needs to be bound to an IP VPN instance. In this case, route learning and data forwarding are restricted to the IP VPN instance, which is similar to MPLS VPN.

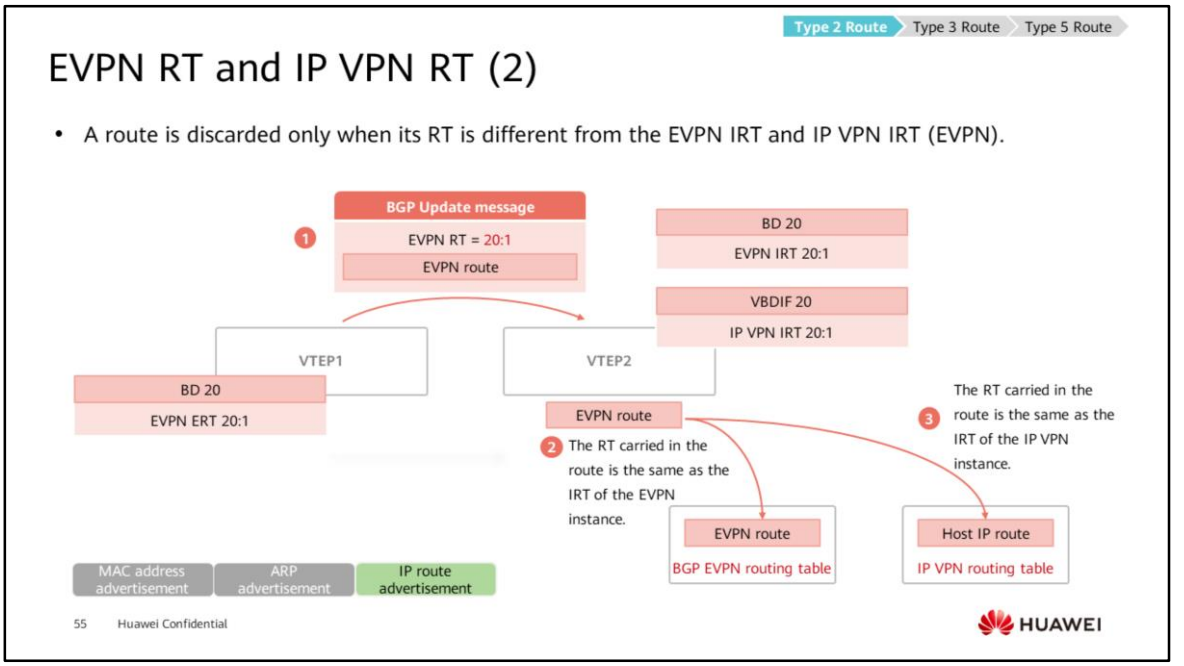


HUAWEI

- On Huawei devices, Symmetric IRB is used.

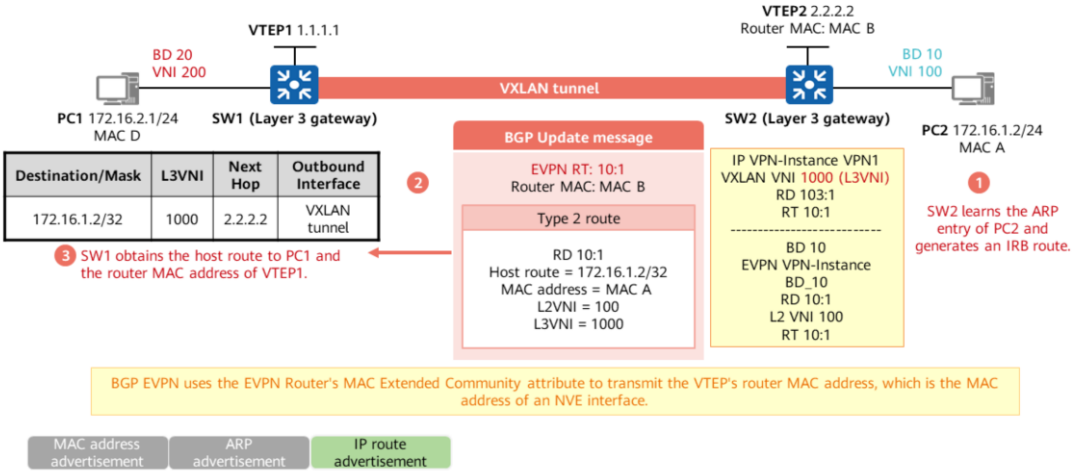


- In a BGP EVPN scenario, if you want to control the sending and receiving of EVPN routes based on the RT value of the IP VPN instance, run the **vpn-target evpn** command to configure the RT value. In this case, the ERT is carried in EVPN routes and sent to the remote BGP EVPN peer, the IRT matches the RT carried in an EVPN route to determine which EVPN routes can be added to the routing table of the local VPN instance address family.
- Note: The RT configured using the **vpn-target evpn** command is called an RT (EVPN), which is different from common RTs.



- VTEP1 sends a Type 2 BGP EVPN route (IRB type), which carries the ERT (20:1) of the EVPN instance bound to the BD.
- After receiving the BGP Update message, VTEP2 checks the RT value (20:1) carried in the BGP Update message and compares it with the IRT in the local EVPN instance and the IRT (EVPN) in the IP VPN instance. VTEP2 finds that the IRT of the EVPN instance bound to BD 20 and IRT of the IP VPN instance bound to VBDIF 20 are the same, adds the EVPN routes to the EVPN routing table of BD 20, and adds the IP routes contained in the EVPN routes to the routing table of the IP VPN instance bound to VBDIF 20.

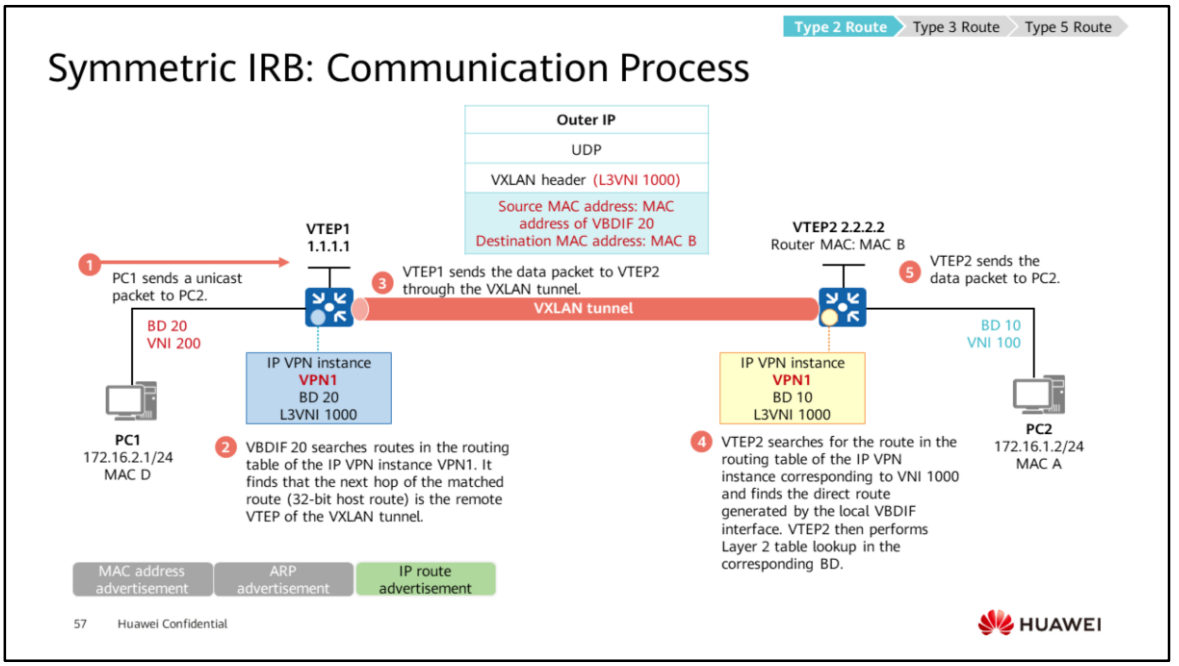
Symmetric IRB: Host IP Route Advertisement (IRB Route)



BGP EVPN uses the EVPN Router's MAC Extended Community attribute to transmit the VTEP's router MAC address, which is the MAC address of an NVE interface.

MAC address advertisement | ARP advertisement | IP route advertisement

HUAWEI



- In symmetric IRB mode, VTEPs transmit 32-bit host routes generated using ARP entries. Therefore, VTEP1 matches the 32-bit host routes transmitted from VTEP2 during route lookup. Even if VTEP1 has the direct route generated by VBDIF 10, it still forwards packets based on 32-bit host routes according to the longest match rule.
- In step 4, VTEP2 decapsulates the VXLAN packet and finds that the destination MAC address of the inner data of a packet is the router MAC address (MAC B) of VTEP2. VTEP2 determines that it needs to forward the packet based on the routing table, finds the corresponding IP VPN instance based on VNI 1000, and searches the routing table of the IP VPN instance for the route, finds the direct route generated by VBDIF 10, searches the local MAC address table, and sends the packet to PC2.

Type 2 Route | Type 3 Route | Type 5 Route

Type 3 Route

- Type 3 route (inclusive multicast route)
  - This type of route is used on the VXLAN control plane for automatic VTEP discovery and dynamic VXLAN tunnel establishment.
  - VTEPs that function as BGP EVPN peers exchange inclusive multicast routes to transfer L2VNIs and VTEPs' IP addresses.
  - The Originating Router's IP Address field identifies the local VTEP's IP address; the MPLS Label field identifies the L2VNI.

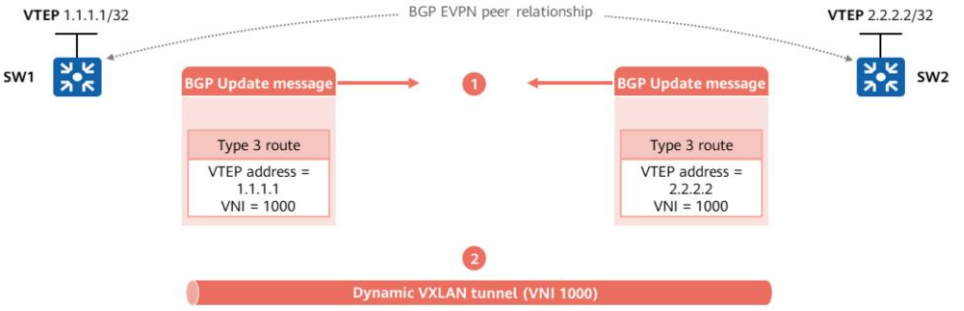
NLRI format	Route Distinguisher (8 bytes)	RD value of an EVPN instance.
	Ethernet Tag ID (4 bytes)	VLAN ID on the device. The value is all 0s in this type of route.
	IP Address Length (1 byte)	Mask length of the local VTEP's IP address carried in the route.
	Originating Router's IP Address (4 or 16 bytes)	Local VTEP's IP address carried in the route.
PMSI attribute	Flags (1 byte)	This field is inapplicable in VXLAN scenarios.
	Tunnel Type (1 byte)	The value can only be 6, representing Ingress Replication in VXLAN scenarios.
	MPLS Label (3 bytes) = L2VNI	L2VNI carried in the route.
	Tunnel Identifier (variable)	This field is the local VTEP's IP address in VXLAN scenarios.

HUAWEI

- The Provider Multicast Service Interface (PMSI) is an optional transitive BGP attribute. In VXLAN scenarios, Tunnel Type has a fixed value of 6, which is used to carry the VTEP's IP address and L2VNI of the sender.

VXLAN Tunnel Establishment

- VTEPs exchange L2VNIs and VTEP IP addresses through Type 3 routes. If there are reachable routes between the local and remote VTEPs' IP addresses, a VXLAN tunnel is established between the VTEPs. Additionally, if the local and remote VNIs are the same, an ingress replication list is created for BUM packet forwarding.



HUAWEI

Type 5 Route

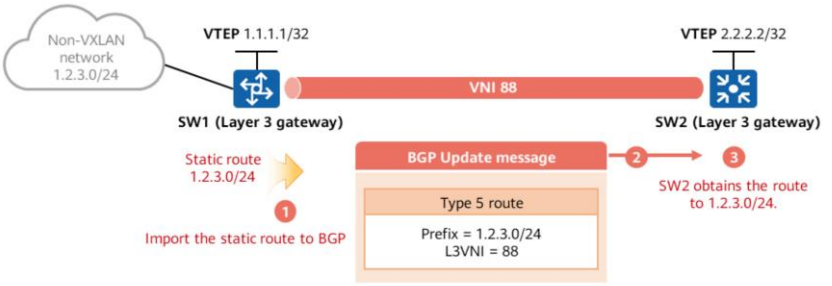
- Type 5 route (IP prefix route)
  - The IP Prefix Length and IP Prefix fields in an IP prefix route can identify a host IP address or network segment.
  - If the IP Prefix Length and IP Prefix fields in an IP prefix route identify a host IP address, the route is used for IP route advertisement in distributed VXLAN gateway scenarios, which functions the same as an IRB route on the VXLAN control plane.
  - If the IP Prefix Length and IP Prefix fields in an IP prefix route identify a network segment, the route allows hosts on a VXLAN to access external networks.

Format of an IP prefix route	Field description
Route Distinguisher (8 bytes)	RD value of an EVPN instance.
Ethernet Segment Identifier (10 bytes)	Unique ID for defining the connection between local and remote devices.
Ethernet Tag ID (4 bytes)	VLAN ID configured on the device.
IP Prefix Length (1 byte)	Length of the IP prefix carried in the route.
IP Prefix (4 or 16 bytes)	IP prefix carried in the route.
GW IP Address (4 or 16 bytes)	Default gateway address. This field is inapplicable in VXLAN scenarios.
MPLS Label (3 bytes)	L3VNI carried in the route.

HUAWEI

Application Scenarios of IP Prefix Route Advertisement

- For a non-VXLAN network, a VTEP can advertise external routes to the entire VXLAN network through Type 5 routes, instructing hosts on the VXLAN network to access the external network.



HUAWEI

- Similar to Type 2 IRB routes, Type 5 routes carry the router MAC address of the VTEP through the EVPN router's MAC extended community attribute during route transmission. In addition, Type 5 routes carry only the L3VNI. Therefore, the forwarding process is also called IRB forwarding.

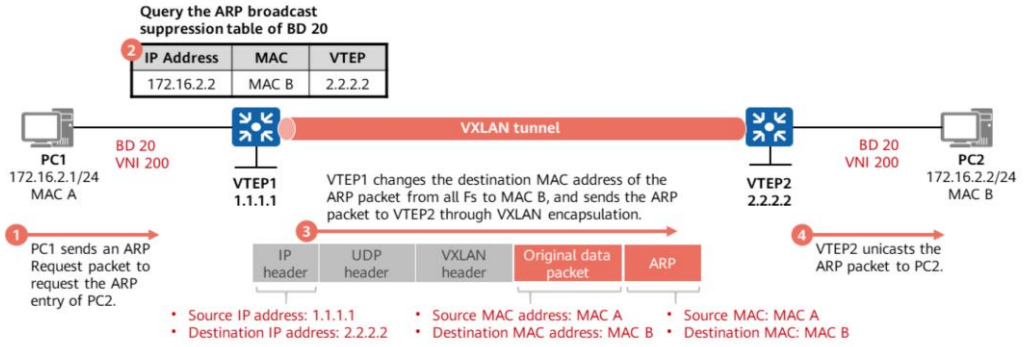
Contents

1. Overview of VXLAN and Campus Network Virtualization
2. Basic Concepts and Fundamentals of VXLAN
3. **BGP EVPN**
   - Basic Concepts
   - BGP EVPN Routes
   - BGP EVPN Features
4. Campus Network Virtualization

HUAWEI

ARP Broadcast Suppression

- BGP EVPN Type 2 routes enable VTEPs to learn MAC addresses without depending on communication between hosts. However, ARP entries between hosts still need to be flooded and forwarded on the VXLAN overlay, which consumes a large number of network resources.
- To reduce broadcast traffic, configure ARP broadcast suppression using BGP EVPN routes.

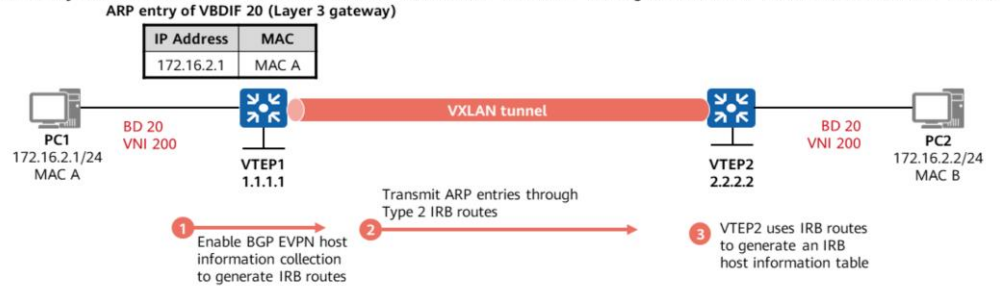


HUAWEI

- ARP broadcast suppression is an effective method to relieve the burden of a gateway in processing ARP packets. When receiving an ARP Request packet, the gateway searches the ARP broadcast suppression table for the mapping between the IP address and MAC address of the destination device. If the ARP Request packet matches an entry in the table, the gateway replaces the broadcast MAC address in the ARP Request packet with the MAC address of the destination device. Then, the gateway sends the ARP Request packet through the interface corresponding to the destination MAC address.

Host Information Collection

- When ARP broadcast suppression is enabled on a device, the device generates the ARP broadcast suppression table. ARP broadcast suppression entries are originated from Type 2 routes (IRB routes and host ARP advertisement) carried in BGP EVPN packets.
- By default, a Layer 3 gateway does not generate BGP EVPN routes based on local ARP information. You need to manually enable BGP EVPN host information collection. The VTEP then generates IRB routes based on ARP entries.

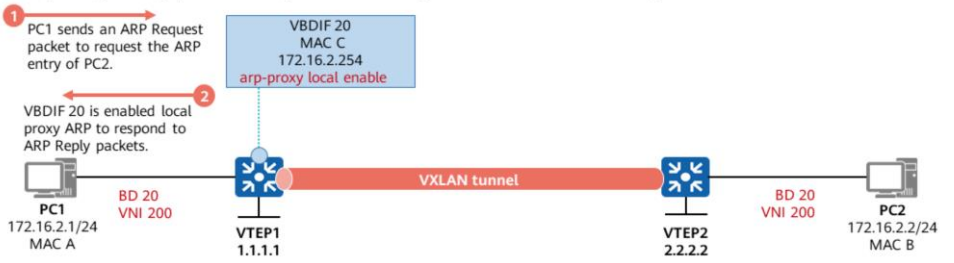


HUAWEI

- An ARP route carries the following valid information: host MAC address, host IP address, and L2VNI. An IRB route carries the following valid information: host MAC address, host IP address, L2VNI, and L3VNI. As a result, an IRB route includes an ARP route and can be used to advertise both the host IP route and host ARP entry.

Local Proxy ARP (1)

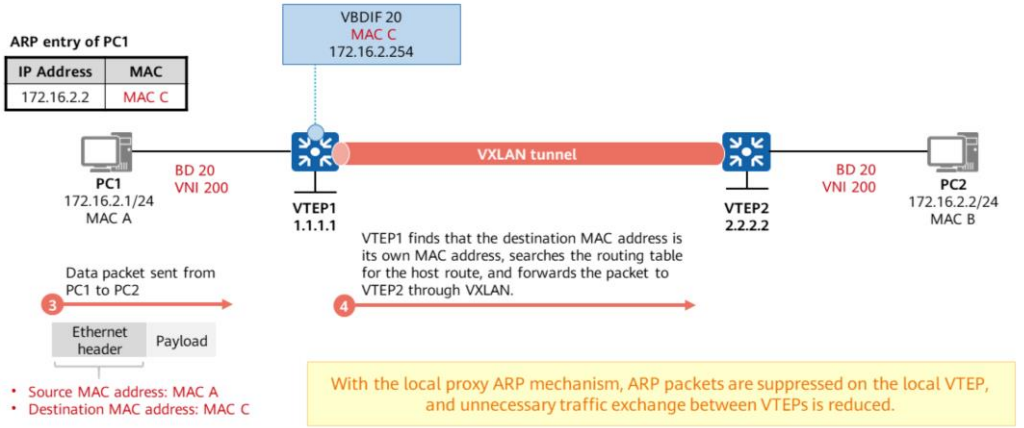
- After BGP EVPN host information collection is enabled on the entire network, the Layer 3 gateway learns the 32-bit host routes of all hosts. In this way, the Layer 3 gateway can use the symmetric IRB mode to forward traffic between hosts in the same BD.
- In this case, you can enable local proxy ARP on the VBDIF interface of the Layer 3 gateway. The VBDIF interface responds to the ARP Request packet sent by a downstream host to an IP address on the same network segment. Then the Layer 3 gateway performs Layer 3 forwarding on the same network segment.



HUAWEI

- On a VXLAN network, a bridge domain (BD) is a Layer 2 broadcast domain. After a VTEP receives BUM packets, it broadcasts the packets in the BD. To reduce broadcast traffic, a network administrator usually configures access-side isolation or port isolation to isolate access users in a BD. However, as services become more diverse and keep increasing, users have growing needs for intra-BD communication. To allow isolated users in a BD to communicate, configure local proxy ARP on the VBDIF interface of the BD.

Local Proxy ARP (2)



With the local proxy ARP mechanism, ARP packets are suppressed on the local VTEP, and unnecessary traffic exchange between VTEPs is reduced.

HUAWEI

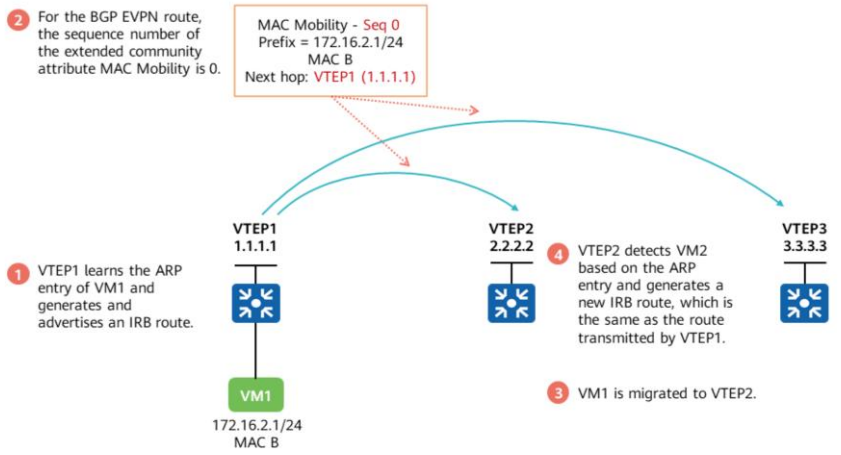
Distributed Gateway

- When local proxy ARP is enabled, a VTEP only needs to maintain local ARP entries. ARP information transmitted by other VTEPs through BGP EVPN routes is not used during forwarding. In this case, the VTEP does not need to maintain ARP entries learned from other VTEPs.
- After distributed gateway is enabled, the VTEP processes only the ARP packets received from the user-side host and deletes the learned network-side ARP entries.

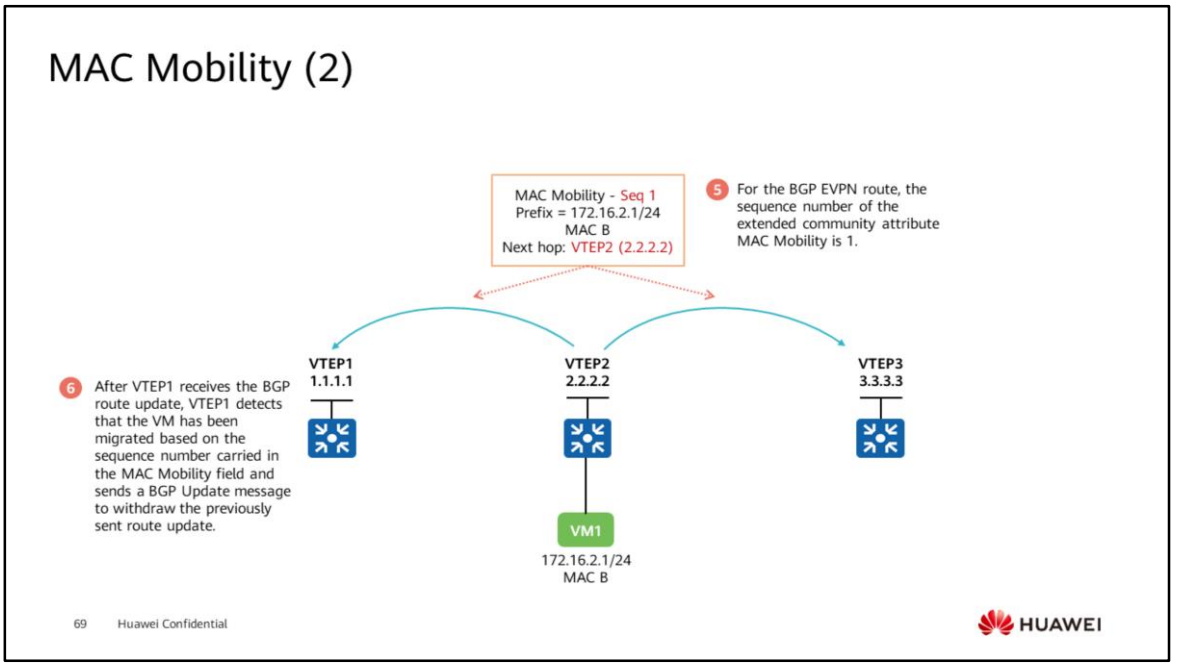
HUAWEI

- Generally, the same MAC address is configured for VBDIF interfaces with the same interface number on different VTEPs. After the distributed gateway function is enabled, VBDIF interfaces with the same IP address and MAC address do not report ARP conflicts. In addition, when hosts and VMs are migrated to different VTEPs, the gateway does not need to resolve ARP entries again.

MAC Mobility (1)



HUAWEI



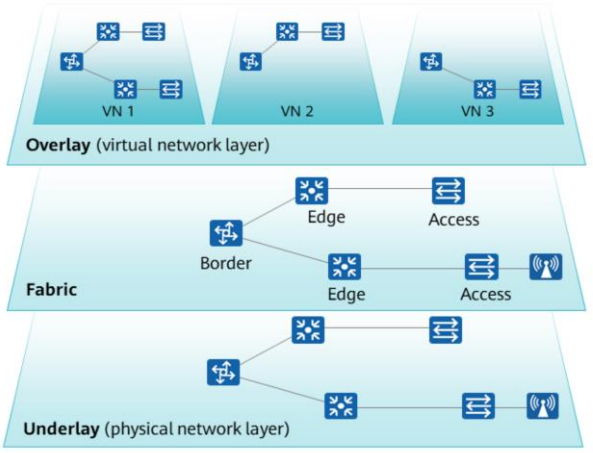
- The MAC Mobility extended attribute is used to announce the location change of a host or VM when the host or VM is migrated from one VTEP to another VTEP.

Contents

1. Overview of VXLAN and Campus Network Virtualization
2. Basic Concepts and Fundamentals of VXLAN
3. BGP EVPN
4. **Campus Network Virtualization**

HUAWEI

Layers and Concepts of a VXLAN-based Virtualized Campus Network



- Multiple VNs can be created based on service requirements to implement service isolation.
- VXLAN is used to implement Layer 2 and Layer 3 communication.

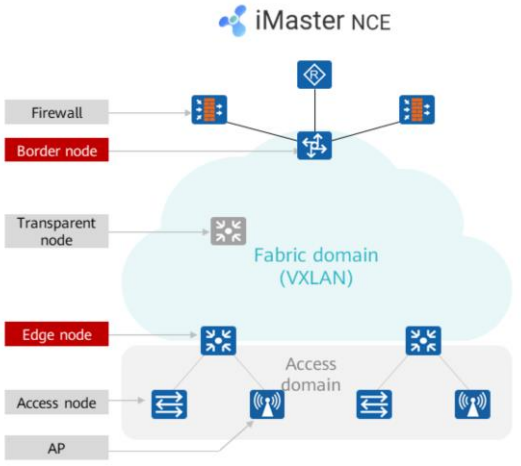
- A virtualization technology is used to construct a logical topology based on any physical topology.
- A service network is created in the fabric and decoupled from the physical network.

- Physical network established by physical devices.
- Interoperability for all services on the campus network.
- Basic bearer network for service data forwarding.

HUAWEI

- VXLAN allows a virtual Layer 2 or Layer 3 network (overlay network) to be built over a physical network (underlay network). The overlay network transmits packets between different sites through Layer 3 forwarding paths provided by the underlay network.
- In technical applications, different overlay networks are created for different services. Services are aware of the overlay network only. The underlay network is transparent to services.

Architecture Overview: Network Nodes



- **Firewall:** is required when L4-L7 security policies are deployed. It can be deployed in off-path mode or at the campus egress.
- **Border node:** used to implement communication between the fabric and external networks. It is usually a core switch.
- **Edge node:** a fabric edge device that connects user-side devices to the fabric. Data packets from wired users are encapsulated into VXLAN packets on edge nodes.
- **Transparent node:** a transparent transmission node in the fabric. It does not need to support VXLAN.
- **Access node:** is also known as an extended node. It is an access node for wired users and is optional. It provides network access for users and does not need to support VXLAN.
- **AP:** wireless access node, through which users access the wireless network and finally access the fabric.

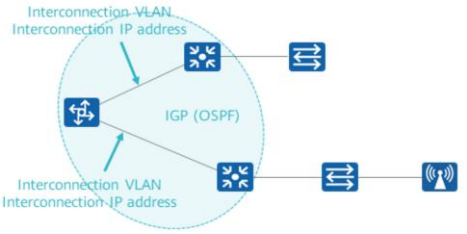
HUAWEI

- L4-L7 security policy: The firewall supports security control from Layer 4 to Layer 7.

What Are an Underlay Network and a Fabric?

Underlay network

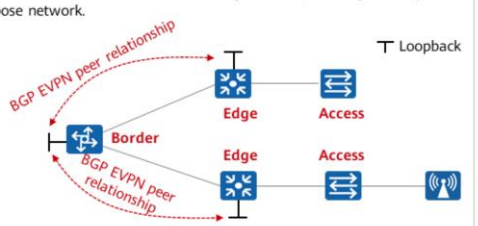
The underlay network is the foundation of the entire virtualized campus network. It is a physical network consisting of physical network devices, which provides interoperability for all services on a campus network.



- Multi-layer architecture (core, aggregation, and access layers)
- Multiple topologies (such as tree, ring, and mesh topologies)
- Underlay information includes device interconnection VLANs, interconnection IP addresses, and an IGP.

Fabric

A campus fabric is a network resource pool that is abstracted from physical network devices on the underlay network, creating a multiple-purpose network.

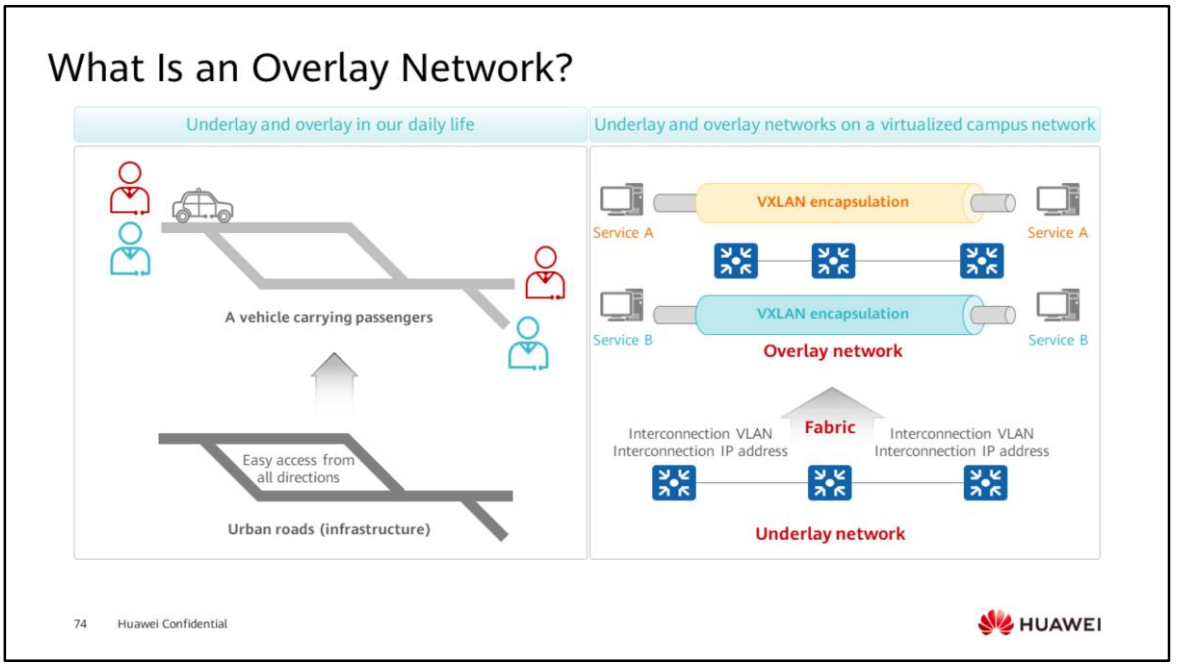


The fabric consists of the following resource pools:

- Overlay network resource pool (BD IDs and VNIs) for terminal access
- VLAN ID pool for terminal access
- Access point pool (switch ports for wired access or SSIDs for wireless access) for terminal access

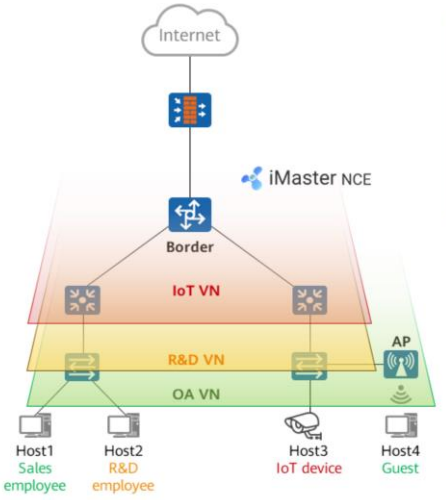
These resource pools can be used to create multiple VNs.

HUAWEI



- An overlay network is a logical network that is constructed based on a physical network using a tunneling technology and with separated forwarding and control planes.

What Is a VN?



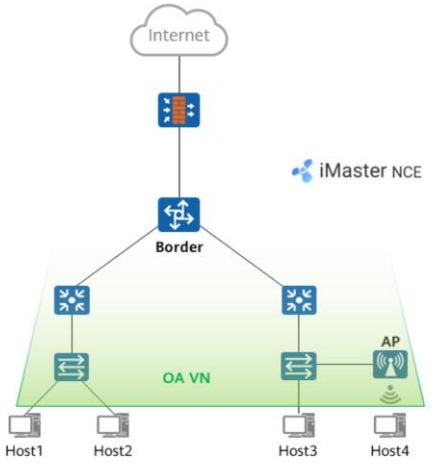
- VXLAN can be used to construct multiple VNs on the underlay network.
- A VN is considered as an overlay network.
- Service data of each VN is encapsulated using VXLAN in the fabric so that the VN services are isolated on the forwarding plane. The control plane uses BGP EVPN to establish VXLAN tunnels and exchange overlay routing information.

Each VN has the following parameters:

- Network service resources (such as the DHCP server and third-party RADIUS/Portal server)
- (Optional) External networks
- User IP address segments, VLANs, and gateway address
- Wired access ports and/or wireless access points (APs)
- Other parameters

HUAWEI

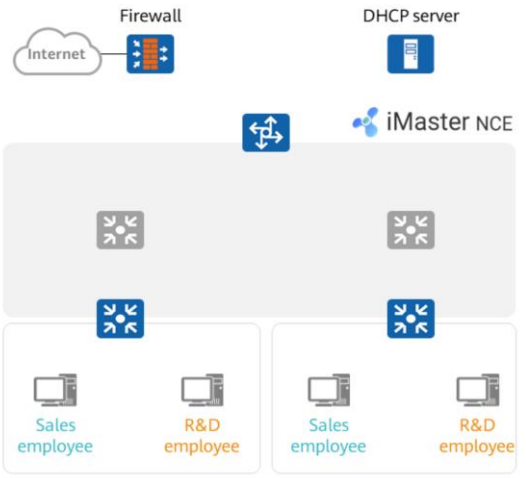
Creating a VN



1. **VN settings**
   - User gateway location
   - External network
   - Network service resources
   - User subnet and gateway
2. **Wired access**
   - Access sites, devices, and ports
   - Authentication mode
   - VLAN information
3. **Wireless access**
   - Access sites and devices

HUAWEI

Typical Case Analysis: Requirements



Fabric requirements:

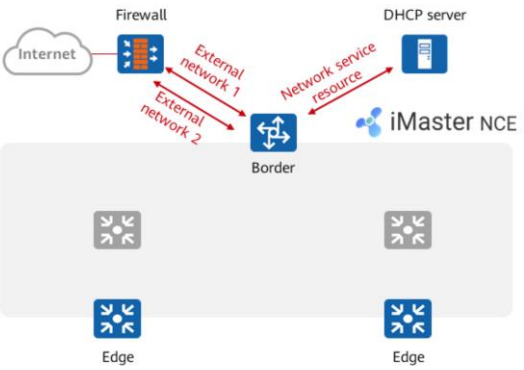
- Build a fabric based on the physical network.
- Use the distributed gateway solution.

VN requirements:

- Create two VNs, one for OA and the other for R&D.
- By default, the two VNs are completely isolated, and intra-subnet and inter-subnet communication can be implemented in each VN.
- Devices in both VNs can access the external network connected to the firewall.
- Terminals in the two VNs can obtain IP addresses from the DHCP server.

HUAWEI

Typical Case Analysis: Fabric Management (1)

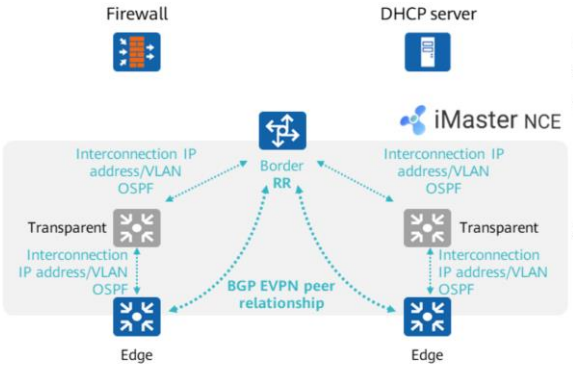


Fabric creation and configuration:

1. You can add physical devices (core, aggregation, and access switches) to the fabric based on service requirements.
2. You can specify a switch as a border node or an edge node.
3. iMaster NCE-Campus automatically specifies a border node as an RR to optimize the logical network architecture and BGP peer relationship model.
4. You can predefine two external networks for two VNs to access the Internet.
5. You can define a network service resource, which is used by terminals to obtain IP addresses from the DHCP server in the resource.

HUAWEI

Typical Case Analysis: Fabric Management (2)

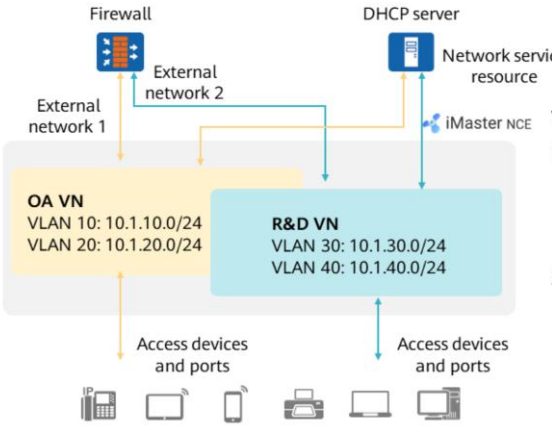


Automatic deployment of the fabric and underlay network:

1. Based on the discovered physical network topology and the user-defined fabric, iMaster NCE-Campus automatically orchestrates the network. (You can select multiple OSPF areas or a single OSPF area.)
2. iMaster NCE-Campus automatically delivers underlay network configurations to devices based on the network orchestration result so that the devices have reachable routes to each other's IP address.
3. iMaster NCE-Campus automatically delivers fabric configurations to devices, and BGP EVPN peer relationships are established between devices.

HUAWEI

Typical Case Analysis: VN Management

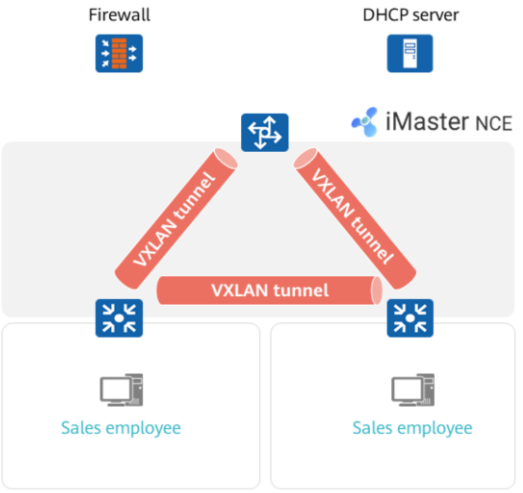


VN creation:

1. You can create OA and R&D VNs and specify the IP network segments/VLANs, gateway addresses, associated external networks and network service resources, and terminal access points for the VNs.
2. iMaster NCE-Campus translates user intents into configurations and delivers the configurations to network devices.

HUAWEI

Typical Case Analysis: Tunnel Establishment

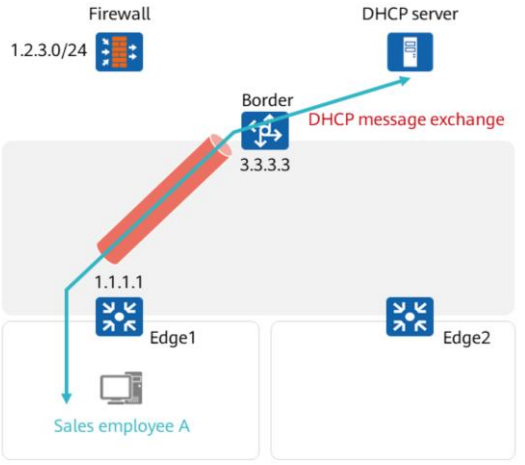


Automatic VXLAN tunnel establishment

1. BGP EVPN advertises information used to establish VXLAN tunnels between peers.
2. VXLAN tunnels are established between devices to prepare for subsequent data forwarding.

HUAWEI

Typical Case Analysis: Address Obtaining

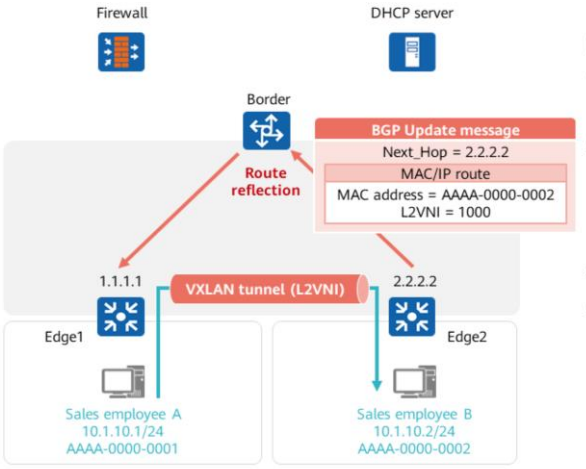


Address obtaining:

1. After sales employee A accesses the network, the user is authenticated. After the authentication succeeds, the authentication point Edge1 obtains the authorization result of the user and assigns the user to the corresponding VLAN.
2. Host A sends a DHCP Request message. After receiving the Request message, the gateway Edge1 relays the Request message to the border node through the VXLAN tunnel.
3. The border node decapsulates the VXLAN packet and forwards the DHCP Request message to the DHCP server.
4. The DHCP server assigns an IP address to host A.

HUAWEI

Typical Case Analysis: Intra-Subnet Communication

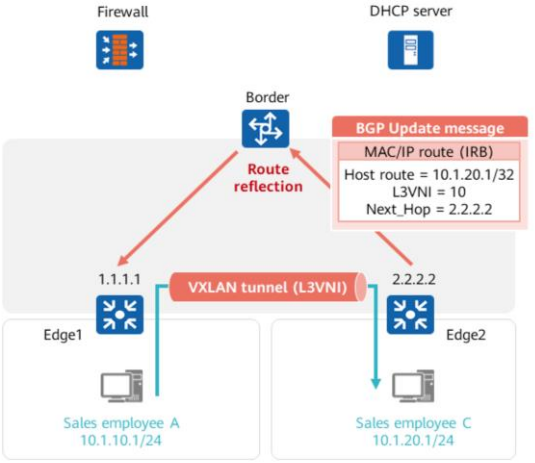


Intra-subnet communication on the same VN:

1. Sales employees A and B pass the access authentication and access the campus network.
2. For example, Edge2 advertises the MAC address of host B to the border node through a BGP Update message, and the border node (RR) advertises the MAC address to Edge1.
3. Edge1 learns the MAC address AAAA.0000.0002.
4. When host A sends data packets to host B, Edge1 encapsulates the data packets into VXLAN packets and forwards them to Edge2. Edge2 performs VXLAN decapsulation and sends the decapsulated packets to the destination.

HUAWEI

Typical Case Analysis: Inter-Subnet Communication

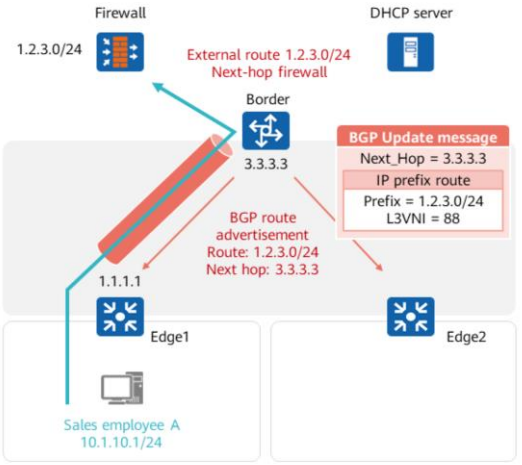


Inter-subnet communication on the same VN:

1. Sales employee C passes the access authentication and accesses the campus network.
2. Edge2 advertises the host route of host C to the border node through a BGP Update message. The border node (RR) advertises the host route to Edge1.
3. Edge1 learns the route to 10.1.20.1/32. The next hop of the route is 2.2.2.2, and the outbound interface is the VXLAN tunnel interface.
4. When host A sends data packets to host C, Edge1 encapsulates the data packets into VXLAN packets and forwards them to Edge2. Edge2 performs VXLAN decapsulation and sends the decapsulated packets to the destination.

HUAWEI

Typical Case Analysis: Accessing the External Network



Accessing the external network:

1. After you associate the external network (destination network segment 1.2.3.0/24) with the OA VN, iMaster NCE-Campus redistributes the external route to BGP and advertises the route to Edge1 and Edge2.
2. When host A sends data packets to 1.2.3.0/24, Edge1 encapsulates the data packets into VXLAN packets and sends them to the border node. The border node decapsulates the VXLAN packets and forwards the IP packets to the firewall.

HUAWEI

Quiz

1. (True or false) BGP EVPN Type 2 host IP routes can be used to transmit ARP information. ( )
   A. True
   B. False
2. (Single-answer question) Which of the following statements about BGP EVPN is false? ( )
   A. MP_REACH_NLRI is used to carry routes.
   B. The extended community attribute is used to carry routes.
   C. MP_REACH_NLRI is used to carry L2VNIs and L3VNIs.
   D. The Next_Hop attribute is used to carry routes.

HUAWEI

1. A
2. D

Summary

- VXLAN uses a Layer 3 routed network as the underlay network and uses tunnels to build an overlay network, supporting large-scale tenant networks.
- VXLAN does not define the control plane. To prevent BUM traffic flooding, use other protocols on the control plane to optimize BUM traffic forwarding.
- BGP EVPN defines several new types of BGP EVPN routes by extending BGP. These BGP EVPN routes can be used to transmit VTEP addresses, host information, and routing information, effectively controlling flooding of BUM traffic.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Network Admission Control

HUAWEI

Foreword

- Campus networks are faced with an increasing number of information security threats, such as viruses, Trojan horses, spyware, and malicious attacks. On a traditional campus network, the intranet is considered secure and threats come from the extranet. However, research shows that roughly 80% of security threats come from the intranet. Network faults caused by the intranet threats will lead to a broad range of serious damage. Even worse, the service system and network will break down.
- The Network Admission Control (NAC) solution integrates terminal security with access control and takes check, isolation, hardening, and audit measures to improve the proactive protection capabilities of terminals. This solution ensures security of each terminal and the entire campus network.

HUAWEI

Objectives

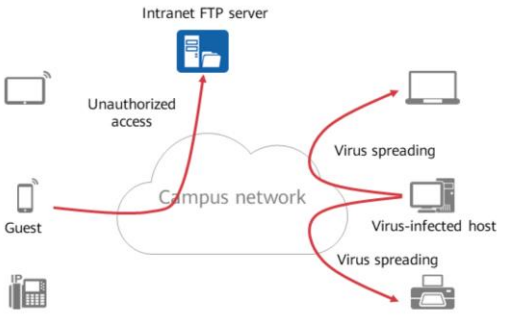
- On completion of this course, you will be able to:
  - Describe basic concepts in NAC.
  - Learn about typical authentication technologies and their working mechanisms and application scenarios.
  - Configure basic user access authentication functions.
  - Learn about the functions and working mechanism of policy association.

Contents

1. **Overview of NAC**
2. User Authentication Technologies
3. User Authorization and Logout
4. NAC Configuration
5. Policy Association

HUAWEI

Technical Background of NAC



- Unauthorized users may access a campus network, which compromises information security on the campus network.
- Various types of terminals are connected to a campus network, and it is difficult to control user behaviors on the campus network.
- For security purposes, a campus network cannot grant access rights to all terminals. Authentication must be performed based on user identities and terminal status. The terminals that do not meet certain conditions are not allowed to access the campus network.

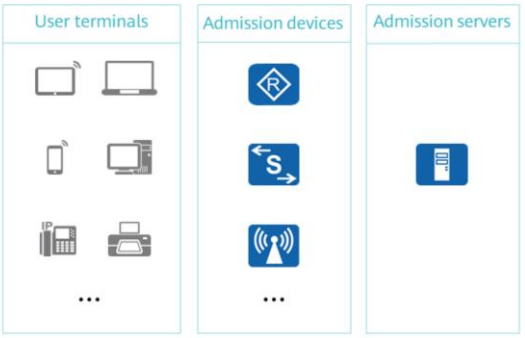
HUAWEI

- Currently, the intranets of most campus networks are faced with the following security issues:
  - Antivirus software is not managed in a centralized manner, and patch management is disordered. Even if enterprises purchase antivirus software, it is difficult to ensure that the virus signature databases of all terminals are the latest. As a result, once a terminal is infected with viruses or malicious code, the virus will soon spread on the intranet.
  - User identities cannot be verified. Therefore, unauthorized access cannot be prevented, posing security risks to the network.
  - Access control for terminals is not implemented. A user can access all network resources as long as the user successfully connects to the network.

Overview of NAC

- An important design idea of NAC is to allow only authorized users and secure terminals to access networks. NAC provides user authentication, permission management, security check, repair, and upgrade functions to improve the overall terminal security protection capabilities on a campus network.

NAC system architecture

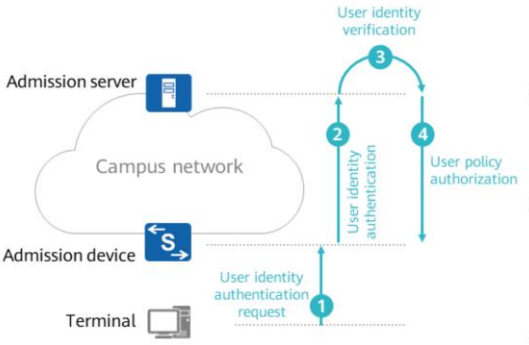


- User terminal
  - Various terminals such as PCs, mobile phones, printers, and cameras that access the network.
- Admission device
  - An admission device is an authentication control point that authenticates access users and executes campus security policies to implement access control (for example, allowing or denying user access).
  - Admission devices can be switches, routers, APs, VPN gateways, or security devices.
- Admission server
  - An admission server authenticates and authorizes users. It verifies the user identities of terminals that attempt to access the network, and grants network access rights to authenticated terminals.
  - An admission server is typically an authentication server (such as a RADIUS server) or a user data source server that stores user identity information.

HUAWEI

- A terminal agent (also known as client software) is usually installed on a user terminal. It works with an admission server to implement user identity authentication, terminal security check, system repair and upgrade, and terminal behavior monitoring and audit.
- Admission devices, which can be switches, routers, APs, or other network devices, provide the following functions:
  - User identity verification.
  - User authentication. Admission devices can implement the commonly used 802.1X authentication, MAC address authentication, and Portal authentication by working with the client software and admission server.
  - User permission control.
- Admission servers include the security control server, security management server, virus signature database server, and patch server.
  - A security control server authenticates users, performs security audit, executes security policies, and works with admission devices to deliver user permissions.
  - A security management server manages user information (including adding, deleting, or modifying user permissions and departments), and defines and manages security policies.
  - A virus signature database server controls automatic update of the virus signature database in antivirus software on terminals.
  - A patch server controls patch installation and update for terminals' operating systems and applications.

Fundamentals of NAC

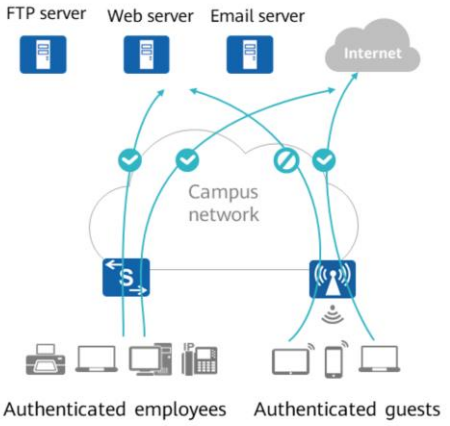


1. User identity authentication request
   - An admission device exchanges packets with a terminal to obtain the user credential.
2. User identity authentication
   - The admission device sends the user credential to the admission server for authentication.
3. User identity verification
   - The admission server verifies the identity of the terminal and delivers the verification result and corresponding policy to the admission device.
4. User policy authorization
   - The admission device executes the policy based on the authorization result received from the admission server.

HUAWEI

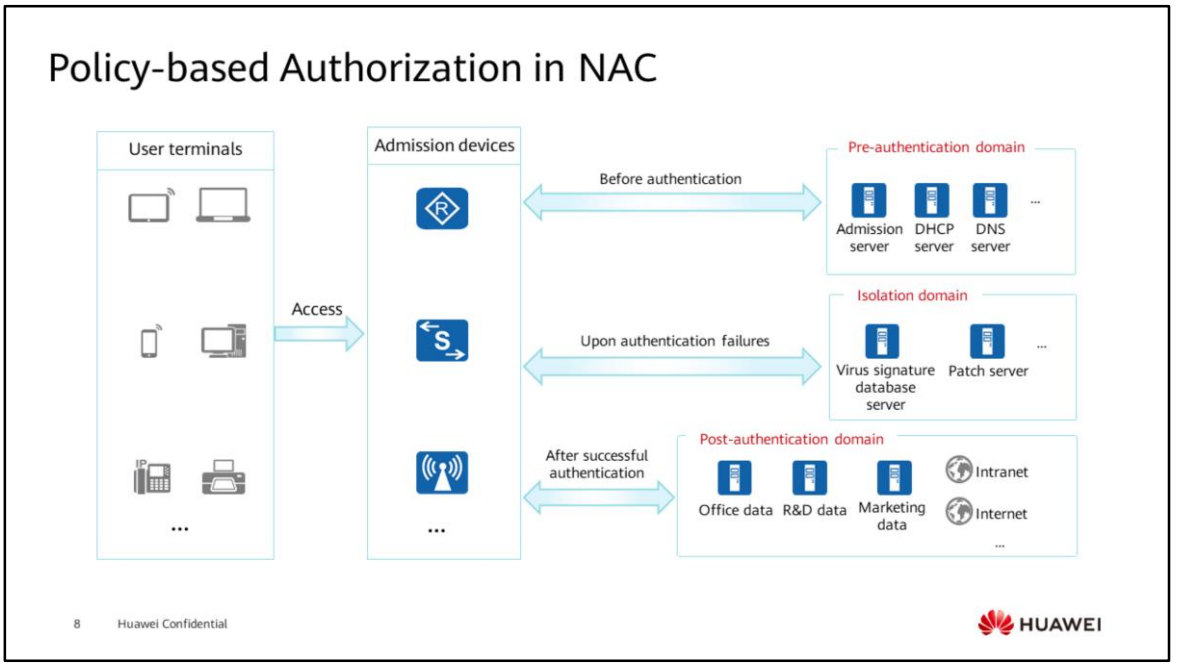
- User identity authentication request: A terminal sends the user credential to an admission device.
- User identity authentication: The admission device sends the user credential to the admission server for authentication.
- User identity verification: The admission server stores user identity information and provides user management functions. After receiving the user credential, the admission server determines whether the terminal identity is valid and delivers the verification result and corresponding policy to the admission device.
- User policy authorization: The admission device executes the policy based on the authorization result received from the admission server. For example, the admission device permits or denies access from the terminal. The admission device can also perform more complex policy-based control on the terminal, for example, increasing or decreasing the forwarding priority or limiting the network access rate.

Policy-based Control in NAC



- Policy-based control
  - Even if a terminal is authenticated and successfully accesses the network, it does not mean that the terminal can access all resources on the network. Rather, policies are used to grant terminals different network access rights based on user identities.

HUAWEI



- The basic NAC process is as follows:
    1. User can access the network and have the pre-authentication domain network permission before authentication, including access to the access control server, DHCP, and DNS.
    2. After terminals are successfully authenticated, the admission server delivers network access rights to the admission device to allow users to access resources in the post-authentication domain. For authorized but insecure users, they are granted the rights to access resources in the isolation domain. Only after network vulnerabilities on the terminals are fixed, the admission server grants the rights to access resources in the post-authentication domain to the users. In the isolation domain, users can access resources in the pre-authentication domain to install and upgrade terminal agent software, patches, and antivirus software.
    3. Unauthorized users and users who have not completed authentication are allowed to access resources only in the pre-authentication domain or isolation domain.

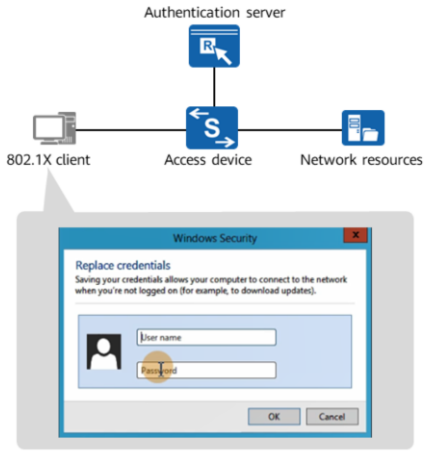
Contents

1. Overview of NAC
2. **User Authentication Technologies**
3. User Authorization and Logout
4. NAC Configuration
5. Policy Association

HUAWEI

802.1X MAC Portal Multi-mode

Overview of 802.1X Authentication



Introduction

- 802.1X authentication is a port-based network access control technology. That is, user identities are verified and network access rights are controlled on ports of access devices.
- 802.1X authentication uses the Extensible Authentication Protocol over LAN (EAPoL) to exchange authentication information between the client, access device, and authentication server.

Networking mode

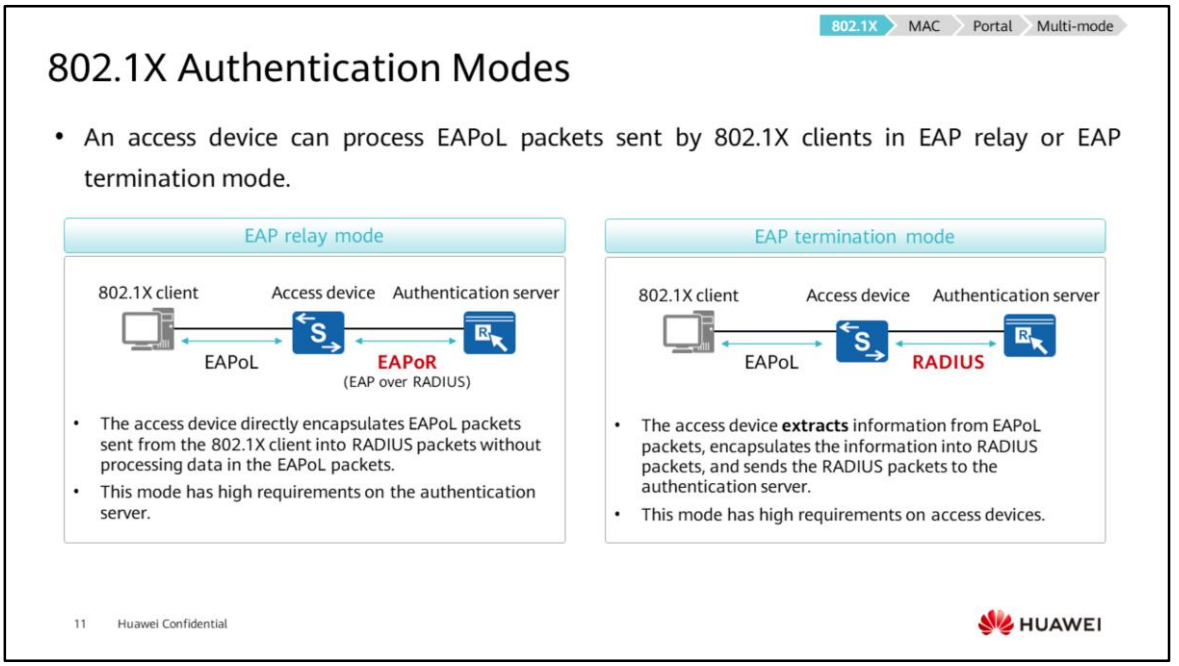
- 802.1X clients are usually user terminals. A user triggers 802.1X authentication using client software.
- An access device is usually a network device that supports 802.1X authentication. It provides a physical or logical interface for clients to access the LAN.
- An authentication server, which is typically a RADIUS server, carries out authentication, authorization, and accounting on users.

Application scenario

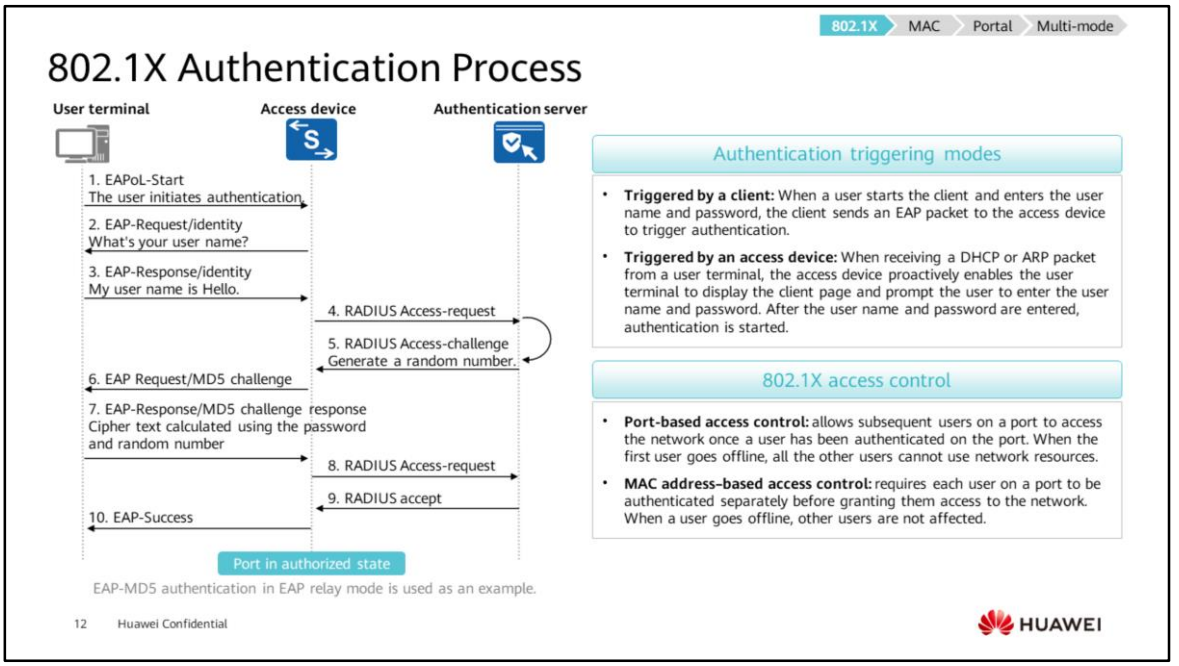
802.1X authentication applies to office users who have high security requirements.

HUAWEI

- The EAP packets transmitted between the client and access device are encapsulated in EAPoL format and transmitted across the LAN.
- Users can determine the authentication mode between the access device and authentication server based on the client support and network security requirements.
    - EAP termination mode: The access device terminates EAP packets and encapsulates them into RADIUS packets. The authentication server then uses the standard RADIUS protocol to implement authentication, authorization, and accounting.
    - EAP relay mode: The access device directly encapsulates the received EAP packets into EAP over RADIUS (EAPoR) packets, and then transmits these packets over a complex network to the authentication server.
- EAPoL defines EAP encapsulation on IEEE 802 (such as 802.3 and 802.11) networks. EAPoL only transmits EAP packets between 802.1X clients and access devices, and does not implement authentication.
- Typical EAP authentication protocols include Extensible Authentication Protocol Transport Layer Security (EAP-TLS), EAP Tunneled Transport Layer Security (EAP-TTLS), EAP Protected Extensible Authentication Protocol (EAP-PEAP), and EAP Message-Digest Algorithm 5 (EAP-MD5).



- EAP relay mode
  - This mode simplifies the processing on the access device and supports various authentication methods. However, the authentication server must support EAP and have high processing capability.
  - The commonly used authentication modes include EAP-TLS, EAP-TTLS, and EAP-PEAP. EAP-TLS has the highest security because it requires a certificate to be loaded on both the client and authentication server. EAP-TTLS and EAP-PEAP are easier to deploy since the certificate needs to be loaded only on the authentication server, but not the client.
- EAP termination mode
  - This mode is advantageous in that mainstream RADIUS servers support PAP authentication and CHAP authentication, eliminating the need of server upgrade. However, the workload on the access device is heavy because it needs to extract the client authentication information from the EAP packets sent by the client and encapsulate the information using the standard RADIUS protocol. In addition, the access device does not support other EAP authentication methods except MD5-Challenge.
  - The major difference between PAP and CHAP is that passwords in CHAP authentication are transmitted in cipher text, whereas passwords in PAP authentication are transmitted in plain text. In this aspect, CHAP provides higher security and is recommended.



- EAP relay authentication process:
    1. When a user needs to access an external network, the user starts the 802.1X client, enters the applied and registered user name and password, and initiates a connection request. The client then sends an authentication request packet (EAPoL-Start) to the access device to start the authentication process.
    2. After receiving the authentication request packet, the access device returns an EAP-Request/Identity packet, requesting the client to send the previously entered user name.
    3. In response to the request sent by the access device, the client sends an EAP-Response/Identity packet containing the user name to the access device.
    4. The access device encapsulates the EAP-Response/Identity packet into a RADIUS Access-Request packet and sends the RADIUS packet to the authentication server.
    5. After receiving the user name forwarded by the access device, the RADIUS server searches the user name table in the local database for the corresponding password, encrypts the password with a randomly generated MD5 challenge, and sends a RADIUS Access-Challenge packet containing the MD5 challenge to the access device.

Basic 802.1X Authentication Configuration (1)

1. Create an 802.1X access profile and enter the 802.1X access profile view.

   [Huawei] **dot1x-access-profile name** *access-profile-name*

   The device uses 802.1X access profiles to uniformly manage all 802.1X access configurations. By default, the device has a built-in 802.1X access profile named **dot1x_access_profile**.

2. Configure an authentication mode for 802.1X users.

   [Huawei-dot1x-access-profile-ProfileName] **dot1x authentication-method** { **chap** | **pap** | **eap** }

   In 802.1X authentication, the device supports two authentication modes:

   - EAP termination: The device parses EAP packets, encapsulates user authentication information into RADIUS packets, and sends the packets to the RADIUS server for authentication. You can specify the **chap** or **pap** parameter to configure the authentication mode between the device and RADIUS server.
   - EAP relay: The device directly encapsulates the received EAP packets containing user authentication information into RADIUS packets, and sends the RADIUS packets to the RADIUS server. This mechanism is also known as EAP over RADIUS (EAPoR).

HUAWEI

Basic 802.1X Authentication Configuration (2)

3. Configure the type of packets that can trigger 802.1X authentication.

   [Huawei-dot1x-access-profile-ProfileName] **authentication trigger-condition** { **dhcp** | **arp** | **dhcpv6** | **nd** | **any-l2-packet** }

   After 802.1X authentication is enabled, DHCP, DHCPv6, ND, and ARP packets can trigger 802.1X authentication by default.

4. Configure Layer 2 transparent transmission of 802.1X authentication packets.

   [Huawei] **l2protocol-tunnel user-defined-protocol** *protocol-name* **protocol-mac** *protocol-mac* **group-mac** *group-mac*

   [Huawei-GigabitEthernet0/0/1] **l2protocol-tunnel user-defined-protocol** *protocol-name* **enable**

   In the 802.1X authentication scenario, if there is a Layer 2 switch between the 802.1X-enabled access device and users, Layer 2 transparent transmission of 802.1X authentication packets must be enabled on the Layer 2 switch; otherwise, users cannot be successfully authenticated.

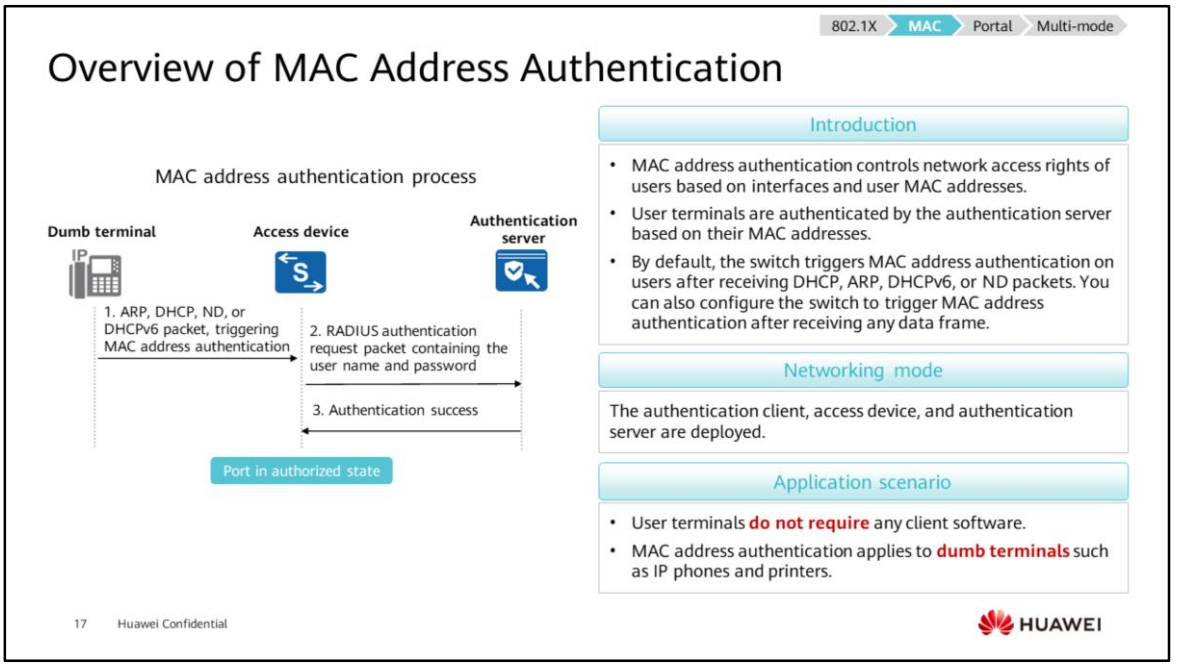
   - The **l2protocol-tunnel user-defined-protocol** command defines Layer 2 protocol packets, including the protocol name, multicast destination MAC address (specified by **protocol-mac**), and replacement multicast MAC address (**group-mac**). For 802.1X authentication, values of the **protocol-mac** and **group-mac** parameters are 0180-c200-0003 and 0100-0000-0002, respectively.
   - The **l2protocol-tunnel enable** command enables Layer 2 protocol tunneling on an interface.

HUAWEI

Summary of 802.1X Authentication

- 802.1X authentication is a port-based network access control technology that encapsulates protocol packets in EAPoL format.
- You can set 802.1X authentication parameters based on the actual network requirements or device capabilities:
  - 802.1X access control: interface-based or MAC address–based
  - Authentication mode: EAP termination or EAP relay
  - Authentication triggering mode: triggered by a client or access device
  - Authentication protocol: EAP-TLS, EAP-TTLS, EAP-PEAP, or EAP-MD5

HUAWEI



- Dumb terminal: Compared with other terminals, dumb terminals have limited functions and simple interaction modes. In this document, dumb terminals refer to terminals whose authentication information such as user names and passwords cannot be entered.
- By default, a MAC address without hyphens (-) is used as the user name and password for MAC address authentication, for example, 0005e0112233.
- Passwords of MAC address authentication users can be processed using PAP or CHAP. The following MAC address authentication process uses PAP as an example:
    1. When a terminal accesses the network, the access device detects and learns the MAC address of the terminal, triggering MAC address authentication.
    2. The access device generates a random value (MD5 challenge), arranges the user MAC address, password, and random value in sequence, encrypts them using the MD5 algorithm, encapsulates the encryption results into a RADIUS authentication request packet, and sends the packet to the RADIUS server.
    3. The RADIUS server arranges the user MAC address, password saved in the local database, and received random value in sequence, and uses the random value to encrypt them using the MD5 algorithm. If the encrypted password is the same as that received from the access device, the RADIUS server sends an authentication accept packet to the access device, indicating that MAC address authentication is successful and the terminal is allowed to access the network.
- Different from PAP, CHAP arranges CHAP ID，the user MAC address, and random value in sequence, encrypts them using the MD5 algorithm.

MAC Address Authentication Configuration (1)

1. Create a MAC access profile and enter the MAC access profile view.

   [Huawei] **mac-access-profile name** *access-profile-name*

   The device uses MAC access profiles to uniformly manage access configurations of MAC address authentication users. By default, the device has a built-in MAC access profile named **mac_access_profile**.

2. Configure an authentication mode for MAC address authentication users.

   [Huawei-mac-access-profile-ProfileName] **mac-authen authentication-method** { **chap** | **pap** }

   By default, PAP authentication is used.

   During MAC address authentication, the access device and authentication server exchange RADIUS packets in either PAP or CHAP mode. CHAP is more secure than PAP.

3. Configure the type of packets that can trigger MAC address authentication.

   [Huawei-mac-access-profile-ProfileName] **authentication trigger-condition { dhcp | arp | dhcpv6 | nd | any-l2-packet }**

   By default, DHCP, ARP, DHCPv6, and ND packets can trigger MAC address authentication.

HUAWEI

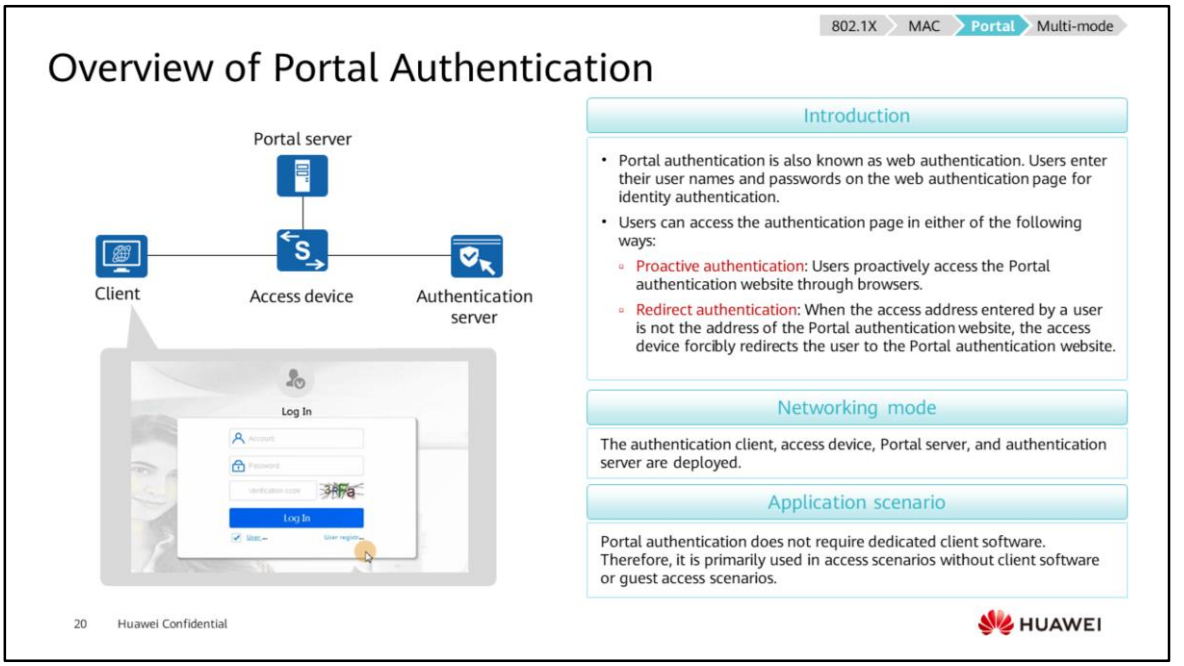
MAC Address Authentication Configuration (2)

4. Configure the user name format for MAC address authentication.

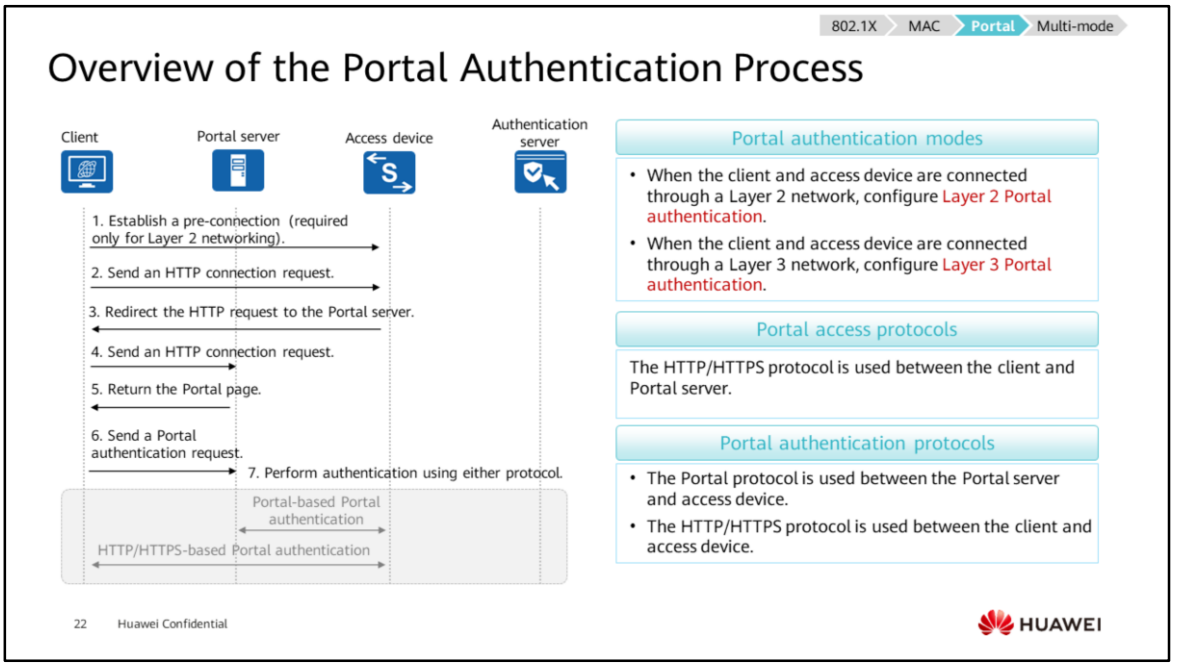
[Huawei-mac-access-profile-ProfileName] **mac-authen username** { **fixed** *username* [**password cipher** *password*] | **macaddress** [**format** { **with-hyphen** | **without-hyphen** } [**password cipher** *password*]]}

- **fixed** *username* specifies a fixed user name for MAC address authentication. If no password is configured, users can log in without a password, which is not recommended.
- **macaddress** specifies the MAC addresses of users as user names for MAC address authentication. If no password is configured, the MAC address of a user is used as the password. When local authentication is used, a password must be configured.
  - **with-hyphen**: indicates that the MAC address contains delimiters.
  - **without-hyphen**: indicates that the MAC address does not contain delimiters.

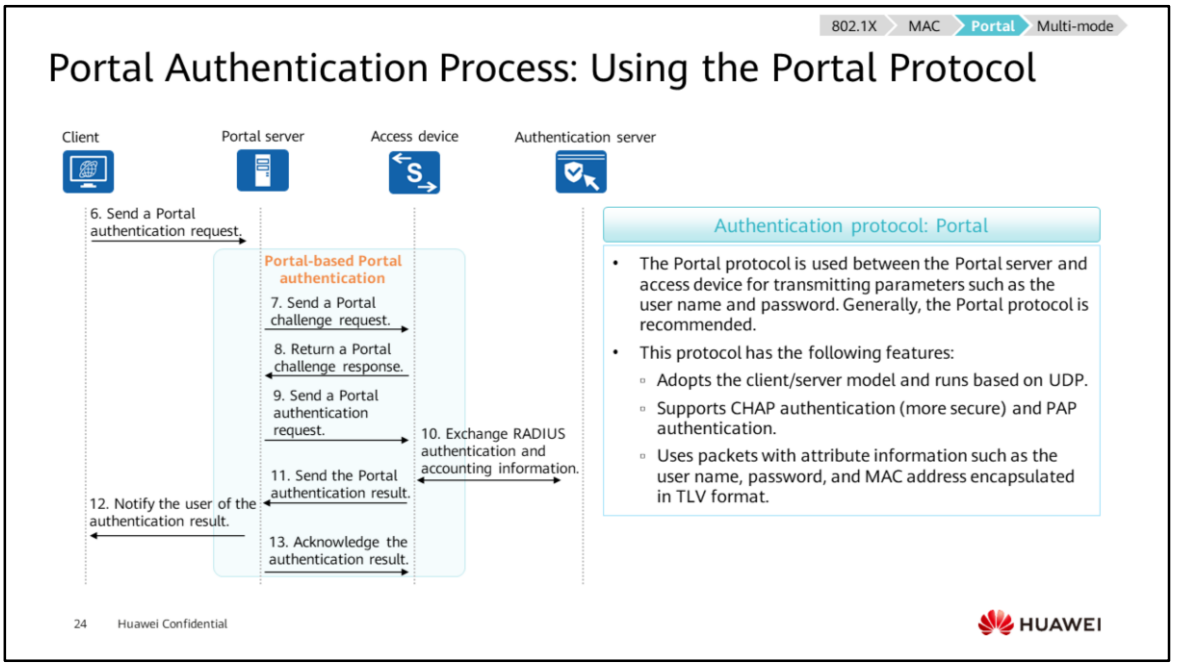
HUAWEI



- Client: In most cases, a client is a host where an HTTP/HTTPS-capable browser is installed.
- Access device: a network device such as a switch or router, which provides the following functions:
  - Redirects all HTTP and HTTPS requests of users on authentication subnets to the Portal server before authentication is performed.
  - Interacts with the Portal server and authentication server to implement user authentication, authorization, and accounting.
  - Grants users access to specified network resources upon successful authentication.
- Portal server: a server system that receives authentication requests from clients, provides Portal services and authentication pages, and exchanges client authentication information with access devices.
- Authentication server: interacts with access devices to implement user authentication, authorization, and accounting.



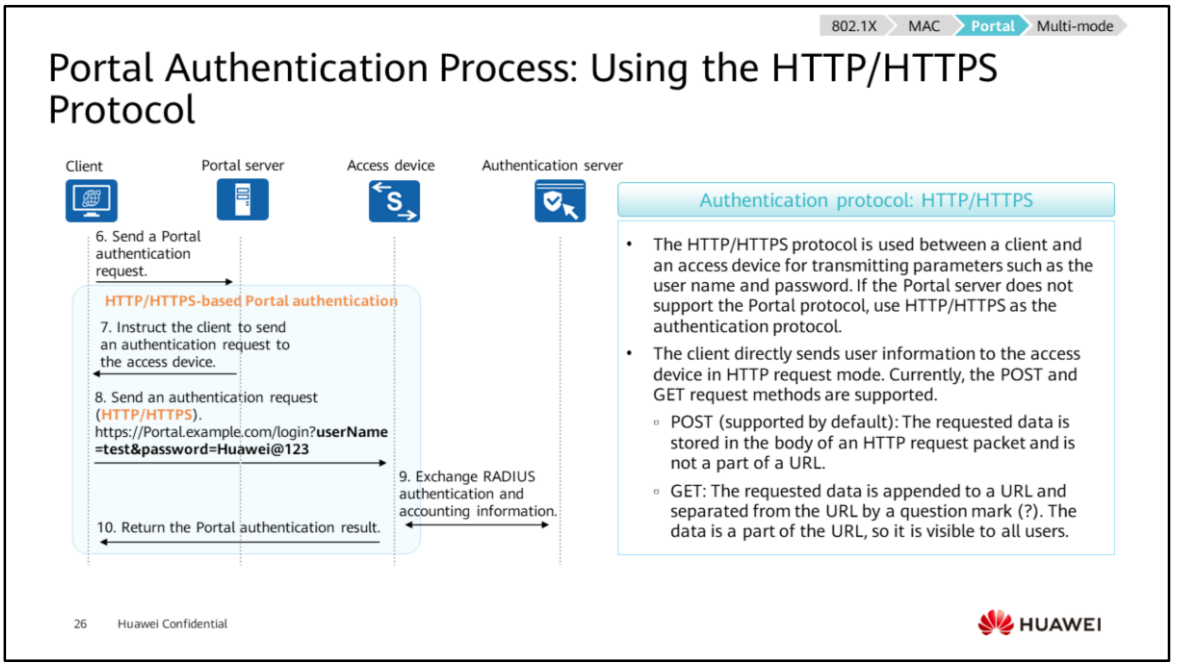
- Select a Portal authentication mode based on the actual network requirements.
  - When Layer 2 authentication is used, the device can learn users' MAC addresses and identify the users based on their MAC addresses and IP addresses. Layer 2 authentication provides a simple authentication process while ensuring high security. However, users must be in the same network segment as the access device, causing inflexible networking.
  - When Layer 3 authentication is used, the device cannot obtain the MAC address of a client, so it identifies the user based only on the client IP address. Layer 3 authentication allows for flexible networking and facilitates remote control. However, users can only be identified based on their IP addresses, leading to poor security.
- The Portal authentication process is as follows:
  1. Before authentication, the client establishes a pre-connection with the access device. The access device creates a user online entry for the client and grants the client access to certain network resources. The Layer 3 authentication process is similar to the Layer 2 authentication process, except that no pre-connection is established between the client and access device.
  2. The client initiates an HTTP connection request.



- The following uses CHAP authentication as an example to describe the Portal-based Portal authentication process:

    7. After receiving the Portal authentication request, the Portal server sends a Portal challenge request packet to the access device. This step is performed only when CHAP authentication is used between the Portal server and access device. If PAP authentication is used, steps 7 and 8 are not performed.
    8. The access device sends a Portal challenge response packet to the Portal server.
    9. The Portal server encapsulates the entered user name and password into a Portal authentication request packet and sends the packet to the access device.
    10. The access device and RADIUS server exchange user information to authenticate the user, including:

        - The access device encapsulates the entered user name and password into a RADIUS authentication request packet and sends the packet to the RADIUS server.



- HTTPS is a secure HTTP and also known as HyperText Transfer Protocol over Transport Layer Security (HTTP over TLS) or HyperText Transfer Protocol over Secure Socket Layer (HTTP over SSL). HTTPS uses HTTP for communication and SSL/TLS for data encryption.
- A URL is a concise representation of the location and access method of a resource that can be obtained from the Internet. It is the address of a standard resource on the Internet. Each file on the Internet has a unique URL. The URL contains information about the location of the file and how a browser should process the file.
- When HTTP/HTTPS-based Portal authentication is used, the authentication process is as follows:
    1. The Portal server instructs the client to send a Portal authentication request to the access device.
    2. The client sends a Portal authentication request to the access device.
    3. After receiving the Portal authentication request, the access device parses the packet according to parameter names to obtain parameters such as the user name and password, and then sends the obtained user name and password to the RADIUS server for authentication. The process is similar to the Portal-based Portal authentication.
    4. The access device returns the Portal authentication result to the client and adds the user to the local online user list.
- As shown in the figure, an HTTP request is sent in Get mode: https://Portal.example.com/login?userName=test&password=Huawei@123. You can see that the user name and password are in plain text and are separated from the URL by a question mark (?).

Portal Authentication Configuration - Configuring an External Portal Server (1)

1. Enables Portal interconnection using the HTTP/HTTPS protocol.

   [Huawei] **portal web-authen-server** { **http** | **https ssl-policy** *policy-name* } [**port** *port-number*]

   When HTTP/HTTPS-based Portal authentication is used, you need to enable Portal interconnection using the HTTP/HTTPS protocol on the device. If HTTPS is used, you also need to specify an SSL policy name.

2. Create a Portal server template and enter the Portal server template view.

   [Huawei] **web-auth-server** *server-name*

   A Portal server template is used to uniformly configure parameters of an external Portal server, such as the IP address and URL of the Portal server.

3. Configure the protocol used for Portal authentication.

   [Huawei-web-auth-server-ServerName] **protocol** { **http** [**password-encrypt** { **none** | **uam** }] | **portal** }

   You can configure the HTTP/HTTPS or Portal protocol for Portal authentication based on your network requirements.

HUAWEI

Portal Authentication Configuration - Configuring an External Portal Server (2)

4. Configure parameters for the Portal server when the Portal protocol is used for Portal authentication.

```
[Huawei-web-auth-server-ServerName] server-ip server-ip-address &<1-10>
```

```
[Huawei-web-auth-server-ServerName] source-ip ip-address
```

```
[Huawei-web-auth-server-ServerName] shared-key cipher key-string
```

```
[Huawei-web-auth-server-ServerName] url url-string
```

Command functions:

- **server-ip**: configures the IP address of the Portal server.
- **source-ip**: configures the IP address used by the device to communicate with the Portal server.
- **shared-key**: configures the shared key used by the device to exchange information with the Portal server.
- **url**: configures a redirect URL or pushed URL to be sent to access terminals.

5. Configure parameters for the Portal server when the HTTP/HTTPS protocol is used for authentication.

```
[Huawei-web-auth-server-ServerName] url url-string
```

Only the URL of the Portal server needs to be configured.

HUAWEI

Portal Authentication Configuration - Configuring a Portal Access Profile

1. Create a Portal access profile and enter the Portal access profile view.

   [Huawei] **portal-access-profile name** *access-profile-name*

   The device uses Portal access profiles to uniformly manage all access configurations of Portal users.

   By default, the device has a built-in Portal access profile named **portal_access_profile**.

2. Configure a Portal server template used by the Portal access profile.

   [Huawei-portal-acces-profile-ProfileName] **web-auth-server** *server-name* { **direct** | **layer3** }

   When a user in the Portal access profile accesses restricted network resources, the HTTP request of the user is forcibly redirected to the authentication page of the Portal server for Portal authentication.

   - If there is no Layer 3 forwarding device between the user and device (authentication point), specify the **direct** parameter to enable Layer 2 Portal authentication. In this mode, the device identifies users based on their IP addresses and MAC addresses.
   - When a Layer 3 forwarding device exists between the user and device, specify the **layer3** parameter to enable Layer 3 Portal authentication. In this mode, the device identifies users based only on their IP addresses.

HUAWEI

Comparison Between User Authentication Technologies

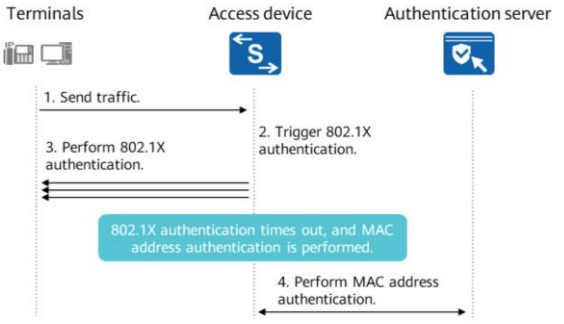
- Huawei NAC solution supports various user authentication technologies, including 802.1X authentication, MAC address authentication, and Portal authentication. Authentication points can be flexibly deployed on network devices such as access switches, aggregation switches, access controllers (ACs), routers, and firewalls to implement NAC by working with iMaster NCE-Campus.

Item	802.1X Authentication	Portal Authentication	MAC Address Authentication
Client software	Required	Not required	Not required
Advantage	High security	Flexible deployment	Client software not required
Disadvantage	Inflexible deployment	Low security	MAC address registration required, making management complicated
Application scenario	New network with concentrated users and high requirements for security	Scenario with scattered and moving users	Access authentication of dumb terminals such as printers and fax machines

HUAWEI

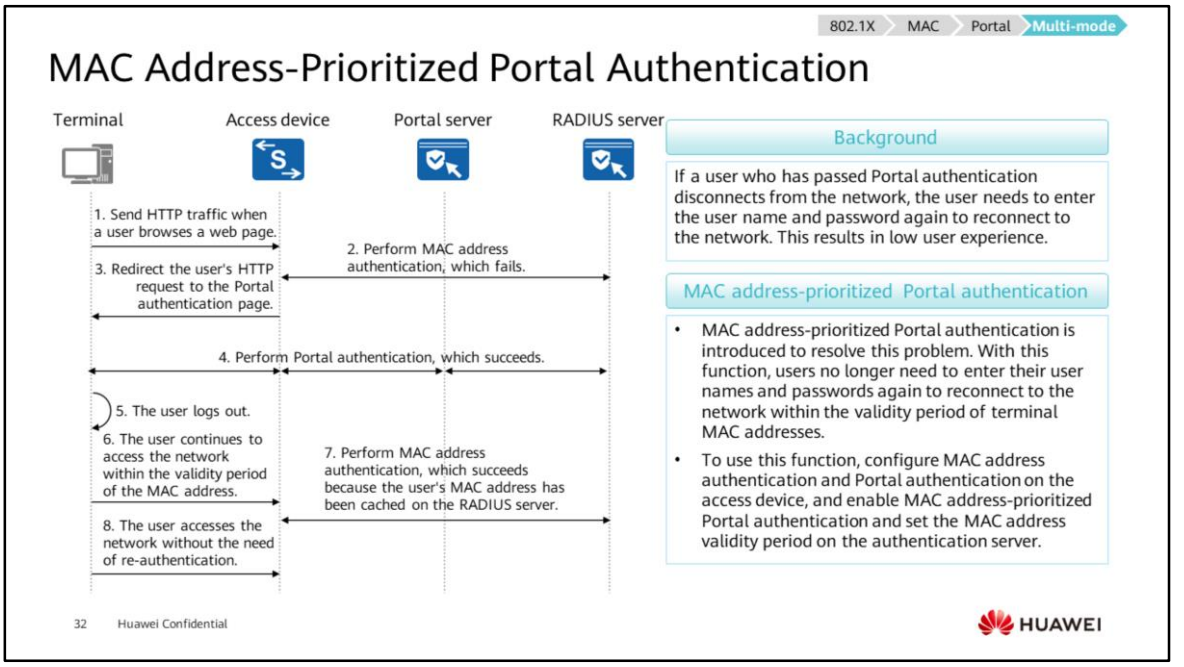
MAC Address Bypass Authentication

- 802.1X authentication, MAC address authentication, and Portal authentication have their own characteristics. You can use multi-mode authentication to meet authentication requirements in different scenarios.



- Dumb terminals such as printers and fax machines do not support 802.1X authentication. When both PCs and dumb terminals are connected to an interface of an access device, you can configure MAC address bypass authentication to allow the dumb terminals to access the network using MAC address authentication.
- MAC address bypass authentication takes a longer period of time than MAC address authentication because it has an 802.1X authentication stage additionally.

HUAWEI

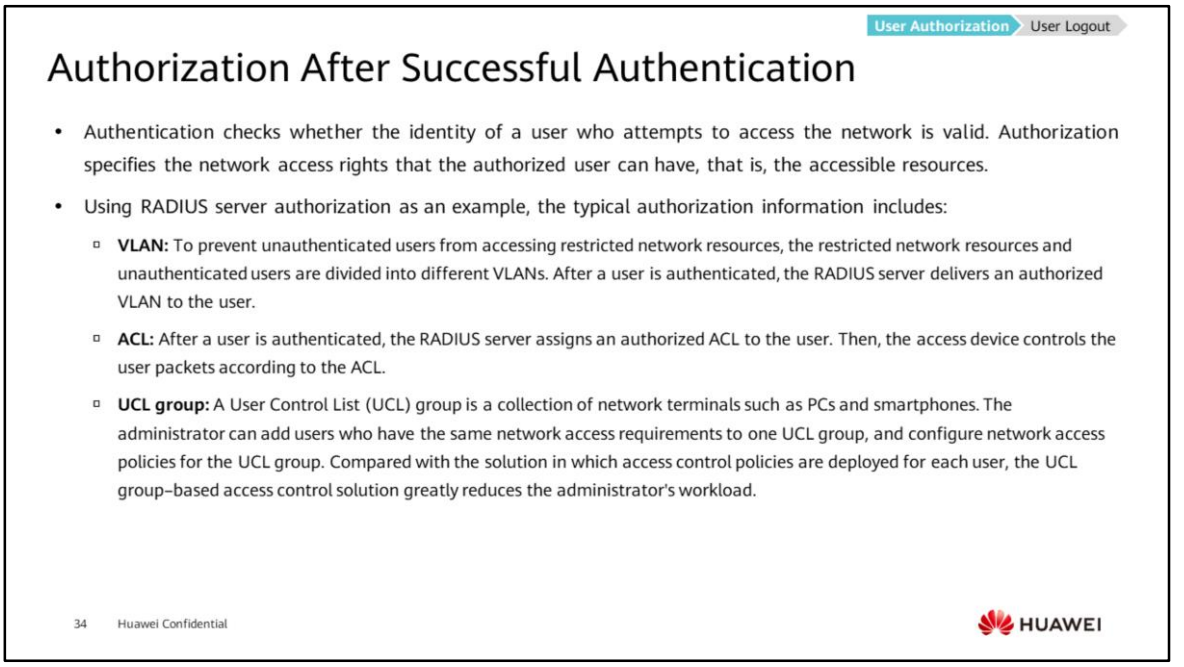


- After passing Portal authentication, terminals may be disconnected from the wireless network when they move from one wireless signal coverage area to another or when the wireless signal is unstable. In this case, users need to enter their user names and passwords for identity authentication every time the terminals go online, leading to poor network access experience. MAC address-prioritized Portal authentication is used to resolve this problem. Generally, there is no need to enable MAC address-prioritized Portal authentication on wired networks that provide stable signals.

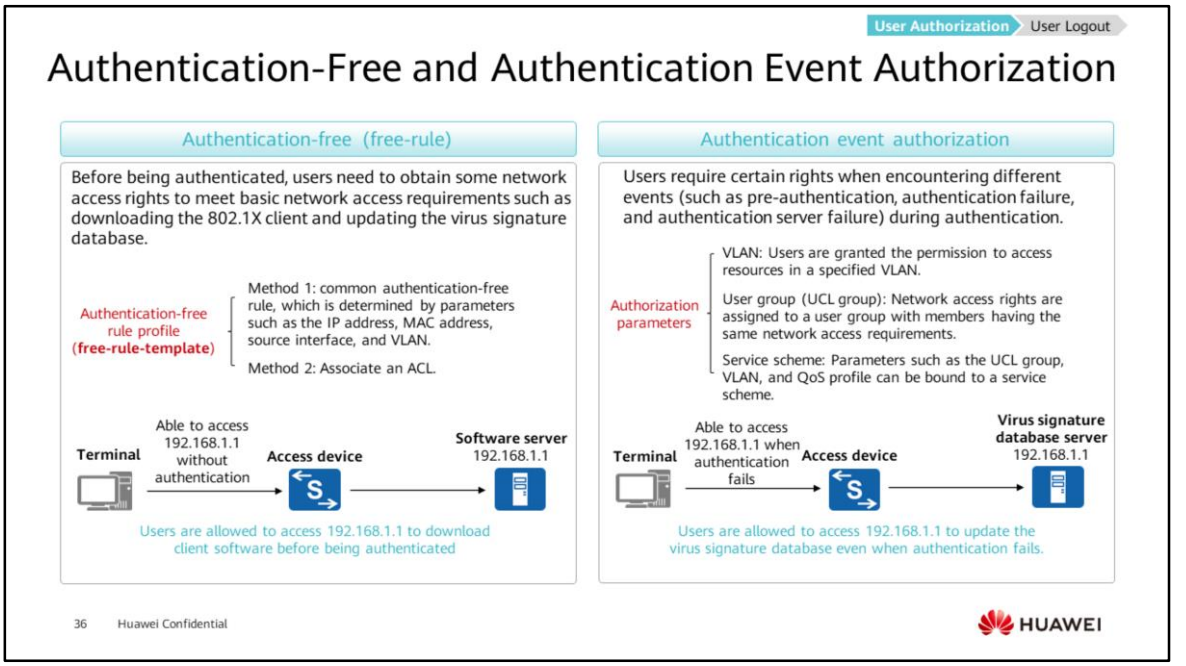
Contents

1. Overview of NAC
2. User Authentication Technologies
3. **User Authorization and Logout**
4. NAC Configuration
5. Policy Association

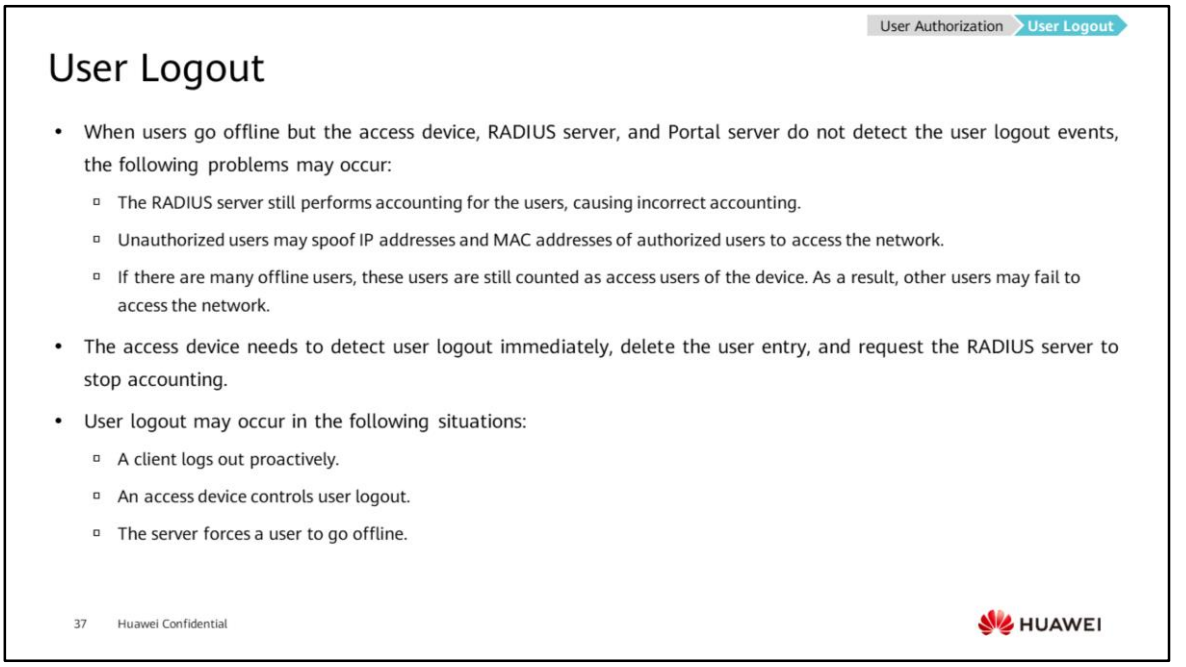
HUAWEI



- When the RADIUS server is used, the authentication accept packet also contains user authorization information because RADIUS authorization is combined with authentication.
- VLAN-based authorization by the RADIUS server:
  - After a user is authenticated, the RADIUS server delivers an authorized VLAN to the user. The access device then changes the VLAN to which the user belongs to the authorized VLAN, with the interface configuration remaining unchanged. The authorized VLAN has a higher priority than the VLAN configured on the interface. That is, the authorized VLAN takes effect after the authentication succeeds, and the configured VLAN takes effect when the user is offline.
- The RADIUS server can assign an authorized ACL to a user in either of the following modes:
  - Static ACL assignment: The RADIUS server uses the standard RADIUS attribute **Filter-Id** to assign an ACL ID to the user. In this mode, the ACL and corresponding rules are configured on the access device in advance.
  - Dynamic ACL assignment: The RADIUS server uses the Huawei extended RADIUS attribute **HW-Data-Filter** to assign an ACL ID and corresponding rules to the user. In this mode, the ACL ID and ACL rules are configured on the RADIUS server.



- When an authentication-free rule is configured using an ACL, the ACL number is in the range from 6000 to 6031.
- The NAC escape mechanism grants specified network access rights to users when the authentication server is Down or to users who fail the authentication or are in pre-connection state. The escape solutions vary according to the authentication modes. Some escape solutions are shared by all authentication modes, while some are supported only in specific authentication modes. For details, see "NAC Escape Mechanism" in the product documentation.

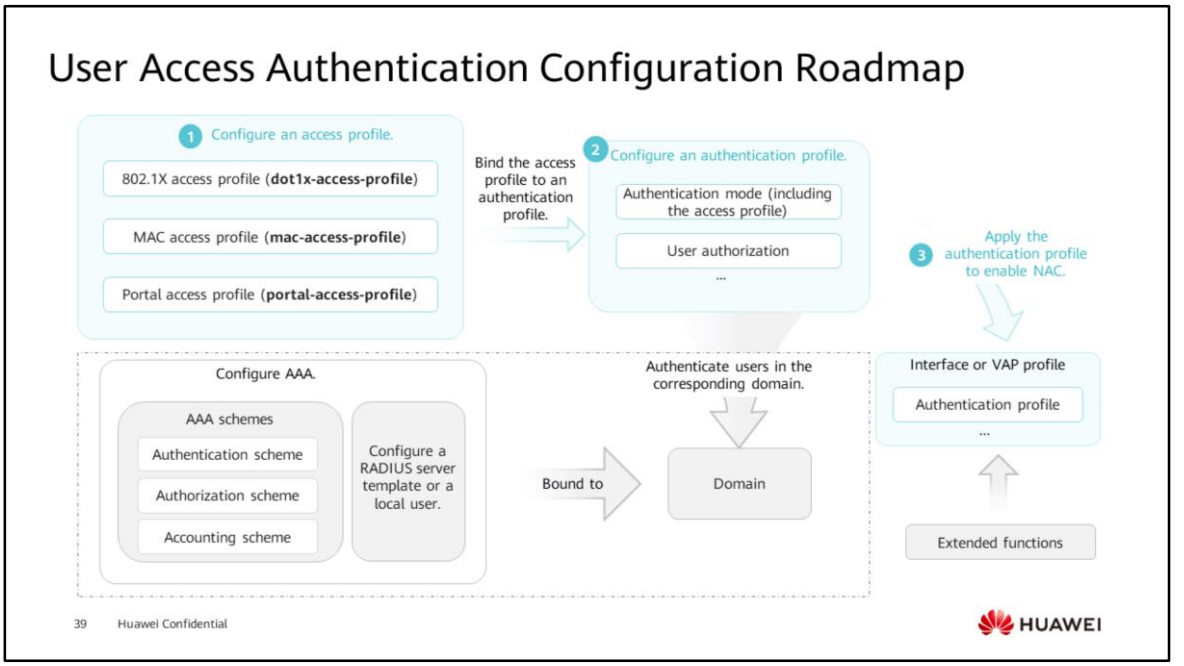


- Note:
  - MAC address authentication supports only user logout control by the access device and server.
  - Portal authentication allows both the authentication server and Portal server to control user logout.

Contents

1. Overview of NAC
2. User Authentication Technologies
3. User Authorization and Logout
4. **NAC Configuration**
5. Policy Association

HUAWEI



- VAP indicates a virtual access point.
- VAP profile: Configure WLAN parameters in a VAP profile, and bind the VAP profile to an AP group or AP. Then, a VAP is generated on the AP to provide wireless access services for STAs.

Authentication Profile Configuration (1)

1. Create an authentication profile and enter the authentication profile view.

   [Huawei] **authentication-profile name** *authentication-profile-name*

   You can configure parameters in an authentication profile to implement varying access control for different users on the device. After configuring an authentication profile, apply it to an interface or a VAP profile to enable NAC.

2. Bind an access profile to an authentication profile.

   [Huawei-authen-profile-ProfileName] **dot1x-access-profile** *access-profile-name*

   [Huawei-authen-profile-ProfileName] **mac-access-profile** *access-profile-name*

   [Huawei-authen-profile-ProfileName] **portal-access-profile** *access-profile-name*

   You can bind an access profile of the required type to an authentication profile. When more than one access profile is bound to an authentication profile, multi-mode authentication is used. There is no limitation on the sequence of binding access profiles. The device triggers authentication of a specific type based on the authentication packet received. You can bind one 802.1X access profile, one MAC access profile, and one Portal access profile to the same authentication profile at most.

HUAWEI

Authentication Profile Configuration (2)

3. Enable MAC address bypass authentication.

   [Huawei-authen-profile-ProfileName] **authentication dot1x-mac-bypass**

   MAC address bypass authentication combines 802.1X authentication and MAC address authentication. Before enabling this function in an authentication profile, ensure that an 802.1X access profile and a MAC access profile have been bound to the authentication profile.

4. Configure a default domain or forcible domain for users.

   [Huawei-authen-profile-ProfileName] **access-domain** *domain-name* [**dot1x** | **mac-authen** | **portal**] * [**force**]

   The device manages users in domains. **force** configures a forcible domain. If this parameter is not specified, a default domain is configured. The **dot1x**, **mac-authen**, and **portal** parameters specify the type of users for whom a default domain or forcible domain is configured.

   - When a default domain is configured, the device authenticates users based on the domain names contained in the user names. If a user name does not contain any domain name, the device authenticates the user in the default domain.
   - When a forcible domain is configured, the device authenticates users in the forcible domain, regardless of whether the user names contain domain names.

HUAWEI

Extended Function: Configuring Static Users

- During network deployment, static IP addresses are assigned to dumb terminals such as printers, which can be configured as static users for flexible authentication.

1. Configure parameters such as the IP address range, domain name, connected interface, and VLAN for a static user.

```
[Huawei] static-user start-ip-address [ end-ip-address ] [ domain-name domain-name | interface interface-type interface-number | mac-address mac-address | vlan vlan-id ] *
```

2. Configure parameters such as the user name and password of the static user.

```
[Huawei] static-user username macaddress format { with-hyphen [ normal ] [ colon ] | without-hyphen } [ uppercase ] [ password-with-macaddress ]
```

```
[Huawei] static-user username format-include { ip-address | mac-address | system-name }
```

```
[Huawei] static-user password cipher password
```

- After a static user is configured, you can enable 802.1X authentication, MAC address authentication, or Portal authentication on the interface connected to the user. Then, the device uses static user information such as the user IP address as the user name for authentication.

HUAWEI

- After a static user is configured, the device preferentially uses the user name and password of the static user to authenticate the user when detecting that the user information matches the parameters such as the IP address range and domain name configured for the static user. If the authentication fails, the device performs 802.1X, MAC address, or Portal authentication on the user.
- You can run the **static-user username macaddress format** command to specify the MAC address of a terminal as the user name and password for authentication, as well as the user name format. This command has a higher priority than the **static-user username format-include and static-user password cipher password** commands.
- The **static-user username format-include** and **static-user password** commands are used to configure the user name and password of a static user respectively.
- By default (the S5731 is used as an example):
  - The user name of a static user is a combination of [system-name] and [ip-address]. For example, when the system name of an access device is **huawei** and the user IP address is 1.1.1.1, then the static user name is **huawei1.1.1.1**.
  - The password of a static user is not configured.

Applying NAC

- An authentication profile is used to manage NAC configurations in a unified manner. To enable NAC, apply an authentication profile to an interface or VAP profile. This implements access control on users connected to the corresponding interface or VAPs.
- The authentication method for users is determined by the access profile bound to the authentication profile.

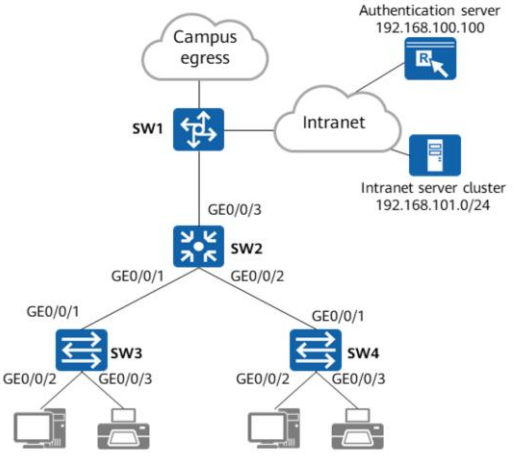
1. Enter the interface view or VAP profile view.
2. Apply the configured authentication profile. The following uses GE0/0/1 as an example.

[Huawei-GigabitEthernet0/0/1] **authentication-profile** *authentication-profile-name*

The **authentication-profile** command applies an authentication profile to an interface or VAP profile.
By default, no authentication profile is applied to an interface or VAP profile. Different types of interfaces or VAP profiles may support different NAC functions. For details, see the product documentation.

HUAWEI

Configuration Example: Networking Requirements and Authentication Planning



Networking requirements

To enhance network security, an enterprise requires that all terminals (such as PCs, printers, and IP phones) be authenticated before accessing the network.

The enterprise network has the following characteristics:

- Access switches do not support 802.1X authentication.
- The enterprise network is small in scale and does not have any branches.
- The enterprise has no more than 1000 employees. A maximum of 2000 terminals, including guest terminals, access the network every day.
- Dumb terminals, such as IP phones and printers, are connected to the enterprise network.

Authentication planning

- Deploy 802.1X authentication control points on the aggregation switch.
- Deploy MAC address authentication for access of dumb terminals.
- The authenticated users can access the intranet server cluster.

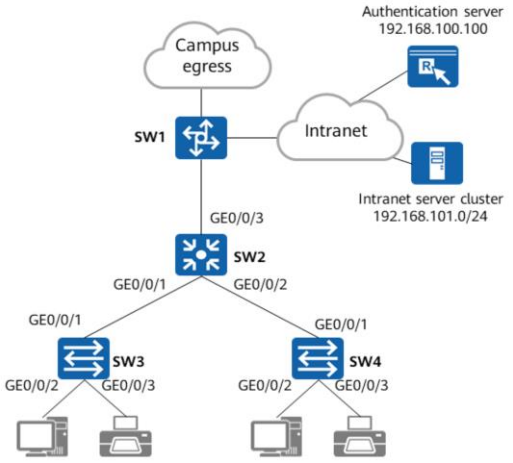
HUAWEI

Configuration Example: Data Plan

Item	Data
RADIUS server	IP address: 192.168.100.100
Intranet server cluster	Network segment: 192.168.101.0/24
Aggregation switch (SW2)	• VLAN to which the uplink interface GE0/0/3 belongs: VLAN 100 • VLAN to which downlink interfaces GE0/0/1 and GE0/0/2 belong: VLAN 200
Access switches (SW3 and SW4)	User VLAN ID: 200
RADIUS scheme	• Authentication server IP address: 192.168.100.100 • Authentication server port number: 1812 • Accounting server IP address: 192.168.100.100 • Accounting server port number: 1813 • Shared key of the RADIUS server: Huawei@123 • Authentication domain: nac
ACL number of the post-authentication domain	3001

HUAWEI

Configuration Example: Procedure (1)



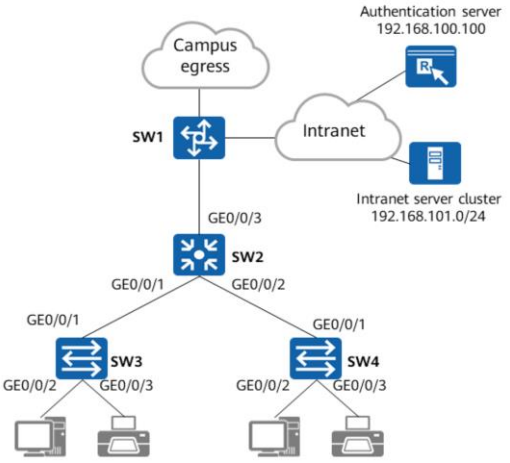
1. Enable network connectivity.
2. Configure transparent transmission of 802.1X packets. SW3 is used as an example.

```
[SW3] l2protocol-tunnel user-defined-protocol 802.1X protocol-mac 0180-c200-0003 group-mac 0100-0000-0002
[SW3] interface gigabitethernet 0/0/1
[SW3-GigabitEthernet0/0/1] l2protocol-tunnel user-defined-protocol 802.1X enable
[SW3-GigabitEthernet0/0/1] bpdu enable
[SW3-GigabitEthernet0/0/1] quit
[SW3] interface gigabitethernet 0/0/2
[SW3-GigabitEthernet0/0/2] l2protocol-tunnel user-defined-protocol 802.1X enable
[SW3-GigabitEthernet0/0/2] bpdu enable
[SW3-GigabitEthernet0/0/2] quit
[SW3] interface gigabitethernet 0/0/3
[SW3-GigabitEthernet0/0/3] l2protocol-tunnel user-defined-protocol 802.1X enable
[SW3-GigabitEthernet0/0/3] bpdu enable
[SW3-GigabitEthernet0/0/3] quit
```


HUAWEI

- In this example, SW3 and SW4 are deployed between the authentication switch SW2 and users. Therefore, transparent transmission of 802.1X packets must be configured on SW3 and SW4 so that SW2 can perform 802.1X authentication on users.

Configuration Example: Procedure (2)



3. Configure a RADIUS server template.

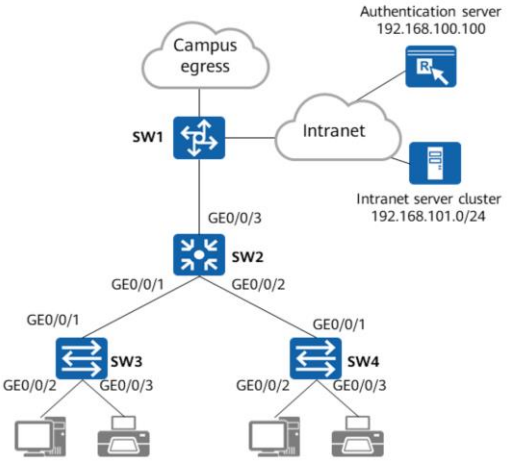
```
[SW2] radius-server template rd1
[SW2-radius-rd1] radius-server authentication 192.168.100.100 1812
[SW2-radius-rd1] radius-server accounting 192.168.100.100 1813
[SW2-radius-rd1] radius-server shared-key cipher Huawei@123
[SW2-radius-rd1] quit
```

4. Configure an AAA authentication scheme, accounting scheme, and an authentication domain.

```
[SW2] aaa
[SW2-aaa] authentication-scheme a1
[SW2-aaa-authen-a1] authentication-mode radius
[SW2-aaa-authen-a1] quit
[SW2-aaa] accounting-scheme a2
[SW2-aaa-accounting-a2] accounting-mode radius
[SW2-aaa-accounting-a2] quit
[SW2-aaa] domain nac
[SW2-aaa-domain-nac] authentication-scheme a1
[SW2-aaa-domain-nac] accounting-scheme a1
[SW2-aaa-domain-nac] radius-server rd1
[SW2-aaa-domain-nac] quit
[SW2-aaa] quit
```


HUAWEI

Configuration Example: Procedure (3)



5. Configure an 802.1X access profile.

```
[SW2] dot1x-access-profile name d1
[SW2-dot1x-access-profile-d1] dot1x authentication-method eap
[SW2-dot1x-access-profile-d1] quit
```

6. Configure a MAC access profile.

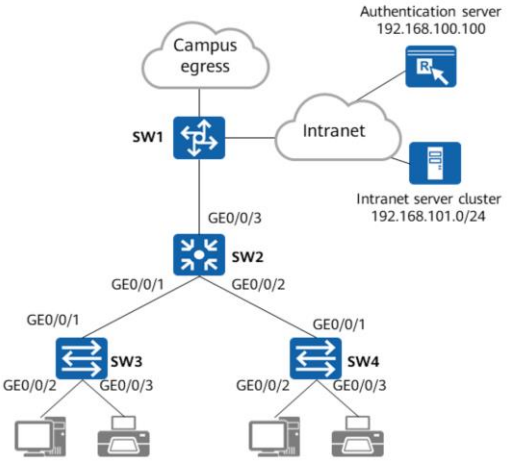
```
[SW2] mac-access-profile name m1
[SW2-mac-access-profile-m1] mac-authen username macaddress
[SW2-mac-access-profile-m1] quit
```

7. Configure an authentication profile.

```
[SW2] authentication-profile name p1
[SW2-authen-profile-p1] mac-access-profile m1
[SW2-authen-profile-p1] dot1x-access-profile d1
[SW2-authen-profile-p1] access-domain nac force
[SW2-authen-profile-p1] quit
```


HUAWEI

Configuration Example: Procedure (4)



8. Apply the authentication profile to interfaces.

```
[SW2] interface gigabitethernet 0/0/1
[SW2-Gigabitethernet0/0/1] authentication-profile p1
[SW2-Gigabitethernet0/0/1] quit
[SW2] interface gigabitethernet 0/0/2
[SW2-Gigabitethernet0/0/2] authentication-profile p1
[SW2-Gigabitethernet0/0/2] quit
```

9. Configure the authorized ACL to be delivered upon successful user authentication.

```
[SW2] acl 3001
[SW2-acl-adv-3001] rule 1 permit ip destination 192.168.101.0 0.0.0.255
[SW2-acl-adv-3001] rule 2 deny ip destination any
[SW2-acl-adv-3001] quit
```

10. Create user accounts on the RADIUS server and configure corresponding rights.

HUAWEI

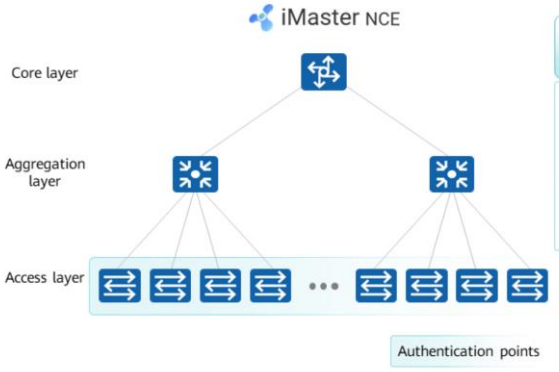
Contents

1. Overview of NAC
2. User Authentication Technologies
3. User Authorization and Logout
4. NAC Configuration
5. **Policy Association**

HUAWEI

Technical Background (1)

- Deploying authentication at the access layer helps implement fine-grained permission management and high network security. As the network scale expands, some problems emerge.



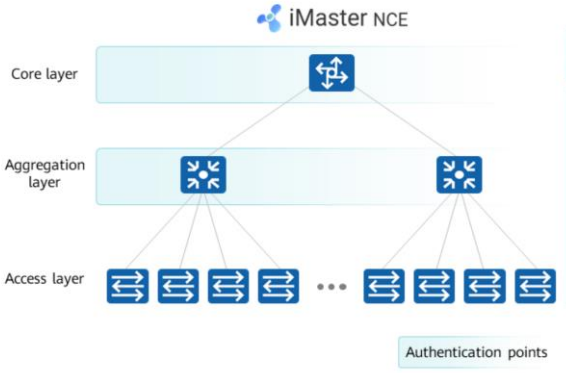
Problems faced when access devices are used as authentication points

- A large number of devices are deployed at the access layer, causing heavy configuration workload and difficult O&M.
- The large number of access devices increases the load of the AAA server.
- Users must access the network at fixed locations.

HUAWEI

Technical Background (2)

- One solution is to move user authentication points from the access layer to the aggregation or core layer. In this way, authentication information on the entire network is centralized, the number of authentication points is greatly reduced, and the configuration and maintenance workload is reduced.



Problems faced when authentication points are moved upwards

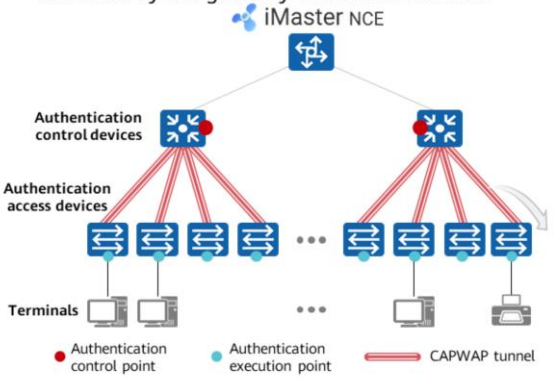
- Devices at the access layer must transparently transmit BPDUs. Otherwise, 802.1X authentication fails.
- Authentication points cannot control mutual access between users in the same VLAN on an access device.
- An administrator does not know the access positions of users, making fault locating difficult.
- The gateway cannot detect user logout in real time.

HUAWEI

- Devices at the access layer must transparently transmit BPDUs. Otherwise, 802.1X authentication fails. The reason is as follows:
  - EAP packets transmitted in 802.1X authentication are BPDUs. By default, Huawei switches do not perform Layer 2 forwarding for BPDUs. If a Layer 2 switch exists between the 802.1X-enabled device and users, Layer 2 transparent transmission must be enabled on the Layer 2 switch. Otherwise, the EAP packets sent by users cannot reach the 802.1X-enabled device, causing authentication failures.

Overview of Policy Association

- Policy association provides a solution to contradiction between policy strengths and complexity on large campus networks. In the solution, user access policies are centrally managed on the gateway and executed by the gateway and access devices.



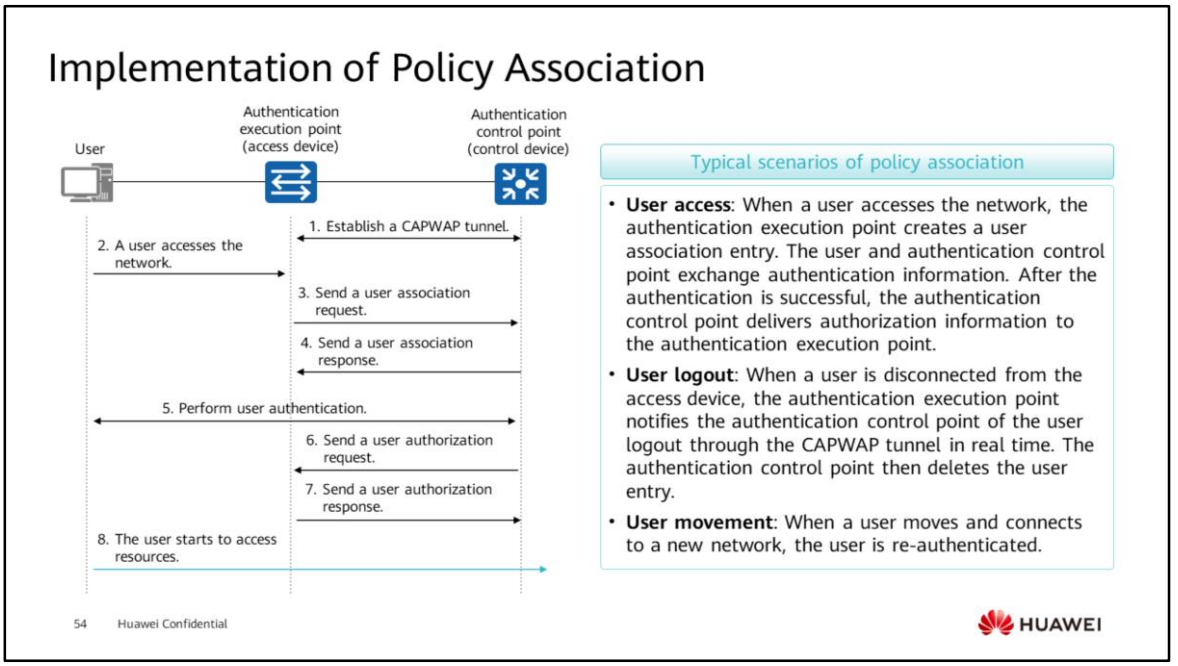
Roles in the policy association solution

- **Terminals**: provide human-machine interfaces for user authentication and resource access. The terminals include PCs, laptops, tablets, and dumb terminals.
- **Authentication access device**: an authentication execution point that executes network access policies for users.
- **Authentication control device**: an authentication control point that authenticates users and controls their access policies.

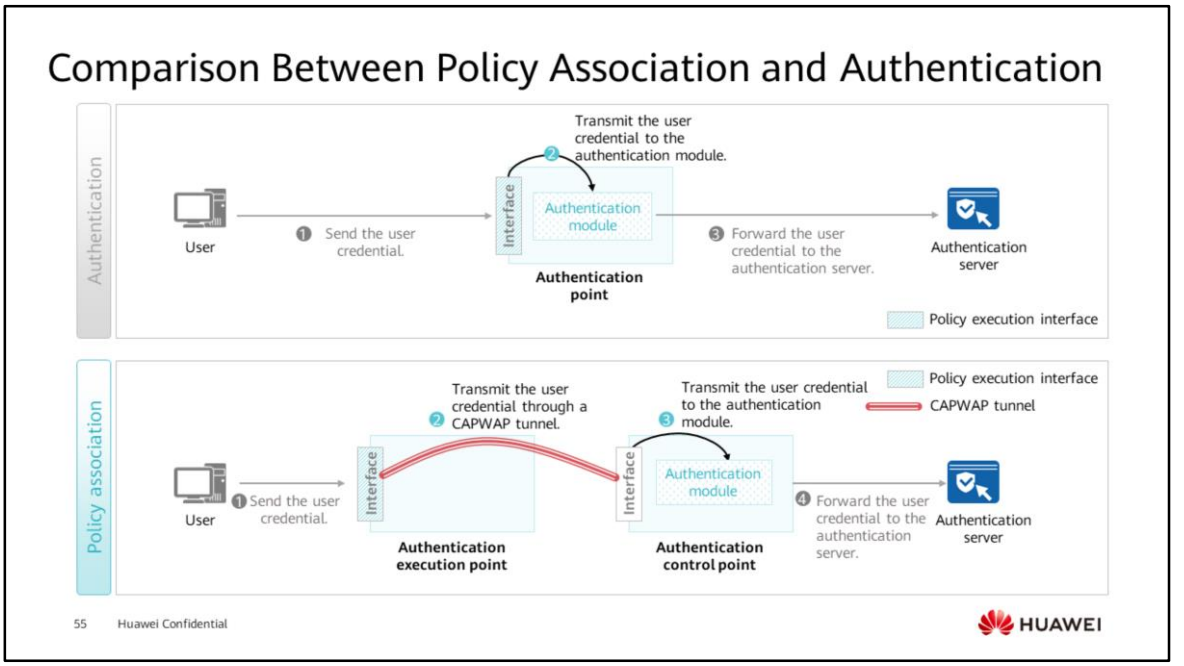
- A Control And Provisioning of Wireless Access Points (CAPWAP) tunnel is established between an authentication control point and an authentication execution point.
- Authentication control devices and authentication access devices use CAPWAP tunnels to associate users, transmit messages, deliver user authorization policies, and synchronize user information.

HUAWEI

- After policy association is configured, access devices can transparently transmit BPDUs and report user logout and user access positions in real time. In addition, authentication control devices request authentication access devices to execute user access policies, thus controlling user access to the network.
- Roles involved in policy association:
  - Authentication access device: an authentication execution point that executes network access policies for users. When an authenticated user accesses the network, the device executes the corresponding access policy, such as a policy defining the VLAN, ACL, or UCL group to which the user belongs.
  - Authentication control device: an authentication control point that authenticates users and controls their access policies. The device authenticates the identities of users who attempt to access the network and specifies their network access rights, that is, accessible resources.



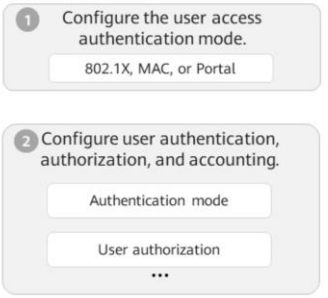
- User login process:
    1. The control device and access device establish a CAPWAP tunnel with each other.
    2. When detecting the access of a new user, the access device creates a user association entry to record basic information such as the user and access interface.
    3. The access device sends a user association request to the control device.
    4. The control device creates a user association entry to save the mapping between the user and access device, and returns a user association response to notify the access device of successful association.
    5. The user initiates an authentication request to the control device. The access device forwards the authentication packets between the user and control device.
    6. The control device deletes the user association entry. When the authentication succeeds, the control device generates a complete user entry, and sends a user authorization request to the access device, and delivers the network access policy of the user to the access device.
    7. The access device updates the user association entry, grants the specified network access rights to the user, and sends a user authorization response to the control device.
    8. The user accesses the specified network resources.



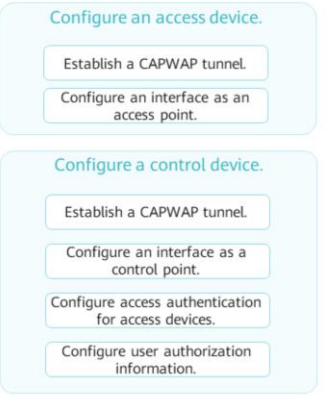
- Authentication:
  - The user exchanges information with the authentication point. The authentication point directly exchanges authentication information with the authentication server. When the authentication succeeds, the authentication server delivers the user rights to the authentication point. The interface on the authentication point then executes the corresponding user policy.
- Policy association:
  - The user exchanges information with the authentication execution point. The authentication execution point transmits the user credential to the authentication control point through a CAPWAP tunnel. The authentication control point exchanges authentication information with the authentication server.
  - When the authentication succeeds, the authentication server delivers user rights to the authentication control point, which further forwards the user rights to the authentication execution point through the CAPWAP tunnel. Finally, the interface on the authentication execution point executes the corresponding user policy.

Policy Association Configuration Roadmap

- Policy association controls network access of users. Before configuring policy association, complete the following tasks:



- Policy association needs to be configured on both control devices and access devices in any sequence.



HUAWEI

Policy Association Configuration: Configuring an Access Device

1. Establish a CAPWAP tunnel.

   [Huawei] **as access interface vlanif** *vlan-id*

   The source interface for establishing a CAPWAP tunnel is configured on the access device.

   [Huawei] **as access controller ip-address** *ip-address*

   The IP address of a control device is configured on the access device.

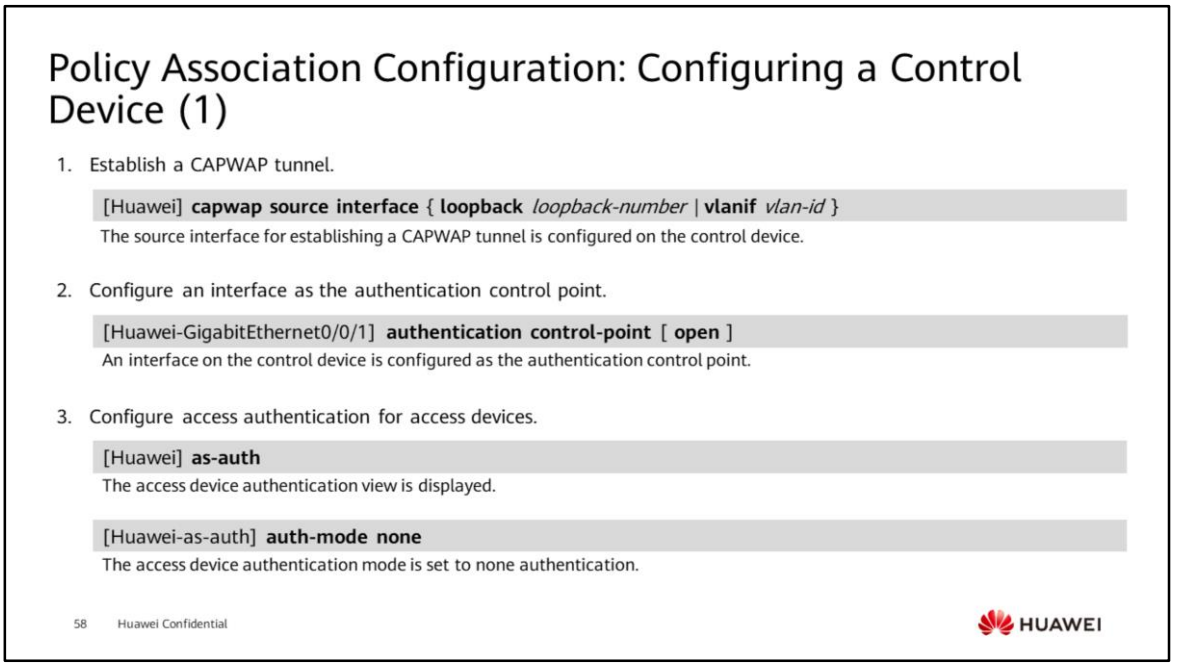
2. Configure an interface as the access point (authentication execution point).

   [Huawei-GigabitEthernet0/0/1] **authentication access-point** [**open**]

   Remote access control is enabled on an interface of the access device.

HUAWEI

- [Huawei-GigabitEthernet0/0/1] **authentication access-point [open]**
  - open: Disables right control of the access point.



- [Huawei-GigabitEthernet0/0/1] **authentication control-poin [open]**
  - open: Enables the forwarding function of the control point.
- Configure access authentication for access devices.
  - By default, access devices can connect to a control device only after passing authentication. The control device authenticates access devices using a blacklist and whitelist. Blacklisted access devices cannot connect to the control device, whereas whitelisted access devices can. The control device does not authenticate access devices out of the blacklist and whitelist, and you need to manually specify allowed access devices. You can also configure none authentication for access devices. As a result of this configuration, an access device can connect to the control device regardless of whether the access device is in the blacklist or whitelist.
  - For details about how to configure this function, see the product documentation.

Policy Association Configuration: Configuring a Control Device (2)

4. Configure user authorization information to be delivered to the access device and control device.

[Huawei-aaa] **service-scheme** *service-scheme-name*

A service scheme is created and the service scheme view is displayed.

[Huawei-aaa-service-huawei] **remote-authorize** { **acl** | **car** | **ucl-group** }

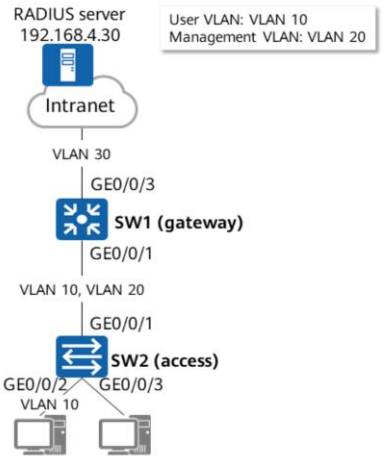
User authorization information to be delivered to the access device is configured.

[Huawei-aaa-service-huawei] **local-authorize** { **none** | { **acl** | **car** | **priority** | **ucl-group** | **vlan** } }

User authorization information to be delivered to the control device is configured.

HUAWEI

Configuration Example: Networking Requirements and Authentication Planning



Networking requirements

Large campus networks have many access devices, so user access policy deployment is time-consuming and the policies are difficult to modify.

The customer requires that NAC authentication and user access policies be configured on the gateway and the access policies be executed on access devices to simplify device deployment at the access layer.

Authentication planning

- Configure the gateway SW1 as the control device and SW2 as the access device.
- Configure the control device to authenticate users and the access device to execute user access policies.
- Configure VLAN 10 as the user VLAN and VLAN 20 as the management VLAN of the CAPWAP tunnel. In this example, 802.1X authentication is used.

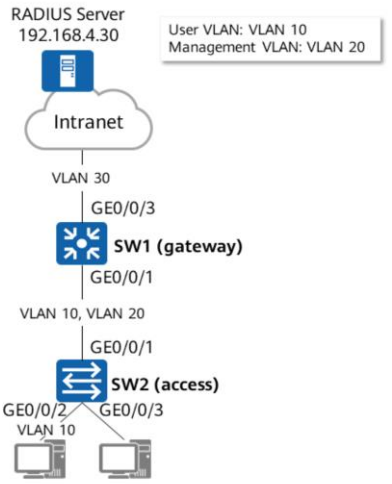
HUAWEI

Configuration Example: Data Plan

Item	Data
RADIUS server	IP address: 192.168.4.30
Gateway switch (control device, SW1)	• VLAN to which the uplink interface GE0/0/3 belongs: VLAN 30 • VLAN to which downlink interface GE0/0/1 belongs: VLAN 10 (user VLAN), VLAN 20 (management VLAN)
Access switch (SW2)	VLAN to which users belong: VLAN 10
RADIUS scheme	• Authentication server IP address: 192.168.4.30 • Authentication server port number: 1812 • Shared key of the RADIUS server: Huawei@123 • Authentication domain: nac
ACL number of the post-authentication domain	3001
Access permission	Prohibit access to resources on the 192.168.5.0/24 network segment.

HUAWEI

Configuration Example: Procedure (1)

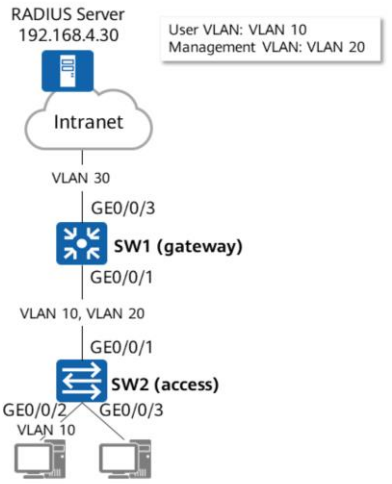


1. Create VLANs and configure the allowed VLANs on interfaces.
2. On the control device, configure an interface address pool on VLANIF 10 to assign IP addresses to users.

```
[SW1] dhcp enable
[SW1] interface vlanif 10
[SW1-Vlanif10] ip address 192.168.1.1 255.255.255.0
[SW1-Vlanif10] dhcp select interface
[SW1-Vlanif10] quit
```


HUAWEI

Configuration Example: Procedure (2)



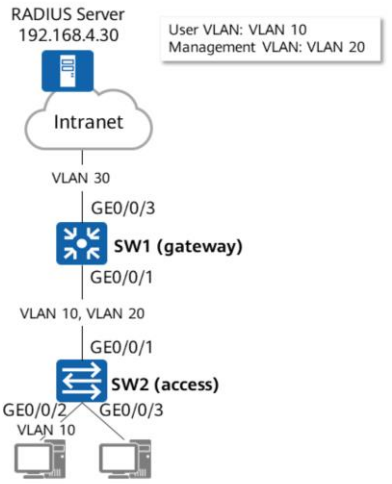
3. Configure the control device and access device to establish a CAPWAP tunnel.

```
[SW1] interface vlanif 20
[SW1-Vlanif20] ip address 192.168.2.1 255.255.255.0
[SW1-Vlanif20] dhcp select interface
[SW1-Vlanif20] dhcp server option 43 ip-address 192.168.2.1
[SW1-Vlanif20] quit
[SW1] capwap source interface vlanif 20
[SW1] as-auth
[SW1-as-auth] auth-mode none
[SW1-as-auth] quit
```

```
[SW2] interface vlanif 20
[SW2-Vlanif20] ip address dhcp-alloc
[SW2-Vlanif20] quit
[SW2] as access interface vlanif 20
```


HUAWEI

Configuration Example: Procedure (3)



4. On the control device, create and configure a RADIUS server template, an AAA authentication scheme, and an authentication domain.

```
[SW1] radius-server template rd1
[SW1-radius-rd1] radius-server authentication 192.168.4.30 1812
[SW1-radius-rd1] radius-server shared-key cipher Huawei@123
[SW1-radius-rd1] quit
```

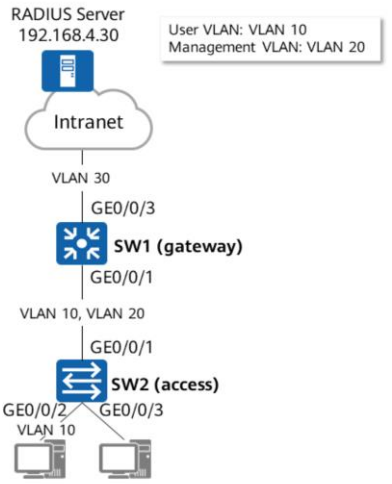
```
[SW1] aaa
[SW1-aaa] authentication-scheme abc
[SW1-aaa-authen-abc] authentication-mode radius
[SW1-aaa-authen-abc] quit
```

```
[SW1] domain nac
[SW1-aaa-domain-isp1] authentication-scheme abc
[SW1-aaa-domain-isp1] radius-server rd1
[SW1-aaa-domain-isp1] quit
[SW1-aaa] quit
```

```
[SW1] domain nac
                # Configure a global default domain named nac.
```


HUAWEI

Configuration Example: Procedure (4)



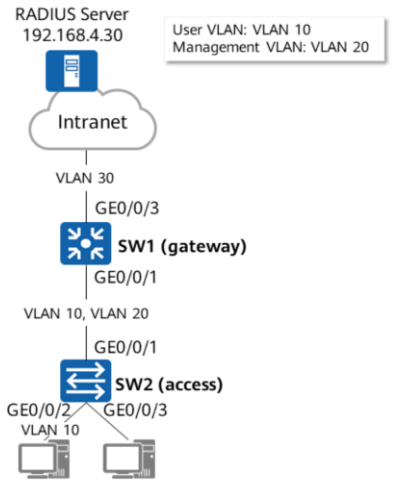
5. Configure the control device as the control point and the access device as the access point.

```
[SW1] interface gigabitethernet 0/0/1
[SW1-GigabitEthernet0/0/1] authentication control-point
[SW1-GigabitEthernet0/0/1] quit
```

```
[SW2] interface gigabitethernet 0/0/2
[SW2-GigabitEthernet0/0/2] authentication access-point
[SW2-GigabitEthernet0/0/2] quit
[SW2] interface gigabitethernet 0/0/3
[SW2-GigabitEthernet0/0/3] authentication access-point
[SW2-GigabitEthernet0/0/3] quit
```


HUAWEI

Configuration Example: Procedure (5)

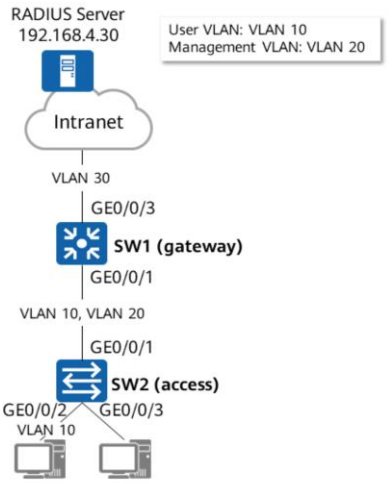


6. Configure the control device to deliver ACL-based authorization information to the access device, and bind the AAA service scheme **asd** to the authentication domain **nac**.

```
[SW1] aaa
[SW1-aaa] service-scheme asd
[SW1-aaa-service-asd] remote-authorize acl
[SW1-aaa-service-asd] quit
[SW1-aaa] domain nac
[SW1-aaa-domain-nac] service-scheme asd
[SW1-aaa-domain-nac] quit
[SW1-aaa] quit
```


HUAWEI

Configuration Example: Procedure (6)



7. Configure ACLs and ACL rules for authorization on the control device and access device.

```
[SW1] acl 3001
[SW1-acl-adv-3001] rule deny ip destination 192.168.5.0 0.0.0.255
[SW1-acl-adv-3001] quit
```

```
[SW2] acl 3001
[SW2-acl-adv-3001] rule deny ip destination 192.168.5.0 0.0.0.255
[SW2-acl-adv-3001] quit
```

8. Configure 802.1X authentication on the control device and access device, and configure an authentication-free rule on the control device to allow packets from the management VLAN of the CAPWAP tunnel to pass through. (The 802.1X authentication configuration is not mentioned here.)

```
[SW1] free-rule-template  name default_free_rule
[SW1-free-rule-default_free_rule]  free-rule 1 source vlan 20
[SW1-free-rule-default_free_rule]  quit
```


HUAWEI

Quiz

1. (Single-answer question) When PCs and dumb terminals such as printers and fax machines are connected to an interface of an access device, which authentication mode can be used to ensure network security and allow dumb terminals to access the network? ( )

   A. MAC address authentication

   B. 802.1X authentication

   C. Portal authentication

   D. MAC address bypass authentication

HUAWEI

1. D

Quiz

2. (Multiple-answer question) Based on which of the following can a Huawei device deliver user access rights to a successfully authenticated user? ( )

   A. VLAN

   B. IP address

   C. ACL

   D. UCL group

HUAWEI

2. ACD

Summary

- User access control is the first line of defense to protect a network, where you can deploy user authentication, such as MAC address authentication, 802.1X authentication, and Portal authentication, to ensure security.
- The implementation modes and application scenarios of these technologies are different. You need to select and deploy them based on network characteristics and requirements.
- iMaster NCE-Campus can be used together with network products to provide an advanced NAC solution that implements powerful functions such as terminal type identification and policy automation.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Free Mobility
HUAWEI

Foreword

- On a campus network, different network access policies can be deployed for users on access devices to meet diverse network access requirements.
- On a traditional campus network, users' network access rights are controlled using the Network Admission Control (NAC) technology, in conjunction with VLAN and ACL technologies. However, this solution has many defects. For example, the association between ACLs and users takes effect only on authentication points, resulting in poor flexibility; VLANs and ACLs need to be configured on a large number of switches that function as authentication points in advance, causing a heavy deployment and maintenance workload.
- Mobile working requires that these defects be removed and employees access the network from any place, any VLAN, or any IP network segment with controlled network access rights. To implement these, the free mobility solution is introduced.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the requirements of large-scale campus networks for policy control.
  - Describe the differences between free mobility and traditional technologies or solutions.
  - Describe the basic functions and working mechanism of free mobility.
  - Understand the relationship between free mobility and campus network access authentication.
  - Describe the typical application solution of free mobility.

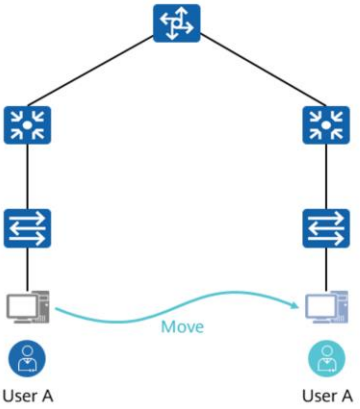
HUAWEI

Contents

1. **Technical Background and Basic Concepts of Free Mobility**
2. Working Mechanism of Free Mobility
3. Free Mobility Solution Design

HUAWEI

Mobility Requirements of Users on the Network



Scenario description

1. User mobility requirements, such as mobile working, are common on large-scale campus networks.
2. To ease campus network management, it is required that users obtain the same network access policy regardless of their locations and IP addresses.

Solution challenges

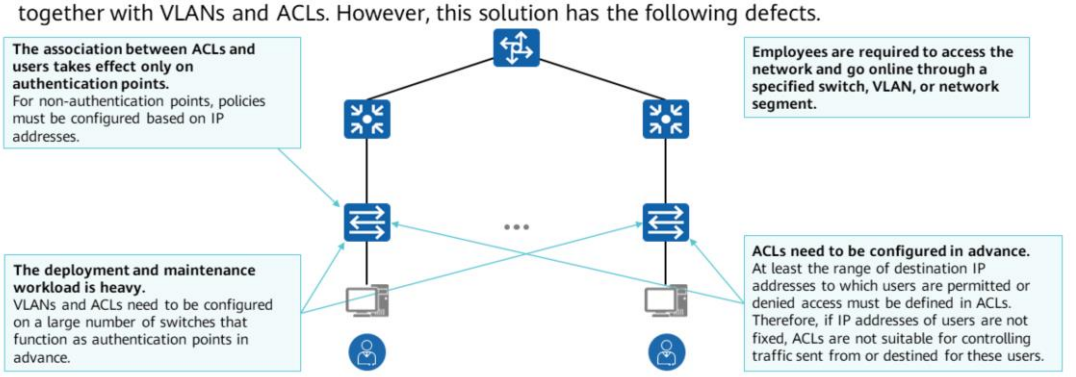
1. How to ensure that users can access the campus network from any place?
2. How to ensure the security of user access to the campus network?
3. Can user information be managed in a centralized manner?
4. Are the network configuration and configuration delivery simple enough?
5. Is it easy to adjust policies?

HUAWEI

- With the construction and promotion of wireless networks, the boundaries of enterprise campus networks are disappearing, and office locations of enterprise employees become more flexible.
- The large-scale movement of employees' access locations improves enterprise production efficiency, but also brings challenges to enterprise network management and security. Mobile working causes the IP addresses of employees' hosts to change frequently. However, traditional campus networks cannot adapt to this change because employees' rights are controlled based on their IP addresses.

Traditional Solution: Using NAC Together with VLANs and ACLs

- On a traditional campus network, users' network access rights are controlled using the NAC technology together with VLANs and ACLs. However, this solution has the following defects.

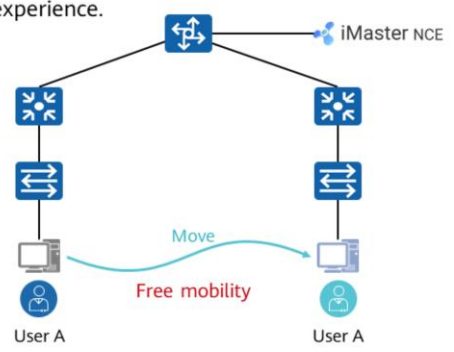


HUAWEI

- Mobile working requires that these defects be removed and employees obtain consistent network access rights when accessing the network from any place, any VLAN, or any IP network segment. In addition, administrators want to have a simple policy control approach that is decoupled from network topologies and IP addresses.

Overview of Free Mobility

- The free mobility solution allows a user to obtain the same network access policy regardless of the user's location and IP address on a campus network. iMaster NCE-Campus and switches work together to enable network access rights to automatically move with users, improving mobile working experience.



Free mobility resolves the problems faced by traditional campus networks in three aspects

1. Decoupling of service policies from IP addresses.
2. Centralized management of user information.
3. Centralized policy management.

Advantages of free mobility

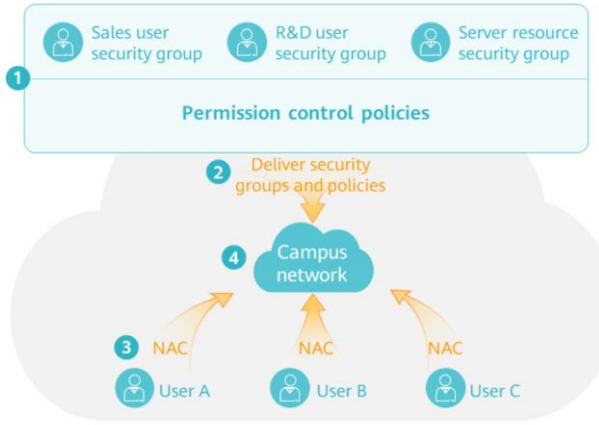
1. Simplified network planning: Administrators no longer need to consider user IP addresses when configuring policies.
2. Enhanced control capability: User authentication information can be synchronized between network devices.
3. Improved management efficiency: Administrators no longer need to configure devices one by one.

HUAWEI

- Decoupling of service policies from IP addresses
  - Using iMaster NCE-Campus, administrators can divide users and resources on the entire network into different security groups based on diverse dimensions. In addition, network devices in the free mobility solution use an innovative software and hardware design. Such a device can match source and destination security groups based on the source and destination IP addresses of packets, and then find the matching inter-group policy based on these security groups.
  - Through the innovative designs, all the user- and IP address–based service policies on traditional networks can be transformed into security group–based policies. When predefining service policies, administrators no longer need to consider user IP addresses. This achieves decoupling of service policies from IP addresses.
- Centralized management of user information
  - Administrators can use iMaster NCE-Campus to centrally manage user authentication and login information and obtain the mappings between network-wide users and their IP addresses.
- Centralized policy management
  - iMaster NCE-Campus is not only the authentication center on a campus network, but also the management center of service policies. Administrators can use iMaster NCE to centrally manage service policies on policy enforcement devices on the entire network. After being configured for once, service policies can be automatically delivered to policy enforcement devices on the campus network. These policies include permission control policies (for example, group A is forbidden to access group B) and experience guarantee policies (for example, the bandwidth and priority for forwarding traffic of group A are controlled).

Implementation of Free Mobility

- Free mobility implements policy management and permission control based on security groups.



1. Security groups are defined, each representing a group of users who have the same network access requirements.
2. Permission control policies are defined based on security groups and are delivered to network devices.
3. Users obtain authorized security groups after they are successfully authenticated.
4. After user traffic enters a network, network devices enforce policies based on the source and destination security groups of the traffic.

HUAWEI

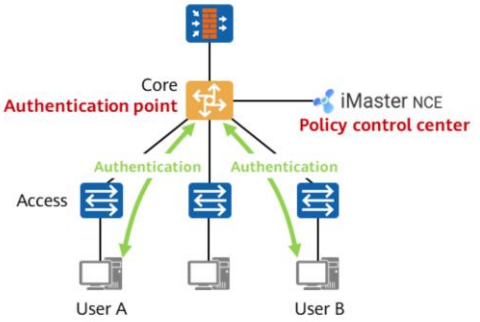
- Free mobility introduces the concept of security group. Security groups are related only to user identities and are completely decoupled from network information such as user VLANs and IP addresses.
- User policy management and permission control are implemented based on security groups.

Traditional Solution vs. Free Mobility Solution

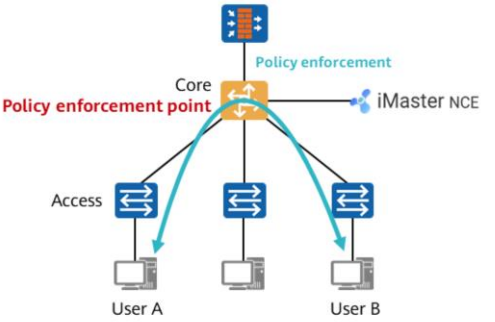
Customer Requirement	Traditional Solution	Free Mobility Solution
Consistent network access rights in the mobile working scenario	The "dynamic VLAN + static ACL" solution is used. The configuration planning is complex, the pre-configuration workload is heavy, ACL maintenance is difficult, and consistent network access rights cannot be ensured for users.	Network access policies are decoupled from IP addresses. Administrators divide security groups based on user identities on iMaster NCE-Campus, and use a policy control matrix to manage policies in a unified manner. This simplifies configuration and maintenance, and ensures that users can obtain the same network access rights as they move.
Service-based user isolation	The "Layer 2 VLAN isolation + Layer 3 ACL isolation" solution is used. In the mobile working scenario, the configuration planning is complex, the pre-configuration workload is heavy, ACL maintenance is difficult, and user isolation is hard to implement.	The "Layer 2 port isolation + inter-group isolation based on security groups" solution is used. Inter-group isolation policies are decoupled from IP addresses, simplifying configuration and maintenance and effectively achieving user isolation.
Simplified configuration and management	A large number of rules such as VLANs and ACLs need to be planned based on IP addresses during network design. These rules are complex to configure and difficult to understand, causing inconvenience in subsequent maintenance.	Based on security groups (UCL groups), only several user groups and inter-group policies need to be defined, greatly simplifying planning and configuration. In addition, the policy control matrix is easy to understand and maintain.

HUAWEI

Basic Concepts in Free Mobility: Authentication and Policy Enforcement



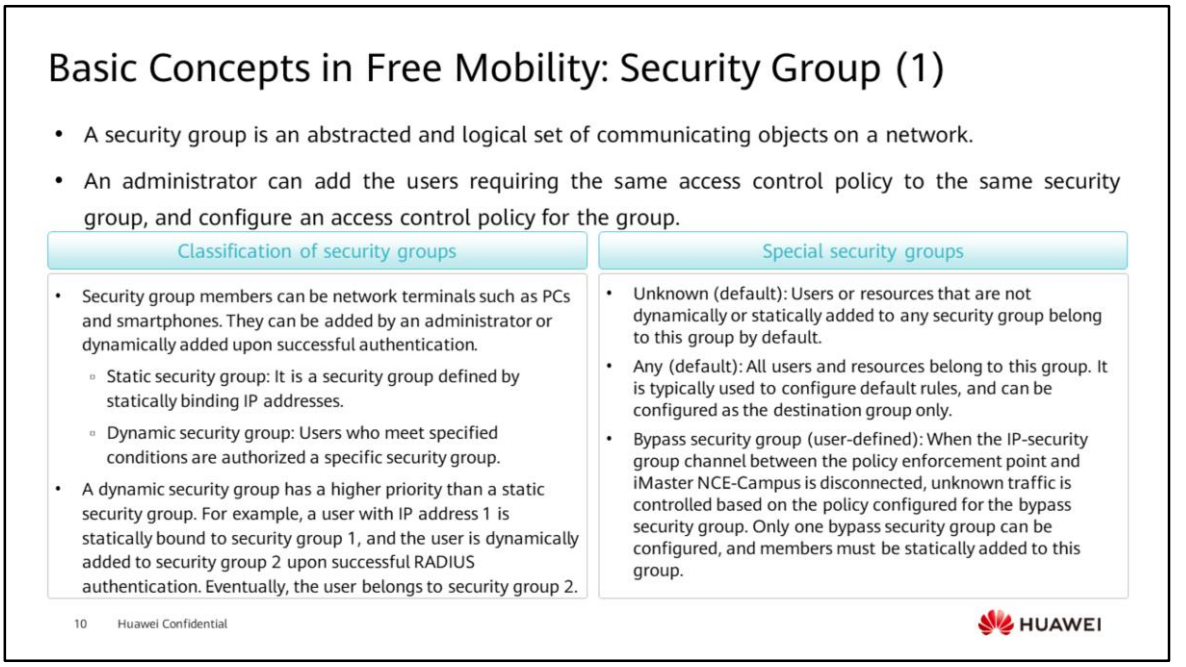
- **Authentication point:** authenticates terminals and obtains authorization results (containing the security groups to which terminals belong) from iMaster NCE-Campus.
- **iMaster NCE-Campus:** serves as the authentication server and the policy control center to maintain the policy control matrix, which defines policies for controlling access between network-wide security groups.



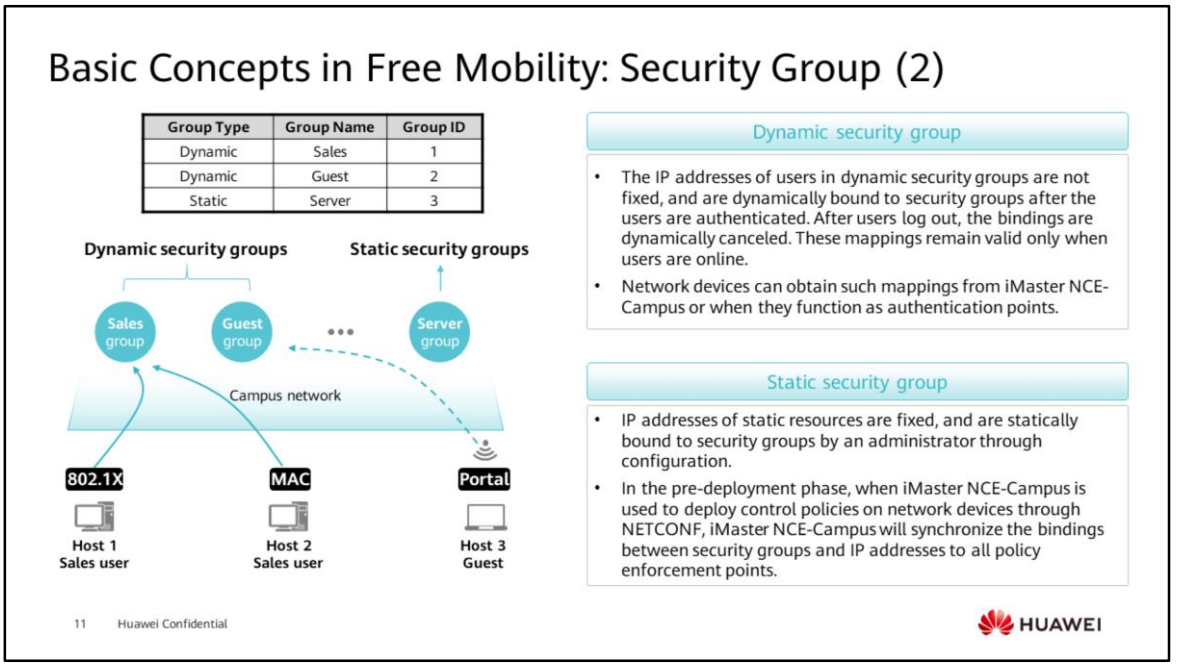
- **Policy enforcement point:** obtains the policy control matrix from iMaster NCE-Campus, and enforces policies based on the source and destination security groups to which the source and destination IP addresses of the received traffic belong. If the policy allows the traffic to pass through, the policy enforcement point forwards the traffic. Otherwise, the traffic is dropped.

HUAWEI

- Additional information about the policy enforcement point:
  - The policy enforcement point is responsible for enforcing security group–based service policies. To enforce these policies, the policy enforcement point must be able to identify the source and destination security groups of packets. It can obtain the mappings between IP addresses and security groups during the authentication process or from iMaster NCE-Campus.
  - The authentication point and policy enforcement point are two device roles. Based on the administrator's configuration and device capabilities, a physical device can play either or both of the two roles.



- Security group:
  - An administrator can define security groups on iMaster NCE-Campus to describe and organize the sources or destinations of network traffic, and then configure policies to control mutual access between the security groups.
  - An administrator can add network objects that have the same access requirements to the same security group, and configure a policy for this security group. In this way, these network objects obtain the same permissions as configured for the security group. For example, an administrator can define the following security groups: R&D group (a collection of individual hosts), printer group (a collection of printers), and database server group (a collection of server IP addresses and ports). Compared with the solution in which access control policies are deployed for each user, the security group–based access control solution greatly reduces the administrator's workload.



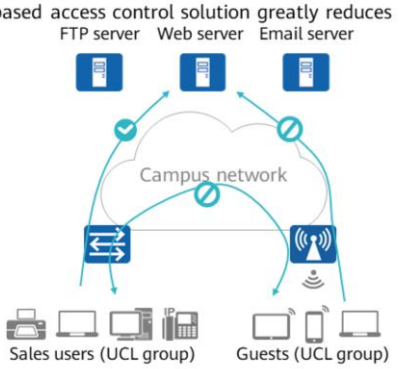
- Classification of security groups:
  - A security group can be both a dynamic security group and a static security group. That is, it is bound to both multiple authorization rules to represent dynamic users and multiple IP addresses or IP network segments to represent static resources. The differences are as follows:
    - The IP addresses of users in dynamic security groups are not fixed, and are dynamically bound to security groups after the users are authenticated. After users log out, the bindings are dynamically canceled. These mappings remain valid only when users are online. Network devices can obtain the mappings between IP addresses and security groups (also known as IP-security group entries) by acting as authentication points, or iMaster NCE-Campus delivers the mappings to network devices.
    - The IP addresses of static resources are fixed. An administrator needs to bind these IP addresses to static security groups on iMaster NCE-Campus, which then synchronizes the bindings to policy enforcement points. Static security group members can be any network objects that use fixed IP addresses, such as terminals, servers, interfaces of network devices, and authentication-free users.
  - If an IP address is added to different groups both in static mode and upon successful authentication, the dynamic security group assigned during authentication takes effect.
  - An IP address can belong to only one security group.

Basic Concepts in Free Mobility: UCL Group

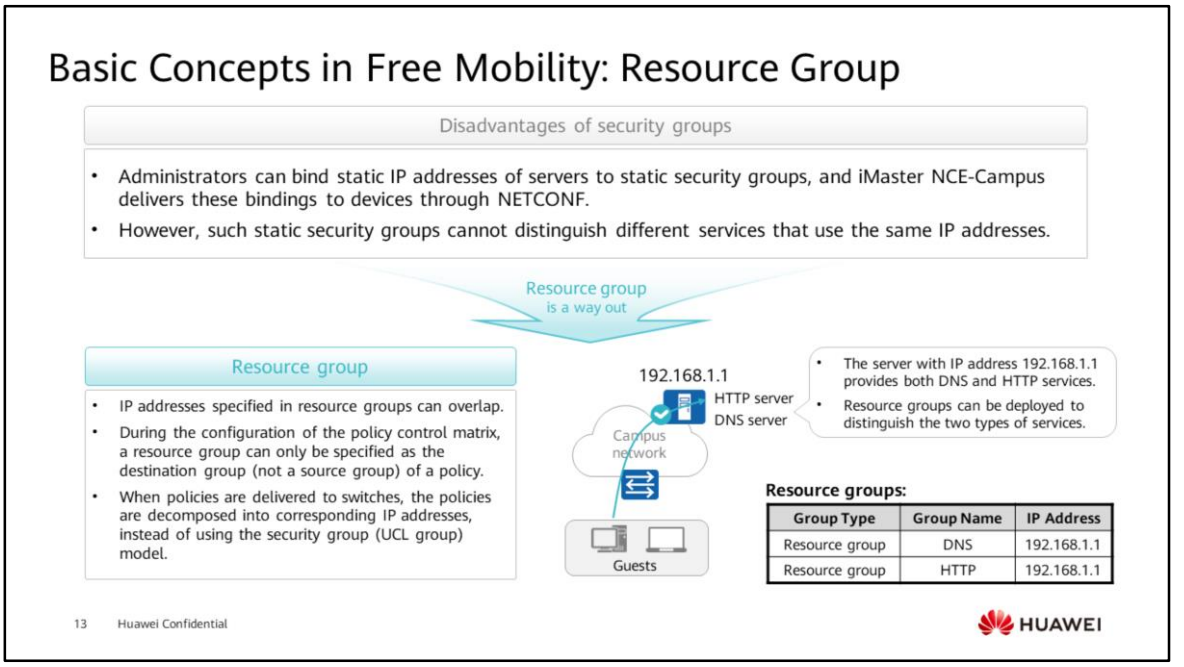
- User Control List (UCL) group:
  - UCL groups identify user types. An administrator can add the users requiring the same network access policy to the same UCL group, and configure a network access policy for the group. Compared with the solution in which network access policies are deployed for each user, the UCL group–based access control solution greatly reduces the administrator's workload.
  - A UCL group is called a security group on iMaster NCE-Campus.
- UCL group configuration on a switch:
  - Create a UCL group.

    [Huawei] **ucl-group** *group-index* [**name** *group-name*]
  - Configure a static UCL group.

    [Huawei] **ucl-group ip** *ip-address* { *mask-length* | *ip-mask* } { *group-index* | **name** *group-name* }



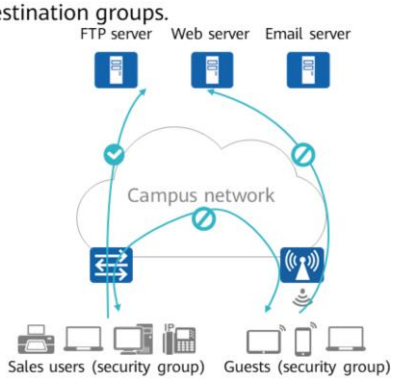
HUAWEI



- Dumb terminals, data center server resources, and authentication-free users do not need to be authenticated when accessing the network. Therefore, they cannot be authorized by the AAA server. In this case, you can bind their static IP addresses to resource groups.
- Note: When resource groups are used, a policy enforcement point generates a policy by IP address, instead of based on each resource group. As such, the device may have a large number of policies.

Basic Concepts in Free Mobility: Policy Control Matrix

- After security groups are defined, administrators can define inter-group policies on the entire network.
- A policy control matrix is used to configure inter-group policies to control access from source groups to destination groups.

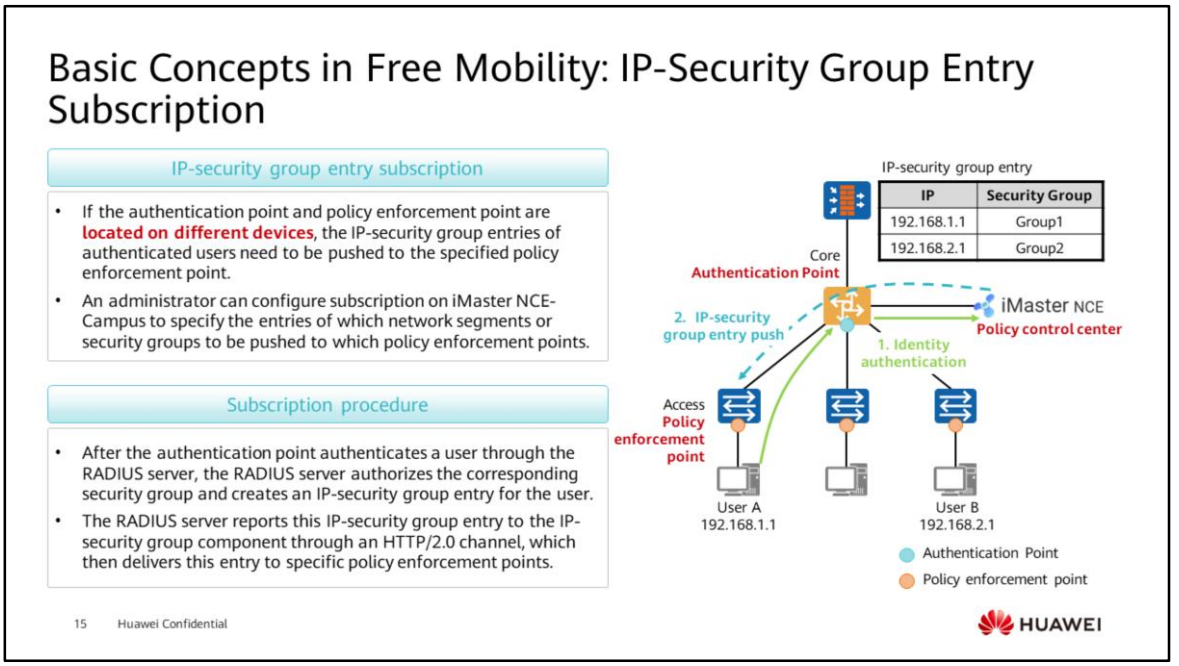


Policy control matrix

Source Group \ Destination Group	Sales	Guest	Server
Sales	✓	✗	✓
Guest	✗	✓	✗
Server	✓	✓	✓

HUAWEI

- When multiple policies are configured to control access from a source security group to multiple destination groups, an administrator needs to configure priorities of the policies to determine the sequence in which policies are matched. For example, if the destination groups are resource groups with overlapping IP addresses, the administrator can set a high priority for a policy so that the policy can be matched preferentially.
- For unknown users:
  - If a policy enforcement device does not find any security group corresponding to an IP address, it considers that the IP address belongs to the default security group named **unknown**, and enforces the matching security group policy (default policy: permit).
- The following uses the traffic from the sales group to the server group as an example to describe policy matching in the policy control matrix:
  - The device (policy enforcement point) first searches for the policy of controlling access from the sales group to the server group. If no such inter-group policy is found in the policy control matrix, the device continues matching policies.
  - The device then searches for the policy of controlling access from the sales group to the **any** group. If no such inter-group policy is found in the policy control matrix, the device continues matching policies.
  - Finally, the device matches traffic with the policy of controlling access from the **any** group to the **any** group. By default, this policy exists in the policy control matrix and defines the permit action. That is, traffic is permitted by default if no policy is matched.

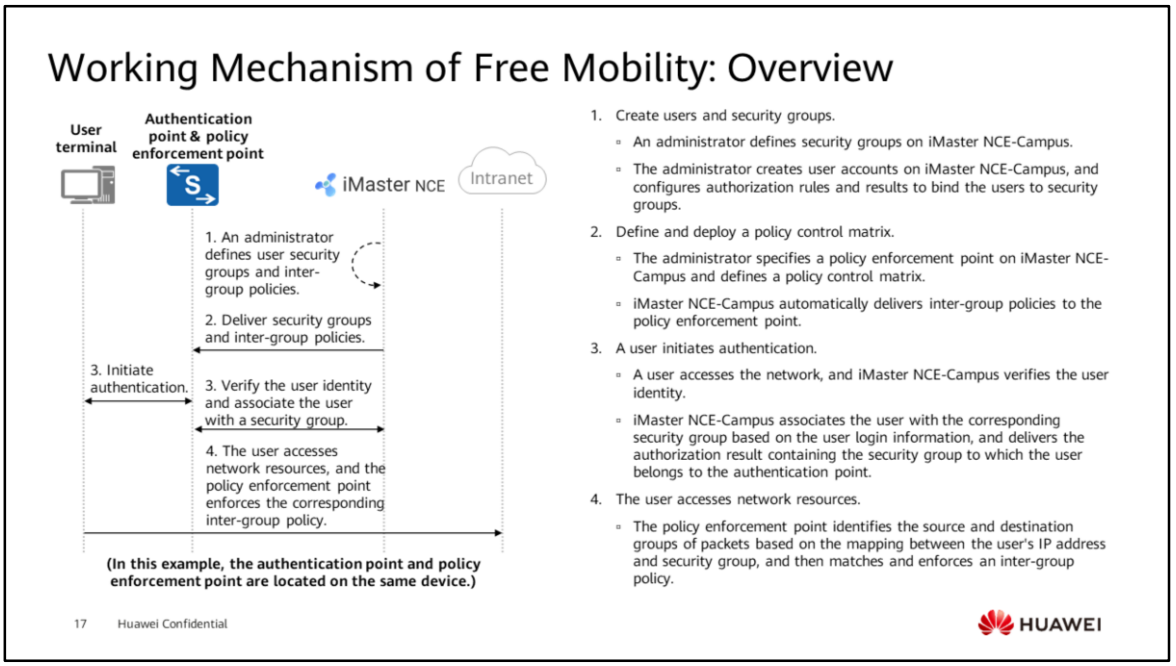


- A policy enforcement point can obtain IP-security group entries in either of the following ways:
  - The policy enforcement point obtains IP-security group entries during user authentication when it is located on the same device as the authentication point.
  - iMaster NCE-Campus pushes IP-security group entries to the policy enforcement point through an HTTP/2.0 channel. This scenario is known as IP-security group entry subscription.
- In the IP-security group entry subscription scenario, iMaster NCE-Campus must have already generated IP-security group entries during user authentication. Therefore, IP-security group entry subscription has the following requirements:
  - iMaster NCE-Campus functions as the RADIUS server to perform end user authentication. In this way, it can generate IP-security group entries during user authentication, and synchronize the entries to the free mobility component.
  - iMaster NCE-Campus functions as a RADIUS relay agent when a third-party RADIUS server is used for user authentication. In this way, iMaster NCE-Campus can obtain security group information during the authentication process to generate IP-security group entries.

Contents

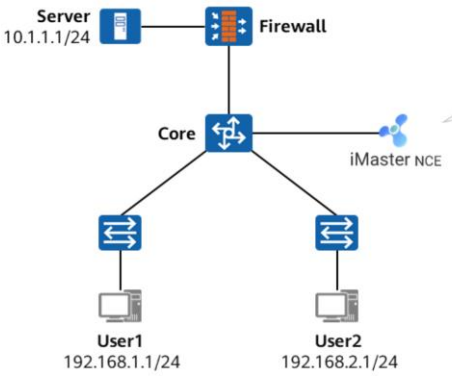
1. Technical Background and Basic Concepts of Free Mobility
2. **Working Mechanism of Free Mobility**
3. Free Mobility Solution Design

HUAWEI



- Free mobility deployment based on iMaster NCE-Campus:
  - Create users and security groups.
    - An administrator can define users and security groups on iMaster NCE-Campus in a unified manner. Security groups can be defined based on network services for configuring inter-group control policies.
  - Define and deploy a policy control matrix.
    - The administrator defines inter-group policies in a policy control matrix on iMaster NCE-Campus, and delivers the policy control matrix to policy enforcement points.
  - A user initiates authentication.
    - When a user is being authenticated, iMaster NCE-Campus associates the user with a security group based on the user login information. After the user is authenticated successfully, iMaster NCE-Campus delivers the authorization result containing the security group to which the user belongs to the authentication point. During 802.1X authentication, if a terminal has not obtained an IP address, the authentication point automatically detects the actual IP address of the user after the user is successfully authenticated and obtains an IP address, and reports the user's actual IP address to iMaster NCE-Campus. iMaster NCE-Campus collects IP addresses of all online users and synchronizes all user information to policy enforcement points.

Working Mechanism of Free Mobility (1)



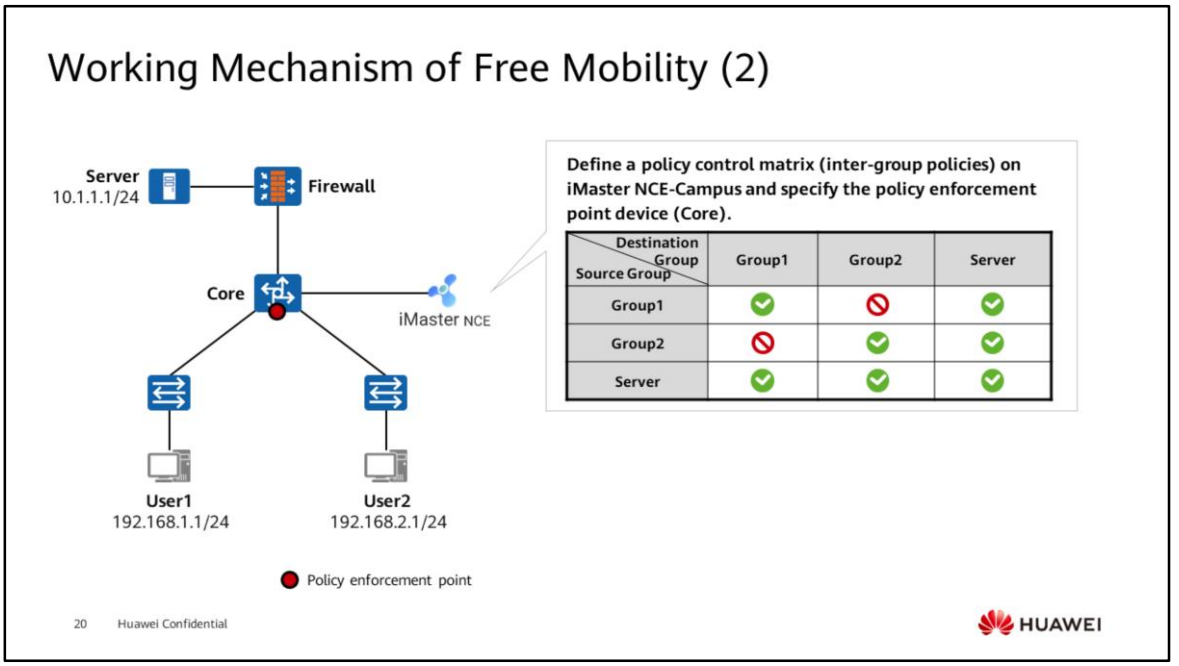
Create users on iMaster NCE-Campus.

User	Password
User1	***
User2	***

Create security groups, bind the server IP address to the server security group, and configure authorization rules and results to bind User1 and User2 to security groups Group1 and Group2, respectively.

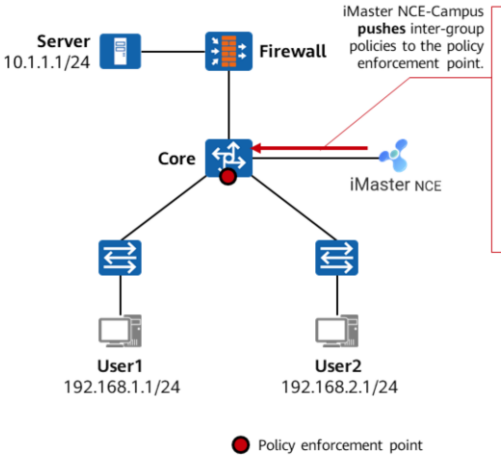
Group Type	Group Name	Group ID	
Dynamic	Group1	1	Dynamic authorization (User1)
Dynamic	Group2	2	Dynamic authorization (User2)
Static	Server	3	Static binding <- 10.1.1.1/32 (server)

HUAWEI



- To manually specify a device as a policy enforcement point managed by iMaster NCE-Campus, enable the free mobility function by running the following command:
- [HUAWEI] **group-policy controller** i*p-address1* [*port-number1*] **password** *password* [**src-ip** *ip-address2*]
    - *ip-address1* [*port-number1*]: specifies the IP address of a controller and the port number used by the device to exchange packets with the controller. If no port number is configured, the default port number 5222 is used.
    - **password** *password:* specifies the password used by the device to connect to the controller. The password configured on the device must be the same as that configured on the controller.
    - **src-ip** *ip-address2*: specifies the source IP address used by the device to communicate with the controller. If this parameter is not configured, the device selects one of its own IP addresses to communicate with the controller.

Working Mechanism of Free Mobility (3)

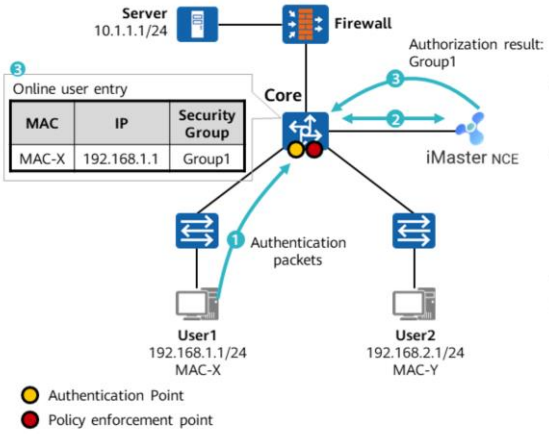


Policy control matrix

Source Group \ Destination Group	Group1	Group2	Server
Group1	✓	✗	✓
Group2	✗	✓	✓
Server	✓	✓	✓

HUAWEI

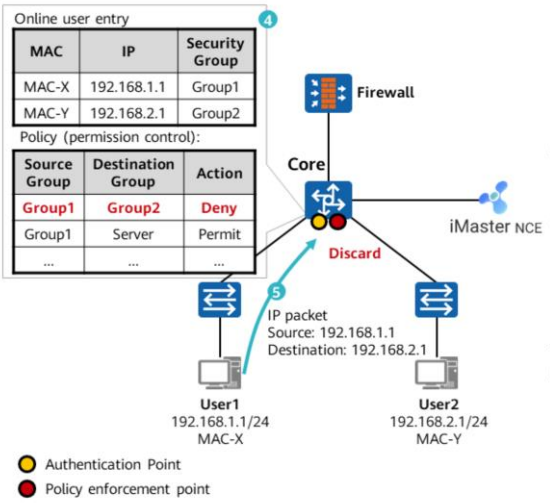
Working Mechanism of Free Mobility (4)



1. User1 attempts to access the network. The core switch functioning as the authentication point exchanges user authentication information with iMaster NCE-Campus.
2. iMaster NCE-Campus checks the login information of User1 and associates User1 with the corresponding security group (Group1) in the authorization policy.
3. After User1 is authenticated, iMaster NCE-Campus associates the user IP address with Group1 and records the association in an IP-security group entry. In addition, the iMaster NCE-Campus notifies the authentication point of the security group to which the user belongs, and the core device at the authentication point generates an online user entry.
4. The preceding process also applies to User2.
5. After receiving a service packet from a terminal, the core switch that also functions as the policy enforcement point identifies the security group that matches the source and destination IP addresses of the packet and enforces the corresponding inter-group policy.

HUAWEI

Working Mechanism of Free Mobility (5)



1. User1 attempts to access the network. The core switch functioning as the authentication point exchanges user authentication information with iMaster NCE-Campus.
2. iMaster NCE-Campus checks the login information of User1 and associates User1 with the corresponding security group (Group1) in the authorization policy.
3. After User1 is authenticated, iMaster NCE-Campus associates the user IP address with Group1 and records the association in an IP-security group entry. In addition, the iMaster NCE-Campus notifies the authentication point of the security group to which the user belongs, and the core device at the authentication point generates an online user entry.
4. The preceding process also applies to User2.
5. After receiving a service packet from a terminal, the core switch that also functions as the policy enforcement point identifies the security group that matches the source and destination IP addresses of the packet and enforces the corresponding inter-group policy.

HUAWEI

Contents

1. Technical Background and Basic Concepts of Free Mobility
2. Working Mechanism of Free Mobility
3. **Free Mobility Solution Design**

HUAWEI

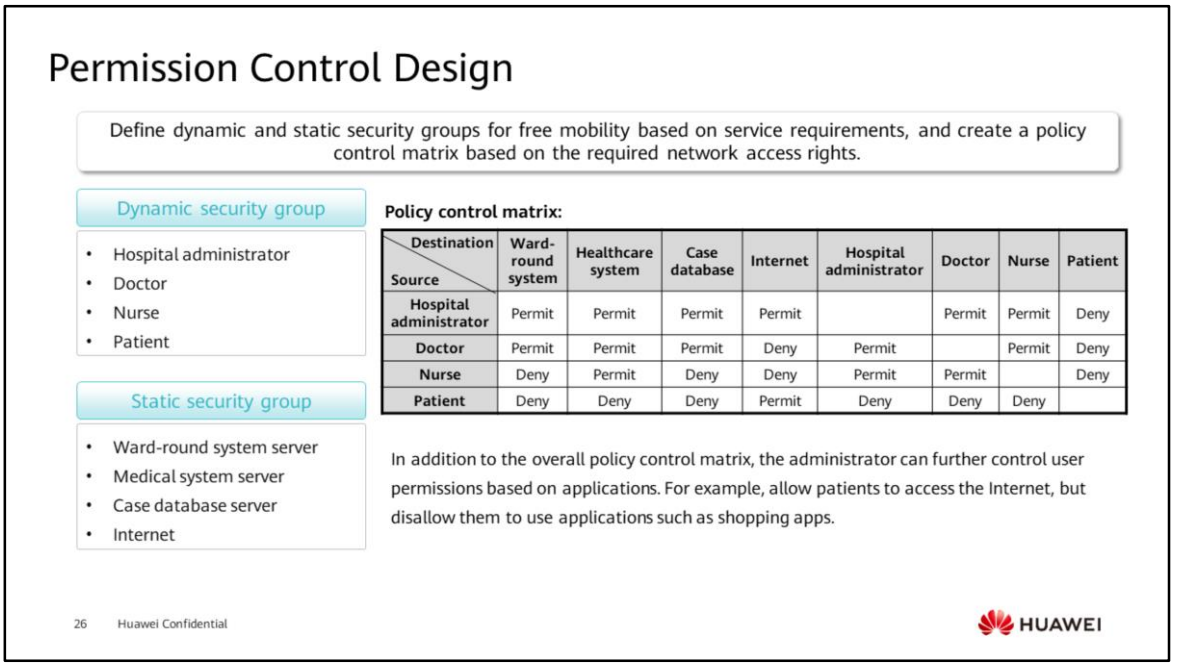
Security Group Design

Dynamic security group (users)	Static security group (network resources)
Users and terminals that can access the network only after being authenticated. In most cases, dynamic security groups can be defined based on user identities, such as students and teachers in schools. In addition, dynamic security groups can be customized based on **5W1H** conditions.	Terminals or network segments that use fixed IP addresses, such as servers and the Internet. Static security groups are defined based on service types provided by network resources. For example, the school web server, the data server for storing student information, and the exam resource server can be divided into different static security groups.

- **Who**: identity of a user, for example, a corporate executive, employee, or guest;
- **Where**: user access location, for example, local (within a campus) or remote access;
- **What**: type of the access terminal, for example, a mobile phone, PC, or laptop;
- **When**: time range when a user accesses the network, for example, in the daytime or at night;
- **Whose**: device owner, for example, a company-issued terminal or BYOD terminal;
- **How**: user access mode, for example, wired or wireless access.

The most important rule for defining a dynamic security group is to add users with the same network access requirements to one security group.

HUAWEI



- A security group policy reflects whether two security groups can communicate with each other. The administrator can configure policies to permit or deny communication between every two security groups in a policy control matrix on iMaster NCE-Campus. If no control policy is created for a source group and a destination group, they can communicate with each other by default.
- When planning security group policies, pay attention to the direction of policies. Generally, packets are transmitted in both directions between two terminals.
    - For Huawei switches, traffic from switch A to switch B and traffic from switch B to switch A match different policies. Whether traffic is permitted or denied depends on the source and destination groups of the traffic. If the permit action is configured for the A-to-B traffic and the deny action for the B-to-A traffic, all packets sent from switch A to switch B are allowed to pass through, but the packets sent from switch B to switch A are discarded, regardless of which device initiates the request. If no matching policy is found, a switch performs the default action — permit.
- Generally, network access involves bidirectional communication. Therefore, to simplify management, you only need to consider the access permissions from user security groups to other users and servers when designing permission control policies.
    - To prevent users from accessing a security group, you only need to configure a unidirectional deny rule.
    - To allow users to access a security group, you only need to configure a unidirectional permit rule.

User Authentication Design

- The free mobility solution involves the same user authentication process as the traditional solution.
- After a user is successfully authenticated, authorization information such as ACLs and VLANs is delivered in the traditional solution, whereas security group (UCL group) information is delivered in the free mobility solution.

Authentication modes supported by free mobility

Authentication Mode		Recommended Authentication Scenario
802.1X authentication		Wired user authentication
Portal authentication	Portal authentication	Wireless user authentication
	MAC address-prioritized Portal authentication	
MAC address authentication		Authentication for dumb terminals (such as printers)
PPPoE authentication		Authentication for education campus users

Recommended locations of authentication points

Device Role	Location
User authentication point	Core switch
	Aggregation switch

HUAWEI

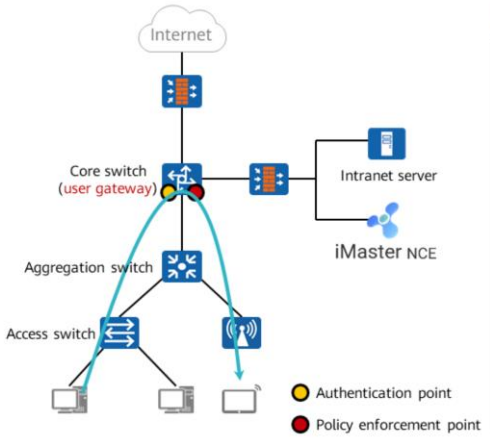
Locations of Authentication Points and Policy Enforcement Points

- Typically, a user gateway functions as both the authentication point and policy enforcement point for the following major reasons:
  - There are a large number of access switches on a network. Configuring the authentication function on each access switch requires a heavy workload and leads to difficulties in management.
  - iMaster NCE-Campus needs to synchronize permission control policies to policy enforcement points. If access switches act as both authentication points and policy enforcement points, there will be a great number of policy enforcement points. This increases the workload and difficulties of device management on iMaster NCE-Campus and prolongs the policy synchronization time.
- To prevent users on a Layer 2 network connected to an upstream user gateway from communicating with each other, configure Layer 2 isolation. In this way, communication traffic of the users must pass through the user gateway.

HUAWEI

- If the authentication point and policy enforcement point are located on different devices, the IP-security group entries of authenticated users need to be pushed to the specified policy enforcement point. An administrator can configure subscription on iMaster NCE-Campus to specify the entries of which network segments or security groups to be pushed to which policy enforcement points.

Typical Free Mobility Solution



(Recommended) Core switch functioning as the user gateway, authentication point, and policy enforcement point

- The core switch is used as the Layer 3 gateway and authentication point for wired and wireless users, as well as the policy enforcement point in the free mobility solution.
- As the policy enforcement point is deployed at the upper layer, configure port isolation on access and aggregation switches to prevent traffic from being directly transmitted through the access or aggregation switches without passing through the core switch.
- The policy enforcement points control mutual access between users and access from users to network resources such as servers, and are deployed on core switches.
- When 802.1X authentication is used, configure transparent transmission of 802.1X packets on the access and aggregation switches if the authentication point is located on a core switch.

HUAWEI

Quiz

1. (True or false) In the free mobility solution, a policy enforcement point must be an authentication point and is typically deployed on a user gateway. ( )
   A. True
   B. False
2. (True or false) Free mobility implements policy management and permission control based on security groups. Rules such as VLANs and ACLs need to be planned during network design. ( )
   A. True
   B. False

HUAWEI

1. B
2. B

Summary

- Free mobility transforms IP-based policies into security group–based policies, which are easier to understand and decouple policies from IP addresses. This allows network administrators to implement policy control between security groups without considering the IP addresses of users.
- Network objects of the same type and with the same permissions are added to one security group. Members in a security group can be PCs, mobile phones, printers, and servers. After network objects are divided into different security groups, administrators can define security group policies to determine the network services that each security group can use, including access rights and application control.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Large- and Medium-Sized Virtualized Campus Network Design

HUAWEI

Foreword

- Huawei CloudCampus 3.0 Solution (hereafter referred to as Huawei CloudCampus Solution) adopts a new philosophy of Intent-Driven Network (IDN), and introduces big data analytics and Artificial Intelligence (AI) technologies into cloud and software-defined networking (SDN). It helps enterprises build intelligent, simplified, converged, open, and secure networks.
- This course describes the planning and design process for large- and medium-sized virtualized campus networks that use the CloudCampus Solution, including the network architecture design, underlay network design, fabric and overlay network design, admission control and free mobility design, WLAN design, egress network design, network security and QoS design, as well as O&M design.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the network layers and network architecture of large- and medium-sized virtualized campus networks.
  - Complete the network architecture design and underlay network design for a virtualized campus network based on actual requirements.
  - Complete the fabric and overlay network design for a virtualized campus network based on actual requirements.
  - Complete the admission control design, free mobility design, WLAN design, egress network design, network security and QoS design, and O&M design for a virtualized campus network based on actual requirements.

HUAWEI

Contents

1. **CloudCampus Solution and Virtualized Campus Network Overview**
2. Network Architecture Design
3. Underlay Network Design
4. Fabric and Overlay Network Design
5. Admission Control and Free Mobility Design
6. WLAN Design
7. Egress Network Design
8. Network Security and QoS Design
9. O&M Design

HUAWEI

Large- and Medium-Sized Campus Networks

- Different from the wide area network (WAN) and data center network, a campus network typically refers to the internal network of an enterprise or organization. A campus network is built for more efficient running of key enterprise services.
- In terms of network scale, campus networks can be classified into large- and medium-sized campus networks as well as small- and medium-sized campus networks. A large- or medium-sized campus network typically has:
  - Over 2,000 end users
  - Over 100 network elements (NEs)
- Some enterprises have branches dispersed across different areas, and each branch network can be considered a single campus network.

HUAWEI

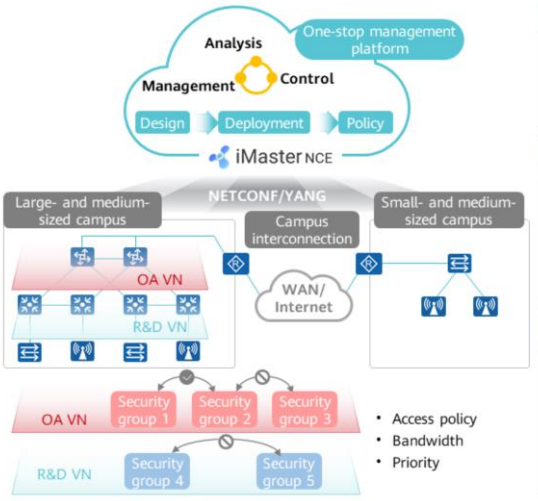
Service Requirements and Challenges of Large- and Medium-Sized Campus Networks

- Enterprises' campus networks are the cornerstone of their digital transformation. Nowadays, mobile office, cloud computing, SDN, Internet of Things (IoT), artificial intelligence (AI), and big data are gaining momentum. Driven by this, new technologies and applications are constantly emerging and are making inroads to enterprise campus networks, which poses many new challenges for the campus networks.

Converged network	Automated network deployment	User experience awareness
Requirements: Diversified access terminals and services require a converged campus network. **Challenges:** • Wi-Fi and IoT services are independently planned, deployed, and managed, resulting in high network construction costs. • The network management and O&M workload is heavy.	**Requirements:** Network deployment should be automated to address the growing complexity in deployment and policies due to the surge in applications and services. **Challenges:** • Manual configuration is repetitive, complex, and labor-intensive. • New service rollout requires configuring devices one by one, which is time-consuming and costly. • The workload of network policy deployment and adjustment is heavy.	**Requirements:** Network O&M should become automated and intelligent, with insights into user experience anytime, anywhere. **Challenges:** • Service faults cannot be detected in a timely manner. • Fault locating heavily relies on the O&M skills and experience of professionals, so faults cannot be quickly located. • The network cannot be self-optimized.

HUAWEI

Huawei CloudCampus 3.0 Solution (CloudCampus Solution)



Rapid network provisioning, improving deployment efficiency

- **Device plug-and-play**: simplified device deployment, scenario-specific guided configuration, template-based configuration
- **Simplified network deployment**: network resource pooling, multi-purpose network, automatic service provisioning

Fast service provisioning, improving user experience

- **Free mobility**: GUI-based policy configuration; access anytime anywhere, with consistent permission and experience during roaming
- **Intelligent terminal identification**: anti-spoofing for terminal access, with the accuracy of intelligent terminal identification reaching over 95%
- **Intelligent HQoS**: application-based traffic scheduling and shaping and fine-grained bandwidth management ensure service experience of VIP users

Quick intelligent O&M, improving network performance

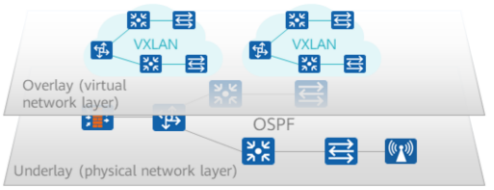
- **Real-time experience visualization**: uses Telemetry for network experience visualization at any moment, for any user, and in any area
- **Precise fault analysis**: proactively identifies typical network issues and provides suggestions; compares and analyzes real-time data to predict faults
- **Intelligent network optimization**: predictive optimization of wireless networks based on historical data, improving overall network performance

HUAWEI

CloudCampus Highlights: Simplified Deployment

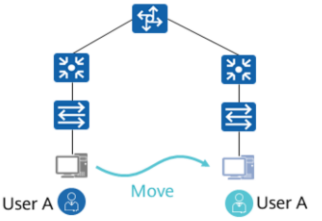
Physical and virtual network automation

- **Automated deployment of a physical network: Devices can be pre-configured for plug-and-play.**
  - iMaster NCE-Campus provides a GUI for automated deployment of devices, as well as route orchestration and interworking configuration of underlay networks.
- **Automated provisioning of virtual networks: Virtual networks can be automatically created to achieve "one network for multiple purposes".**
  - With iMaster NCE-Campus, fabric networks can be deployed, and VXLAN tunnels can be automatically set up based on the BGP EVPN control plane, so as to achieve automated virtual network construction, centralized service configuration, and automatic service provisioning.



User policy automation

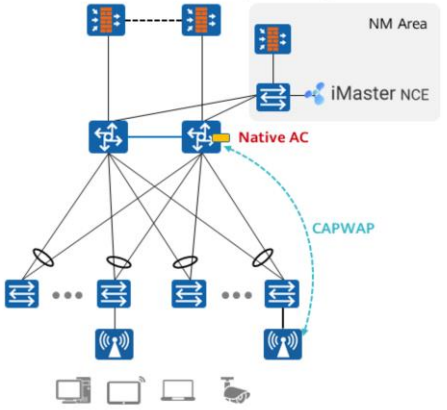
- **Centering on users, services, and experience, policies can migrate with users, ensuring a consistent service experience.**
  - The free mobility solution uses iMaster NCE-Campus to plan security groups and inter-group policies and automatically deliver policies to network devices.
  - When an authenticated user accesses the network using different terminals at different locations, iMaster NCE-Campus automatically identifies the user and delivers relevant user policies to the corresponding policy enforcement device on the network. This achieves consistent and assured user access experience. A user can have consistent policies and service experience, irrespective of the access location.



HUAWEI

CloudCampus Highlights: Wired and Wireless Convergence

- Huawei campus switches integrate the WLAN access controller (AC) functionality to implement wired and wireless convergence, providing unified management and experience for wired and wireless users.



Unified forwarding

Wired and wireless traffic is centrally processed by the core switch before being forwarded.

Unified authentication

The core switch functions as the unified authentication point and Layer 3 gateway for both wired and wireless users.

Unified policy execution

The core switch is the policy enforcement point for both wired and wireless users.

HUAWEI

- By integrating the user authentication, user management, and policy association functions, the CloudCampus Solution provides unified authentication and access policy control for both wired and wireless users. Administrators can have a consistent user management experience, with simplified O&M for wired and wireless networks.

CloudCampus Highlights: WLAN and IoT Convergence

- Huawei APs integrate IoT modules to provide functions of IoT base stations, implementing Wi-Fi and IoT convergence as well as simplified management.

Scenarios and challenges

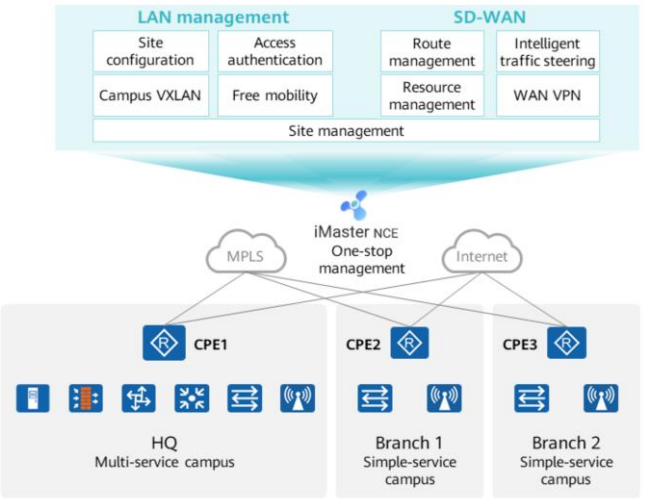
- Scenarios: retail, healthcare, education, enterprise, and other campuses where innovative digital services need to be provided based on IoT.
- Challenges: Wi-Fi and IoT (such as Bluetooth and RFID) networks are deployed separately. Numerous wireless networks are deployed, resulting in high costs and inflexible service expansion. There is also radio interference between these wireless networks, affecting service experience.

Huawei IoT AP

- Wi-Fi & IoT converged architecture.
- Converged site for the AP and IoT base station, reducing auxiliary resources (such as access and power supply management) by 50%.
- Cloud-based management and plug-and-play, facilitating service configuration.
- Wi-Fi and IoT configuration association, allowing automatic Wi-Fi channel switching when a conflict is detected.

HUAWEI

CloudCampus Highlights: LAN and WAN Convergence (1)



Integrated GUI, improving deployment and O&M efficiency

Integrated deployment, integrated policy, integrated O&M

Diverse SD-WAN policies, improving application experience

Identification of 6000+ applications, application-based intelligent traffic steering, application-specific fine-grained QoS scheduling

One platform, reducing user investment

iMaster NCE-Campus can be deployed in a single-node system or a cluster.

HUAWEI

- CPE: Customer Premises Equipment
- GUI: Graphical User Interface

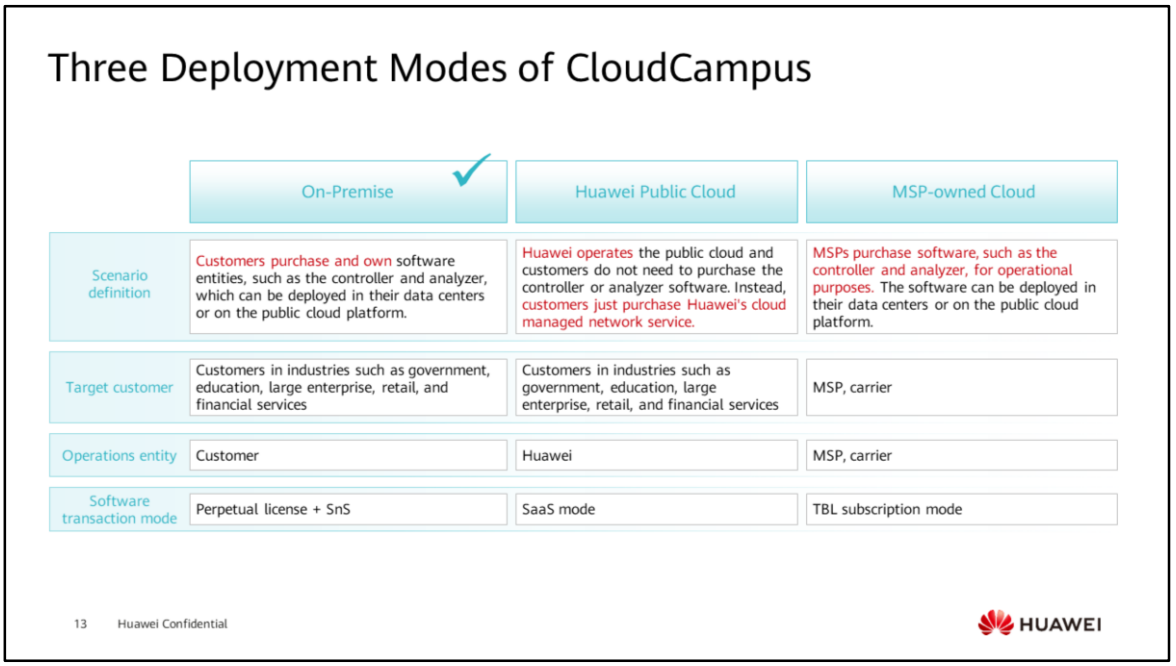
CloudCampus Highlights: LAN and WAN Convergence (2)

LAN-WAN interconnection solution

- The CloudCampus Solution provides two technologies for campus egress interconnection: **IPsec VPN** and **SD-WAN EVPN**.
  - IPsec VPN is a static VPN technology that creates a VPN tunnel by setting up an IPsec tunnel between sites. It diverts traffic to a VPN tunnel based on the configured static network segments so that the sites can access each other through the VPN tunnel.
  - SD-WAN EVPN is a dynamic VPN technology that establishes tunnels between sites on demand and dynamically advertises routes. EVPN creates VPN tunnels between sites by setting up GRE tunnels between them. It supports IPsec encryption on the GRE tunnels, securing data transmission.
- Application experience optimization policies provided on the WAN side:
  - Application identification, intelligent traffic steering, QoS, NAT policy, application optimization, etc.

HUAWEI

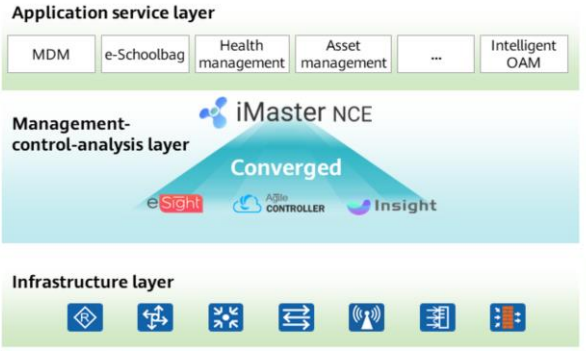
- Application identification:
  - Precise identification of applications on a network is the prerequisite and basis for network services such as intelligent traffic steering, QoS, application optimization, and security. Service policies can be applied in subsequent service processes only after applications are identified.
  - Two application identification methods are available in the SD-WAN EVPN Interconnection Solution: first packet inspection (FPI) and service awareness (SA).
- Intelligent traffic steering:
  - The SD-WAN EVPN Interconnection Solution supports traffic steering based on the application quality, load balancing, application priority, and bandwidth.



- Managed service provider (MSP): delivers and manages network-based services, applications, and devices. The serviced objects include enterprises, residential areas, and other service providers.
- Perpetual license + SnS: The perpetual license is sold together with SnS services, such as software patches, software upgrades (including new features of new releases), and remote support. In the perpetual license + SnS mode, a customer needs to pay SnS fee for a certain period of time, in addition to purchasing the license upon the first purchase. If the customer does not renew the SnS annual fee after it expires, the customer can only use functions provided in the license for the current version and cannot use the service functions covered in the SnS annual fee.
- SaaS mode: MSPs are responsible for deploying or leasing hardware infrastructure, and O&M and management of the hardware and software. Software is provided for customers as cloud services and customers need to periodically pay for the cloud services.
- Term Based License (TBL) mode: This mode differs from the perpetual license + SnS mode in that the licenses purchased by customers have limited validity periods. If a customer does not renew the subscription after the license expires, the customer can no longer use the software product.
- SnS: refers to Subscription and Support. It consists of two parts: software support and software subscription. The complete software charging mode consists of the annual software SnS fee and software license fee.
- Note: This course uses the on-premise deployment as an example.

CloudCampus Solution Components: iMaster NCE-Campus

- iMaster NCE-Campus is the configuration and management platform used in the CloudCampus Solution. It provides a portal for CloudCampus service configuration, O&M, and monitoring.



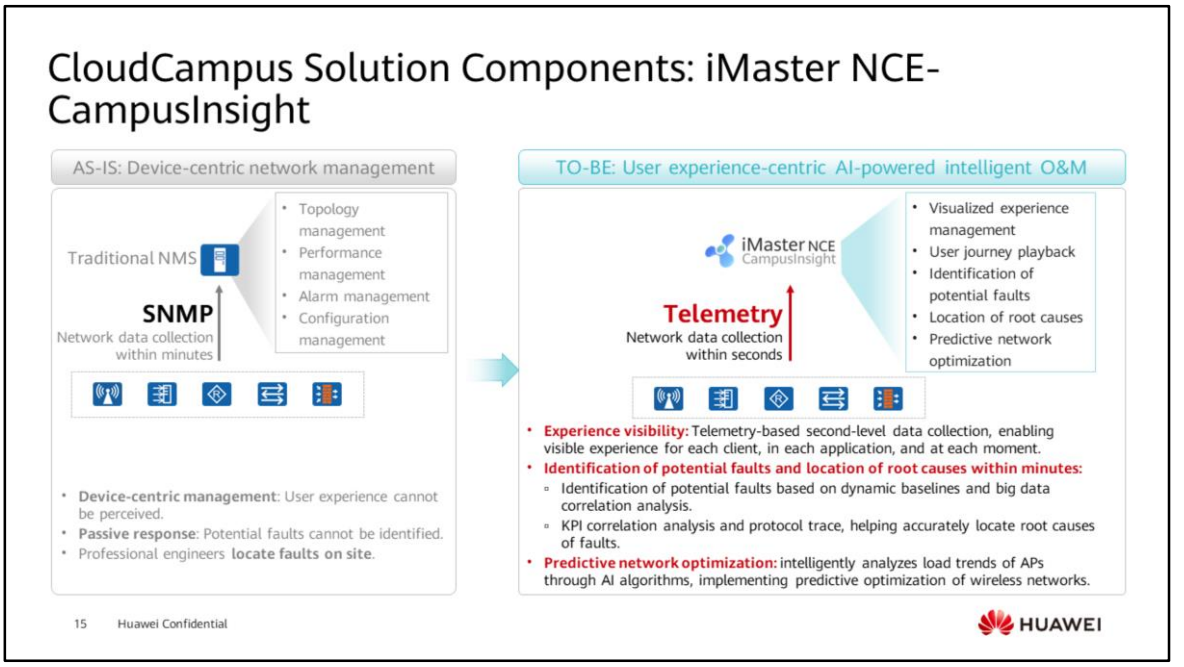
Automated + Intelligent
- SDN-based automatic service configuration/deployment
- AI-powered intelligent analysis/prediction/troubleshooting

Manage + Control + Analyze
- Unified data base
- Centralized detection/locating/processing

Plan + Construct + Maintain + Optimize
- Full lifecycle management
- Simulation/Verification/Monitoring/Optimization

iMaster NCE-Campus, an autonomous driving campus network management and control system

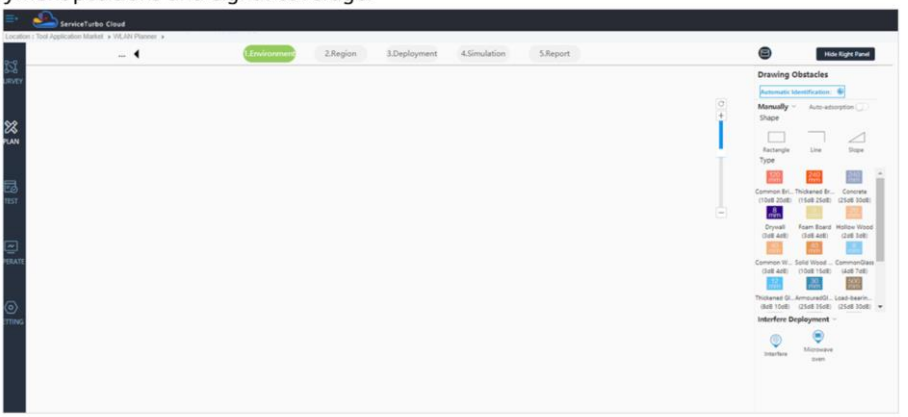
HUAWEI



- Using algorithms to improve efficiency and leveraging scenario-specific continuous learning and expert experience, intelligent O&M frees O&M personnel from nerve-wracking alarms and noises, making O&M automated and intelligent.

CloudCampus Solution Components: WLAN Planner

- WLAN Planner is an efficient WLAN network planning tool. It simulates signals to determine APs' deployment positions and signal coverage.



HUAWEI

- Efficiently supports WLAN network planning:
  - No installation: Cloud-based tool, without the need of software installation.
  - All-scenario: All scenarios supported, including indoor, outdoor, agile distributed, and high-density scenarios.
  - Efficient: Improved simulation efficiency when compared with the traditional standalone deployment.
  - High quality: Supports tens of thousands of WLAN projects.
- For more information about the WLAN, such as WLAN planning, SSID planning, and radio calibration, see the Small- and Medium-Sized Cloud-Managed Campus Network Design or HCIX-WLAN series courses.

CloudCampus Solution Components: Network Hardware Products

CloudEngine S series switches

High-performance campus switches

NetEngine AR routers

Campus egress routers
Integrate routing, switching, Wi-Fi, 5G, and security functions

AirEngine Wi-Fi 6 APs

Campus APs
Models that apply to all scenarios: indoor settled, indoor wall plate, agile distributed, and outdoor AP models

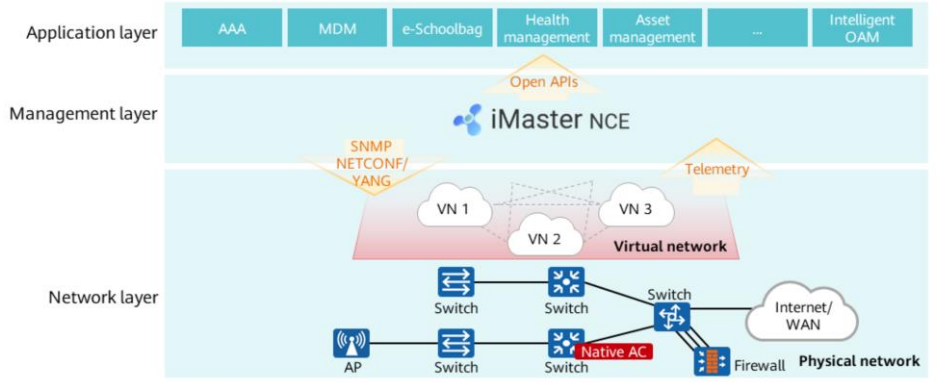
HiSecEngine AI firewalls

Firewalls
Provide comprehensive and integrated network security protection capabilities

HUAWEI

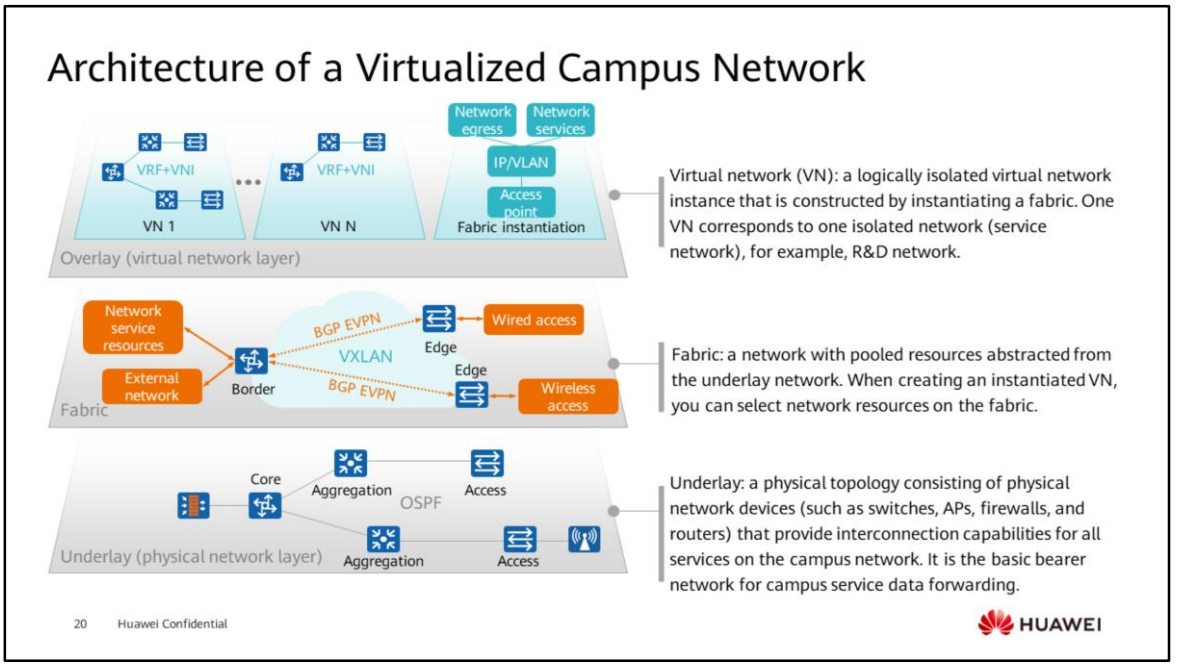
Network Architecture of the CloudCampus Solution (VXLAN-based Virtualized Campus Network)

- The following figure shows the typical network architecture of the CloudCampus Solution, which consists of the network layer, management layer, and application layer.



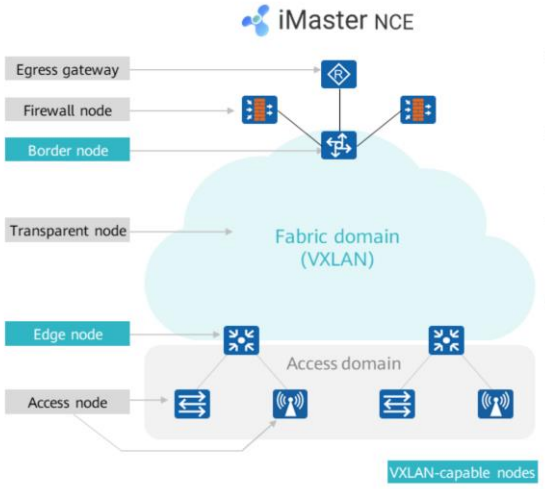
HUAWEI

- Network layer
  - Virtualization technologies are introduced to divide the network layer into a physical network and a virtual network.
    - Physical network: is also called the underlay network and provides basic connection services for a campus network. To meet access requirements of multiple types of terminals, the physical network provides converged access for three networks, allowing simultaneous access of wired, wireless, and IoT terminals.
    - Virtual network: is also called the overlay network. Virtualization technology is used to construct one or more overlay networks on top of the underlay network. Service policies are deployed on the overlay networks and are separated from the underlay network, decoupling services from networks. Multiple overlay networks can serve different services or customer segments.



- On a large- or medium-sized campus network, the virtualization solution may need to be used to decouple services from the network, so as to build a multi-purpose network and achieve flexible, fast service deployment without changing the basic network infrastructure. Using such a solution means that the virtualized campus network architecture must be different from the traditional network architecture.
- This slide presents the virtualized campus network architecture. The underlay is the physical network layer, and the overlay is the virtual network layer constructed on top of the underlay using the Virtual Extensible LAN (VXLAN) technology.

Network Nodes on a Virtualized Campus Network



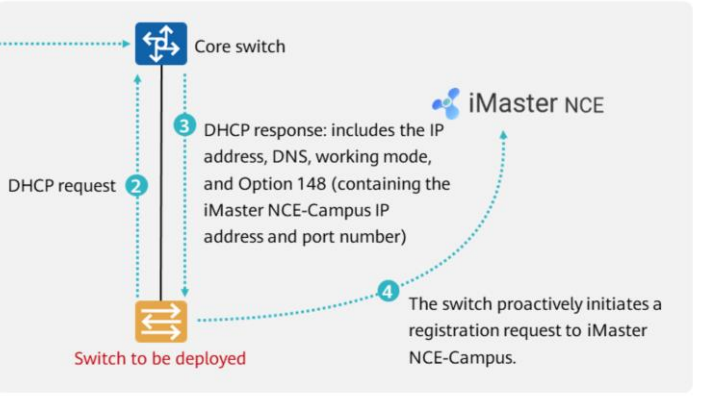
- **Egress gateway**: is an egress device of the campus network, which can be an AR router or a firewall.
- **Firewall node:** This node is required when Layer 4 to Layer 7 security policies are deployed. It can be deployed in off-path mode or at the campus egress.
- **Border node**: implements communication between the fabric and external networks. It is typically a core switch.
- **Transparent node:** It does not need to support VXLAN.
- **Edge node**: is a fabric edge device that connects user-side devices to the fabric. Data packets from wired users enter the VXLAN network through edge nodes.
- **Access node**: is typically an access switch (wired access node) or an AP (wireless access node). Wired access nodes can function as edge nodes, that is, VXLAN is deployed across core and access nodes. If wired access nodes do not need to support VXLAN, aggregation nodes can function as edge nodes — that is, VXLAN is deployed across core and aggregation nodes — in which scenario policy association can be deployed on wired access nodes and edge nodes.

HUAWEI

- On a fabric network, a VXLAN tunnel endpoint (VTEP) can function as either a border or edge node:
  - Border node: It corresponds to a physical network device and forwards data between the fabric and external networks. In most cases, border nodes are VXLAN-capable core switches.
  - Edge node: It corresponds to a physical network device. User traffic enters the fabric from the edge node. Typically, edge nodes are VXLAN-capable access or aggregation switches.
- Policy association:
  - Policy association provides a solution to contradiction between policy strengths and complexity on large campus networks. In the solution, user access policies are centrally managed on the gateway devices and enforced by gateway and authentication access devices.
  - After policy association is configured, authentication access devices can transparently transmit BPDUs and report user logout and user access positions in real time. In addition, the authentication control device requests authentication access devices to enforce user access policies, thus controlling user access to the network.

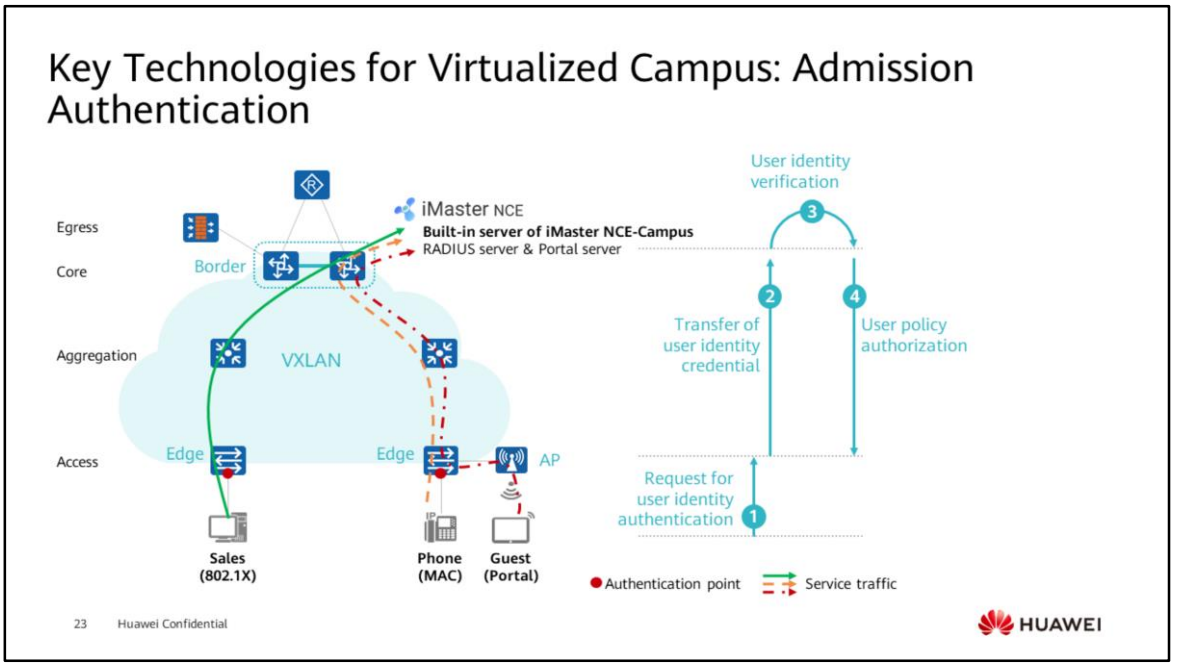
Key Technologies in CloudCampus: DHCP-based Plug-and-Play of Network Devices

- The network administrator has deployed the DHCP service on the network. (It is recommended that the DHCP service be deployed on the core switch.)
- In addition to delivering an IP address to the device to be deployed, the DHCP server uses DHCP Option 148 to notify the device of the iMaster NCE-Campus IP address and port number.



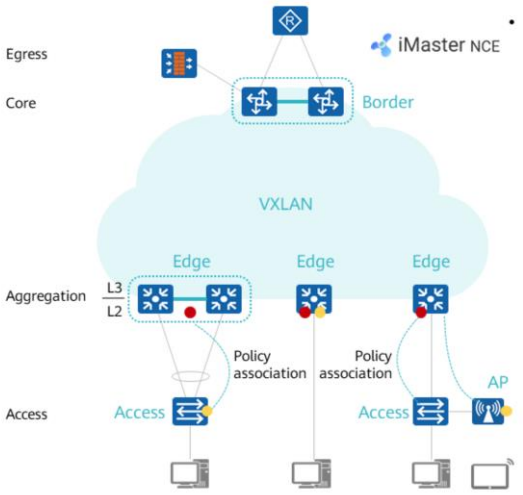
Devices supported: AR routers, switches, APs

HUAWEI



1. Request for user identity authentication: The terminal sends its identity credential to the admission device.
2. User identity authentication: The admission device sends the identity credential to the admission server for identity authentication.
3. User identity verification: The admission server stores user identity information and manages users. After receiving the identity credential of the terminal, the admission server verifies the identity of the terminal, determines whether the terminal identity is valid, and delivers the verification result and policy to the admission device.
4. User policy authorization: As a policy enforcement device, the admission device implements policy control over the terminal based on the authorization result provided by the admission server, for example, permitting or denying network access, or performing more complex policy control on the terminal. Complex policy control can be increasing or decreasing the forwarding priority of the terminal, or restricting the network access rate of the terminal.

Key Technologies for Virtualized Campus: Policy Association



- If access switches do not support VXLAN, policy association can be deployed between access and aggregation switches (gateways). The gateway manages user access policies in a unified manner, and the access switch enforces user access policies.
  - Authentication control points and enforcement points are connected through CAPWAP tunnels (which are the management tunnels used by policy association).
  - CAPWAP tunnels implement user association, message transmission, user authorization policy delivery, user information synchronization, and other functions.
  - After policy association is configured, authentication enforcement points can transparently transmit BPDUs, report user logout and access locations in real time, and enforce user access policies, thereby controlling user access to the network.

Authentication control point Authentication enforcement point

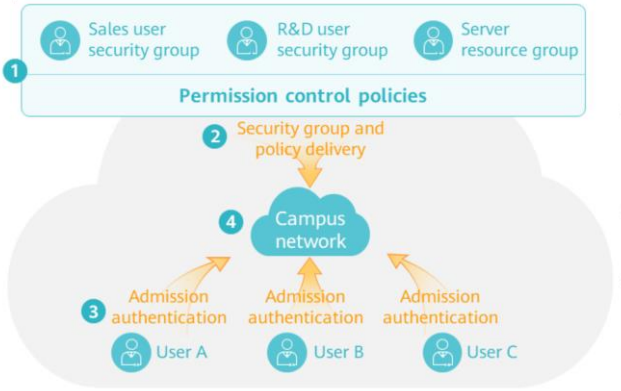
CAPWAP tunnel Link aggregation CSS/iStack

HUAWEI

- CAPWAP: Control and Provisioning of Wireless Access Points

Key Technologies for Virtualized Campus: Free Mobility

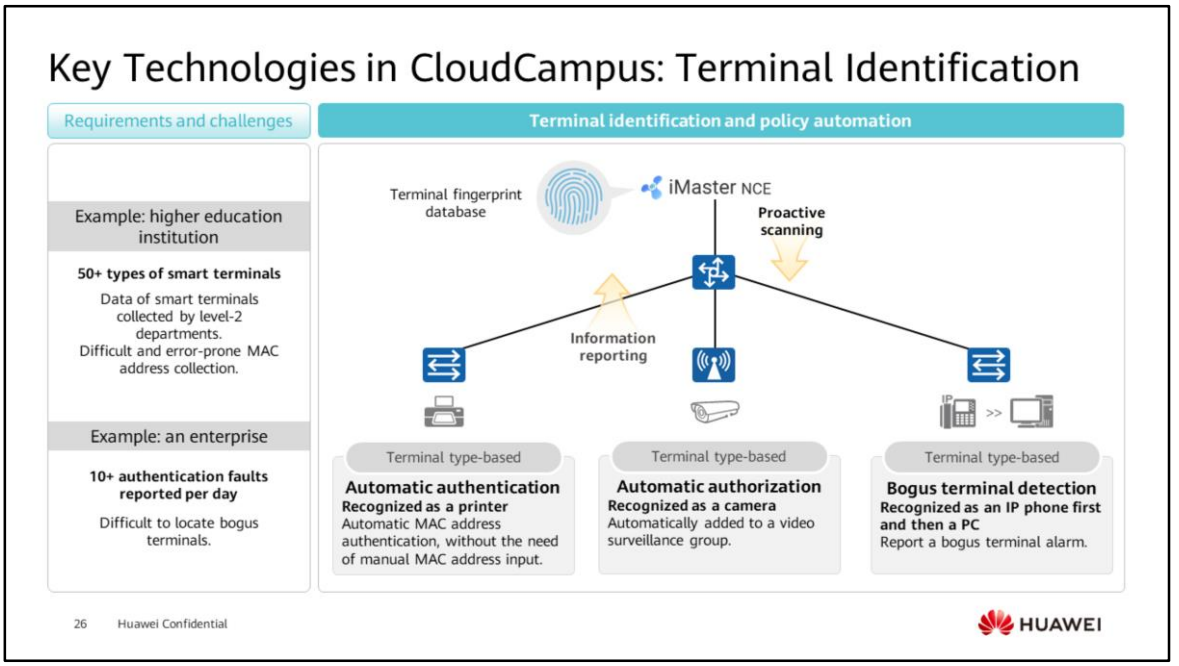
- Free mobility manages policies and controls permissions based on security groups. Regardless of user locations and IP addresses in use, this technology ensures that users can obtain consistent network permissions and corresponding policies can be enforced on the users.



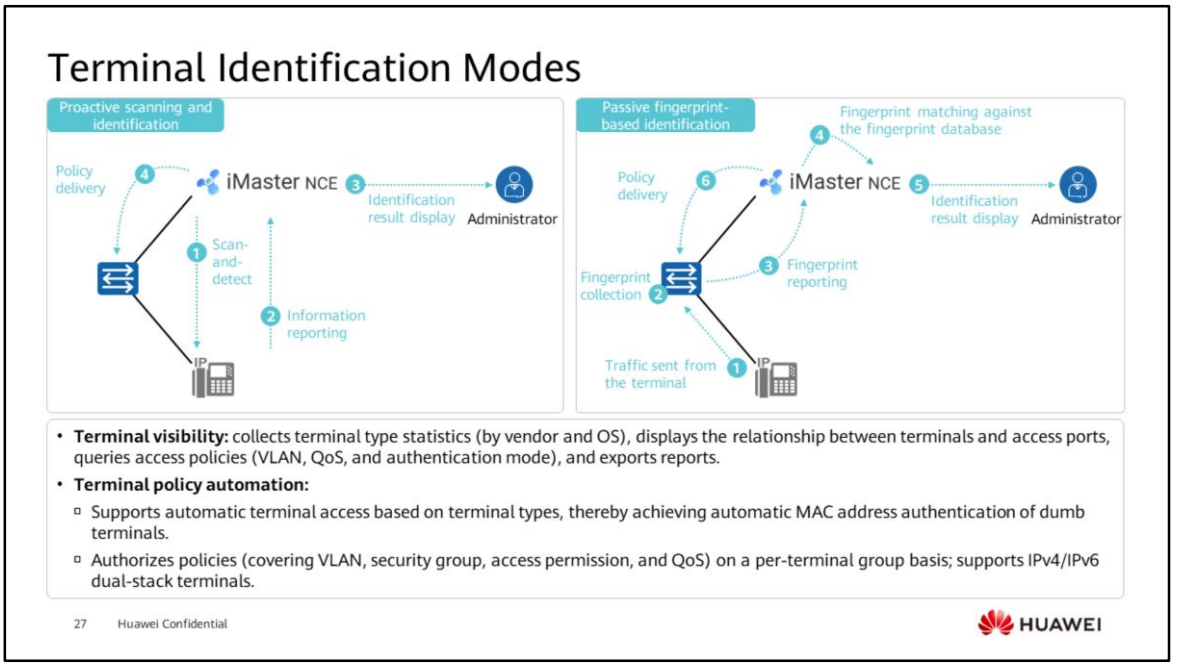
1. Create security groups. A security group is a group of users or network resources that have the same network access policies.
2. Define permission control policies based on security groups and deliver the policies to network devices.
3. Authorized security groups are assigned to the users who pass admission authentication.
4. After user traffic enters the network, network devices enforce policies based on the corresponding source and destination security groups of the traffic.

HUAWEI

- Free mobility introduces the concept of security group. Security groups are related only to user identities and are completely decoupled from network information such as user VLANs and IP addresses.
- User policy management and permission control are performed based on security groups.



- On large- and medium-sized campus networks, access terminals include smart terminals (such as PCs and mobile phones) and dumb terminals (such as IP phones, printers, and IP cameras). Currently, terminal management on campus networks faces the following challenges:
    - The network management system (NMS) can only display the IP and MAC addresses of access terminals, but cannot identify the specific terminal type. As a result, the NMS cannot provide refined management for network terminals.
    - Network service configurations and policies vary according to the terminal type. Consequently, administrators need to manually configure different services and policies for each type of service terminals, complicating service deployment and operations.
- To address these challenges, Huawei provides the automatic terminal identification and policy delivery solution, which delivers the following functions:
    - iMaster NCE-Campus can display the network-wide terminal types and operating systems, for example, dumb terminals including printers, IP cameras, smart all-in-one card, and access control system. iMaster NCE-Campus can also collect statistics and display traffic by terminal type.
    - Administrators do not need to manually configure different services and policies for different types of dumb terminals such as IP phones, printers, and IP cameras on the campus network. Instead, iMaster NCE-Campus can automatically identify these terminals and deliver the corresponding admission policies and service configurations to them.



- Passive fingerprint-based identification: Network devices collect fingerprints of packets sent by terminals and report the fingerprints to iMaster NCE-Campus, which matches the fingerprints against its fingerprint database for terminal type identification. In this mode, terminals can be identified through a MAC organizationally unique identifier (OUI), HTTP User-Agent, DHCP Option, LLDP, and multicast DNS (mDNS).
- Proactive scanning and identification: iMaster NCE-Campus proactively detects or scans terminals, and identifies terminal types based on fingerprint information of the terminals. In this mode, terminals can be identified through SNMP query or network mapper (Nmap).

Terminal Identification Methods

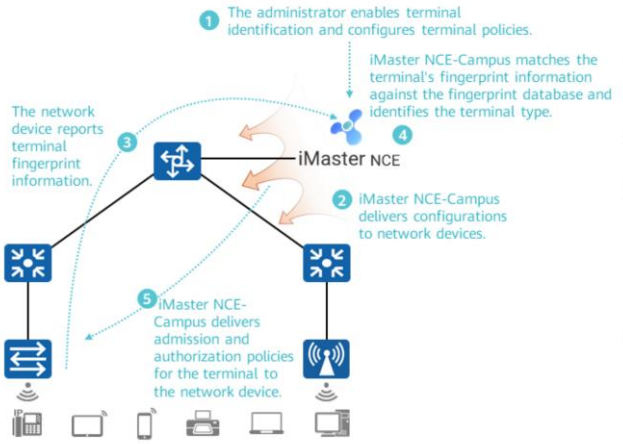
- The terminal management function of iMaster NCE-Campus can help identify terminals and display the terminal type, operating system, and manufacturer.
- The following table describes terminal identification methods.

Type	Identification Method	Method Description	Application Scenario
Passive fingerprint-based identification (information reporting)	MAC OUI	The first three bytes of a MAC address represent the manufacturer.	Identifies the device manufacturer only.
	HTTP User-Agent	A browser's User-Agent string contains the manufacturer, terminal type, operating system, browser type, and other information.	Mobile phones, tablets, PCs, workstations, smart audio and video terminals.
	DHCP Option	Some options of a terminal's DHCP packets can be used to classify terminals, for example, DHCP options 55, 60, and 12.	Mobile phones, tablets, PCs, workstations, IP cameras, IP phones, printers, etc.
	LLDP	LLDPDUs carry device model information.	IP phones, IP cameras, network devices, etc.
	mDNS	mDNS packets contain terminal model and service information.	Printers, IP cameras, etc.
Proactive scanning and identification	SNMP Query	Obtains identification information by querying device information-related SNMP MIB objects.	Network devices, printers, etc.
	Nmap	Scans the OS and services of terminals to detect the terminal model and OS information.	PCs, workstations, printers, phones, IP cameras, etc.

HUAWEI

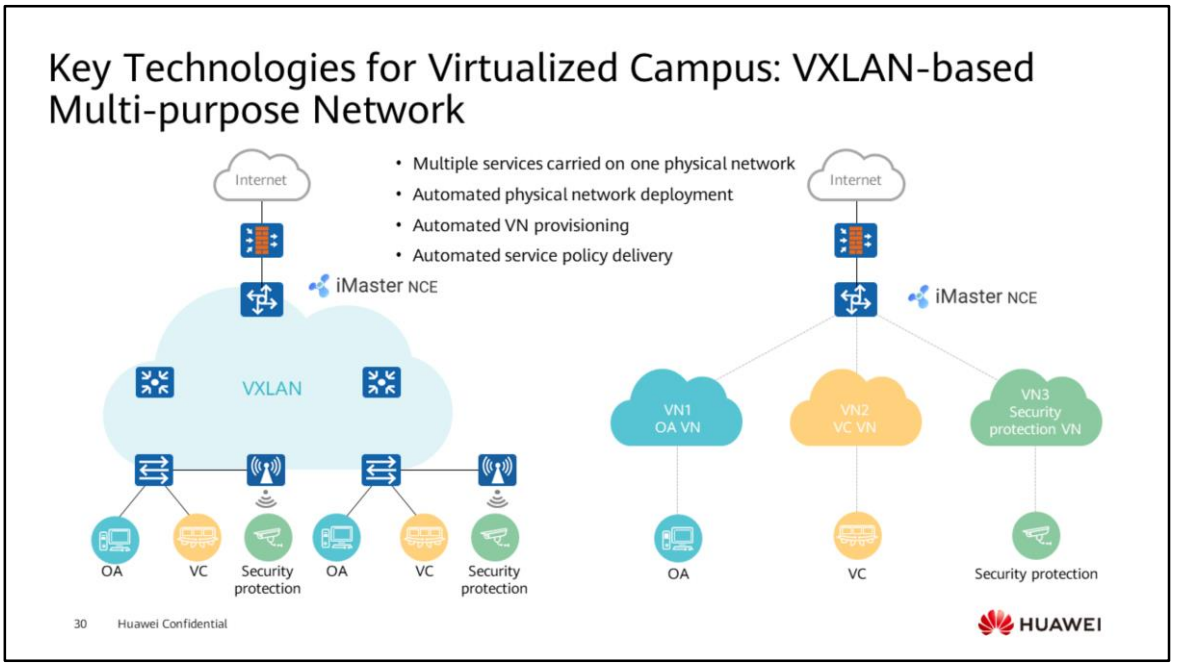
- When terminals access the network, network devices can collect terminal information and report the information to iMaster NCE-Campus. Alternatively, iMaster NCE-Campus can proactively scan terminals to identify the terminal type, OS, and manufacturer.

Process of Automatic Policy Delivery Based on Terminal Types



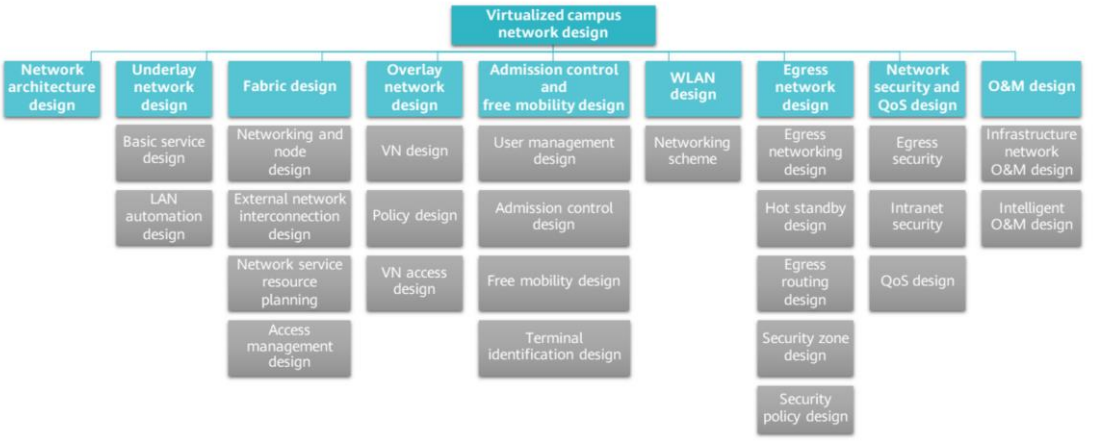
1. On the iMaster NCE-Campus GUI, an administrator enables the terminal identification function, selects terminal types, and specifies the corresponding policies.
2. iMaster NCE-Campus delivers terminal identification configurations to network devices.
3. When terminals access the network, network devices collect the fingerprint information of the terminals and report the information to iMaster NCE-Campus.
4. iMaster NCE-Campus automatically matches the terminals' fingerprint information against the fingerprint database to identify the terminal types.
5. iMaster NCE-Campus automatically delivers admission and authorization policies to the terminals based on the policies defined by the administrator.

HUAWEI



- Using virtualization technologies, multiple virtual networks (VNs) can be created on top of a physical network. Different VNs are used for different services, such as OA, videoconferencing, and security protection.
- VC: Video Conference

Overview of Large- and Medium-Sized Virtualized Campus Network Design



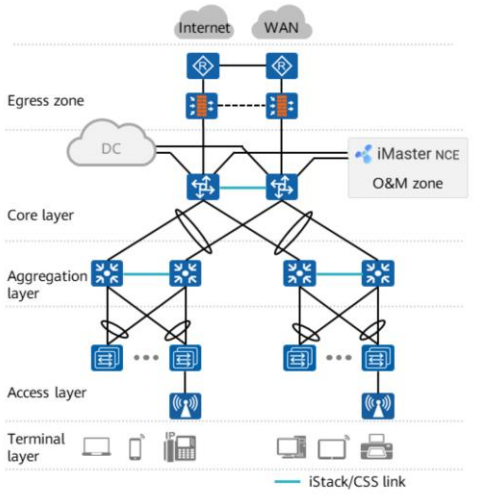
HUAWEI

Contents

1. CloudCampus Solution and Virtualized Campus Network Overview
2. **Network Architecture Design**
3. Underlay Network Design
4. Fabric and Overlay Network Design
5. Admission Control and Free Mobility Design
6. WLAN Design
7. Egress Network Design
8. Network Security and QoS Design
9. O&M Design

HUAWEI

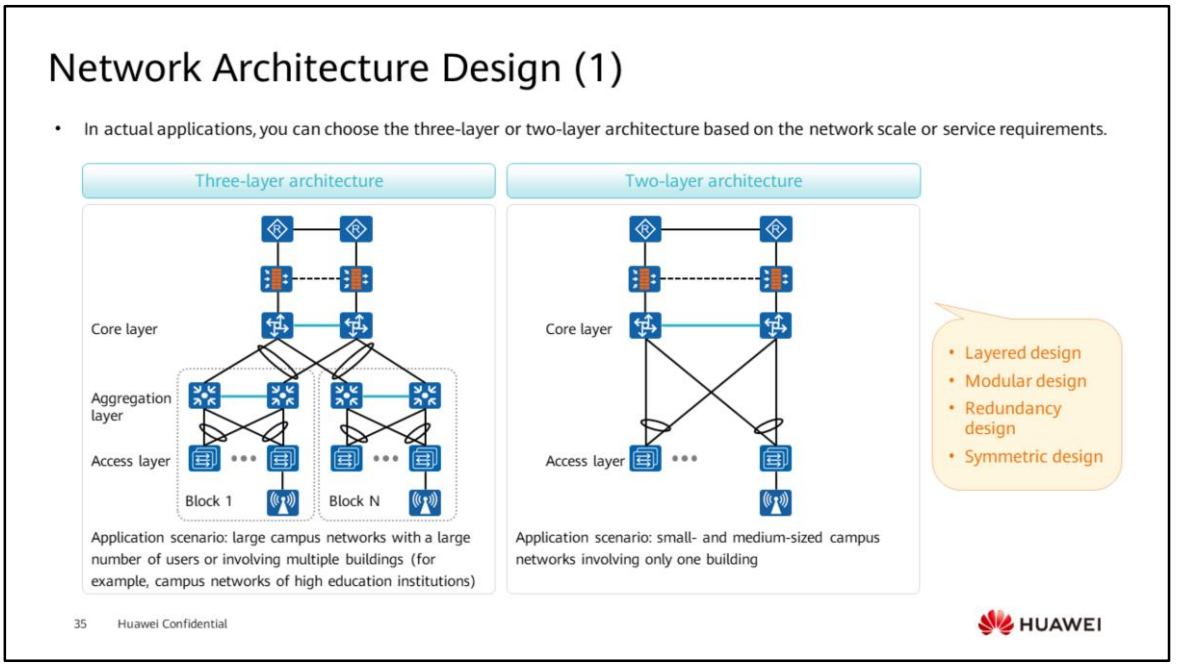
Network Architecture Design Overview



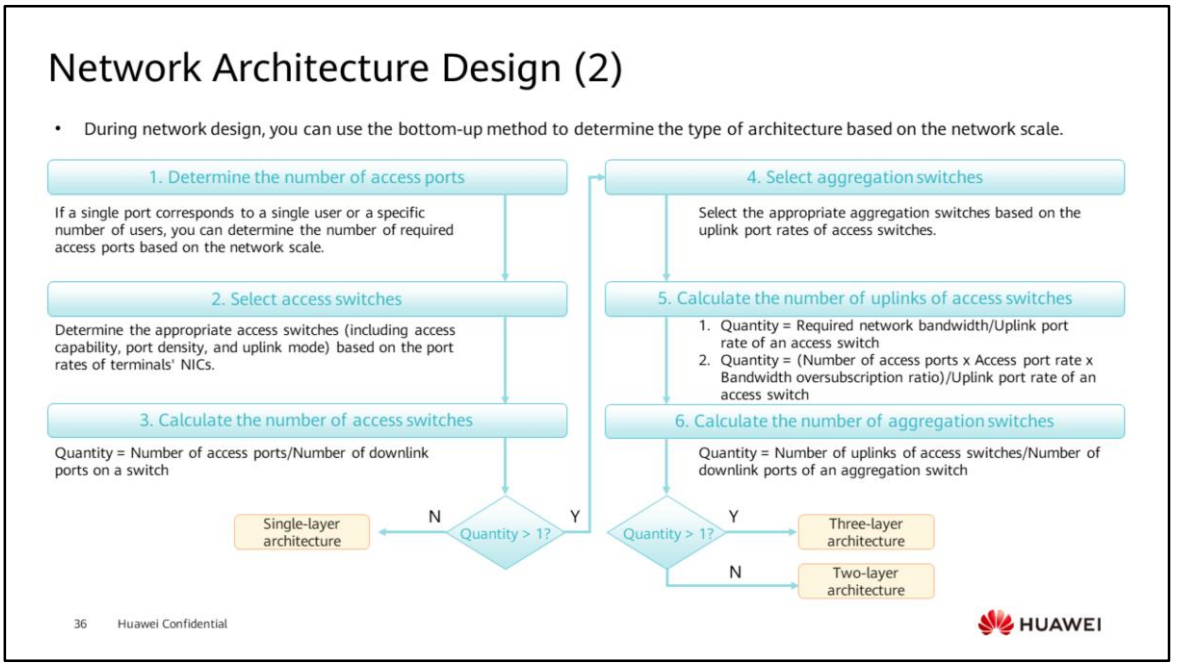
- Overall design principles:
  - Tree and ring(Core devices) network topologies
- Reliability considerations:
  - High reliability of nodes: CSS, iStack, and hot standby (AC, firewall, etc.)
  - High reliability of links: redundant links and Eth-Trunk
- Network layer design principles:
  - Two-layer networking: The network layers are simple, and faults are easy to locate.
  - Three-layer networking: applies to campuses with multiple buildings or regions.

HUAWEI

- Large- and medium-sized campus networks often use the tree topology with the core layer as the root, as shown in the figure. This topology is stable and easy to expand and maintain. A campus network can be divided into the following layers: access layer, aggregation layer, core layer, and multiple zones including the egress zone, DC zone, and O&M zone. Internal changes within a module have limited impact on other modules, facilitating fault location.
- Terminal layer
  - The terminal layer involves various types of terminals that access the campus network, such as PCs, printers, IP phones, mobile phones, and cameras.
- Access layer
  - The access layer provides various access modes for users and is the first network layer for terminals to access a campus network. 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 the terminal layer has wireless terminals, the access layer provides APs that access the network through access switches.
- Aggregation layer
  - The aggregation layer connects the access layer to the core layer. The aggregation layer forwards horizontal traffic between users and forwards vertical traffic to the core layer. It can also function as the switching core for a department or zone and connect the department or zone to the exclusive server zone. In addition, the aggregation layer can further extend the quantity of access terminals.



- Layered design:
  - Each layer can be considered a well-structured module with specific roles and functions. This layered structure is easy to expand and maintain, reducing the design complexity and difficulty.
- Modular design:
  - Each module corresponds to a department, function, or service area. Modules can be expanded flexibly based on the network scale, and adjustment in a department or area does not affect other departments or areas, which facilitates fault locating.
- Redundancy design:
  - Dual-node redundancy design can ensure device-level reliability. Appropriate redundancy improves reliability, but excessive redundancy makes O&M difficult. If dual-node redundancy cannot be implemented, you may consider card-level redundancy, such as dual main control boards or switch fabric units (SFUs), for modular core switches or egress routers. In addition, Eth-Trunk can be deployed for important links to ensure link-level reliability.
- Symmetric design:
  - A symmetric network structure makes the topology clearer and facilitates service deployment, protocol design and analysis.



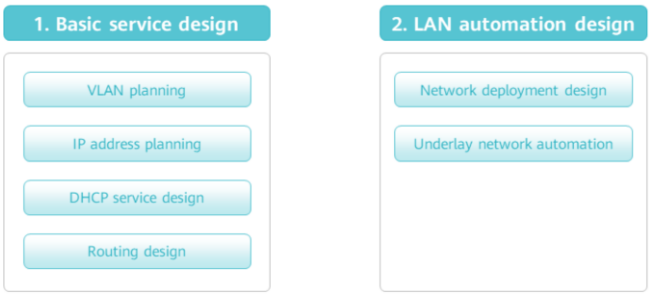
- Design method of the hierarchical model for common network architectures:
    1. Determine the number of access switch ports based on the network scale. Typically, one port corresponds to one terminal or one network access point (for example, AP).
    2. Select switches based on the port rates of terminals' NICs.
    3. Calculate the number of access switches.
        - Number of access switches required = Number of access ports/Number of downlink ports on a switch
        - If the calculation result is greater than 1, aggregation switches need to be deployed. Otherwise, use the single-layer architecture.
    4. Select aggregation switches based on the uplink port rates of access switches.
    5. Calculate the number of uplinks of access switches using either of the following methods:
        - Based on the network bandwidth: Number of uplinks = Network bandwidth/Uplink port rate of an access switch
        - Based on the network scale: Number of uplinks = (Number of access ports x Access port rate x Bandwidth oversubscription ratio)/Uplink port rate of an access switch

Contents

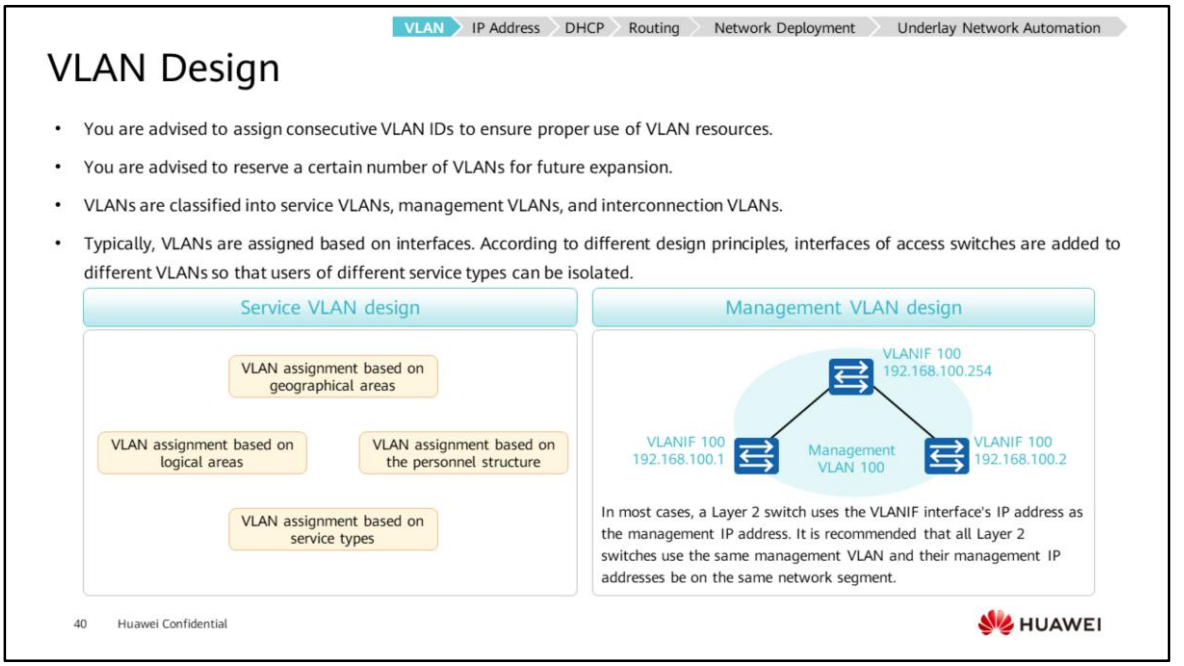
1. CloudCampus Solution and Virtualized Campus Network Overview
2. Network Architecture Design
3. **Underlay Network Design**
4. Fabric and Overlay Network Design
5. Admission Control and Free Mobility Design
6. WLAN Design
7. Egress Network Design
8. Network Security and QoS Design
9. O&M Design

HUAWEI

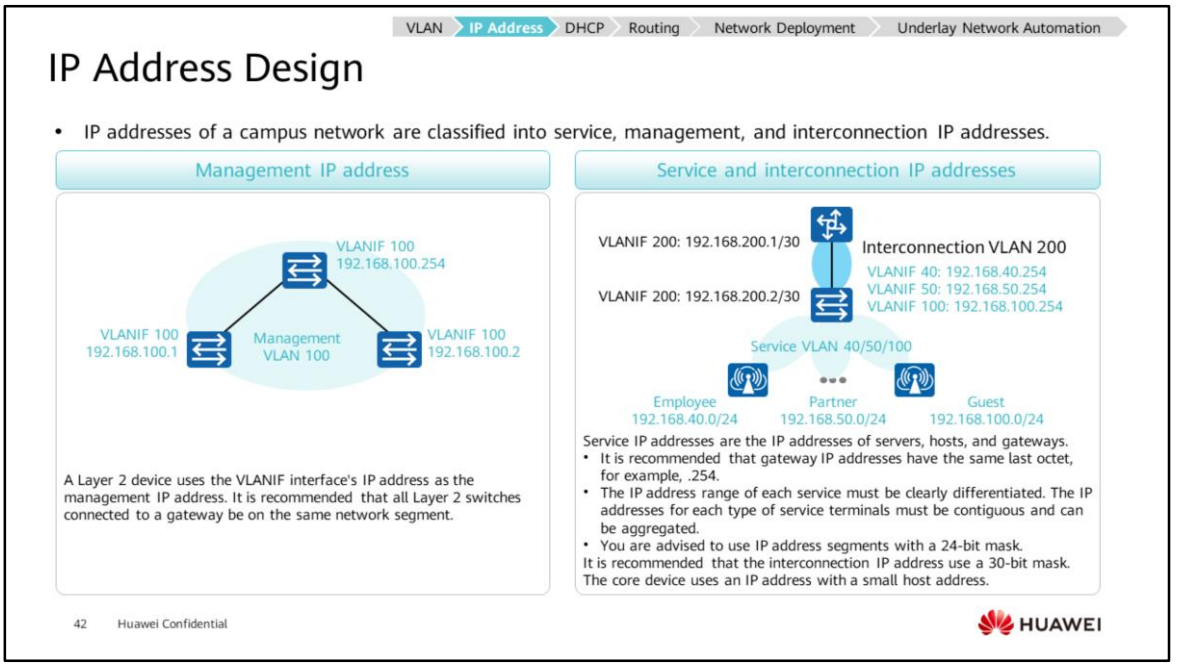
Underlay Network Design Outline



HUAWEI

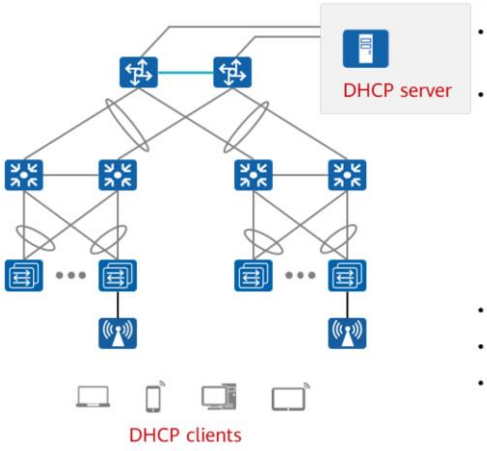


- Service VLAN:
  - Assign VLANs by logical area, geographical area, personnel structure, or service type.
    - Assign VLANs by logical area. For example, VLANs 100 to 199 are used in the core network zone, VLANs 200 to 999 are used in the server zone, and VLANs 2000 to 3499 are used on the access network.
    - Assign VLANs by geographical area. For example, VLANs 2000 to 2199 are used in area A, and VLANs 2200 to 2399 are used in area B.
    - Assign VLANs by personnel structure. For example, department A uses VLANs 2000 to 2009, and department B uses VLANs 2010 to 2019.
    - Assign VLANs by service type. For example, VLANs 200 to 299 are used in the web server zone, VLANs 300 to 399 are used in the app server zone, and VLANs 400 to 499 are used in the database server zone.
  - If users are sensitive to the voice latency, the voice service must be preferentially guaranteed. It is recommended that the voice VLAN be planned for the voice service. Huawei switches can automatically identify voice data, transmit voice data in the voice VLAN, and perform QoS guarantee. When network congestion occurs, voice data can be preferentially transmitted.



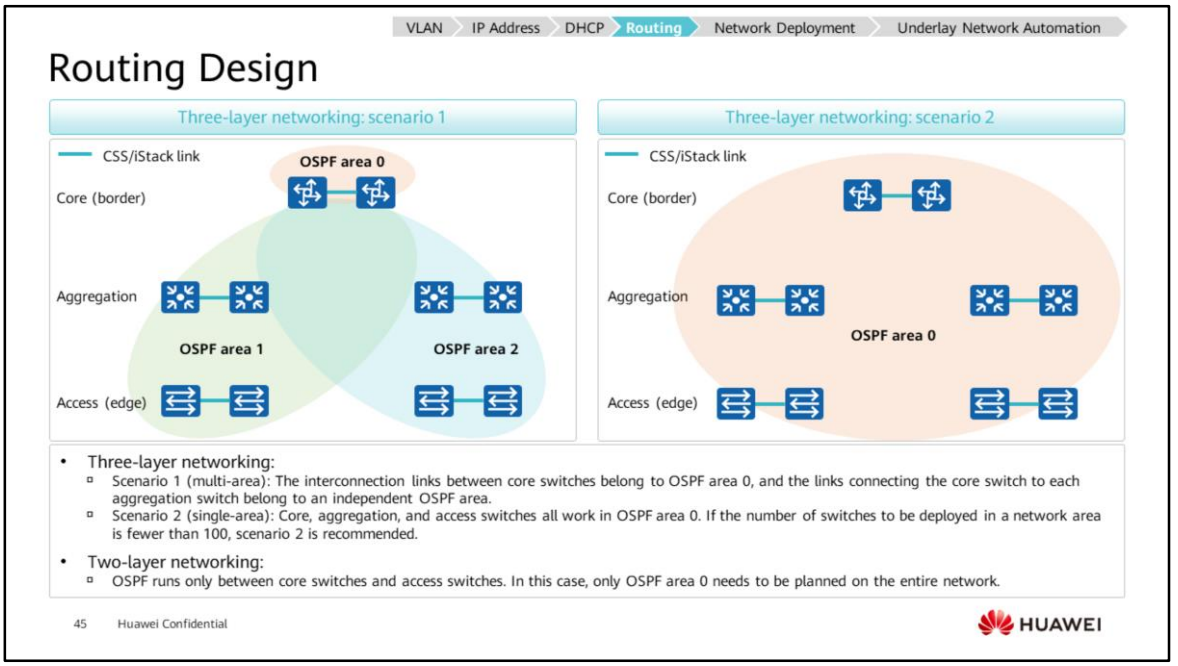
- IP address planning complies with the following principles:
  - Unique: Each host on an IP network must have a unique IP address.
  - Contiguous: Node addresses of the same service must be contiguous to facilitate route planning and summarization. Contiguous addresses facilitate route summarization, thereby reducing the size of the routing table and speeding up route calculation and convergence.
  - Scalable: Some IP addresses need to be reserved at each layer, so that no address segments or routing entries need to be added when the network is expanded.
  - Easy to maintain: Device and service address segments need to be clearly distinguished from each other, facilitating subsequent statistics monitoring and security protection based on address segments. IP addresses can be planned based on VLANs.

Terminal-Oriented DHCP Service Design

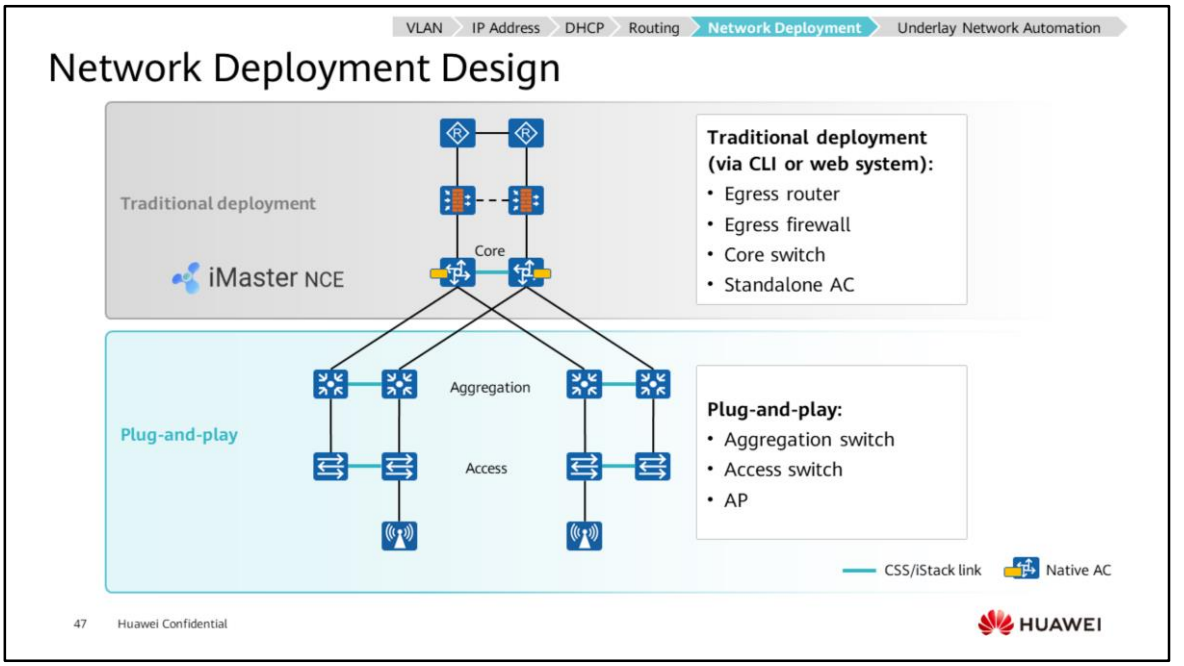


- Plan an independent DHCP server to allocate IP addresses to terminals.
- You are advised to configure DHCP snooping on access devices to defend against attacks.
- Network administrators can use either of the two mechanisms to assign IP addresses to hosts based on network requirements:
  - Dynamic IP address assignment: An IP address is assigned to a host with a lease. This mode applies to scenarios where hosts require temporary access or IP addresses are insufficient, for example, portable computers of traveling employees and mobile terminals in cafes.
  - Static IP address assignment: Fixed IP addresses are assigned to specified hosts or servers such as DNS servers.
- When planning an address pool, filter out static IP addresses.
- The lease needs to be planned based on the online duration of a client.
- On a large or midsize campus network, the DHCP server and hosts are usually not on the same network segment. Therefore, you need to enable the DHCP relay function on the gateway.

HUAWEI



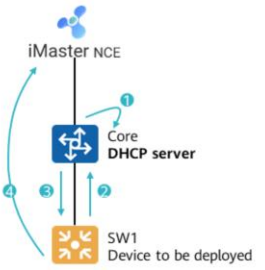
- The routing design includes internal and egress routing design for a campus network.
    - Internal routing design:
        - The routing design must support communication between devices, between terminals, and between devices and terminals on the campus network, as well as communication between these devices/terminals and the external network.
        - It is recommended that you design internal routes based on the gateway location.
            - If gateways are deployed at the aggregation layer, routes need to be deployed at core and aggregation layers. Routing tables can be dynamically updated along with network topology changes, so an Interior Gateway Protocol (IGP), such as Open Shortest Path First (OSPF), is recommended.
            - If gateways are deployed at the core layer, you only need to configure routes at the core layer. It is recommended that static routes be used preferentially.



- Typically, egress and core devices on large- and medium-sized campus networks are centrally deployed in a core equipment room. Services transmitted on these devices are complex and their locations on the network are important. In most cases, network engineers need to commission devices onsite during the deployment. Therefore, you are advised to use the web system or CLI to deploy devices at the core layer and upper layers (including core devices, standalone ACs connected in off-path mode, and egress devices).
- A large number of devices (including aggregation devices, access devices, and APs) are deployed at lower layers of the core layer, and service configurations of these devices are similar. Therefore, you are advised to use the DHCP option mode to achieve plug-and-play of devices, thereby simplifying the deployment.
- Note: Core switches obtain basic configurations such as IP addresses using the CLI. Once they establish management channels with iMaster NCE-Campus, iMaster NCE-Campus will automatically deliver services to them.

DHCP-based Plug-and-Play Deployment Process

- A device to be deployed obtains the NETCONF enabling status and iMaster NCE-Campus address from the DHCP server as follows:
  1. The administrator deploys the DHCP server function on the core device and configures DHCP Option 148, including the NETCONF enabling status of the device and the URL/IP address and port number of iMaster NCE-Campus.
  2. After the device to be deployed (SW1 in the figure) starts with empty configuration, it sends a request packet to the DHCP server through VLAN 1 (VLAN 1 is the PnP VLAN of a switch by default).
  3. Upon receiving the request, the DHCP server sends a DHCP packet containing Option 148 to SW1.
  4. SW1 registers with iMaster NCE-Campus and gets onboarded based on the information carried in Option 148 (NETCONF enabling status, URL/IP address and port number of iMaster NCE-Campus).

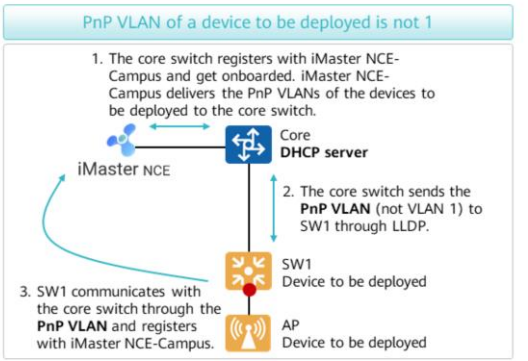


HUAWEI

VLAN > IP Address > DHCP > Routing > Network Deployment > Underlay Network Automation

PnP VLAN

- The Plug-and-Play VLAN (PnP VLAN) is defined for plug-and-play of switches. The default PnP VLAN of a switch is VLAN 1.
- PnP VLANs consist of wired and wireless PnP VLANs, which are uniformly maintained by iMaster NCE-Campus. After a core switch registers with iMaster NCE-Campus, iMaster NCE-Campus automatically delivers the PnP VLANs preconfigured on it to the core switch.
  - The wired PnP VLAN is used to apply for the management IP address of a switch.
  - The wireless PnP VLAN is used to configure the management VLAN of an AP. When a switch has an AP connected, the switch automatically changes the PVID of the interface connected to the AP to the wireless PnP VLAN ID.
  - For a switch, wired and wireless PnP VLANs can be different, but they are negotiated at the same time. If only a wired PnP VLAN is configured, the PVID of the switch interface connected to an AP is changed to the wired PnP VLAN ID.

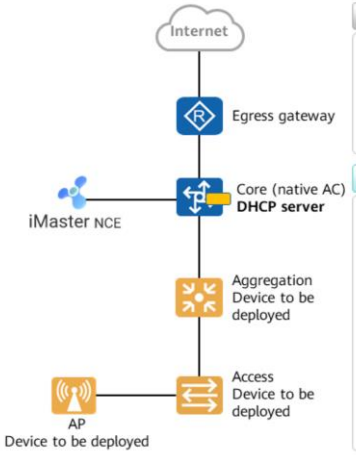


HUAWEI

- The DHCP server pushes PnP VLAN information to its downstream devices through LLDP. Note that:
  - If NETCONF is not enabled on the core switch (DHCP server), the core switch cannot be onboarded on iMaster NCE-Campus. In this case, the administrator needs to manually configure the PnP VLAN on the core switch. Then the switch to be deployed can negotiate with the core switch through LLDP to obtain the configured PnP VLAN.
- Note: The PnP VLAN and management VLAN of a switch can be the same or different.
- LLDP: Link Layer Discovery Protocol

VLAN > IP Address > DHCP > Routing > Network Deployment > Underlay Network Automation

Network Deployment Process: Device Onboarding Before Planning



Preparations

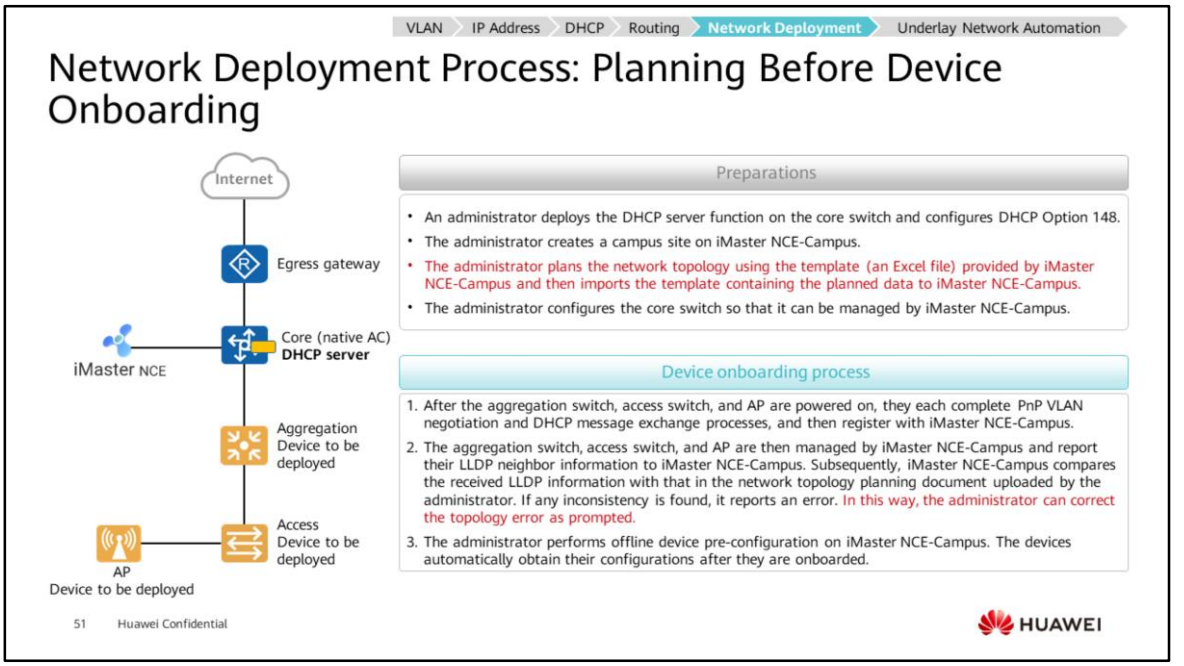
- An administrator deploys the DHCP server function on the core switch and configures DHCP Option 148.
- The administrator creates a campus site on iMaster NCE-Campus.
- The administrator imports switch information (ESN, device model, etc.) to iMaster NCE-Campus. (Batch import is supported.)
- The administrator configures the core switch so that it can be managed by iMaster NCE-Campus.

Device onboarding process

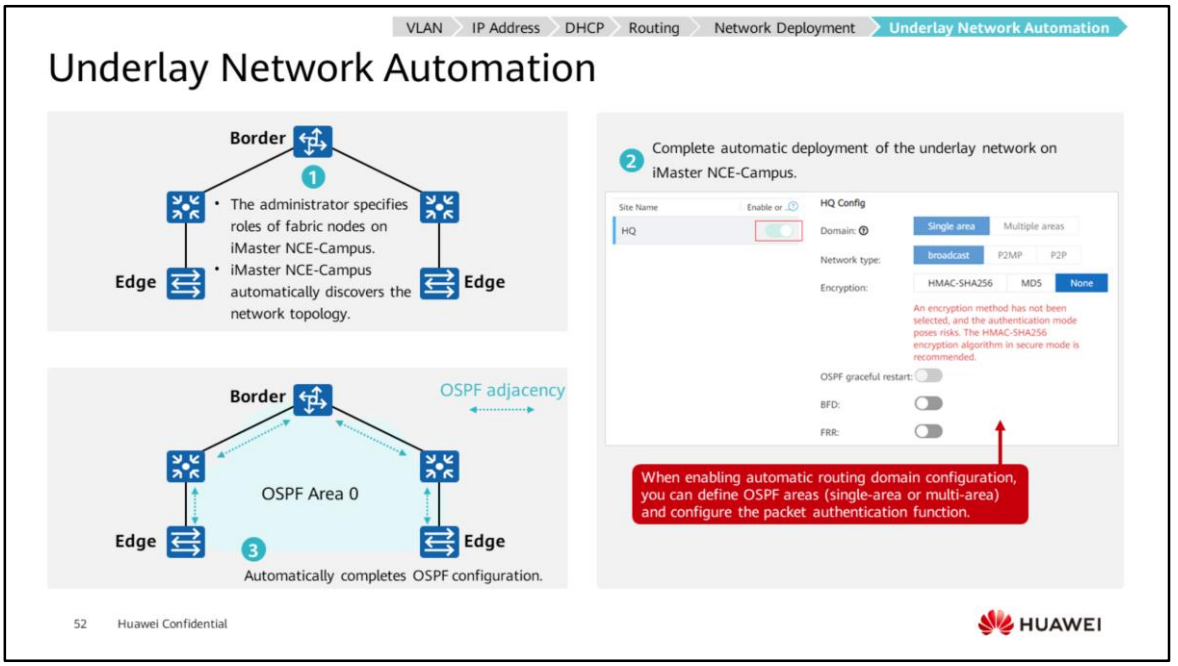
1. After an aggregation switch is powered on, it obtains an IP address and Option 148 from the DHCP server using the PnP VLAN ID (for example VLAN 1). The aggregation switch registers with iMaster NCE-Campus based on the iMaster NCE-Campus information obtained from the DHCP server.
2. After an access switch is powered on, it follows a similar process as the aggregation switch (step 1).
3. The access switch identifies that its downstream device is an AP and changes the PVID of the interface connected to the AP to the wireless PnP VLAN ID. The AP obtains the AC address through DHCP Option 43. After the AP is associated with the AC and VLANIF 1 is configured as the CAPWAP source interface, the AP gets onboarded on the AC.
4. Switches use LLDP to discover the network topology and report their topology information to iMaster NCE-Campus through NETCONF. iMaster NCE-Campus then discovers the network topology based on the received topology information.
5. The administrator performs network planning and configuration provisioning on iMaster NCE-Campus.

HUAWEI

- This deployment process applies to scenarios with scattered installation time.



- This deployment process is applicable to scenarios with concentrated installation time. The administrator is advised to plan the network before onboarding devices. If the network cannot be planned in advance, the administrator can onboard devices first and then determine the network topology.
- Network planning before device deployment and onboarding.
    - During network deployment, the administrator plans the network topology by entering device ESNs and specifying stack members and aggregated links on iMaster NCE-Campus.
    - Alternatively, the administrator can import the preceding planning information in batches using a template. Using a template to import data in batches simplifies operations and is therefore recommended.
    - Then, the administrator uses the recommended method to deploy and onboard devices.
    - After devices register with iMaster NCE-Campus, iMaster NCE-Campus automatically checks whether the actual topology of the devices is the same as the planned one. If cables are incorrectly connected during installation, iMaster NCE-Campus immediately notifies the administrator.
- Note: Parameters can be planned for an Eth-Trunk interface. After a device is onboarded, it automatically obtains the corresponding Eth-Trunk configuration.

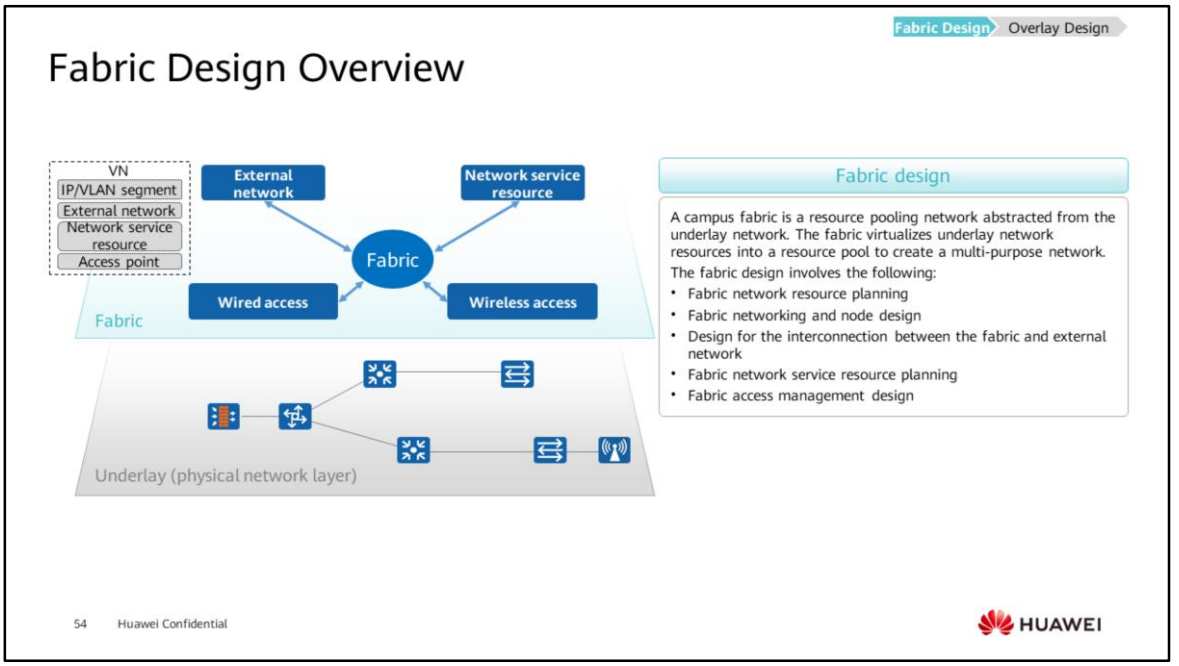


- Automatic routing domain configuration: After this function is enabled, the underlay network can be automatically configured. You can specify sites for automatic routing domain configuration and specify OSPF routing parameters. Currently, the following parameters are supported:
  - Area: In the single-area scenario, all devices belong to Area 0. In the multi-area scenario, border nodes belong to Area 0, and each edge node and its connected border node belong to an area.
  - Network type: You can set the OSPF network type to broadcast, P2MP, or P2P.
  - Encryption: You can set the encryption mode between adjacent devices to HMAC-SHA256, MD5, or none.
  - OSPF GR: You can enable OSPF GR.
- Before enabling the automatic routing domain configuration function, you need to plan network resources required for underlay network automation.
  - Underlay devices (corresponding network scope of the fabric) are connected through VLANIF interfaces at Layer 3, and you need to assign a VLAN to each interconnection link.
  - The VLANIF interfaces for connecting devices are automatically assigned interconnection IP addresses with a 30-bit subnet mask.

Contents

1. CloudCampus Solution and Virtualized Campus Network Overview
2. Network Architecture Design
3. Underlay Network Design
4. **Fabric and Overlay Network Design**
5. Admission Control and Free Mobility Design
6. WLAN Design
7. Egress Network Design
8. Network Security and QoS Design
9. O&M Design

HUAWEI



- **Fabric network resource planning:**
- Before creating virtual networks (VNs), you need to configure global resources, including the VLANs, VXLAN Network Identifiers (VNIs), and bridge domains (BDs). When creating a VN, iMaster NCE-Campus automatically allocates resources from the resource pools.
    - Interconnection VLAN: When creating the external network resource for a fabric, configure an interconnection VLAN to interconnect the fabric with the egress network. When creating the network service resource on a fabric, configure an interconnection VLAN to interconnect the fabric with the network management zone.
    - BD: BDs are used to create Layer 2 broadcast domains in a VN. Typically, BDs have a one-to-one mapping relationship with service VLANs. Therefore, a sufficient number of BD resources need to be planned to match the number of service VLANs. The default BD range is 1-4095.
    - VNI: VNIs are similar to VLAN IDs. They are used to identify VXLAN segments and range from 1 to 4095 by default.

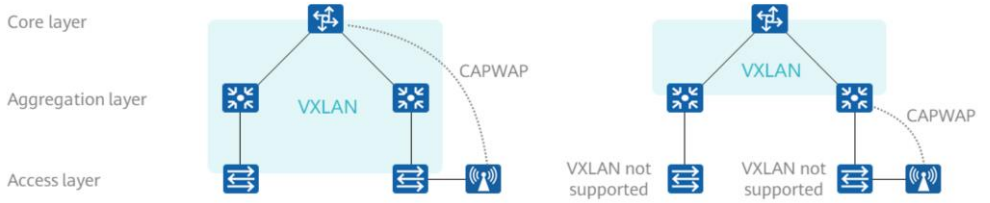
Fabric Networking Scenarios

- In Layer 2 or Layer 3 networking, border or edge nodes can function as gateways.
  - Centralized gateway: A border node functions as the gateway to implement centralized management and simplify O&M.
  - Distributed gateway: Edge nodes function as gateways to facilitate network expansion.
- Recommended networking scenarios:
  - Centralized gateway, VXLAN deployed across core and aggregation layers, native AC deployed on the core device, or standalone AC attached to the core device in off-path mode.
  - Distributed gateways, VXLAN deployed across core and aggregation layers, native AC deployed on the core device, or standalone AC attached to the core device in off-path mode.

Networking	Border Node Location	Edge Node Location	AC Deployment Mode	Recommended
Centralized gateway	Core	Access	Native AC on the core device	No
			Standalone AC attached to the core device in off-path mode	No
		Aggregation	Native AC on the core device	Yes
			Standalone AC attached to the core device in off-path mode	Yes
Distributed gateway	Core	Aggregation	Native AC on the core device	Yes
			Standalone AC attached to the core device in off-path mode	Yes
		Access	Standalone AC attached to the core device in off-path mode	No

HUAWEI

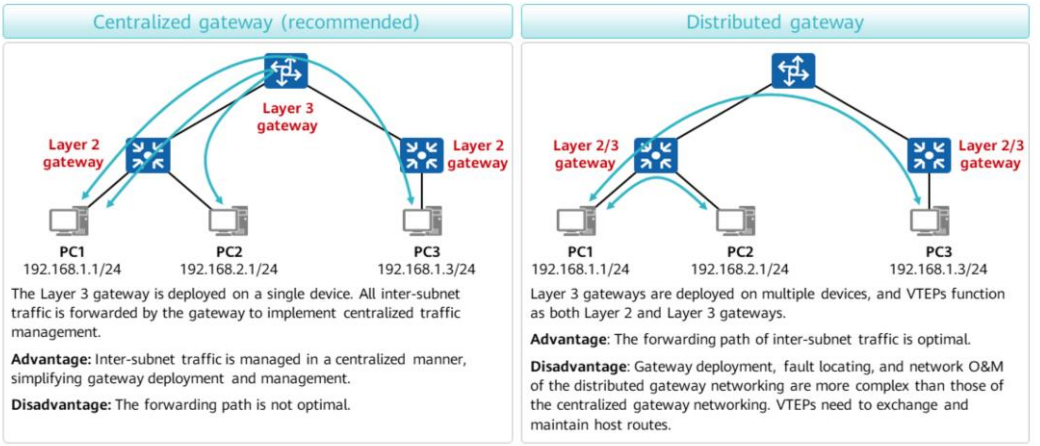
Fabric Networking Design: VXLAN Coverage Range



VXLAN range	VXLAN across core and access layers	VXLAN across core and aggregation layers
Application scenarios	• Automatic end-to-end service deployment is required on the entire network. • Recommended: Number of switch nodes < 1,000	• Existing access switches on the live network need to be reused. Alternatively, low-cost Huawei switches that do not support VXLAN need to be deployed at the access layer to reduce costs. • Recommended: Number of switch nodes < 3,000
Network construction costs	Compared with the VXLAN solution from the core layer to the aggregation layer, this solution has higher network construction costs.	Compared with the VXLAN solution from the core layer to the access layer, this solution has lower network construction costs.

HUAWEI

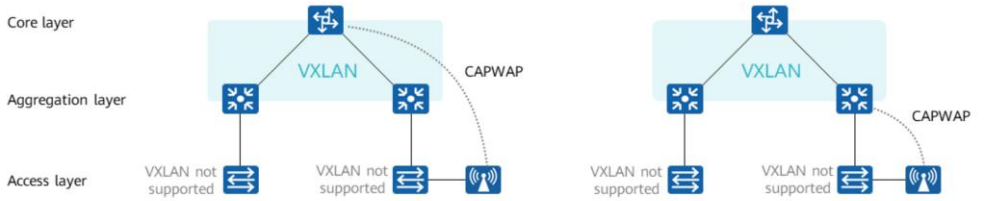
Fabric Networking Design: Centralized and Distributed Gateways



HUAWEI

- In large- and medium-sized campus networks, the virtualization solution is classified into the centralized gateway solution and distributed gateway solution based on the user gateway location. You can select a gateway solution when creating a fabric on iMaster NCE-Campus.
- In the centralized gateway solution, a border node functions as the gateway of all users, and all inter-subnet traffic is forwarded by the border node. In the distributed gateway solution, multiple edge nodes function as user gateways, and inter-subnet traffic is forwarded through these edge nodes.

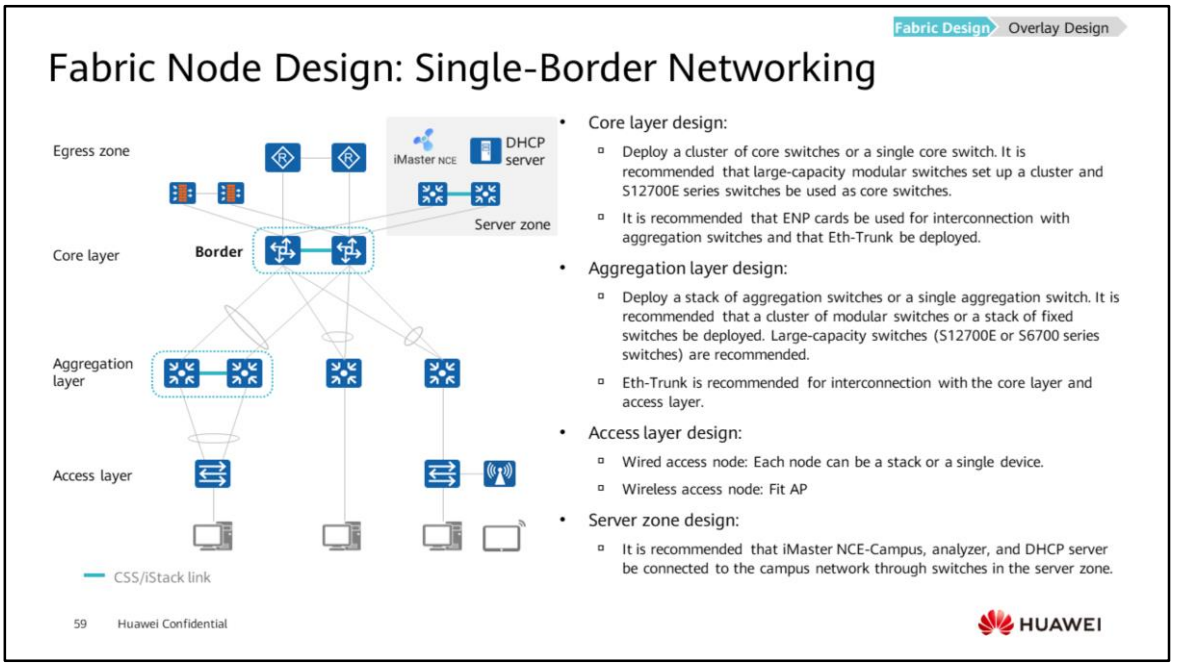
Fabric Networking Design: Gateway Solution Selection



Item	Centralized Gateway	Distributed Gateway
User gateway location	Border	Edge
Recommended edge location	Aggregation switch	Aggregation switch
O&M deployment	The border node functions as the gateway of all users, and the native AC function is typically enabled on the border node to support wireless services.	Multiple edge nodes function as user gateways. This ensures that only part of the network is affected upon the failure of one gateway and facilitates network expansion. The native AC function is enabled on the border node to support wireless services.
Terminal scale	≤ 50000 (calculated based on the number and specifications of centralized gateways). This solution is recommended when the number of terminals does not exceed 50000.	≤ 100000 (calculated based on the number and specifications of distributed gateways). This solution is recommended when the number of terminals exceeds 50000.

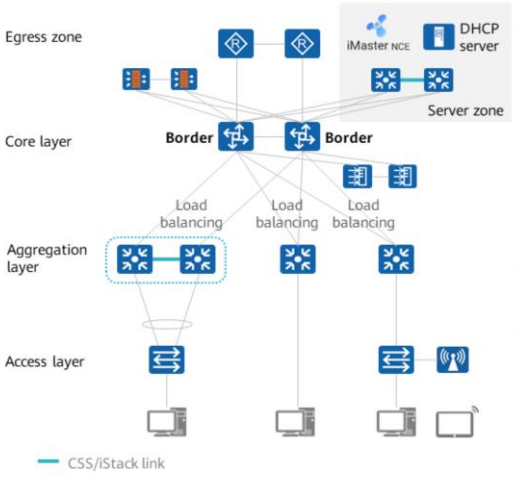
HUAWEI

- When designing the CloudCampus network virtualization solution, first determine the gateway solution to be used. Then you can perform end-to-end design for the entire campus network based on the selected gateway solution.
- The centralized gateway solution supports only one border node, whereas the distributed gateway solution supports multiple border nodes.



- An Ethernet Network Processor (ENP) card is embedded with the Huawei-developed ENP. The card can function as a common LPU to provide data access and switching services and also as a WLAN AC to provide wireless access control functions. In this way, the card achieves wired and wireless convergence.

Fabric Node Design: Dual-Border Networking



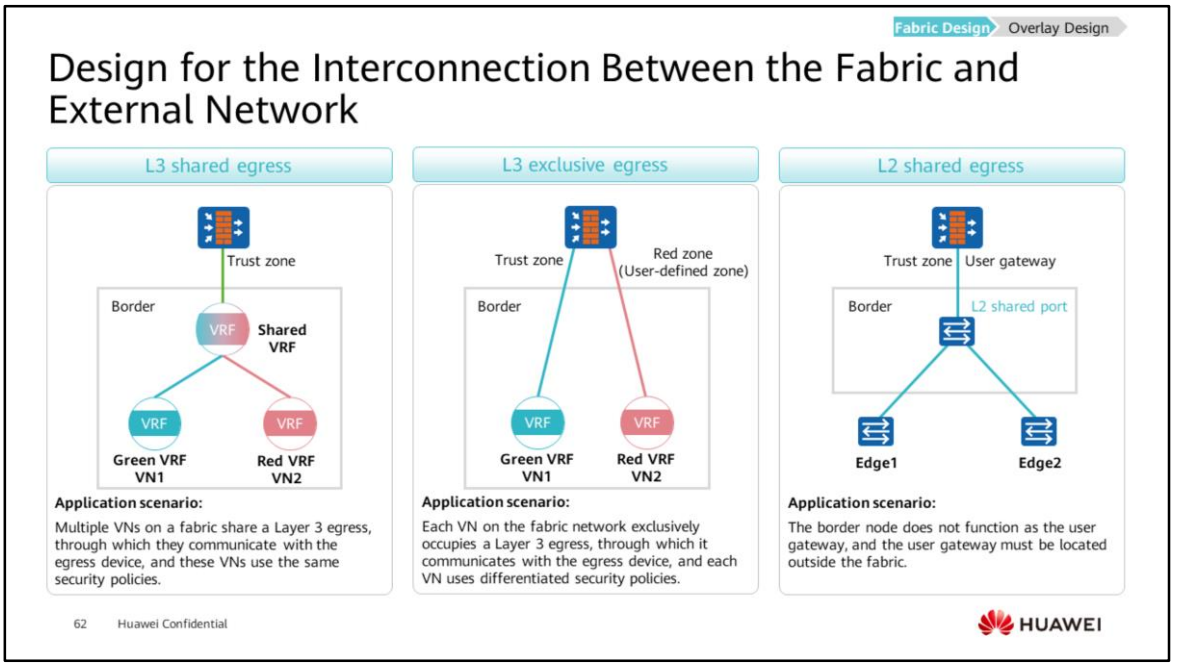
- Core layer design:
  - Core switches are two logical devices. Large-capacity switches (S12700E series switches) are recommended. It is recommended that ENP cards be used for interconnection with aggregation switches and that Eth-Trunk be deployed.
- Aggregation layer design:
  - It is recommended that aggregation switches set up a stack to implement load balancing with core switches.
- Access layer design:
  - The design is similar to that in the single-border networking scenario.
- Server zone design:
  - It is recommended that iMaster NCE-Campus, analyzer, and DHCP server be connected to the campus network through switches in the server zone.

HUAWEI

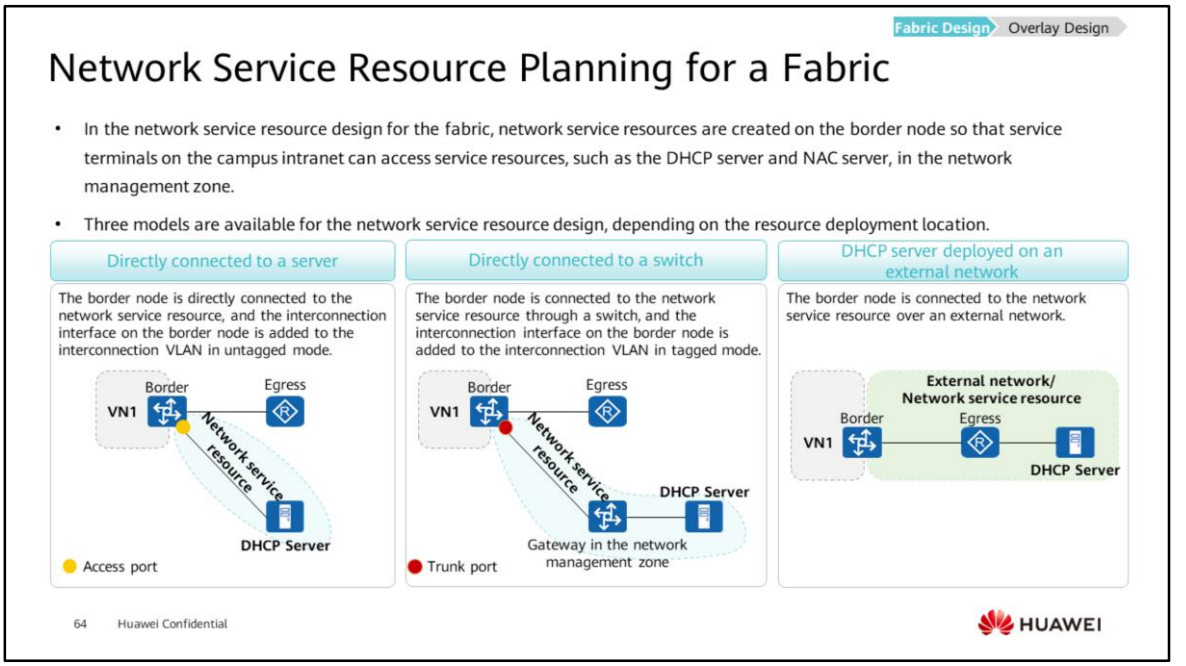
Fabric Node Design: Single/Dual-Border Networking Selection

	Single-Border Networking	Dual-Border Networking
Description	The core switch is a logical device, which can be a physical standalone device or a CSS/iStack system.	Core switches are two logical devices, which can be two standalone physical devices or two CSS/iStack systems.
Characteristics	• Easy management, configuration, and maintenance. • Services may be affected during the device upgrade.	• Complex management, configuration, and maintenance. • The upgrade of core switches has little impact on services.
Application scenario	Simple network management, configuration, and maintenance are required.	• The device upgrade has little impact on services. • Core switches do not support clustering or stacking due to some reasons (such as the deployment distance).
Restrictions	There is no restriction. The single-border networking design is applicable to both centralized and distributed VXLAN gateway networking.	The dual-border networking design is applicable only to distributed VXLAN gateway networking.

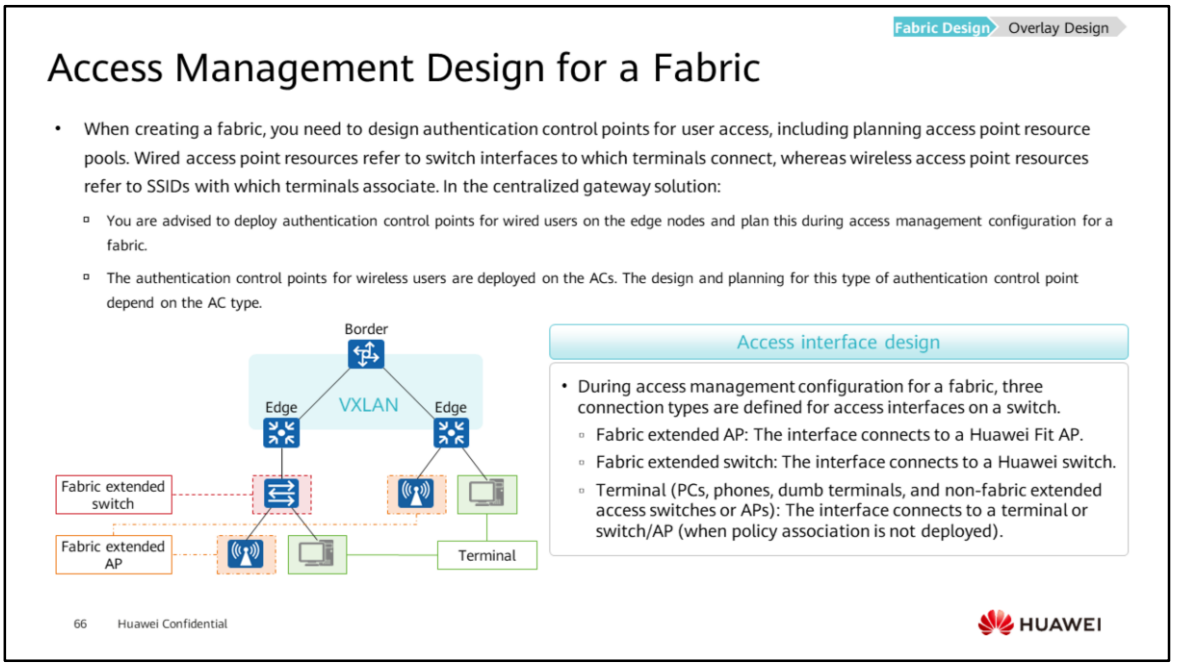
HUAWEI



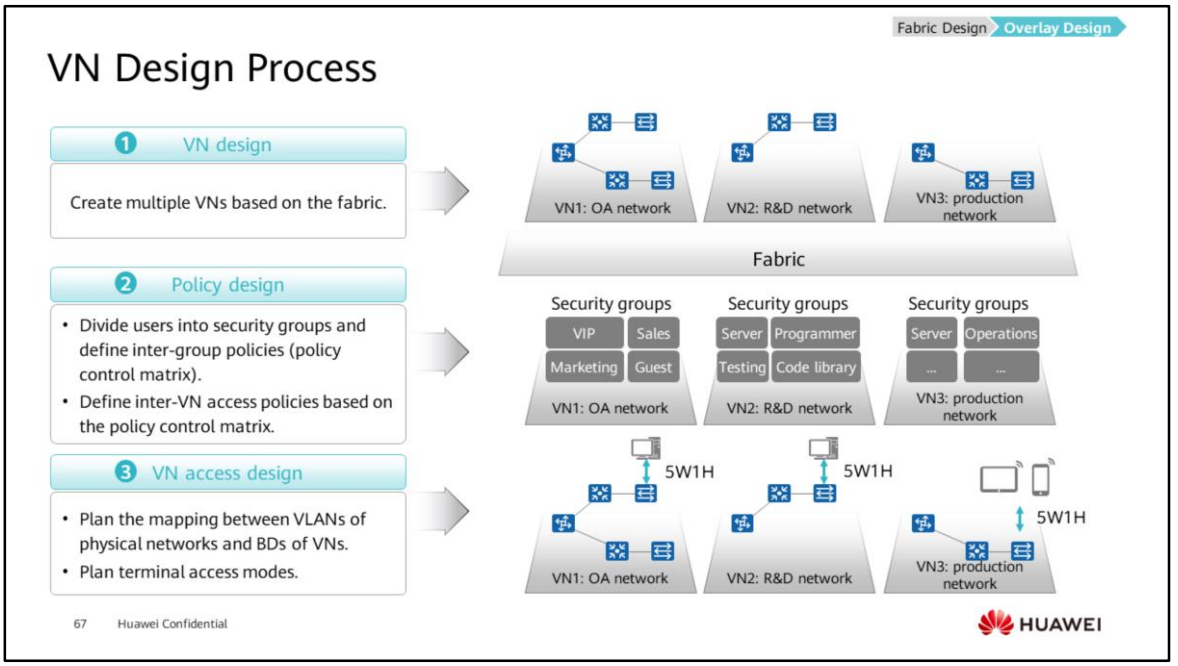
- When the campus intranet needs to communicate with an external network, for example, the Internet, data center, or another branch, traffic must pass through the border node.
- There are three interconnection modes between the fabric network and egress device:
    - L3 shared egress:
        - The external gateway connects to and accesses external networks via VLANIF or VBDIF interfaces. VNs can access the public network or private network specified by another site through the shared VRF egress, and service traffic can be diverted to the firewall through the shared VRF egress. When configuring a multi-border fabric, you can configure multiple core devices in one external network.
        - The L3 shared egress mode is applicable to the scenario where the firewall does not need to perform security check on VNs, there are low requirements on security control policies between VNs, and traffic of all VNs is transmitted in the same security zone.
        - To enable communication between VNs and external networks, you must configure return routes to service subnets on the firewall. As a result, service subnets of different VNs can communicate with each other on the firewall. To isolate VNs on the firewall, configure policies based on service network segments in the VNs.
        - As shown in the figure, a shared VRF is created on the border node, the shared L3 egress is bound to the VRF, and routes are configured to enable the communication with external networks.



- Multiple network service resources can be created, or a network service resource can have access addresses for multiple network service resources.
- When the border node is directly connected to a server:
  - As shown in the figure, a VRF on the border node is allocated to each network service resource. After a network service resource is selected during VN creation, a static route pointing to the network service resource address will be created on the border node, with the next hop being the network service resource address.
- When the border node is directly connected to a switch:
  - If the DHCP server, iMaster NCE-Campus, and other network service resources are all deployed in the network management zone, the border node communicates with these network service resources through the directly connected switch (gateway) in the network management zone. If only a small number of network service resources are deployed, it is recommended that these resources be planned in the same network service resource model. This saves interconnection VLAN and IP address resources and simplifies route configuration in the network management zone.
  - As shown in the figure, a VRF on the border node is allocated to each network service resource. After a network service resource is selected during VN creation, a static route pointing to the network service resource address will be created on the border node, with the next hop being the address of the gateway in the network management zone.

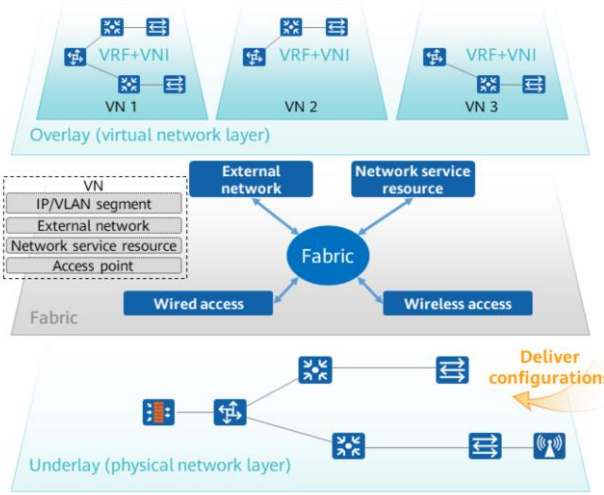


- "Fabric extended AP" and "Fabric extended switch": The two types of connections are used to enable communication between the authentication control point and authentication enforcement point through the policy association management VLAN. In this scenario, the fabric extended switch functions as the authentication enforcement point and can be connected to fabric extended APs and terminals.
- Terminal (PCs, phones, dumb terminals, and non-fabric extended access switches/APs): If this type of interface is connected to a terminal, you can associate an authentication profile specific to that terminal type with the terminal based on the user authentication mode design in the access control design. Binding the authentication profile facilitates access control of the terminal.



- VN design:
  - VNs are generally divided based on services on a campus network. An independent service is assigned a VN, and VNs are isolated from each other by default. For example, on a school campus network, guest, teaching, IoT, and video surveillance services can each be assigned a separate VN. On an enterprise campus network, office network services, production network services, and R&D services can be allocated to different VNs.
- Policy design:
  - Some isolation requirements caused by different user roles can be fulfilled by deploying policy control technologies (such as free mobility).
  - VNs are isolated by default. To enable mutual access between user groups within a VN, additional VN interworking configuration is required.
- VN access design:
  - Service data enters from a physical network to a VN through an edge node. Service data of users enters different VNs depending on the VLANs to which users belong. Therefore, during network design, you need to plan the mappings between VLANs of physical networks and BDs of VNs, and configure VLANs for wired and wireless users.

VN Design



1. Network service abstraction

Physical network resources are pooled through orchestration, and the network is abstracted into Fabric as a Service (FaaS). A VN is a FaaS instance and includes:

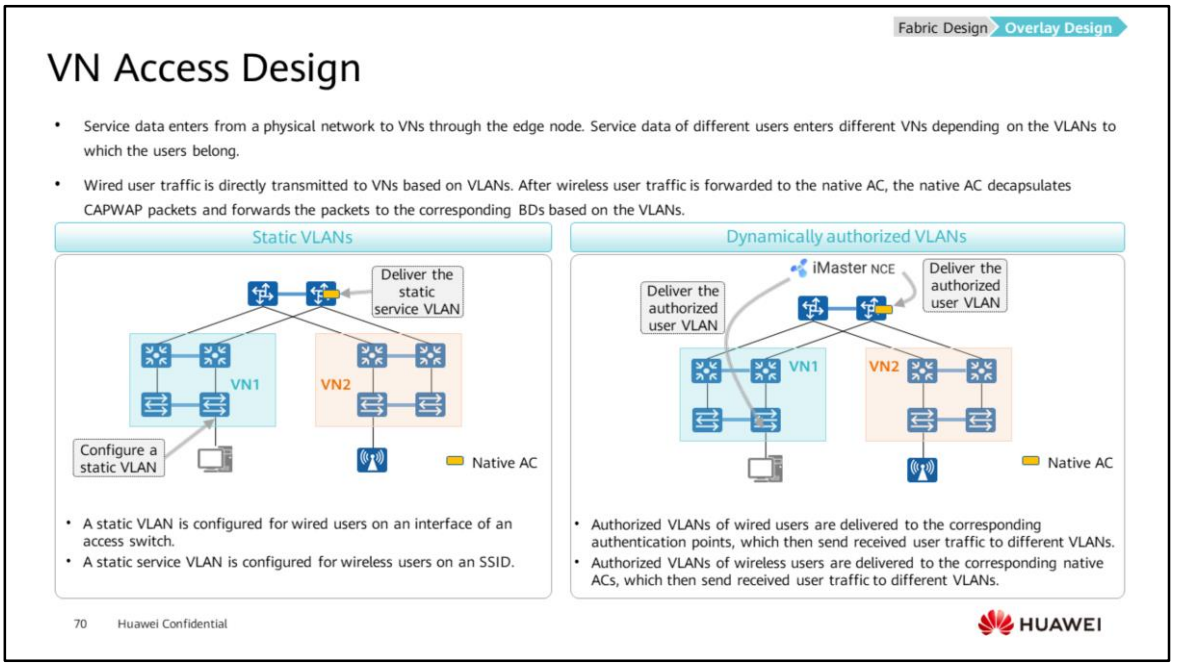
- IP/VLAN segment
- External network
- Network service resource: The IP/VLAN segment is the capability provided by the VN for clients to use network resources.
- Access point: Terminals access VNs through access points.

Deploy VNs

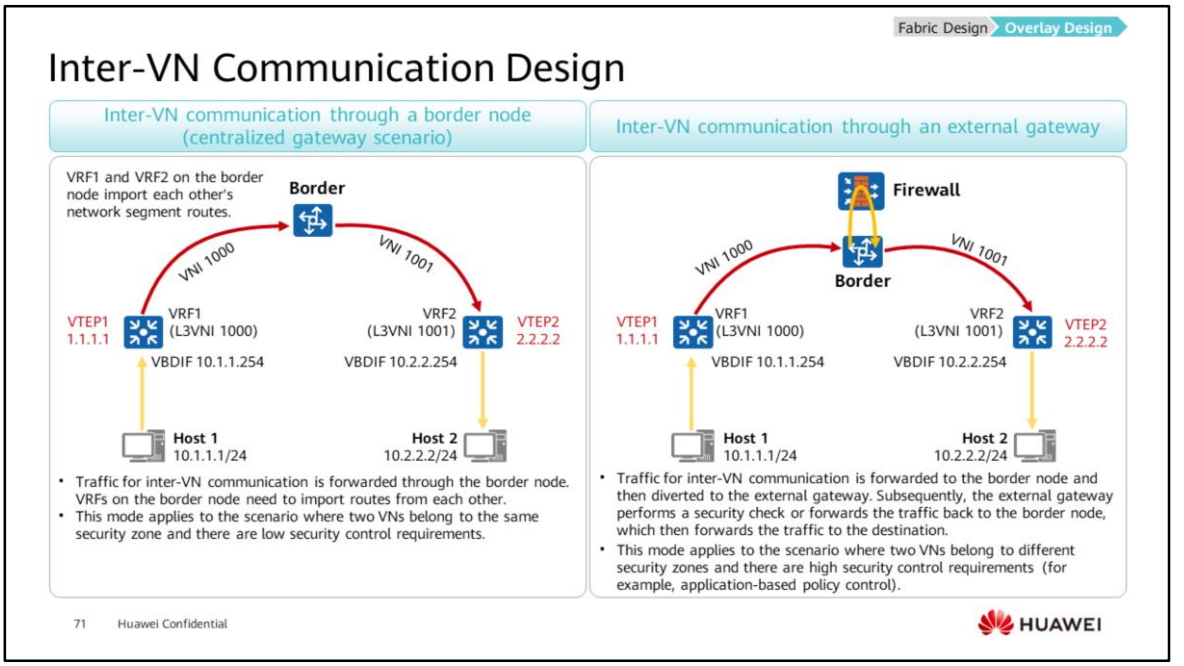
2. Network service orchestration

- Deliver the mappings between VNIs and BDs.
- Deliver the mappings between BDs and VLANs.
- Deliver the IP address segments corresponding to VBDIF interfaces.
- Deliver VRFs and bind them to VBDIF interfaces.

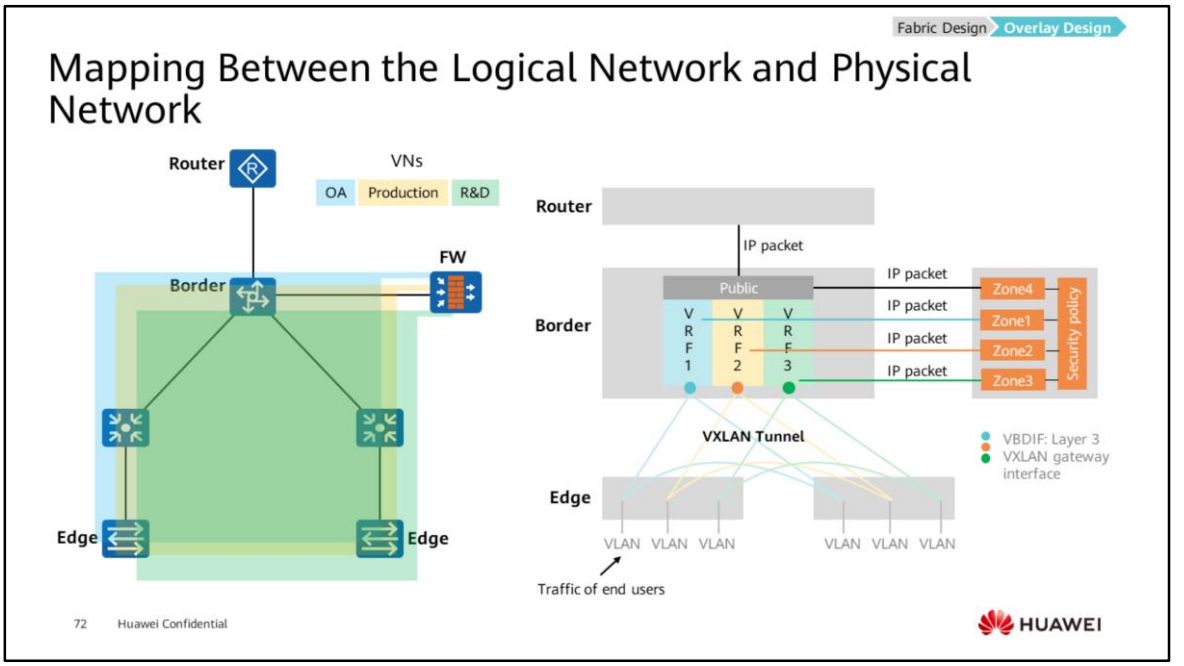
HUAWEI



- Application scenarios:
  - The static VLAN mode applies when terminals access the network at fixed locations and do not need to be authenticated. This access mode is more secure but lacks flexibility. When the locations of terminals change, you need to configure static VLANs for them again.
  - The dynamically authorized VLAN mode applies when terminals need to access the network from any place and need to be authenticated based on the VLAN information delivered during user authentication. This access mode is flexible and the configuration does not need to be modified when the locations of terminals change. Dynamic access is more automated, easy to manage and use, and is therefore recommended.
- VN access design for wireless users:
  - If distributed gateways are used and the border nodes have the native AC function deployed, traffic of wireless users is forwarded to the native AC through CAPWAP tunnels. After the native AC decapsulates CAPWAP packets, the decapsulated packets enter different VNs depending on the VLANs to which the wireless users belong.
  - If centralized gateways are used and the border nodes have the native AC function deployed, it is recommended that traffic of wireless users be directly forwarded to the native AC through CAPWAP tunnels. After the native AC decapsulates CAPWAP packets, the decapsulated packets enter different VNs depending on the VLANs to which the wireless users belong. The administrator needs to configure different VLAN ranges for wired and wireless terminals. The VLANs of wired terminals are bound to BDs on the edge nodes, whereas the VLANs of wireless terminals are bound to BDs on the border nodes.
- Note: 802.1X authentication and MAC address authentication both support dynamically authorized VLANs.



- Inter-VN communication can be implemented through a border node or an external gateway.
  - Through a border node
    - If two VNs belong to the same security zone and have low security control requirements, devices on the two VNs can directly communicate with each other through a border node. In addition, permission control can be implemented based on the free mobility policy. To implement communication between VNs, the border node needs to import their respective network segment routes that are reachable to each other.
  - Through an external gateway
    - If two VNs belong to different security zones and have high security control requirements, it is recommended that devices on the two VNs communicate through an external gateway (a firewall) and that a security zone policy be configured on the firewall for permission control.



- Each VN corresponds to a VRF.
- Each VN corresponds to multiple subnets, and each subnet corresponds to a BD and VNI.
- Traffic from a terminal enters the VN based on the VLAN to which the terminal belongs.

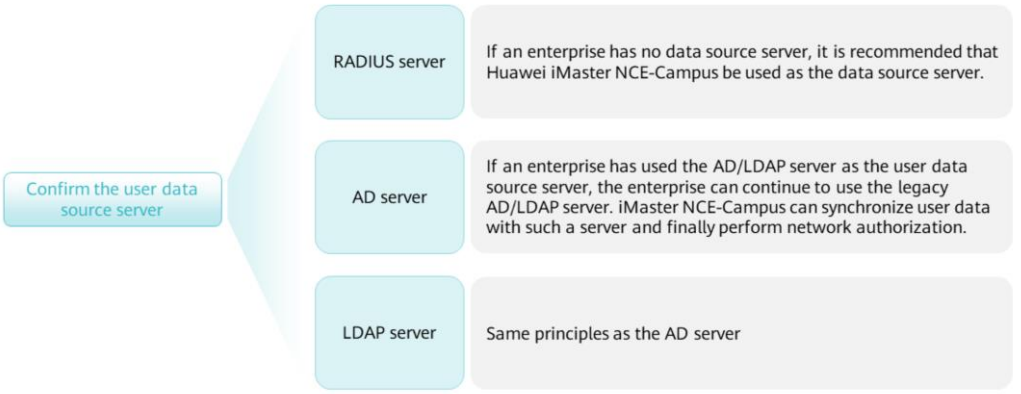
Contents

1. CloudCampus Solution and Virtualized Campus Network Overview
2. Network Architecture Design
3. Underlay Network Design
4. Fabric and Overlay Network Design
5. **Admission Control and Free Mobility Design**
6. WLAN Design
7. Egress Network Design
8. Network Security and QoS Design
9. O&M Design

HUAWEI

User Management Solution Design

- User management solution design involves confirming the user data source server. Typical user data source servers of enterprises include RADIUS, AD, and LDAP servers.



HUAWEI

User Management > User Authentication > Policy Control > Terminal Identification

User Authentication Technology Selection

- Common authentication technologies include 802.1X, MAC address, and Portal authentication. The following table compares these authentication methods.

Item	802.1X Authentication	MAC Address Authentication	Portal Authentication
Client required or not	Required	Not required	Not required
Advantage	High security	No need to install a client	Flexible deployment
Disadvantage	Inflexible deployment	MAC addresses need to be registered, complicating management	Low security
Application scenario	Network authentication of office users with high security requirements	Access authentication of dumb terminals such as printers and fax machines	Network authentication of guests who move frequently and use different types of terminals

- On large- and medium-sized campus networks, 802.1X authentication is recommended for employees, Portal authentication for guests, and MAC address authentication for dumb terminals.
- If a customer wants to use more than one authentication method on the same access point, a combination of authentication methods (hybrid authentication) can be used. After hybrid authentication is configured, terminals can access the network after passing any authentication in the combination. This mode is applicable to scenarios where one port provides access for multiple types of users. For example, if a PC is connected upstream to an IP phone, you can configure hybrid authentication (MAC address authentication + 802.1X authentication). In this way, the IP phone uses MAC address authentication, and the PC uses 802.1X authentication.

HUAWEI

- If 802.1X or MAC address authentication (Layer 2 authentication technologies) is used, the authentication point must be on the same network segment as the user host. It is recommended that the access device function as the authentication point.
    - If a large number of access devices are deployed, the aggregation or core device needs to be deployed as the authentication point. You can also deploy policy association. The gateway functions as the authentication control point, and the access device functions as the authentication enforcement point, thereby simplifying policy deployment.
- If Portal authentication is used, the authentication point can be deployed anywhere as long as it is routable to the user host. Layer 2 Portal authentication is recommended.

Association Between Authenticated Users and VNs

1. Create VNs on iMaster NCE-Campus.

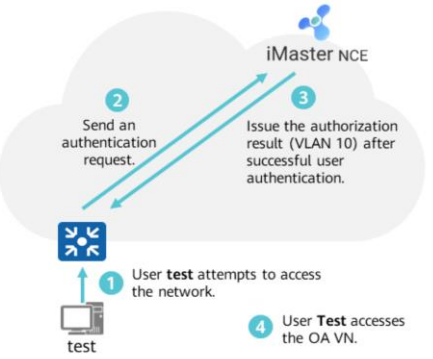
The administrator creates VNs and specifies the IP network segments and VLANs for the VNs.

OA VN
VLAN 10: 10.1.10.0/24
VLAN 20: 10.1.20.0/24

R&D VN
VLAN 30: 10.1.30.0/24
VLAN 40: 10.1.40.0/24

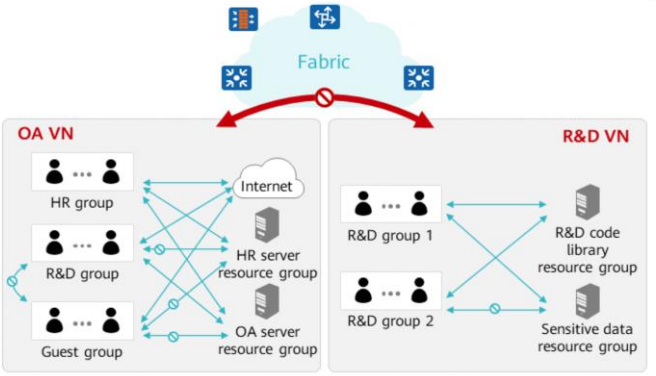
The administrator configures user authentication rules, authorization results, and authorization rules, so that users can obtain corresponding VLANs after being authenticated. For example, the administrator configures user **test** to obtain the authorized VLAN 10 after passing 802.1X authentication.

2. Authenticated users obtain authorized VLANs and are then associated with VNs based on the VLANs.



HUAWEI

Multi-Level Design for Access Policies



- Network access policies can be logically divided into two levels:
  - VN-level access policy: By default, VNs cannot communicate with each other, and service data of each VN is isolated. The CloudCampus Solution enables VNs to communicate within a fabric or through an external network.
  - Security group-level access policy: Users and network resources are divided into different security groups for efficient management and control of inter-group traffic. The free mobility policy enforcement point executes inter-group access policies.

HUAWEI

Policy Design: Create Security Groups

- Dynamic security groups: Terminals and users accessing the network are divided into different security groups through NAC authorization.
  - Define security groups for enterprise employees based on the enterprise organizational structure.

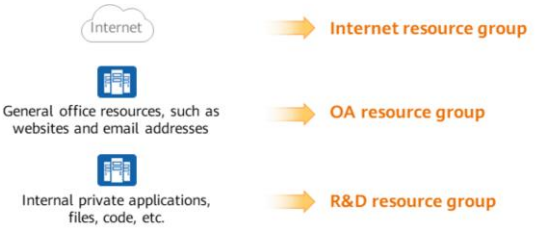
  - Define a security group for a special group of people in the organizational structure, such as the VIP group.
  - Define a security group for non-employees that access the network, such as the guest group.
  - Define a security group for dumb terminals connected to the network, such as printers and access control devices.

Dynamic security groups are authorized based on the 5W1H principle.

HUAWEI

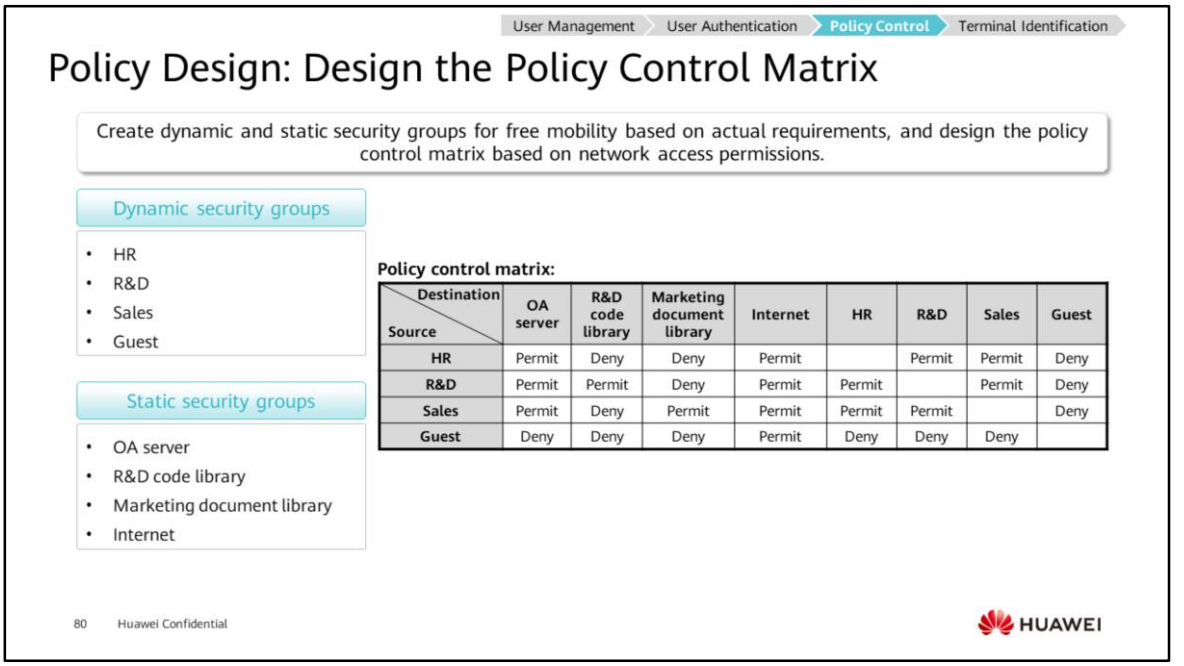
Policy Design: Create Resource Groups

- Static resource groups: Configure a group of IP address resources as a static resource group.
  - Define security groups, such as the OA resource group and R&D resource group, for intranet resources based on security requirements.
  - If Internet access control is required, define an Internet resource group.

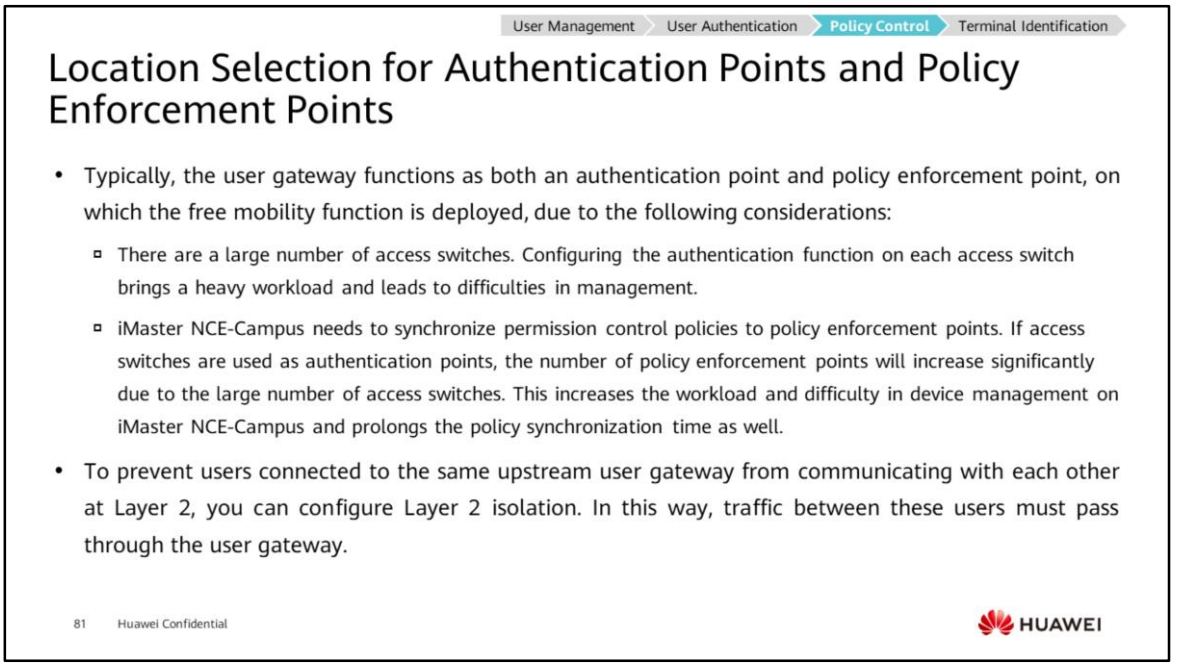


Aggregate the IP addresses of dispersed data center applications into a resource group, thereby simplifying deployment.

HUAWEI



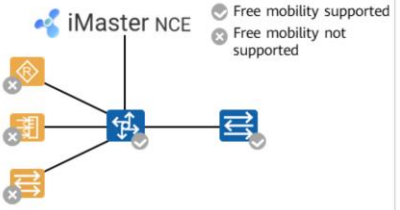
- An inter-group policy directly reflects whether two security groups can communicate with each other. When planning an inter-group policy, simply set the inter-group policy to permit or deny based on whether the two groups can communicate with each other. The administrator can configure inter-group policies through an intuitive policy control matrix on iMaster NCE-Campus. If no inter-group policy is created between source and destination security groups, the default policy between them is permit.
- Take the policy direction into account when planning inter-group policies. Typically, packets are transmitted in two directions between two terminals.
    - Huawei switches consider that traffic from A to B is unrelated to traffic from B to A, so they match policies for the two types of traffic separately and determine whether to forward the traffic based on the corresponding policies. This means that a Huawei switch enforces policies on a packet by considering only the source and destination security groups of that packet. For example, the policy "A ->B permit, B -> A deny" means that all packets sent from A to B will be permitted, whereas all packets sent from B to A will be discarded, regardless of whether A or B initiates the access. The default inter-group policy on a switch is permit.



- If an authentication point and policy enforcement point are deployed on different devices, the IP-security group entries of authenticated users need to be pushed to the specified policy enforcement point. To achieve this, the administrator needs to configure IP-security group entry subscription, that is, specify to which policy enforcement point the IP-security group entries of a certain network segment or security group need to be pushed.

IP-Security Group Entry Synchronization

Technical background

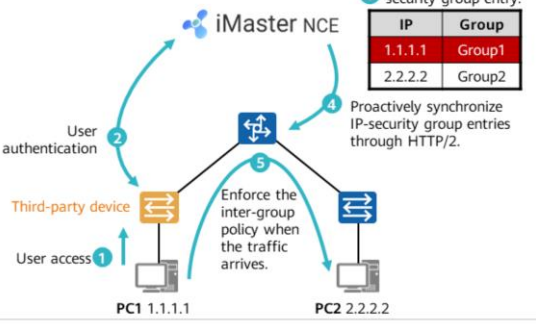


These devices do not support free mobility, so how to realize free mobility if a solution includes these devices?

- Switches that do not support free mobility
- ACs
- Routers
- Third-party devices (non-Huawei)

IP-security group entry synchronization principle

iMaster NCE-Campus synchronizes the mappings between user IP addresses and security groups to the switches functioning as policy enforcement points. In this way, authentication points and policy enforcement points can be separated, thereby implementing flexible networking. In addition, hybrid networking with third-party devices can be easily achieved.



HUAWEI

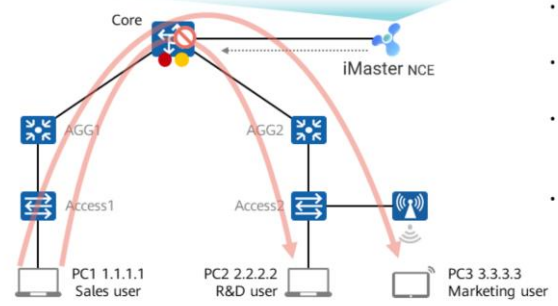
Free Mobility Solution (1)

Security group

Group Name	Group ID
Sales	1
R&D	2
Marketing	3
...	...

Security group-based policy control matrix

	Sales	R&D	Marketing	...
Sales	√	×	√	...
R&D	×	√	√	...
Marketing	√	√	√	...
...	...	...	...	...



Scenario description

- Centralized authentication point + centralized policy enforcement point.
- The authentication point and policy enforcement point are combined on the core device.
- Devices do not support VXLAN.

Scenario characteristics

- The core device functions as the centralized authentication point for both wired and wireless users on the entire network.
- The core device also works as the policy enforcement point for free mobility.
- The core device stores authentication information about all users on the network. After traffic is forwarded to the core device, it enforces policies based on the policy control matrix defined by the administrator.
- The network does not need to support or deploy VXLAN.

● Authentication point ● Policy enforcement point

Security group and policy control matrix delivery

HUAWEI

Free Mobility Solution (2)

Security group

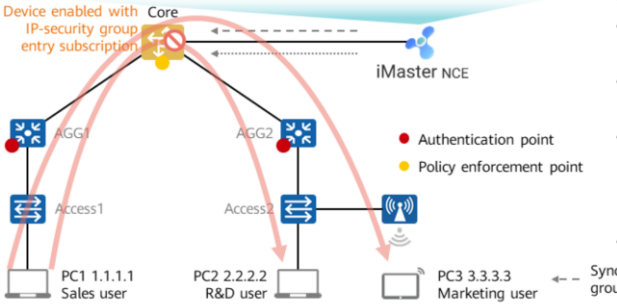
Group Name	Group ID
Sales	1
R&D	2
Marketing	3
...	...

Security group-based policy control matrix

	Sales	R&D	Marketing	...
Sales	√	×	√	...
R&D	×	√	√	...
Marketing	√	√	√	...
...	...	...	...	...

IP-security group mapping table

IP Address	Group ID
1.1.1.1	1
2.2.2.2	2
3.3.3.3	3
...	...



Scenario description

- Distributed authentication points + centralized policy enforcement point.
- Authentication points are separated from the policy enforcement point.
- Devices do not support VXLAN.

Scenario characteristics

- Multiple authentication points exist on the entire network, for example, distributed on aggregation switches.
- The core device functions as the centralized policy enforcement point for free mobility.
- iMaster NCE-Campus synchronizes IP-security group entries to the core device. After traffic is forwarded to the core device, it enforces policies based on the policy control matrix defined by the administrator.
- The network does not need to support or deploy VXLAN.

HUAWEI

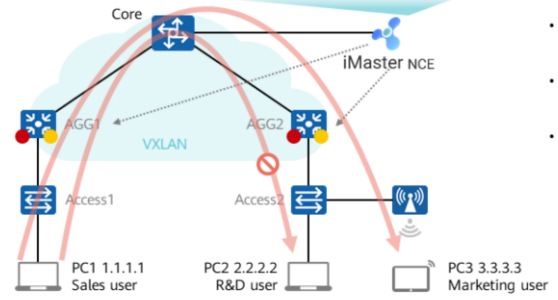
Free Mobility Solution (3)

Security group

Group Name	Group ID
Sales	1
R&D	2
Marketing	3
...	...

Security group-based policy control matrix

	Sales	R&D	Marketing	...
Sales	√	×	√	...
R&D	×	√	√	...
Marketing	√	√	√	...
...	...	...	...	...



Scenario description

- VXLAN-based virtualized campus network, with automatic service provisioning.
- Distributed authentication points + policy enforcement points.

Scenario characteristics

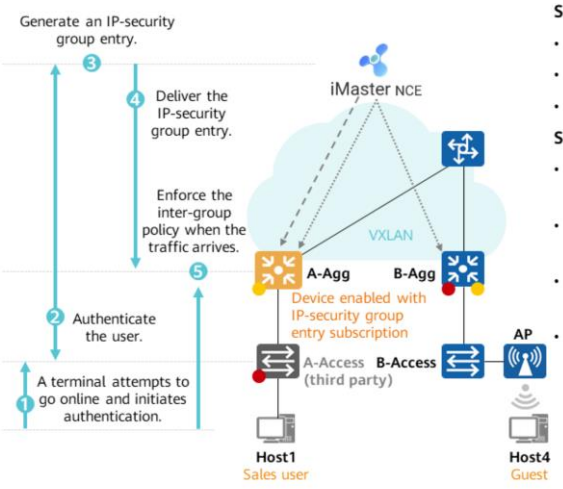
- Multiple authentication points exist on the entire network, for example, distributed on aggregation switches.
- Aggregation switches support VXLAN and function as policy enforcement points.
- The policy for traffic exchanged between users on the same aggregation switch is enforced by the aggregation switch itself.
- Traffic exchanged between users on different aggregation switches is encapsulated using VXLAN. The source security group ID is encapsulated into VXLAN-encapsulated traffic. Policies are enforced on the peer aggregation switch.

● Authentication point ● Policy enforcement point

Security group and policy control matrix delivery

HUAWEI

Free Mobility Solution (4)



Scenario description

- VXLAN-based virtualized campus network.
- Distributed authentication points + policy enforcement points.
- The third-party device functions as a user authentication point.

Scenario characteristics

- A third-party switch exists on the network and functions as a user authentication point.
- Aggregation switches support VXLAN and function as policy enforcement points.
- The policy for traffic exchanged between users on the same aggregation switch is enforced by the aggregation switch itself.
- Traffic exchanged between users on different aggregation switches is encapsulated using VXLAN. The source security group ID is encapsulated into VXLAN-encapsulated traffic. Policies are enforced on the peer aggregation switch.

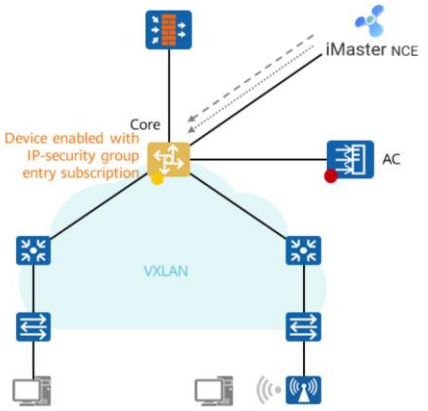
● Authentication point ● Policy enforcement point

Synchronize IP-security group entries

Security group and policy control matrix delivery

HUAWEI

Free Mobility Solution (5)



Scenario description

- VXLAN-based virtualized campus network
- The AC functions as the authentication point for wireless users, and the core switch functions as the gateway for wireless users.

Scenario characteristics

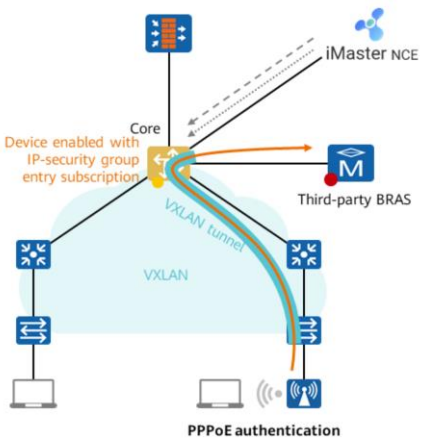
- The AC functions as the authentication point for wireless users and interacts with iMaster NCE-Campus to complete user authentication.
- The core switch functions as the free mobility policy enforcement point, to which iMaster NCE-Campus synchronizes IP-security group entries.

Authentication point
Policy enforcement point
Synchronize IP-security group entries
Security group and policy control matrix delivery

HUAWEI

User Management > User Authentication > Policy Control > Terminal Identification

Free Mobility Solution (6)



Scenario description

- VXLAN-based virtualized campus network.
- The third-party device (BRAS) functions as the user authentication point, and terminals need to access the external network after PPPoE authentication. Free mobility is supported.

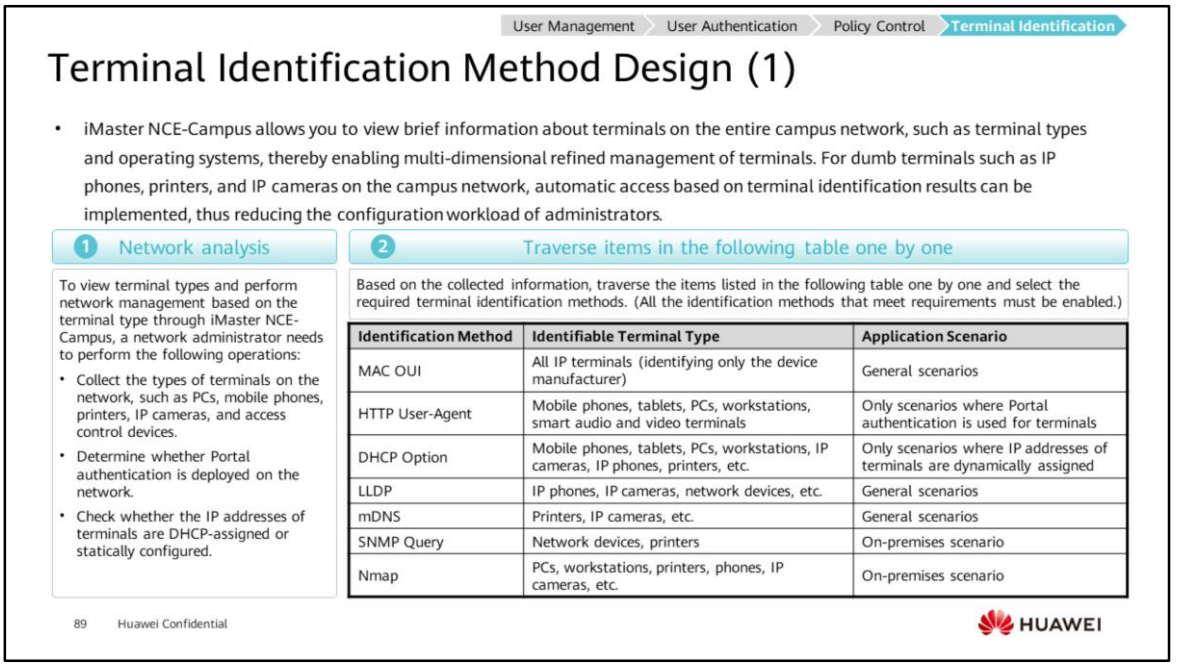
Scenario characteristics

- The third-party BRAS functions as the user authentication point and interacts with iMaster NCE-Campus to complete user authentication.
- iMaster NCE-Campus automatically provisions VXLAN configurations, so that VXLAN tunnels can be created for transmitting user authentication data to the BRAS, thereby enabling Layer 2 communication between terminals and the BRAS.
- The L2 shared egress mode is configured on the core switch through iMaster NCE-Campus for the connection between the fabric and the external network, so that the core switch can communicate with the third-party BRAS.
- The core switch functions as the free mobility policy enforcement point, to which iMaster NCE-Campus synchronizes IP-security group entries.

Authentication point | Policy enforcement point
Synchronize IP-security group entries | Security group and policy control matrix delivery

HUAWEI

- L2 shared egress mode: The border node is not the user gateway, and the user gateway is a device outside the fabric. The border node connects to the egress device through a Layer 2 interface, and the user gateway is deployed on the egress device for access to the external network.



- In this slide, "general scenarios" refer to authentication, non-authentication, and dynamic/static IP address assignment scenarios.
- In non-authentication scenarios, iMaster NCE-Campus can display information about wired terminals only after the ARP snooping function is enabled on access devices.

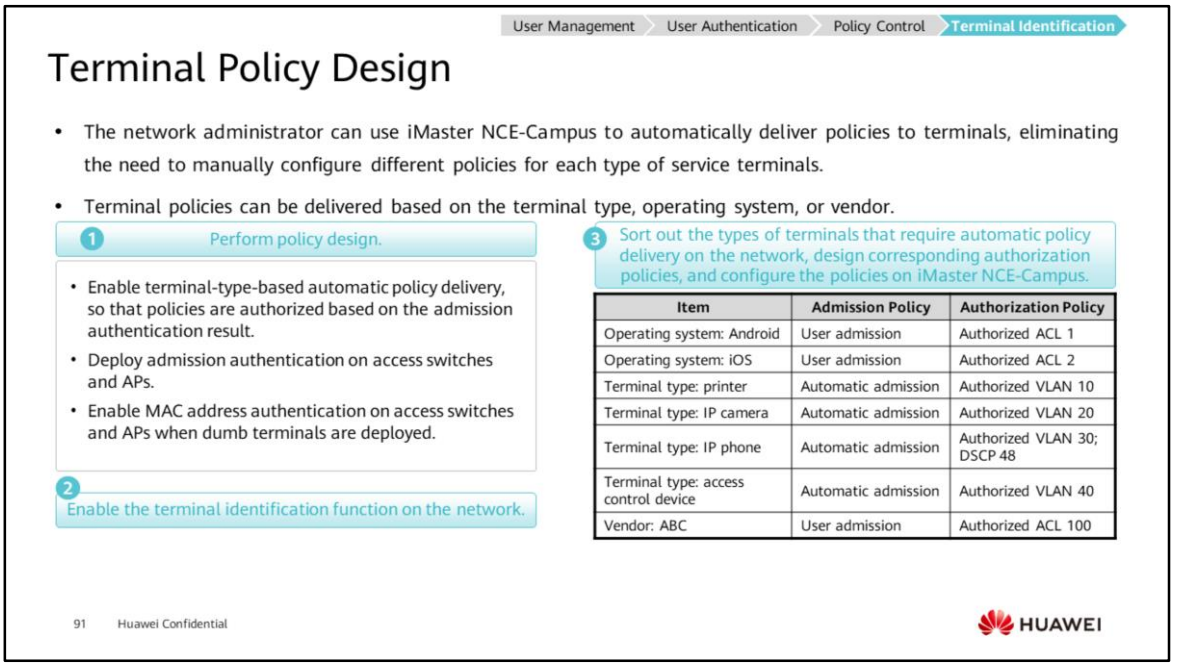
Terminal Identification Method Design (2)

3 Enable the terminal identification function

Identification Method	Operation On iMaster NCE-Campus	Operation on the Network Side
MAC OUI	Enable the terminal identification function.	-
HTTP User-Agent	Enable the terminal identification function.	Enable the terminal identification information reporting function.
DHCP Option	Enable the terminal identification function.	1. Enable the terminal identification information reporting function. 2. Enable DHCP snooping.
LLDP	Enable the terminal identification function.	Terminal identification is enabled by default.
mDNS	Enable the terminal identification function.	1. Enable the terminal identification information reporting function. 2. Enable mDNS snooping.
SNMP Query	1. Enable the terminal identification function. 2. Enable the SNMP scanning function.	-
Nmap	1. Enable the terminal identification function. 2. Install the Nmap plug-in.	-

- If the network administrator cannot determine the terminal identification method to be used, the following passive fingerprint-based identification methods are recommended: MAC OUI, HTTP User-Agent, DHCP Option, LLDP, and mDNS.
- It is recommended that Nmap be disabled by default because this identification takes a long time. If the passive fingerprint-based identification methods cannot meet requirements, enable Nmap.

HUAWEI



- It is recommended that admission and authorization policies be automatically delivered to dumb terminals (such as printers, IP phones, and IP cameras) based on terminal types. This helps implement automatic service provisioning and plug-and-play for dumb terminals.

Contents

1. CloudCampus Solution and Virtualized Campus Network Overview
2. Network Architecture Design
3. Underlay Network Design
4. Fabric and Overlay Network Design
5. Admission Control and Free Mobility Design
6. **WLAN Design**
7. Egress Network Design
8. Network Security and QoS Design
9. O&M Design

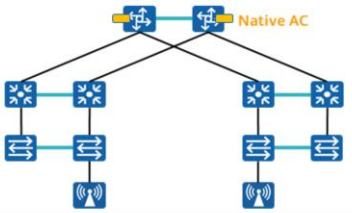
HUAWEI

WLAN Service Solution

- On a large- or medium-sized campus network, the WLAN typically uses the "AC + Fit AP" networking architecture.
- An AC can be deployed in in-path or off-path mode, depending on its location. A native AC (integrated on a switch) must be deployed in in-path mode, whereas a standalone AC can be deployed in either in-path or off-path mode (off-path mode is recommended).

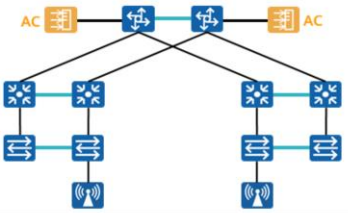
Native AC solution

- Switches provide the native AC function.
- Free mobility is supported. Policies for wired and wireless users are centrally enforced on the switches.
- Unified management of wired and wireless users.



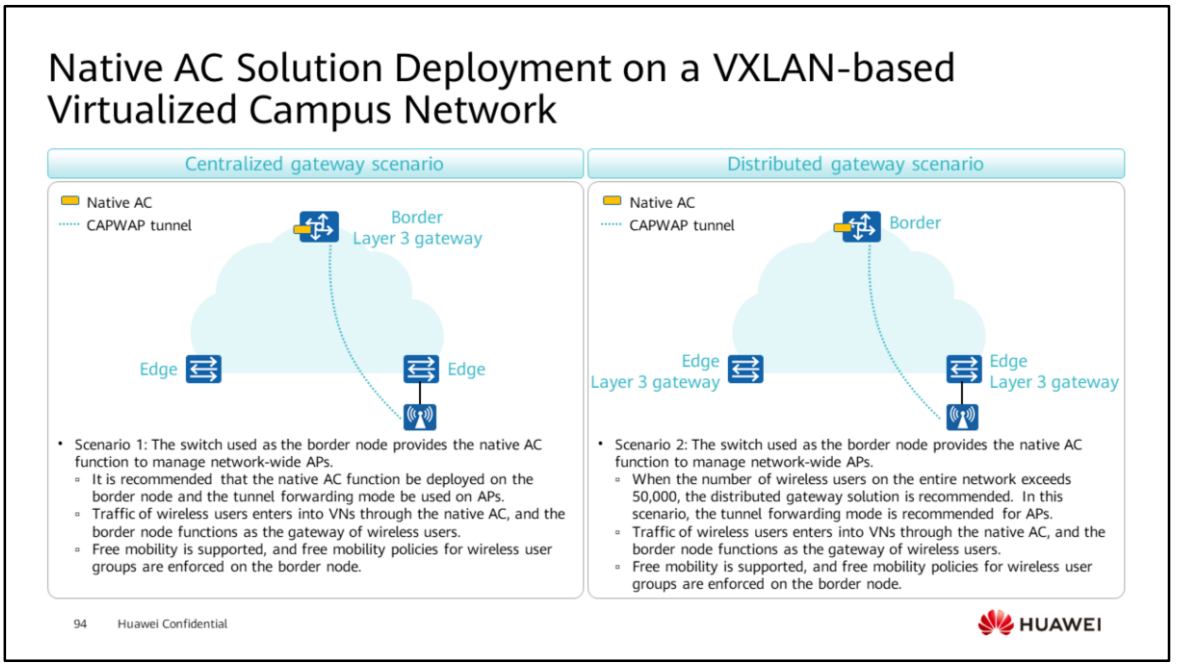
Standalone AC solution

- Standalone AC (connected to the core switch in off-path mode).
- The free mobility solution is supported for wireless users, but it must be used together with the IP-security group entry synchronization solution.
- Separate management for wired and wireless users.

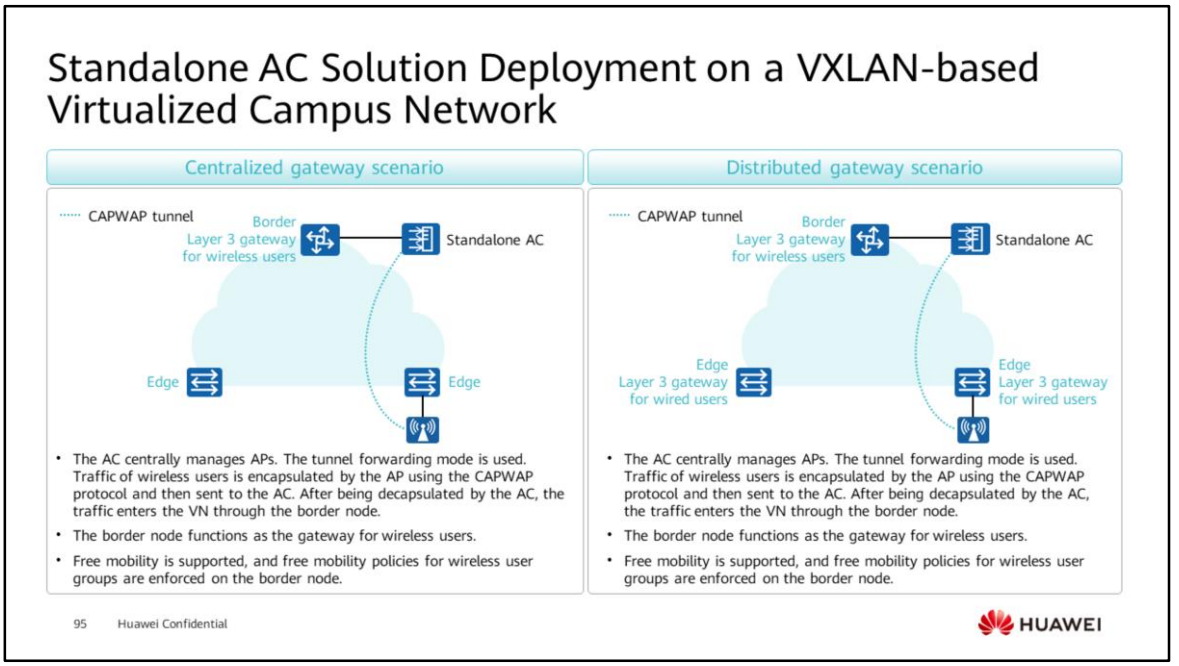


HUAWEI

- Control packets between an AC and AP are forwarded through a CAPWAP tunnel. APs forward service packets of wireless users to the wired side in tunnel forwarding (centralized forwarding) or direct forwarding (local forwarding) mode.
  - Tunnel forwarding:
    - Application scenario: Service packets of wireless users are centrally processed and forwarded by the AC.
    - Advantages: The AC centrally forwards service traffic, ensuring high security and facilitating centralized traffic management and control.
    - Disadvantages: Service packets must be forwarded by the AC, resulting in low forwarding efficiency and a heavy load on the AC.
  - Direct forwarding:
    - Application scenario: Service packets of wireless users are directly forwarded without passing through the AC, saving link bandwidth between the AP and AC.
    - Advantages: Service packets are forwarded without passing through the AC, which is efficient and reduces the load on the AC.
    - Disadvantages: Service data is difficult to manage and control in a centralized manner.
  - The tunnel forwarding mode is recommended for the native AC solution.



- Description of the centralized gateway scenario:
  - User gateway: border node
  - Authentication point: edge node for wired users; border node (native AC) for wireless users
  - Forwarding model:
    - Wired traffic: Traffic enters VNs through the edge node, and free mobility policies are enforced on the edge node.
    - Wireless traffic: Free mobility policies are enforced on the border node. The tunnel forwarding mode is recommended. Traffic enters VNs through the border node (native AC).
- Description of the distributed gateway scenario:
  - User gateway: edge node for wired users; border node for wireless users
  - Authentication point: edge node for wired users; border node (native AC) for wireless users
  - Forwarding model:
    - Wired traffic: Traffic enters VNs through the edge node, and free mobility policies are enforced on the edge node.
    - Wireless traffic: Free mobility policies are enforced on the border node. The tunnel forwarding mode is recommended. Traffic enters VNs through the border node (native AC).



- Description of the centralized gateway scenario:
  - User gateway: border node
  - Authentication point: edge node for wired users; standalone AC for wireless users
  - Forwarding model:
    - Wired traffic: Traffic enters VNs through the edge node, and free mobility policies are enforced on the edge node.
    - Wireless traffic: Free mobility policies are enforced on the border node (the border node needs to subscribe to IP-security group entries). The tunnel forwarding mode is recommended. Traffic enters VNs through the border node (traffic is forwarded from the standalone AC to the border node and then enters VNs).
- Description of the distributed gateway scenario:
  - User gateway: edge node for wired users; border node for wireless users when tunnel forwarding mode is used
  - Authentication point: edge node for wired users; AC for wireless users

Deployment of WLAN Deployment Solutions on a VXLAN-based Virtualized Campus Network

- The following table lists the recommended WLAN deployment solutions for centralized and distributed gateway networking.

Networking Type	AC	Gateway for Wireless Terminals	AP Forwarding Mode	Wireless Traffic Virtualization Location	Number of Terminals Supported	Applicable Scenario
Centralized gateway	Native AC on the border node	Border node	Tunnel forwarding	Border node	≤ 50,000	Recommended when the number of terminals does not exceed 50,000 and centralized forwarding is implemented for wireless terminals.
	Standalone AC connected to the border node in off-path mode	Border node	Tunnel forwarding	Border node	≤ 50,000	Recommended when the wired network is reconstructed first and centralized forwarding is implemented for wireless terminals.
Distributed gateway	Native AC on the border node	Border node	Tunnel forwarding	Border node	50,000~100,000	Recommended when the number of terminals exceeds 50,000 and centralized forwarding is implemented for wireless terminals.

- In new deployment scenarios or wired and wireless network reconstruction scenarios, the native AC is recommended and tunnel forwarding mode is preferentially selected for APs.

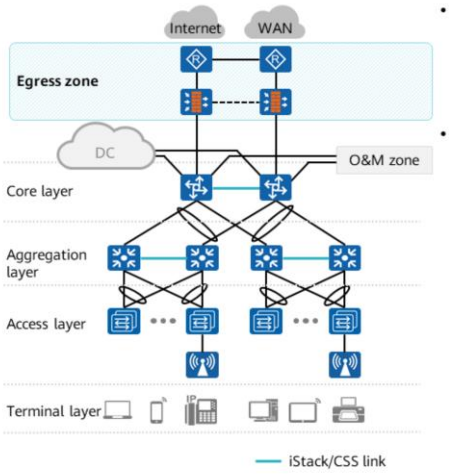
HUAWEI

Contents

1. CloudCampus Solution and Virtualized Campus Network Overview
2. Network Architecture Design
3. Underlay Network Design
4. Fabric and Overlay Network Design
5. Admission Control and Free Mobility Design
6. WLAN Design
7. **Egress Network Design**
8. Network Security and QoS Design
9. O&M Design

HUAWEI

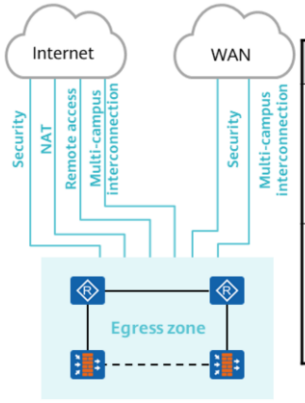
Network Egress Design Overview



- As the boundary between a campus network and external networks (including the Internet and WAN), the egress zone is responsible for communication and security protection between internal and external networks.
- The requirements for the egress zone design are as follows:
  - **Network connectivity:** Internal users can access external networks. If the campus network provides access for external users, external users should be able to access the internal network.
  - **Network security:** To ensure that the campus network is secure and controllable, especially border security, firewalls and the Intrusion Prevention System (IPS) devices need to be configured. Security components are selected based on security requirements and investment scales.
  - **Flexible access modes:** Various access modes are provided, including LAN-side and WAN-side access.
  - **Strong service control capability:** Service deployment and isolation are easy to implement, and various VPN access modes are provided, including IPsec VPN, SSL VPN, and MPLS VPN.

HUAWEI

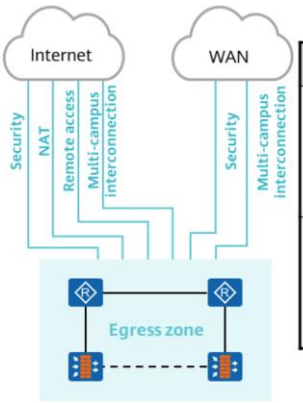
Egress Network Service Scenarios (1)



Service Scenario	Sub-scenario	Description	Applicable Network
Security	Firewall	A firewall uses security zones to divide networks and identify packet forwarding paths. When a packet is transmitted between security zones, the firewall can perform a security check and enforce corresponding security policies.	Internet and leased line
	URL filtering	URL filtering controls internal users' access to external URLs. It allows or forbids users to access certain web page resources, thereby regulating online behaviors.	Internet
NAT	Source NAT	Intranet users usually use private IP addresses to access the Internet. To ensure successful access to Internet services, private IP addresses must be translated into public IP addresses.	Internet
	NAT server	Servers deployed on the campus network, such as the file server, can be mapped to the Internet using public IP addresses or public IP addresses and port numbers through the NAT server, so that they can be accessed from the Internet.	Internet

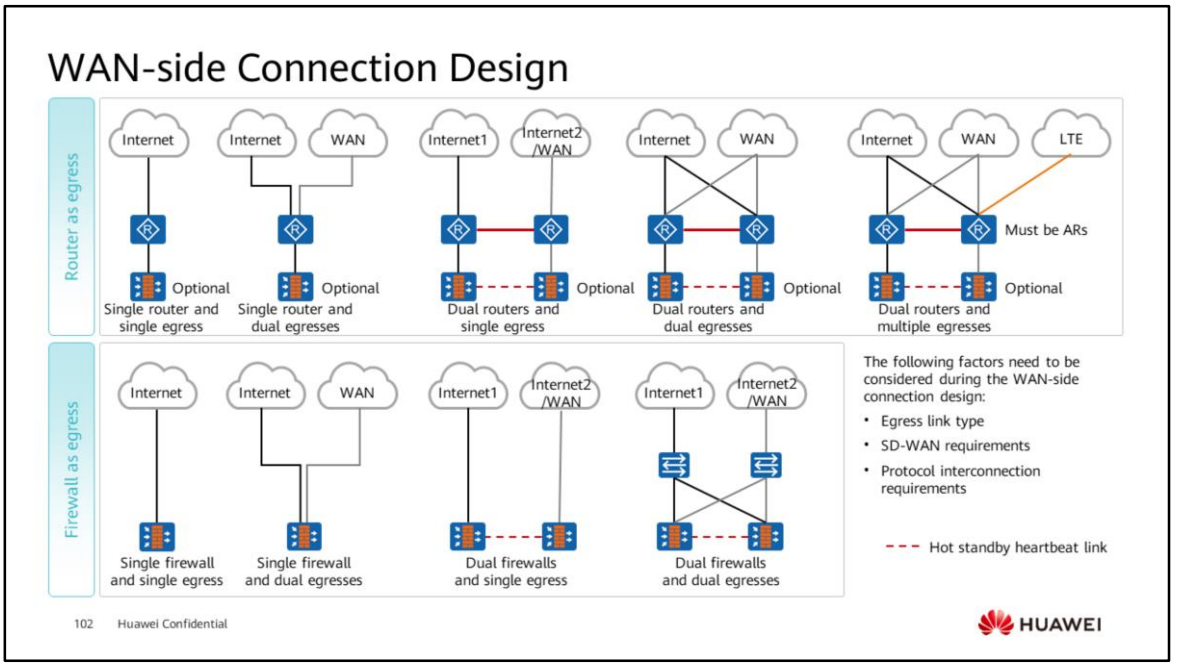
HUAWEI

Egress Network Service Scenarios (2)



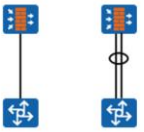
Service Scenario	Sub-scenario	Description	Applicable Network
Remote access	SSL VPN access	The SSL VPN gateway connects to the Internet and provides SSL VPN services for mobile office users (employees on business trips). After a mobile user uses a terminal (such as a laptop or smart phone) to establish an SSL VPN tunnel with the gateway, the user can remotely access intranet resources such as the web server, file server, and mail server through the SSL VPN tunnel.	Internet
Multi-campus interconnection	Traditional MPLS VPN interconnection	CE devices can be connected to the MPLS VPN network.	Leased line
	IPsec VPN interconnection	IPsec VPN tunnels are established between multiple campuses to ensure secure data transmission on a low-security bearer network.	Internet and leased line

HUAWEI

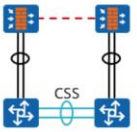


- To ensure reliability, routers and firewalls are usually deployed in redundancy mode. It is recommended that devices be deployed in redundancy mode at the egress of a large- or medium-sized campus network.
- WAN-side connection design:
  - Egress link type
    - Ethernet: Select routers (NE or AR series) or firewalls as egress devices.
    - For non-Ethernet links, such as EI, CE1, and CPOS links, select routers as egress devices.
  - SD-WAN requirements:
    - If there are SD-WAN requirements, select AR routers as egress devices.
    - If there is no SD-WAN requirement, select firewalls or NE routers as egress devices.
  - Protocol interconnection requirements:
    - If BGP runs between egress devices and external networks, it is recommended that routers be used as the egress devices. This is because routers have a larger number of routing table entries and higher performance and many routing policies need to be deployed on egress devices.

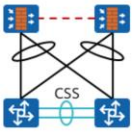
LAN-side Connection Design



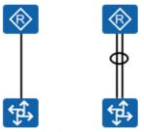
A single firewall is connected to a single core switch



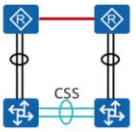
Two firewalls are connected to a CSS of core switches in square-looped mode



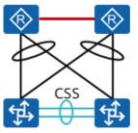
Two firewalls are connected to a CSS of core switches in dual-homed mode



A single router is connected to a single core switch



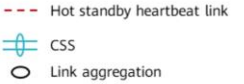
Two routers are connected to a CSS of core switches in square-looped mode



Two routers are connected to a CSS of core switches in dual-homed mode

LAN-side connection design

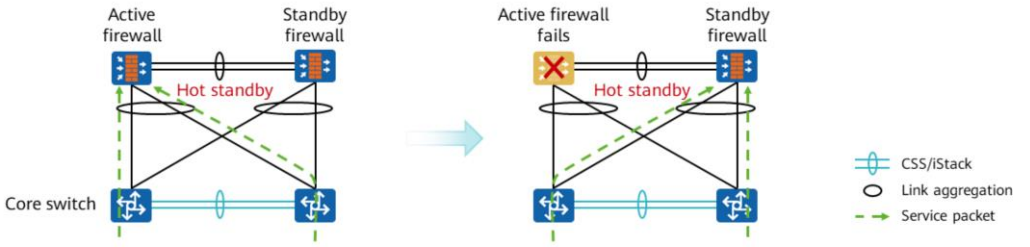
- It is recommended that core switches be deployed in cluster or stack mode to form a logical switch, thereby ensuring reliability of core switches.
- Egress gateways are usually deployed in a two-node cluster to meet reliability requirements. Different types of gateways differ slightly in their two-node cluster deployment capabilities. For example, the link between the active and standby firewalls in a two-node cluster is used only for heartbeat and entry synchronization, and not for traffic forwarding. The link between routers functioning as gateways can be used for service interconnection and for synchronizing and backing up information between them.
- It is recommended that egress gateways and core switches be connected in dual-homed mode and link aggregation be used to enhance the link-level reliability and forwarding bandwidth between devices.



HUAWEI

Firewall Hot Standby

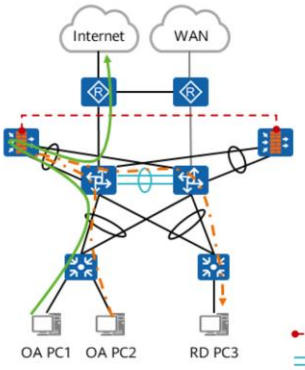
- When firewalls function as egress devices, you are advised to deploy HSB to improve firewall reliability.



- As shown in the figures, the firewalls act as egress devices of the campus network and are directly connected to the core switches. The two firewalls are configured to work in HSB mode, and the Eth-Trunk links connecting the firewalls and core switches work in active/standby mode. When the active firewall is faulty, the standby firewall takes over services from it and forwards service packets.

HUAWEI

Off-Path Deployment of Firewalls



Two firewalls connected to the border node in off-path mode

HSB heartbeat link
CSS/iStack
Link aggregation
Traffic from PC1 to the Internet
Traffic from PC2 to PC3

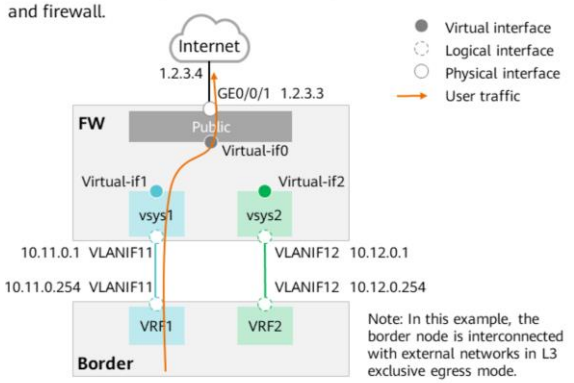
Deploying firewalls in off-path mode

- In addition to in-path deployment, firewalls can be deployed in off-path mode. Generally, off-path deployment of firewalls is selected based on the following principles:
  - Whether traffic passes through a firewall can be flexibly defined.
  - Expansion is easy, for example, in the network reconstruction scenario.
  - A firewall fault has little impact on services.
- On a campus network:
  - If a firewall is used as the egress gateway, the firewall usually provides functions such as security policies, NAT, and VPN gateway. If the north-south traffic must pass through the firewall and if scalability is not considered, in-path deployment is recommended for the firewall.
  - If a router is used as the egress gateway and the firewall only provides advanced security functions, such as intrusion prevention, antivirus, and URL filtering, off-path deployment can be used for the firewall.

HUAWEI

Egress Routing Design: In-Path Firewall Deployment

- Routes between a firewall and core switch include routes from the campus intranet to external networks on the core switch as well as return routes from external networks to the campus intranet on the firewall.
- If a firewall is deployed in in-path mode, you are advised to configure static routes for communication between the core switch and firewall.



Note: In this example, the border node is interconnected with external networks in L3 exclusive egress mode.

Route design for intranet users to access the Internet

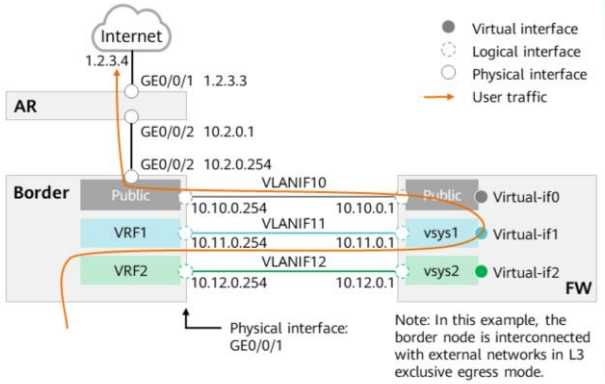
- Routes for traffic from Intranet users to the Internet:
  1. Border: default static VPN route, with the next hop being the IP address of the private network interface VLANIF11 of vsys1.
  2. vsys1: default static route, with the next hop being Public, which directs traffic from intranet users to the Internet to the public system.
  3. FW: default static route, with the next hop being the public network gateway address.
- Return route:
  - vsys1: static route, with the next hop being the IP address of the Layer 3 interface corresponding to VRF1 on the border node, which diverts return traffic from the Internet to users in vsys1 to the intranet.

HUAWEI

- Routes for the traffic from intranet users to the Internet:
  - [Border] ip route-static vpn-instance vpn1 0.0.0.0 0 10.11.0.1
  - [vsys1] ip route-static 0.0.0.0 0 public
  - [FW] ip route-static 0.0.0.0 0 1.2.3.4
- Return route:
  - [vsys1] ip route-static 10.11.0.0 24 10.11.0.254
  - Note: No return route needs to be configured for the return traffic in the public system. After a return packet matches the session table in the public system, the packet is directly forwarded to vsys1 for processing.

Egress Routing Design: Off-Path Firewall Deployment

- In the scenario where the firewall is deployed in off-path mode:
  - It is recommended that static routes be configured for communication between the core switch and firewall.
  - It is recommended that a dynamic routing protocol (OSPF) be configured for communication between the core switch and egress router.



Route design for intranet users to access the Internet

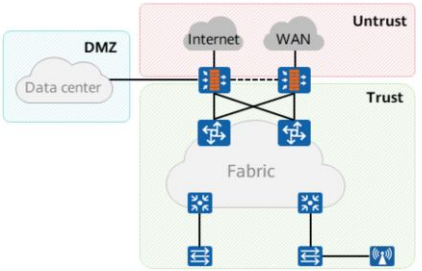
- Routes for traffic from Intranet users to the Internet:
  1. Border: default static VPN route, with the next hop being the IP address of the private network interface VLANIF11 of vsys1.
  2. vsys1: default static route, with the next hop being Public, which directs traffic from intranet users to the Internet to the public system.
  3. FW: default static route, with the next hop being the address of VLANIF10 on the border node.
  4. Border: Deploy a dynamic routing protocol between the border node and AR to implement interworking and enable access the Internet.
- Return routes:
  - Border: static route, with the next hop being the address of VLANIF10 on the FW.
  - vsys1: static route, with the next hop being the IP address of the Layer 3 interface corresponding to VRF1 on the border node, which diverts return traffic from the Internet to users in vsys1 to the intranet.

HUAWEI

- Routes for the traffic from intranet users to the Internet:
  - [Border] ip route-static vpn-instance vpn1 0.0.0.0 0 10.11.0.1
  - [vsys1] ip route-static 0.0.0.0 0 public
  - [FW] ip route-static 0.0.0.0 0 10.10.0.254
- Return routes:
  - [Border] ip route-static 10.11.0.0 24 10.10.0.1
  - [vsys1] ip route-static 10.11.0.0 24 10.11.0.254
  - Note: No return route needs to be configured for the return traffic in the public system. After a return packet matches the session table in the public system, the packet is directly forwarded to vsys1 for processing.

Security Zone Design

- A security zone, also known as a zone, is a collection of networks connected through one or more interfaces, where users have the same security attributes. There are typically three types of security zones: Trust, DMZ, and Untrust.
  - The Trust zone is a security zone with a high security level. It is typically used to define the zone where intranet users are located.
  - The DMZ is a security zone with a medium security level. It is typically used to define the zone where the servers that need to provide services for external networks are located.
  - The Untrust zone is a security zone with a low security level. It is typically used to define insecure networks such as the Internet.



Security zone planning

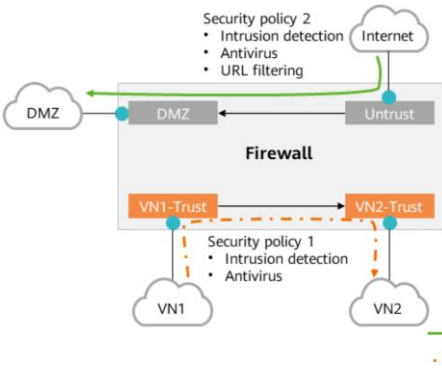
- A campus intranet is considered secure, but is faced with security threats from the outside. Therefore, assign the Internet to the Untrust zone and the campus intranet to the Trust zone. Deploy security devices at the campus network egress to isolate the intranet from the Internet and defend the intranet against external threats. Allocate the data center to the DMZ, and deploy firewalls in the DMZ to isolate traffic between the campus intranet and servers in the data center.

HUAWEI

- On a virtualized campus network, when the user gateways are located inside the fabric, each Layer 3 egress interface for connecting the fabric to an external network corresponds to a Layer 3 logical interface on the firewall. Each logical interface can be bound to a security zone. If the user gateways are located outside a fabric, you need to bind the gateways to security zones based on the security policies of these gateways.
- Most security policies are implemented based on security zones. Each security zone identifies a network, and a firewall connects networks. Firewalls use security zones to divide networks and mark the routes of packets. When packets travel between security zones, security check is triggered and corresponding security policies are enforced. Security zones are isolated by default.

Security Policy Design

- After security zones are created on the firewall, these security zones are isolated from each other by default. To enable communication between security zones (for example, the campus intranet accesses the Internet), you need to configure Layer 3 connectivity and security policies on the firewall.



Recommended security policy design for common zones:

Destination Zone	Traffic Source	Trust Level	Recommended Security Policies
Internet	External users	Untrusted	Intrusion detection, URL filtering, antivirus
	Employees on the go	Medium	
WAN	Enterprise branch	Medium	URL filtering, antivirus
Intranet	Enterprise employees	High	URL filtering, antivirus
	Guests	Low	

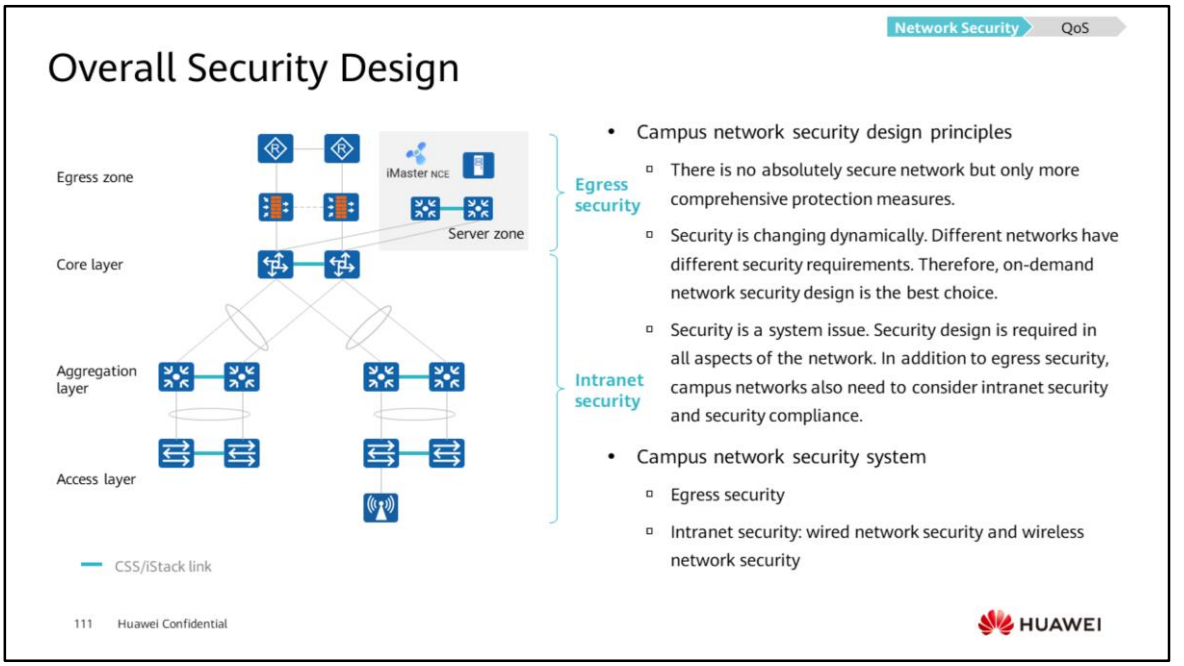
HUAWEI

- As shown in the figure, after security policies are configured, VNs on the intranet can communicate with each other, and the external networks can access servers in the DMZ. In addition, different security protection policies can be applied to traffic in different security zones.

Contents

1. CloudCampus Solution and Virtualized Campus Network Overview
2. Network Architecture Design
3. Underlay Network Design
4. Fabric and Overlay Network Design
5. Admission Control and Free Mobility Design
6. WLAN Design
7. Egress Network Design
8. **Network Security and QoS Design**
9. O&M Design

HUAWEI



- On a traditional campus network, the intranet is often considered secure, and threats mainly come from external networks. Firewalls are often deployed to ensure security on campus borders. As security challenges increase, border defense at the egress cannot meet requirements. The security model needs to shift from passive into proactive and the security scope needs to be expanded from external networks to the intranet to solve security problems from the source (terminals), improving enterprise-wide information security level.

Egress Network Security Design

- Network services provided by the campus intranet to external networks, such as the enterprise website access service and email service, may have potential security risks, threatening the security of the campus intranet. It is recommended that the following security services be deployed on the egress firewall of the campus network to protect the network border:
  - Assign networks for the employees, server networks, and external networks to different security zones, so as to inspect and protect inter-zone traffic.
  - Enable the content security protection functions based on the type of network services provided for external networks by an enterprise. For example, enable antivirus and intrusion prevention for all servers.
  - If employees need to access an external network, enable functions such as URL filtering and antivirus to defend against external threats and prevent information leakage, thereby ensuring enterprise network security.
- Egress security solution design:
  - Dedicated security devices, such as firewalls and intrusion prevention systems, are recommended.
  - Routers with the abovementioned security functions can also be used.
- Security protection functions:

Function	Description
Intrusion prevention	Compares traffic against the intrusion prevention signature database to prevent application-layer attacks, such as buffer overflows, Trojan horses, backdoor attacks, and worms.
Antivirus	Inspects files transmitted on the network for viruses to protect intranets from data breaches and system crashes caused by viruses.
URL filtering	Permits or denies access to URLs to control the online behavior of users.

HUAWEI

Wired Network Security Design

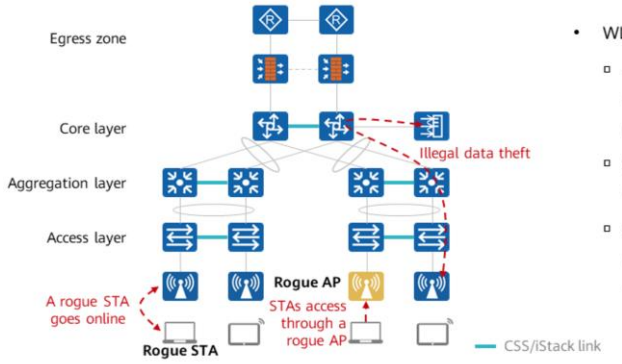
Security Risk	Protection Measure
Broadcast storm	Enable traffic suppression and storm control.
DHCP attack	Enable DHCP snooping and configure uplink interfaces as trusted interfaces.
Impersonating legitimate users	Enable IP source guard (IPSG) and dynamic ARP inspection (DAI).
Mutual access	Enable port isolation.
CPU attack	Enable CPU attack defense.
Denial of service (DoS) attack	Enable attack source tracing.
	Enable port attack defense.
	Enable user-level rate limiting.

HUAWEI

- Security measures:
  - Enable traffic suppression and storm control.
    - Control broadcast, multicast, and unknown unicast packets to prevent broadcast storms. Traffic suppression limits the traffic using the configured threshold, and storm control blocks the traffic by shutting down interfaces.
  - Enable DHCP snooping and configure uplink interfaces as trusted interfaces.
    - DHCP snooping defends against bogus DHCP server attacks, DHCP server DoS attacks, bogus DHCP packet attacks, and other DHCP attacks. DHCP snooping allows administrators to configure trusted and untrusted interfaces, so DHCP clients can obtain IP addresses from authorized DHCP servers. A trusted interface forwards the DHCP packets it receives, whereas an untrusted interface discards the DHCP ACK packets and DHCP Offer packets received from a DHCP server.
    - An interface directly or indirectly connected to the DHCP server trusted by the administrator needs to be configured as a trusted interface, and other interfaces are configured as untrusted interfaces. This ensures that DHCP clients obtain IP addresses only from authorized DHCP servers and prevents bogus DHCP servers from assigning IP addresses to DHCP clients.
  - Enable IPSG and DAI.
    - IPSG prevents unauthorized hosts from using IP addresses of authorized hosts or specified IP addresses to access or attack the network.

Overview of Wireless Network Security Design

- On a WLAN, service data is transmitted through radio signals. Such open channels are vulnerable to service data interception and tampering during transmission, such as rogue STAs, spoofing APs, and DoS attacks of malicious terminals.
- In addition to the security risks of wired networks, wireless networks also face security risks of wireless air interfaces. Therefore, the security risks of wireless networks mainly involve the security of wireless air interfaces.



- WLAN security design involves the following:
  - **Air interface security:** identifies and defends against attacks such as rogue APs, rogue STAs, unauthorized ad-hoc networks, and DoS attacks.
  - **STA access security:** ensures the validity and security of STAs' access to the WLAN.
  - **Service security:** protects service data of authorized users from being intercepted by unauthorized users during transmission.

HUAWEI

Wireless Network Security Design

- Wireless air interface security design:

Security Risk	Defense Measure
Wireless attack and spoofing	WIDS
	WIPS
	Attack detection
Unencrypted management frames	Protected management frame (PMF)

- STA access security design:

Security Risk	Defense Measure
Air interface encryption and access authentication	WPA/WPA2
	WAPI
Terminal access security	STA blacklist and whitelist
Brute-force cracking	Brute-force attack defense and dynamic blacklist

- Wireless service security design:
  - The wired network between APs and ACs also faces common security threats, for example, interception, tampering, and spoofing, on IP networks.
  - To improve data transmission security, CAPWAP tunnels between APs and ACs support Datagram Transport Layer Security (DTLS) encryption, including DTLS encryption of CAPWAP tunnels for management packets, DTLS encryption of CAPWAP tunnels for service data packets, encryption of sensitive information, and integrity check.

HUAWEI

- Wireless air interface security design:
  - The Wireless Intrusion Detection System (WIDS) can detect rogue and interfering APs, bridges, and STAs, as well as ad-hoc devices.
  - The Wireless Intrusion Prevention System (WIPS) can disconnect authorized users from rogue APs, disconnect rogue and interfering devices from the WLAN, and contain such devices.
  - Attack detection: The WIDS and WIPS can also detect attacks such as flood attacks, weak initialization vector (IV) attacks, spoofing attacks, brute force WPA/WPA2/WAPI pre-shared key (PSK) cracking, and brute force WEP shared key cracking in a timely manner. The two systems then record logs, statistics, and alarms to notify network administrators of such attacks. The WLAN device adds devices that initiate flood attacks and brute force key cracking attacks to the dynamic blacklist and rejects packets from such devices within the aging time of the dynamic blacklist.
  - PMF: Management frames on a WLAN are not encrypted, which may cause security problems. The PMF standard is released by the Wi-Fi Alliance based on IEEE 802.11w. It aims to apply security measures defined in WPA2 to unicast and multicast management action frames to improve network trustworthiness.

QoS Design

- In addition to traditional data services such as web, email, and FTP services, large and medium-sized campus networks also transmit services such as video surveillance, video conferencing, voice call, and production scheduling, which have specific requirements on bandwidth, latency, and jitter. For example, video surveillance and video conferencing require high bandwidth, low latency, and low jitter. The voice service does not require high bandwidth, but requires low latency. When congestion occurs, the voice service must be processed first.
- Quality of Service (QoS) is designed to provide different levels of service quality during transmission of data flows to meet requirements of different services for performance indicators such as bandwidth, latency, jitter, packet loss ratio, and throughput. A variety of QoS technologies can be used to improve the network service quality, such as priority mapping, traffic policing, traffic shaping, queue scheduling, and congestion avoidance. These technologies enable the network to deliver an optimal user experience with limited resources.
- The QoS design process consists of requirement survey and analysis, traffic classification design, and scheduling policy design.

HUAWEI

Requirement Survey and Analysis

- Before deploying QoS, you must be familiar with various services as well as the traffic model and QoS requirements of each service. This helps correctly design QoS policies, so that QoS guarantee can be provided for each service.
- Content and purpose of QoS requirement survey:

Requirement Type	Key Points of Requirement Survey	Survey Purpose
Current network situations	Traffic models of various services and E2E forwarding paths of each service, including forwarding paths in normal and abnormal conditions.	To determine the location for QoS policy deployment.
	Bandwidth bottleneck points on the network and interface bandwidth of each bandwidth bottleneck point.	To determine whether to deploy QoS policies or expand the network capacity.
QoS requirements	Control services, multimedia services, and other important services that need special attention.	To understand traffic types and QoS requirements of each traffic type.
	Characteristics of various types of traffic that can be identified by network devices. For example, check whether voice traffic uses a proprietary protocol such as SIP or H.323, whether traffic is originated from or destined for a specific interface of a specific server, and whether traffic is originated from or destined for a specific host network segment.	
	Bandwidth requirements of important services. For example, multimedia services from different vendors at different bit rates require different bandwidths so that the audio or video services can be smoothly transmitted.	
	Processing policies for different multimedia applications. For example, some enterprises require that online videos on intranets be guaranteed and work-irrelevant online videos on the Internet not be guaranteed.	

HUAWEI

Traffic Classification Design (1)

- Services carried on a large- or medium-sized campus network include voice, video, and data services as well as network protocol control signaling required for transmitting the service data.
- QoS counters of these services include the bandwidth, packet loss rate, latency, and jitter. Bandwidth can be controlled by configuring parameters, but the packet loss rate, latency, and jitter cannot.
- Characteristics and QoS requirements of common services:

Service Category	Typical Application or Protocol	Packet Loss Tolerance	Latency Tolerance	Jitter Tolerance
Network protocols	Link-layer loop prevention protocols, routing protocols, multicast group management protocols, and other protocols designed for network connectivity and interoperability.	Low	Low	Allowed
Management protocols	Protocols used by network administrators for monitoring network devices, delivering configurations, and diagnosing faults.	Low	Low	Allowed
VoIP data flow	Real-time voice calls over IP networks. The network must provide low latency and low jitter to ensure service quality.	Very low	Very low	Very low
Voice signaling	Signaling protocols for controlling VoIP calls and establishing communication channels. Signaling protocols have a lower priority than VoIP data flows because intermittent voices are often considered worse than call failure.	Low	Low	Allowed
Multimedia conferencing	Multiple parties share camera feeds and screens over IP networks. In addition, the protocols or applications provide the adaptive bit rate capability. When the network quality is poor, the system automatically decreases the bit rate and image quality to ensure smooth video communication.	Low or medium	Very low	Low

HUAWEI

- Service category examples:
  - Network protocols: STP, OSPF, IGMP, etc.
  - Management protocols: ICMP, SNMP, Telnet, etc.
  - Voice signaling: signaling protocols such as SIP, H.323, H.248, and Media Gateway Control Protocol (MGCP).

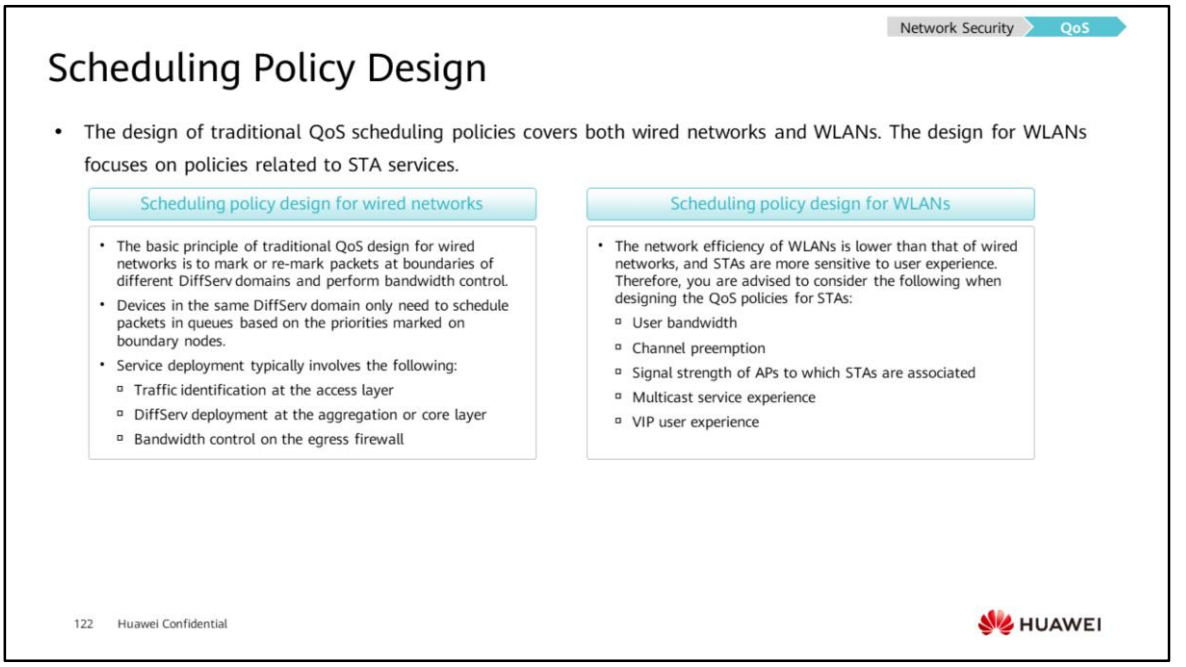
Traffic Classification Design (2)

- Characteristics and QoS requirements of common services (continued):

Service Category	Typical Application or Protocol	Packet Loss Tolerance	Latency Tolerance	Jitter Tolerance
Gaming	The network is required to provide online interactive applications with a low packet loss rate, low latency, and low jitter to ensure fast and accurate response during gaming.	Low	Very low	Low
Streaming media	Online audio and video streaming. Audio and video programs are made in advance and then cached on local terminals before being played. Therefore, they have lower requirements for network delay, packet loss, and jitter.	Low or medium	Medium	Allowed
Online live telecasting	Unlike streaming media, data of online live telecasting is sent and received in real time. Though terminals provide the cache mechanism, the network is required to provide a low packet loss rate and jitter to meet real-time requirements and ensure good experience.	Very low	Medium	Low
Low-latency data services	Data services for which users are waiting for output.	Low	Low or medium	Allowed
Bandwidth-intensive services	Network services that involve the transmission of a large amount of data for a long period of time.	Low	Medium or high	Allowed
Common services	Basic services that are not so important and have no special requirements on enterprise networks.	No special requirements	No special requirements	No special requirements
Low-priority services	Services that of no concern or importance to enterprises.	High	High	Allowed

HUAWEI

- Service category examples:
  - Gaming: for example, online games that transmit operation instructions through RTP or UDP, which poses higher requirements on the network.
  - Low-latency data services: for example, long delay on an online ordering system may reduce the revenue and efficiency of enterprises.
  - Large-volume data services: for example, FTP, database backup, and file dump.
  - Common services: for example, mail and web browsing.
  - Low-priority services: for example, network applications irrelevant to work, such as social networking and entertaining videos.



- Scheduling policy design for wired networks:
  - Traffic identification at the access layer: Access switches function as boundary switches. The switches identify, classify, and mark data flows at the user side. In actual deployments, different interfaces on access switches are connected to different terminals, and different priorities can be allocated to different services on the access switches. Then traffic of the services can be scheduled based on the priorities.
  - DiffServ deployment at the aggregation or core layer: Interfaces on aggregation and core switches are configured to trust DSCP or 802.1p priorities and enforce QoS policies based on priorities marked at the access layer. This ensures that high-priority services are scheduled first. A switch interface trusts 802.1p priorities by default.
  - Bandwidth control on the egress device: Egress devices are also located in the DiffServ domain and are configured to trust DSCP or 802.1p priorities of packets and implement QoS policies. Due to egress bandwidth limits, you need to consider differences when setting bandwidth parameters for WAN interfaces of egress devices. Additionally, QoS policies of egress devices vary according to the enterprise WAN construction mode.

Suggestions on Scheduling Policy Design

- The definition of important data services varies with enterprises. For a portal website, Internet access and gaming traffic is important; for the financial industry, real-time transaction is more important than voice services, and Internet access and gaming traffic is unwanted. Therefore, QoS policies must be designed and deployed based on actual service types and QoS requirements of each enterprise.
- Recommended scheduling policy design:

Application Type	Typical Application or Protocol	CoS	Queue (Priority)	Scheduling Algorithm	Maximum Bandwidth
Signaling and control	• Routing protocols • Network management protocols • Multimedia protocol signaling	CS6	6	PQ	Unlimited
Real-time interactive multimedia	• VoIP • Multimedia conferencing • Online gaming • Desktop cloud	EF	5	PQ	Available interface bandwidth x 30%
On-demand subscription of multimedia or key services	• Online video • Video live telecasting or multicast • Delay-sensitive and mission-critical enterprise services that require real-time interaction, such as online ordering	AF	4	DRR: with weight being 20	Unlimited
Other services	• Common Internet access services such as email and web browsing	BE	0	DRR: with weight being 20	Unlimited

HUAWEI

Contents

1. CloudCampus Solution and Virtualized Campus Network Overview
2. Network Architecture Design
3. Underlay Network Design
4. Fabric and Overlay Network Design
5. Admission Control and Free Mobility Design
6. WLAN Design
7. Egress Network Design
8. Network Security and QoS Design
9. **O&M Design**

HUAWEI

Challenges Facing Campus Network O&M

Precise detection

Traditional O&M is based on SNMP and therefore data collection takes minutes. Once a fault occurs, data generated at the time when the fault occurred cannot be obtained in real time. In addition, no convenient backtracking method is available.

Key requirements: Telemetry-based precise detection

Experience perception

In traditional O&M systems, only device metrics are monitored. User experience may be poor even if the metrics are normal. In addition, traditional O&M systems do not provide correlative analysis for clients and networks.

Key requirements: client & network profiling, experience perception

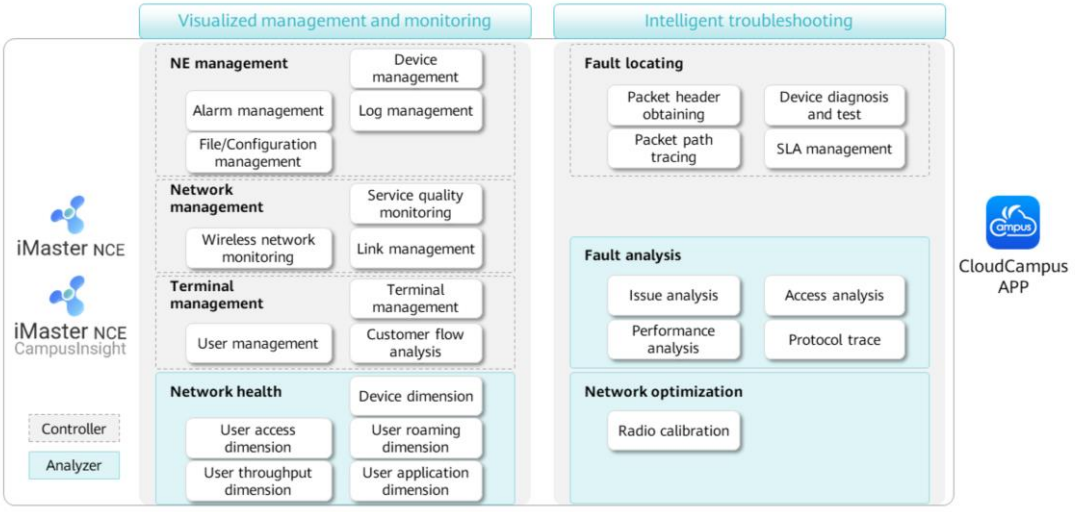
Fault identification

During traditional O&M, network faults can be detected only after receiving clients' complaints. As a result, faults cannot be effectively and proactively identified and analyzed.

Key requirements: intelligent fault identification

HUAWEI

Campus Network O&M Panorama



HUAWEI

Basic Network O&M Design

- In the large- and medium-sized virtualized campus network solution, iMaster NCE-Campus provides comprehensive basic network management, network element (NE) management, service management, and system management functions for managed devices. These functions include user, log, resource, topology, alarm, and performance management. In addition, protocols involved in traditional network O&M can also be applied on iMaster NCE-Campus.

Protocol Name	Configurations Supported on iMaster NCE-Campus
SNMP	Interconnection between network devices and an SNMP server can be configured on iMaster NCE-Campus. iMaster NCE-Campus can manage traditional devices (NETCONF-incapable devices) using SNMP.
NTP	Interconnection between network devices and an NTP server can be configured on iMaster NCE-Campus.
Syslog	Interconnection between iMaster NCE-Campus and a Syslog server can be configured. After they are interconnected, iMaster NCE-Campus uploads the logs obtained from network devices to the Syslog server. Interconnection between network devices and a Syslog server can be configured on iMaster NCE-Campus.
SSH	Local accounts for SSH-based login can be configured on network devices. SSH can be used to log in to the CLI of a network device from iMaster NCE-Campus.
LLDP	LLDP can be enabled for network devices on iMaster NCE-Campus. LLDP is enabled on iMaster NCE-Campus by default. iMaster NCE-Campus can obtain network topology information through LLDP.

HUAWEI

Intelligent O&M Design

- Huawei's intelligent O&M solution uses Telemetry technology to send O&M data (such as device performance indicators and terminal logs) on network devices to Huawei iMaster NCE-CampusInsight. iMaster NCE-CampusInsight then uses big data technology, AI algorithms, and other advanced analysis technologies to digitize user experience on the network, helping customers promptly detect network problems and ultimately improve user experience.
- O&M functions provided by iMaster NCE-CampusInsight:

Category	Function
Network visualization	Wireless network health
	Wired network health
	Integrated topology
Campus service analysis	Issue analysis
	Access analysis
	Performance analysis
	Protocol tracing
Intelligent wireless network	Intelligent radio calibration
	WLAN topology
User application experience	Application analysis

HUAWEI

Deployment Design for the Intelligent O&M Solution

- The deployment process for Huawei intelligent O&M solution involves the following:
  - Interconnection between iMaster NCE-Campus and iMaster NCE-CampusInsight: iMaster NCE-Campus can perform unified O&M and management, and iMaster NCE-CampusInsight can be invoked as a proxy service. iMaster NCE-Campus can synchronize site and device information to iMaster NCE-CampusInsight.
  - Collection of data such as syslogs and performance data of network devices: The network devices must be enabled to collect O&M data and report the data to iMaster NCE-CampusInsight.
- The intelligent O&M solution consists of iMaster NCE-CampusInsight, iMaster NCE-Campus, and devices. Currently, iMaster NCE-CampusInsight can manage Huawei cloud switches and APs, and intelligently analyzes their data.

Network bandwidth design	Deployment location design
Devices need to periodically report data to iMaster NCE-CampusInsight. Therefore, the campus network needs to reserve bandwidth for data reporting. The average bandwidth consumed by each device is 3 kbit/s.	iMaster NCE-CampusInsight and iMaster NCE-Campus can be deployed at different locations. They can collaborate with each other as long as network connectivity is achieved. To avoid instability of the intermediate network, you are advised to deploy them in the same location, for example, a same data center.

HUAWEI

Precautions for Intelligent O&M Design

Network deployment design

During network deployment, ensure that there are reachable routes between network devices and the analyzer iMaster NCE-CampusInsight so that the network devices can send KPI data and log information to iMaster NCE-CampusInsight.

During network deployment, ensure that devices are clock-synchronized with iMaster NCE-CampusInsight. You are advised to deploy an NTP server to synchronize the system clocks on the network.

Intelligent radio calibration function design

Intelligent radio calibration and traditional radio calibration cannot be both deployed on APs in the same region.

Protocol trace function design

To identify DHCP-related connectivity issues and implement protocol trace, you need to use the AC as the DHCP server or enable DHCP snooping on the AC.

Audio and video quality analysis function design

The audio and video quality analysis function requires devices to send related log information to iMaster NCE-CampusInsight. You are advised to configure the same log sending interval for switches and WLAN devices. The maximum difference between the log sending intervals cannot exceed 20s.

HUAWEI

Quiz

1. (Multiple-answer question) For a large-scale campus network with approximately 10,000 terminals, the recommended fabric networking mode is ( ). In this networking mode, the gateway is located on the ( ), and the ( ) functions as the edge node.
   A. Centralized gateway
   B. Distributed gateway
   C. Edge
   D. Border
   E. Access switch
   F. Aggregation switch

HUAWEI

1. ADF

Summary

- To meet requirements of digital transformation, campus networks need to be reconstructed systematically. Based on the IDN concept, Huawei CloudCampus 3.0 Solution helps enterprises build intelligent, simplified, converged, open, and secure networks.
- This course focuses on large- and medium-sized campus networks, illustrates the solution architecture and technical solutions used for such networks, and provides suggestions on engineering design and O&M design. With this information, this course helps trainees understand Huawei CloudCampus 3.0 Solution technologies and design the optimal campus network solution.

HUAWEI

Recommendations

- Huawei's CloudCampus Solution V100R020C10 Product Documentation
  - https://support.huawei.com/hedex/hdx.do?docid=DOC1100797422&lang=en

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Virtualized Campus Network Deployment Guide

HUAWEI

Foreword

- Different from traditional campus networks that focus on standalone devices, a virtualized campus network focuses on the overall service experience of the entire network and uses iMaster NCE-Campus and virtual extensible local area network (VXLAN) technology to flexibly schedule network resources. Virtualization technologies group physical network resources into a network-wide resource pool that can be flexibly adjusted by the service layer and allocated by iMaster NCE-Campus. A physical network is virtualized into multiple logically independent virtual networks that carry various services and have independent network resources. This virtualization decouples services from networks and facilitates service management.
- This course describes the deployment process and typical deployment cases of the VXLAN-based virtualized campus network solution.

HUAWEI

Objectives

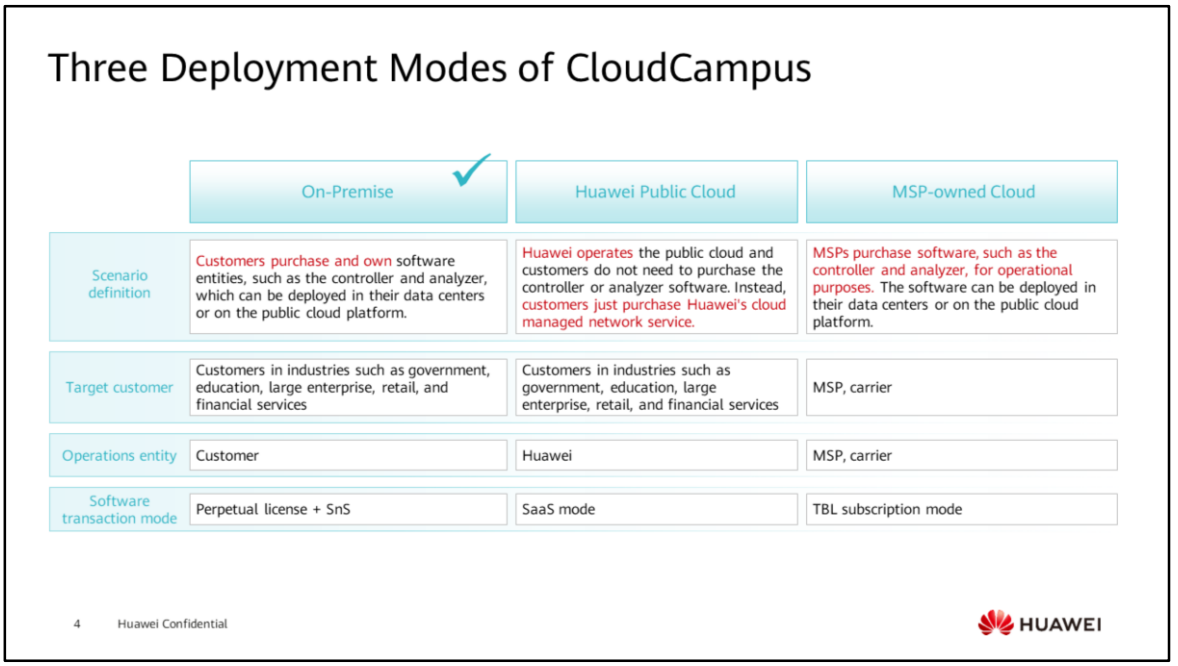
- On completion of this course, you will be able to:
  - Describe the deployment process of the CloudCampus VXLAN-based virtualized campus network solution.
  - Deploy a typical CloudCampus VXLAN-based virtualized campus network solution.
  - Configure iMaster NCE-Campus to manage and maintain the CloudCampus VXLAN-based virtualized campus network.

HUAWEI

Contents

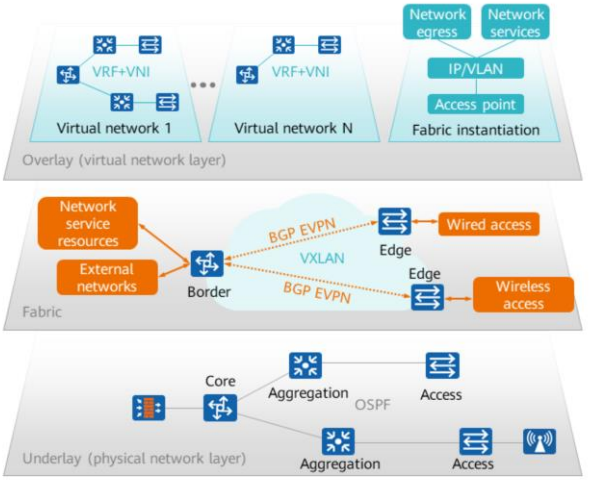
1. **VXLAN-based Virtualized Campus Network Deployment Plan**
2. VXLAN-based Virtualized Campus Network Deployment Process and Guide
   - Deployment Process
   - Preparations for Deployment
   - Deployment Guide

HUAWEI



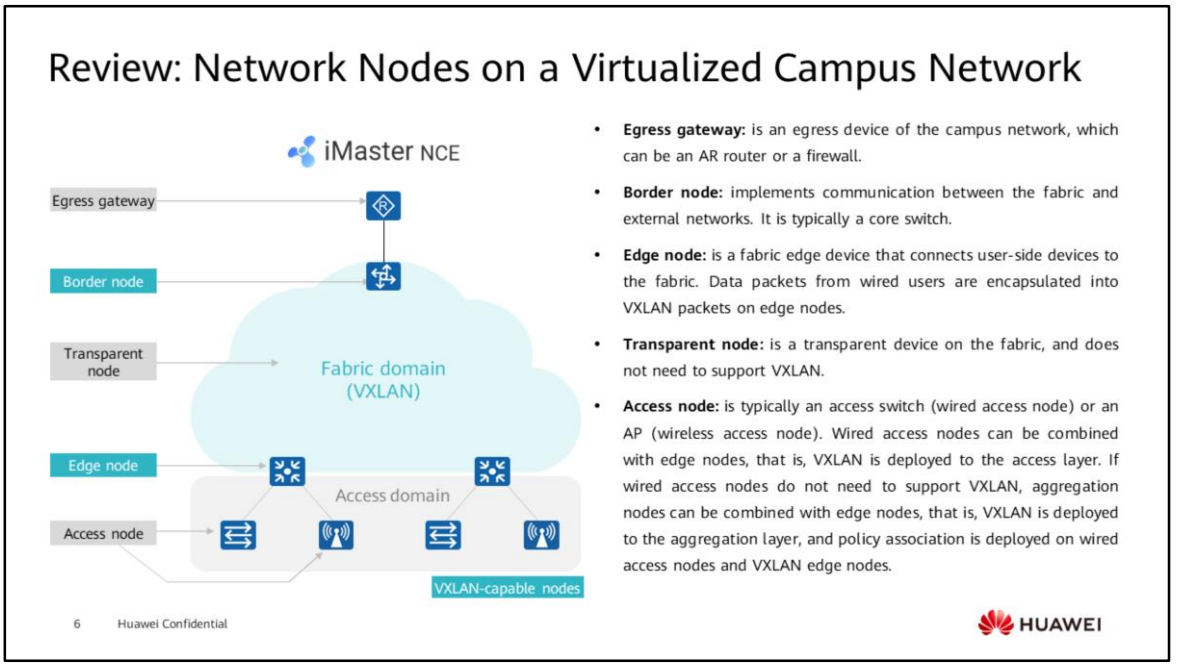
- Managed service provider (MSP): delivers and manages network-based services, applications, and devices. The serviced objects include enterprises, residential areas, and other service providers.
- Perpetual license + SnS: The perpetual license is sold together with SnS services, such as software patches, software upgrades (including new features of new releases), and remote support. In the perpetual license + SnS mode, a customer needs to pay SnS fee for a certain period of time, in addition to purchasing the license upon the first purchase. If the customer does not renew the SnS annual fee after it expires, the customer can only use functions provided in the license for the current version and cannot use the service functions covered in the SnS annual fee.
- SaaS mode: MSPs are responsible for deploying or leasing hardware infrastructure, and O&M and management of the hardware and software. Software is provided for customers as cloud services and customers need to periodically pay for the cloud services.
- Term Based License (TBL) mode: This mode differs from the perpetual license + SnS mode in that the licenses purchased by customers have limited validity periods. If a customer does not renew the subscription after the license expires, the customer can no longer use the software product.
- SnS: refers to Subscription and Support. It consists of two parts: software support and software subscription. The complete software charging mode consists of the annual software SnS fee and software license fee.
- Note: This course uses the on-premise deployment as an example.

Review: Virtualized Campus Network Architecture



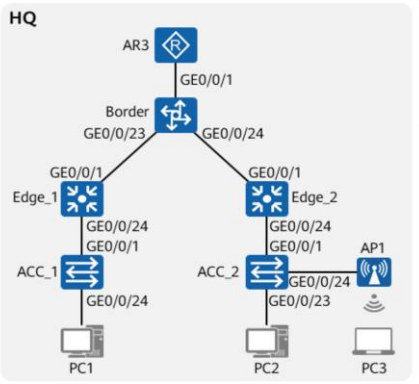
- **Underlay:** a physical topology consisting of physical network devices, such as switches, access points (APs), firewalls, and routers, to provide interconnection capabilities for all services on the campus network, building the basic bearer network for campus service data forwarding.
- **Fabric:** a network with pooled resources abstracted from the underlay network. When creating an instantiated virtual network, you can select the pooled network resources on the fabric.
- **Virtual network:** a logically isolated virtual network instance that is constructed by instantiating a fabric. One virtual network corresponds to one isolated network (service network), for example, research network.

HUAWEI



- On a fabric, VXLAN tunnel endpoints (VTEPs) are further divided into the following roles:
  - Border: is a physical network device and provides data forwarding between the fabric and external networks. In most cases, VXLAN-capable core switches function as border nodes.
  - Edge: is a physical network device. Access user traffic enters the fabric from an edge node. Generally, VXLAN-capable access or aggregation switches function as edge nodes.

Lab: Requirements



- To implement multi-service convergence on a campus network, virtual networks are deployed and configured on the campus network through iMaster NCE-Campus. This enables different virtual networks on the same physical network to be divided based on services.
- Using the VXLAN technology, virtual networks meet the following requirements:
  - Network devices support DHCP-based plug-and-play provisioning.
  - Multiple services on the campus network share the same physical network, but are logically isolated. Mutual access control can be implemented among these services.
  - Service configuration is automated, and virtual network configurations are delivered by iMaster NCE-Campus, removing the need to log in to devices to manually configure them.
  - Users can access virtual networks from anywhere on the campus network, implementing flexible user authentication, onboarding, and free mobility.

HUAWEI

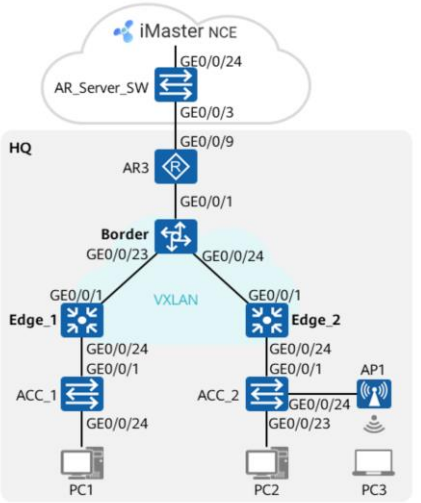
Lab: Gateway Solution Selection

- When designing the virtualized campus network solution, first determine the gateway solution to be used. After the gateway solution is determined, you can perform end-to-end design on the entire campus network based on the selected gateway solution.

Item	Centralized Gateway	Distributed Gateway
User gateway location	Border	Edge
O&M deployment	A border node functions as the gateway of all users, and the native AC function is typically enabled on the border node to support wireless services, simplifying O&M deployment.	Multiple edge nodes function as user gateways. This ensures that only part of the network is affected upon the failure of one gateway and facilitates network expansion. The native AC function is enabled on the border node to support wireless services. The O&M deployment is complex.
Terminal scale	≤ 50,000 (This solution is recommended if the number of terminals does not exceed 50,000.)	50,000 to 100,000 (This solution is recommended when the number of terminals exceeds 50,000.)

HUAWEI

Lab: Physical and VXLAN Networking

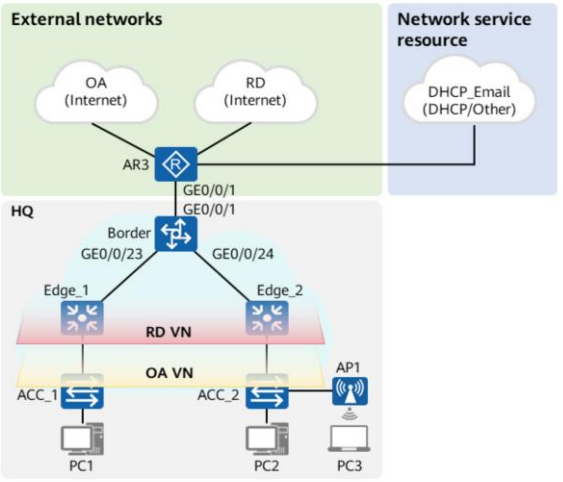


- **HQ:** ACC_1 and ACC_2 function as access devices that connect to wired terminals and provide network services for wired users. AP1 is connected to ACC_2 to provide network services for wireless users. Edge_1 and Edge_2 serve as aggregation devices, and the Border functions as the core device. AR3 works as both the campus egress and the DHCP server, which allocates IP addresses to other devices and user terminals at the HQ. OSPF is used for communication on the underlay network at the HQ.
- **Cloud:** AR_Server_SW is used to simulate the cloud. It connects the HQ and iMaster NCE-Campus, and also functions as the gateway of iMaster NCE-Campus.
- **VXLAN network (fabric):** The network topology for this lab uses a distributed gateway model, in which VXLAN is deployed across core and aggregation layers. Edge_1 and Edge_2 function as the edge nodes of the VXLAN network, whereas Border functions as the border node of the VXLAN network.

HUAWEI

- Note: In this lab, the native AC is deployed on the border node to manage APs. The border node also serves as the DHCP server to allocate IP addresses to APs.

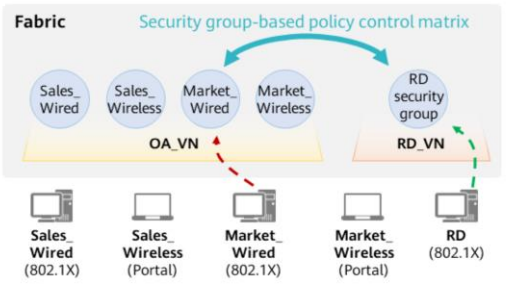
Lab: Virtual Network



- **Virtual network (VN):** Two virtual networks are defined for access of different end users.
  - OA VN: for access of sales personnel (Sales_Wired and Sales_Wireless security groups) and marketing personnel (Market_Wired and Market_Wireless security groups)
  - RD VN: for access of R&D personnel (RD security group)
- **External network:** Two external networks are defined for the two virtual networks, so different end users can access different external networks.
  - OA external network: accessible to sales and marketing personnel.
  - RD external network: accessible to R&D personnel.
- **Network service resource:** One network service resource named DHCP_Email is defined to function as both a DHCP server and an email server. It allocates IP addresses to all end users in the two virtual networks.

HUAWEI

Lab: Security Group and Policy Control Matrix



- **Security group**: Five security groups are defined to identify different end users.
  - Sales_Wired and Sales_Wireless: security groups to which sales personnel are assigned after they pass 802.1X or Portal authentication.
  - Market_Wired and Market_Wireless: security groups to which marketing personnel are assigned after they pass 802.1X or Portal authentication.
  - RD: security group to which R&D personnel are assigned after they pass 802.1X authentication.
- **Resource group**: One resource group is defined, indicating the email server.

Policy control matrix:

Source Group \ Destination Group	Sales_Wired	Sales_Wireless	Market_Wired	Market_Wireless	RD	E_mail
Sales_Wired			Deny	Deny	Deny	
Sales_Wireless			Deny	Deny	Deny	
Market_Wired	Deny	Deny				
Market_Wireless	Deny	Deny			Deny	
RD	Deny	Deny		Deny		Deny

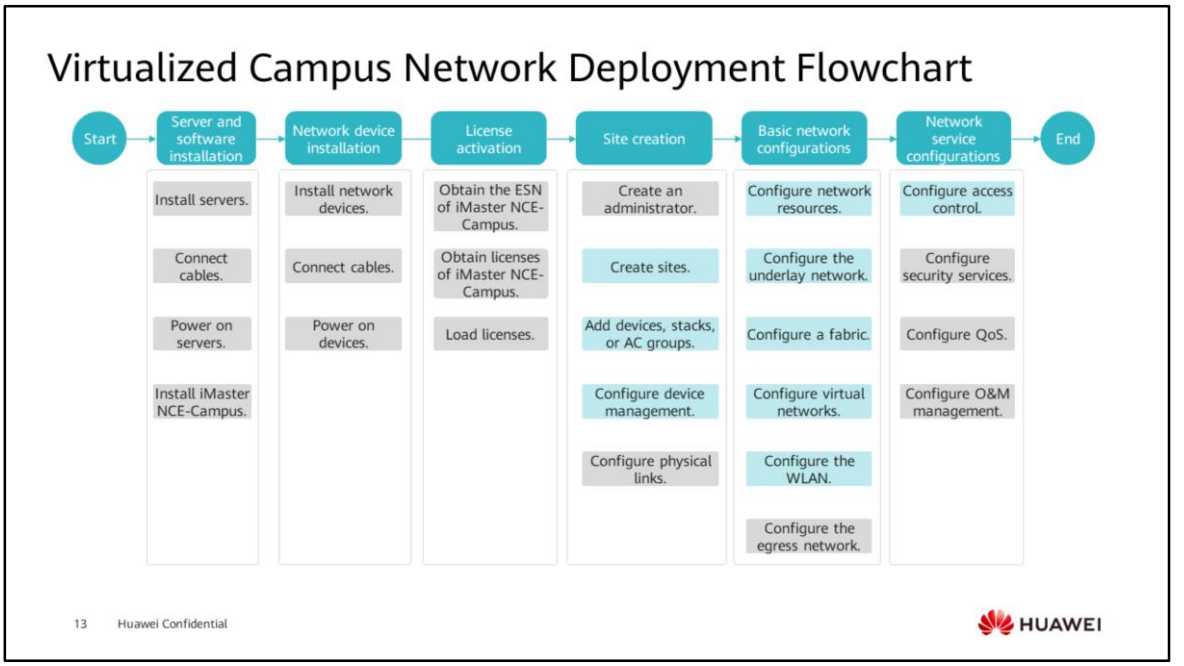
HUAWEI

- Policy Control Matrix:
  - Sales personnel and marketing personnel cannot communicate with each other. Only marketing personnel using wired terminals and R&D personnel can communicate with each other.
  - R&D personnel cannot access the E_mail resource group.
  - In the policy control matrix, only the communication that is allowed should be permitted and other communication should be denied.

Contents

1. VXLAN-based Virtualized Campus Network Deployment Plan
2. **VXLAN-based Virtualized Campus Network Deployment Process and Guide**
   - Deployment Process
   - Preparations for Deployment
   - Deployment Guide

HUAWEI



- VXLAN-based large- and medium-sized virtualized campus networks have complex services. Therefore, the deployment process is complex. The deployment process provided in this slide is the general process for you reference.
- The following part of this course focuses on key operations in the deployment process.

Contents

1. VXLAN-based Virtualized Campus Network Deployment Plan
2. **VXLAN-based Virtualized Campus Network Deployment Process and Guide**
   - Deployment Process
   - Preparations for Deployment
   - Deployment Guide

HUAWEI

Installing Servers and Software (1)

Installing iMaster NCE-Campus

- iMaster NCE-Campus can be deployed in the on-premises scenario, Huawei public cloud scenario, and MSP-owned cloud scenario.
- Large- and medium-sized campuses typically adopt the on-premises deployment mode, in which enterprises install iMaster NCE-Campus by themselves.
- For more information, see iMaster NCE-Campus Product Documentation.

Installing iMaster NCE-CampusInsight

- iMaster NCE-CampusInsight can be deployed independently or integrated with iMaster NCE-Campus.
- It is recommended that iMaster NCE-CampusInsight be integrated with iMaster NCE-Campus in the CloudCampus Solution.
- For more information, see iMaster NCE-CampusInsight Product Documentation.

HUAWEI

Installing Servers and Software (2)

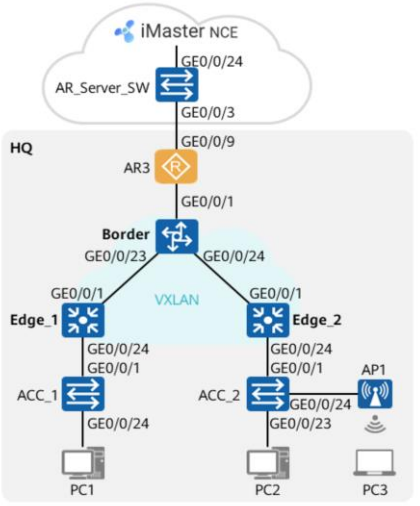
Installing the CloudCampus APP

Huawei AppGallery

- When installing APs, upload information about their actual installation locations to iMaster NCE-Campus for efficient O&M and value-added services based on terminal locations. The CloudCampus APP can record and upload AP information.
- In addition to the deployment function, the CloudCampus APP also provides the Wi-Fi experience testing, speed testing, and video testing.
- The CloudCampus APP can be obtained in either of the following methods:
  - Android system: Search for **CloudCampus** on Huawei AppGallery and install the CloudCampus APP.
  - Scan the QR code to download the CloudCampus APP.

HUAWEI

Preconfiguring Devices



Task	Preconfiguration Plan
Preconfigure AR3.	1. Complete basic configurations such as assigning VLANs, creating VLANIF interfaces, and configuring IP routing. 2. Configure the DHCP server for plug-and-play of devices at the site. 3. Configure the DHCP server for end user access. 4. Simulate the OA and RD external networks.

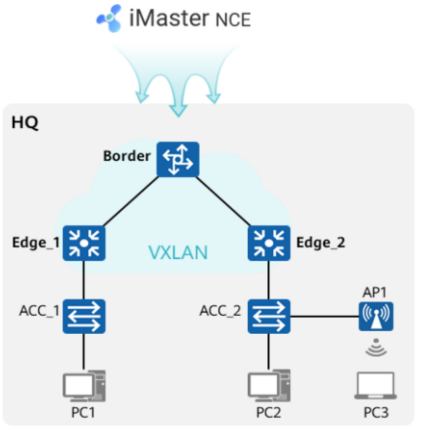
HUAWEI

Contents

1. VXLAN-based Virtualized Campus Network Deployment Plan
2. **VXLAN-based Virtualized Campus Network Deployment Process and Guide**
   - Deployment Process
   - Preparations for Deployment
   - Deployment Guide

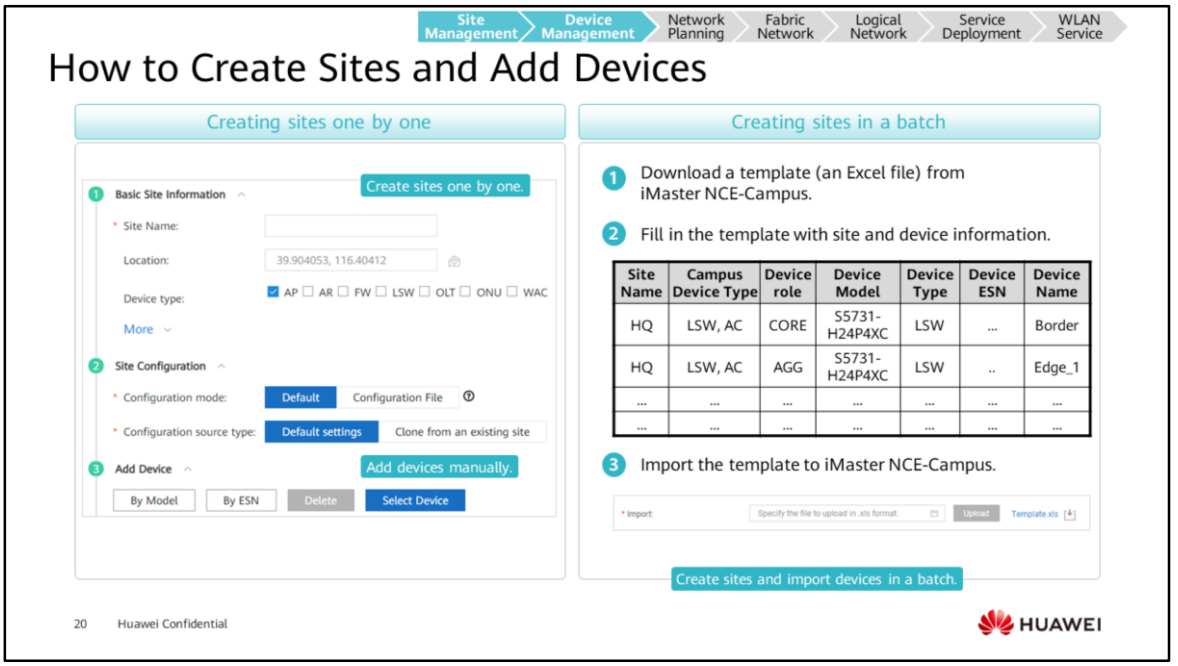
HUAWEI

Creating Sites and Adding Devices

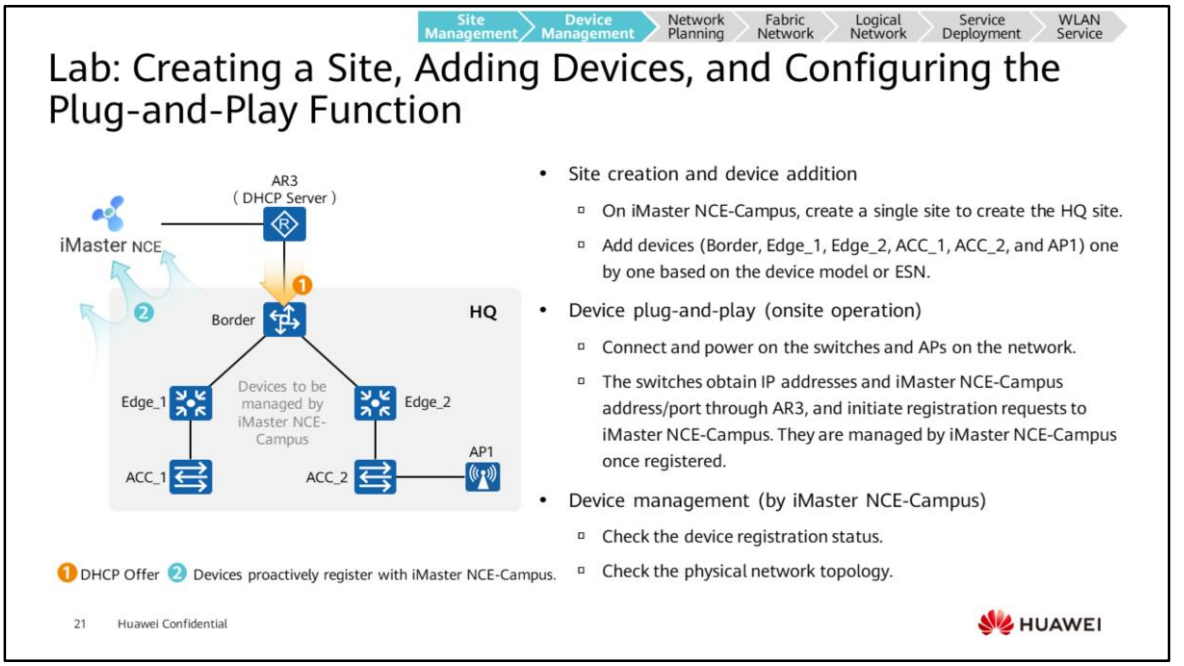


- To enable network devices within the same management scope to be centrally managed on iMaster NCE-Campus, you can create a site on iMaster NCE-Campus and add these network devices to the site.
- You can add devices when creating a site or after a site is created.

HUAWEI



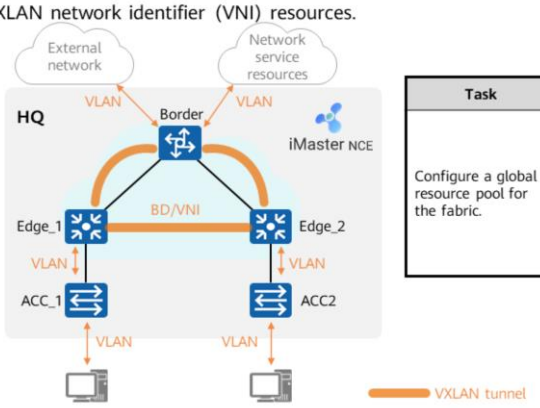
- When creating sites, pay attention to the following points:
  - To facilitate device management and improve service deployment efficiency, add devices on the same network of a tenant to the same site.
  - You can create sites on iMaster NCE-Campus for unified O&M management. Two methods are available to create sites:
    - Create sites one by one: You can create sites one by one when a small number of sites need to be added.
    - Create sites in a batch: You can create sites in a batch when a large number of sites need to be added.
- When adding devices, pay attention to the following points:
  - On a VXLAN-based virtualized campus network, managed devices are typically LSWs (switches) and ACs (wireless access controllers).
  - To add devices manually, select **By Model** or **By ESN**.
  - On large- and medium-sized virtualized campus networks, the WLAN typically adopts the architecture of "AC + Fit AP". In this architecture, Fit APs are centrally managed and configured by the native AC. After the native AC is managed by iMaster NCE-Campus, you can be redirected from iMaster NCE-Campus to the web platform of the native AC to manage Fit APs.



- Device plug-and-play:
  - Customer pain points: In traditional network deployment, engineers need to commission network devices one by one onsite, resulting in heavy configuration workload and low efficiency.
  - Solution:
    - This lab demonstrates the plug-and-play function of network devices in the CloudCampus Solution. The DHCP Option solution is used to implement plug-and-play of switches. In this example, the DHCP service and related parameters must be preconfigured on AR3.
    - In this step, switches on the campus network can be directly deployed using factory settings and get managed by iMaster NCE-Campus, greatly reducing the configuration workload.

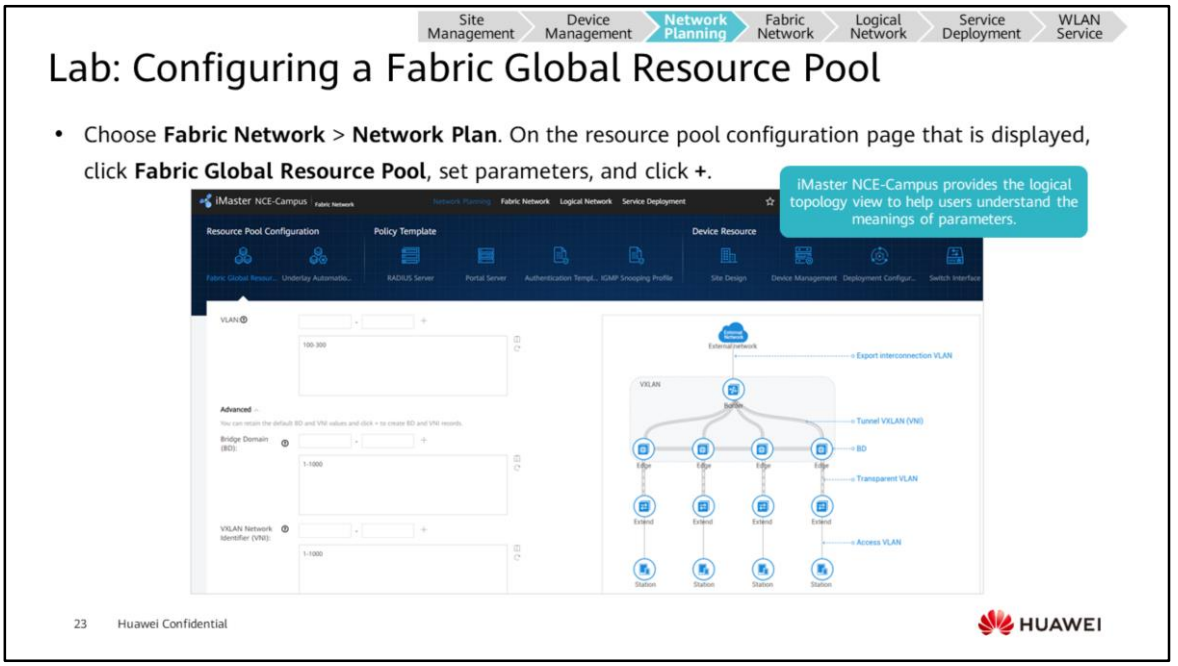
Configuring Resource Pools: Global Resource Pool for the Fabric

- Before creating a fabric, plan a global resource pool for the fabric, which contains VLAN, bridge domain (BD), VXLAN network identifier (VNI) resources.



Task	Resource Pool Plan
Configure a global resource pool for the fabric.	• Configure VLAN resources, including service VLANs (access VLANs for end users) used during virtual network creation, VLANs whose packets are transparently transmitted between the access and aggregation devices, interconnection VLANs used when external networks and network service resources are created for the fabric, and management VLAN for policy association. • Configure BDs and VNIs for Layer 2 broadcast domain isolation in virtual networks.

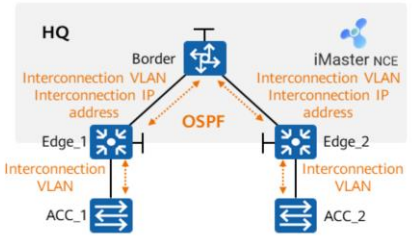
HUAWEI



- Parameters in the fabric global resource pool:
  - **VLAN:** Configure a service VLAN pool when you need to configure VLANs for connecting to external networks, VLANs for connecting to network service resources, CAPWAP management VLAN, and terminal access VLANs.
  - **Bridge Domain (BD):** On a VXLAN network, VNIs can be mapped to BDs in 1:1 mode so that a BD can function as a VXLAN network entity to transmit traffic. A BD is a Layer 2 broadcast domain used to forward data packets on a VXLAN network.
  - **VXLAN Network Identifier (VNI):** A VNI is similar to a VLAN ID and identifies a VXLAN.

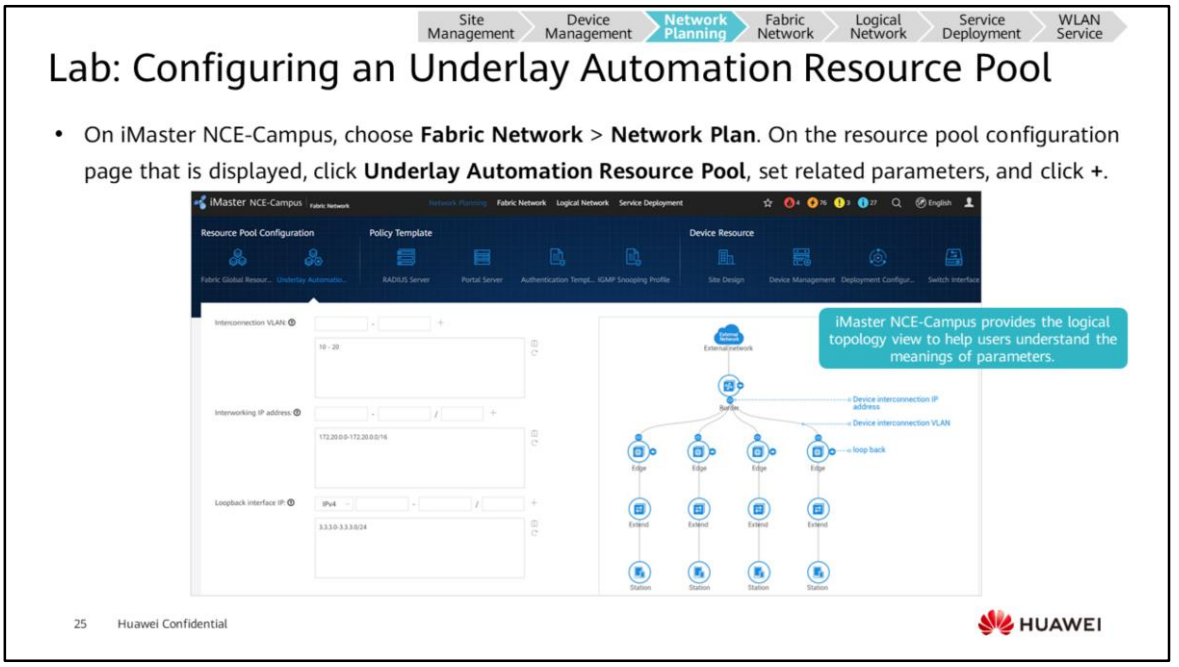
Configuring Resource Pools: Underlay Automation Resource Pool

- The underlay automation resource pool defines the interconnection VLANs, interconnection IP addresses, and loopback interface IP addresses used for routing domain orchestration on the underlay network.



Task	Resource Pool Plan
Configuring an underlay automation resource pool.	1. Configure interconnection VLANs, which cannot conflict with the fabric global resource pool and service VLANs for user access. 2. Configure interconnection IP addresses, which cannot conflict with the fabric global resource pool and service IP addresses for user access. 3. Configure loopback interface IP addresses. As the VXLAN control plane uses BGP EVPN, loopback interface IP addresses are required for establishing BGP EVPN peer relationships between border and edge nodes. The preceding resources are used for automatic orchestration of routing domains on the underlay network.

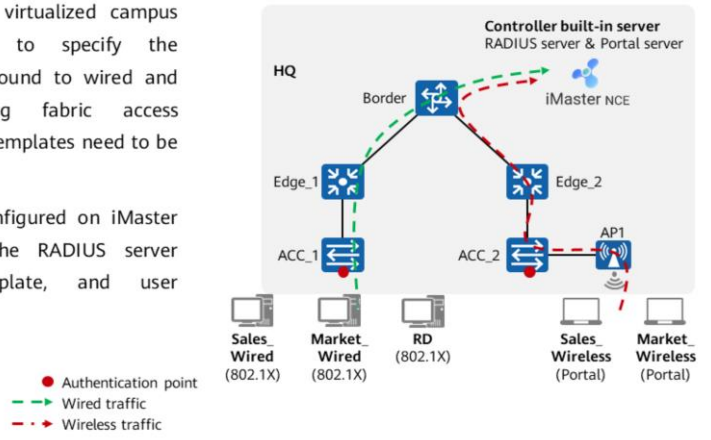
HUAWEI



- Parameters for configuring the underlay automation resource pool:
  - Interconnection VLAN: Configure an interconnection VLAN pool when border and edge nodes on a fabric need to communicate on the underlay network.
  - Interworking IP address: Configure an interconnection IP address pool when border and edge nodes on a fabric need to communicate on the underlay network.
  - Loopback interface IP: If automatic routing domain configuration is required on a fabric, configure an IP address pool for the loopback interfaces.

Configuring Policy Template Resources: User Access Authentication

- In the large- and medium-sized virtualized campus network solution, you need to specify the authentication templates to be bound to wired and wireless access points during fabric access management. The authentication templates need to be planned.
- The policy template resources configured on iMaster NCE-Campus typically include the RADIUS server template, Portal server template, and user authentication template.



HUAWEI

Lab: Configuring User Authentication Templates

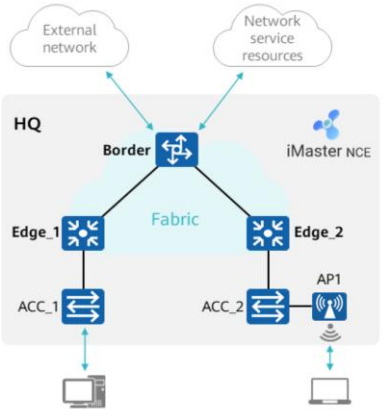
- On iMaster NCE-Campus, choose **Fabric Network** > **Network Planning** > **Policy Template**, and create and configure a RADIUS server template, a Portal server template, and user authentication templates.

Task	Procedure
Create a RADIUS server template.	1. Select a RADIUS server type. The iMaster NCE-Campus built-in RADIUS server is recommended. 2. Configure the RADIUS server IP address. If the iMaster NCE-Campus built-in RADIUS server is used, you do not need to configure the RADIUS server IP address.
Create a Portal server template.	1. Select a Portal server type. The iMaster NCE-Campus built-in Portal server is recommended. 2. Configure the Portal server IP address. If the iMaster NCE-Campus built-in Portal server is used, you do not need to configure the Portal server IP address. 3. Configure the Portal server URL. If the iMaster NCE-Campus built-in Portal server is used, you do not need to configure the Portal server URL.
Create user authentication templates.	1. Select a user authentication mode, which is typically 802.1X authentication, MAC address authentication, or Portal authentication. A template can contain multiple authentication modes. 2. Select the RADIUS server template or Portal server template to be bound.

HUAWEI

Configuring a Fabric

- A fabric is a virtual network built on top of a physical network using the VXLAN technology, and has all resources pooled.



Task	Procedure
Configure a fabric.	1. Create a fabric and complete automatic deployment of an underlay network. 2. Create an external network, that is, configure the connectivity between the fabric and external networks (including the interconnection port, interconnection VLAN, interconnection IP address, peer address, and route). 3. Configure network service resources, that is, configure the connectivity between the fabric and network service resources (DHCP server, RADIUS server, Portal server, and other servers). The connectivity configuration includes the device IP address, interconnection VLAN, interconnection IP address, peer IP address, and interconnection port. 4. Configure access management, that is, configure an interface on an access device to invoke an authentication template.

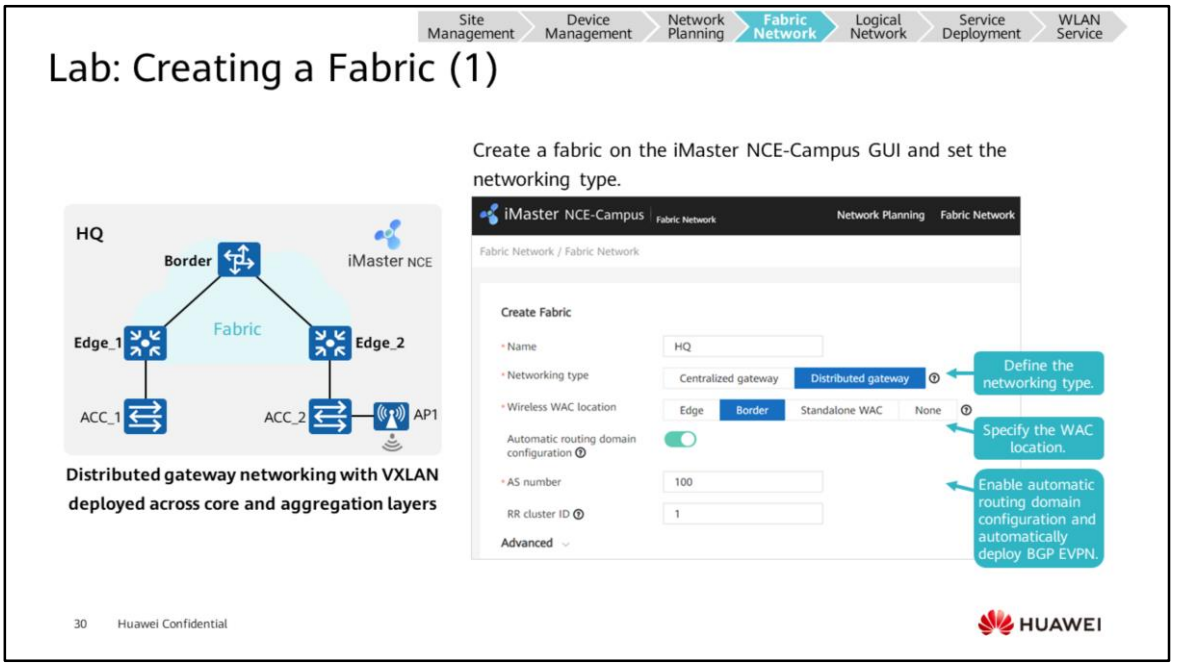
HUAWEI

Configuring a Fabric: Creating a Fabric

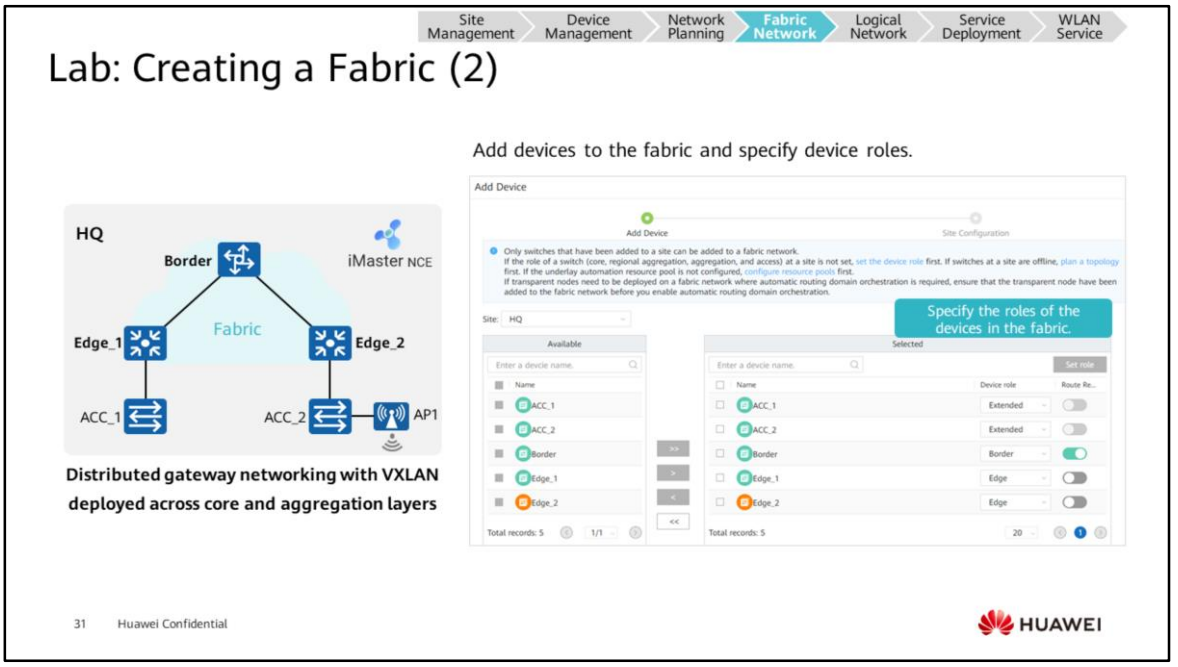
- You need to create a fabric and then complete basic networking configurations for the fabric, including selecting border and edge nodes as well as configuring BGP EVPN of the VXLAN control plane.
- In addition, you need to enable automatic routing domain configuration of the underlay network when creating a fabric.

Task	Procedure
Create a fabric.	1. Select a networking type: centralized or distributed gateway. 2. Select the WAC location: edge node, border node, standalone WAC, or None. 3. Enable the automatic routing domain configuration function of the underlay network. 4. Configure an AS number, that is, configure BGP EVPN and set an RR cluster ID. 5. (Optional) Configure storm suppression on the fabric. To mitigate the risks caused by network storms, you are advised to enable storm suppression. 6. (Optional) Enable the function of reporting terminal identification information. You need to enable this function when the terminal identification function is configured. 7. Add devices to the fabric and specify the devices as border and edge nodes. If VXLAN is deployed across core and aggregation layers on the fabric, you need to configure access switches as extended nodes.

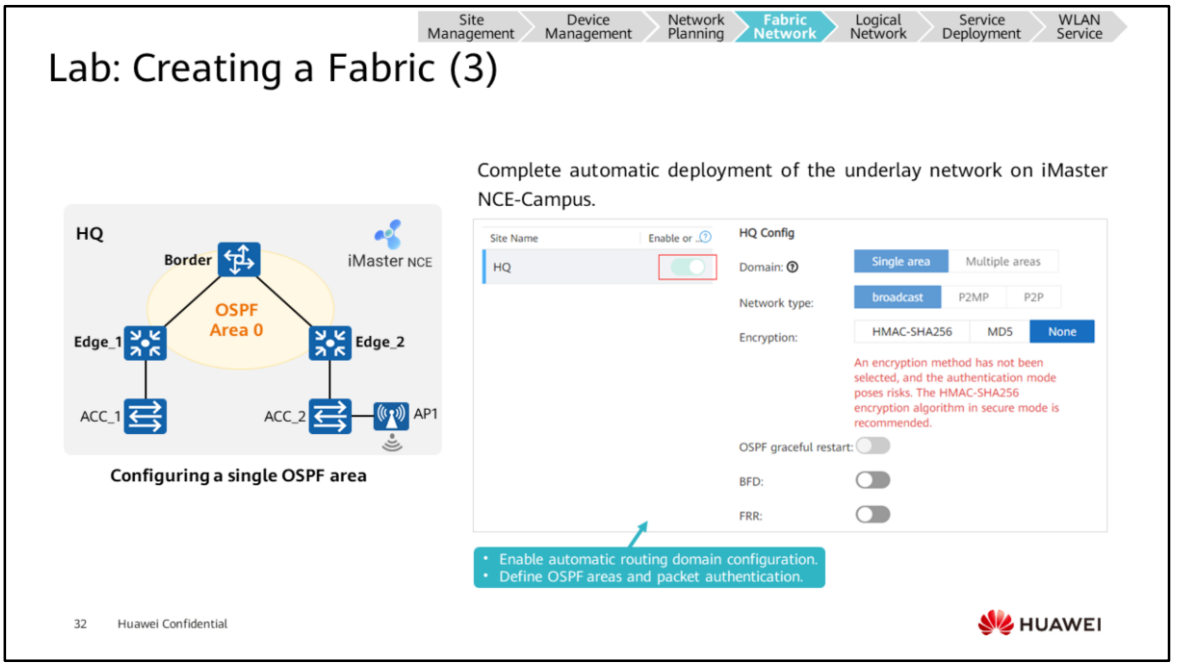
HUAWEI



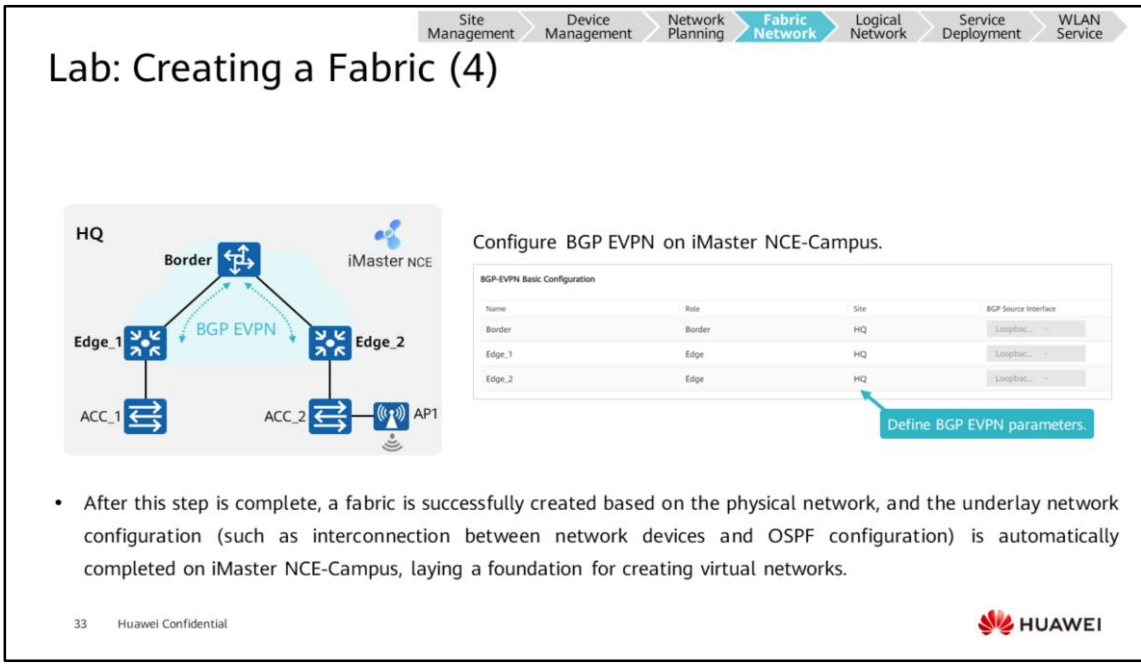
- A fabric consists of core, aggregation, and access devices. An access device supports access of different network services, reducing costs while improving network device utilization.
- A virtualized campus network uses the overlay virtualization technology (VXLAN) to create multiple virtual networks on the same fabric and allow flexible service deployment.
- The networking type specifies the network deployment mode of a fabric:
  - Centralized: The fabric uses a centralized gateway, which transmits traffic accessing external networks and the intranet. In this networking mode, only the border node can function as the centralized gateway.
  - Distributed: The fabric uses distributed gateways. Traffic accessing external networks and the intranet passes through different gateways. In this networking mode, both border nodes and edge nodes can function as gateways.
- iMaster NCE-Campus allows you to specify a BGP route reflector (RR) and automatically deploy BGP EVPN configurations on all edge nodes in a fabric.
  - AS number: specifies the BGP AS number used in a fabric.
  - RR cluster ID: specifies the cluster ID of an RR. If there are multiple RRs in a fabric, for example, if two RRs are configured on a dual-border network, you need to configure a cluster ID for the RRs to prevent BGP routing loops. The value is an integer ranging from 1 to 4294967295 or an IPv4 address.



- Role: specifies the role of a device in the fabric, including the border node, edge node, and extended node. By default, the role of a device is an extended node.
- Route reflector: In a fabric, border devices are typically used as route reflectors, which simplify full-mesh connections required by IBGP and reduce network and CPU loads.



- Automatic routing domain configuration: After this function is enabled, the underlay network is automatically configured. You can specify sites for automatic routing domain configuration and specify OSPF route parameters. Currently, the following parameters are supported:
  - **Domain:** In single-domain mode, all devices belong to area 0. In multi-domain mode, all edge nodes are located in their respective non-backbone areas, and border nodes are interconnected through the backbone area (area 0).
  - **Network type:** You can specify the OSPF network type to broadcast, P2MP, or P2P.
  - **Encryption:** You can set the encryption mode between adjacent devices to HMAC-SHA256, MD5, or None.
  - **Key:** It refers to the authentication key ID used for ciphertext authentication on an interface, and must be consistent with that of the peer device. The value is an integer in the range from 1 to 255.
  - **Password:** It specifies the ciphertext authentication key. The value is a string of 1 to 255 characters and cannot contain spaces.
  - **Confirm password:** You need to enter the ciphertext authentication key again for confirmation.
  - **OSPF graceful:** You can enable OSPF GR.



- After a fabric is created, the underlay network can be automatically deployed, and configurations such as VLANIF interfaces, loopback interfaces, VTEP IP addresses, and routes required for establishing a BGP EVPN can be automatically provisioned. iMaster NCE-Campus allocates resources from the fabric global resource pool and underlay automation resource pool to devices.

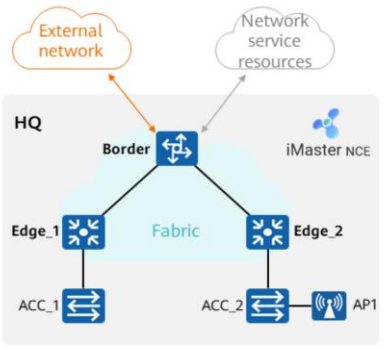
Configuring a Fabric: Creating an External Network

- In the fabric resource model design, external networks are created on border nodes so that terminals on the campus network can access external networks such as the Internet.
- Three types of external network resources are defined: Layer 3 shared egress, Layer 3 exclusive egress, and Layer 2 shared egress. If the user gateway is located in the fabric, the Layer 3 shared egress or Layer 3 exclusive egress is used.
  - Layer 3 shared egress: Multiple virtual networks on the fabric share a Layer 3 egress to communicate with the egress device. The Layer 3 shared egress helps save interconnection VLAN and IP resources and applies to scenarios where there are low requirements on security control policies between virtual networks.
  - Layer 3 exclusive egress: Each virtual network on the fabric exclusively uses a Layer 3 egress to communicate with the egress device. In this case, multiple security zones are typically configured on the firewall, each corresponding to one Layer 3 exclusive egress. The traffic of different virtual networks to the firewall is isolated from each other.

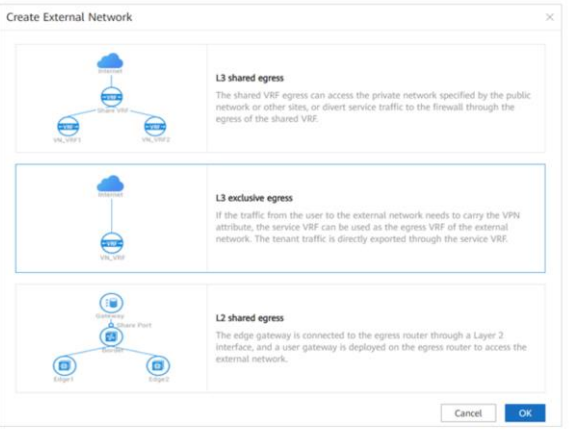
Task	Procedure
Create an external network.	1. Select the type of the connection with the external network. 2. Select an external network resource type. 3. Configure interconnection physical interfaces, VLANs, and IP addresses. 4. Configure a routing protocol for interconnection.

HUAWEI

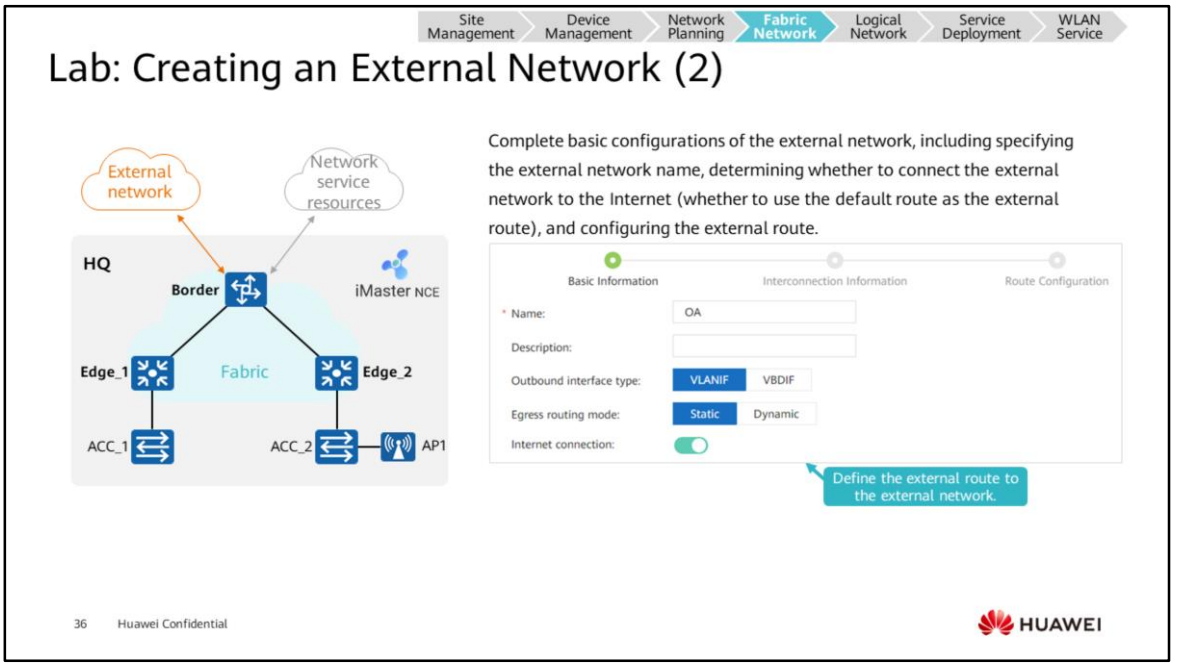
Lab: Creating an External Network (1)



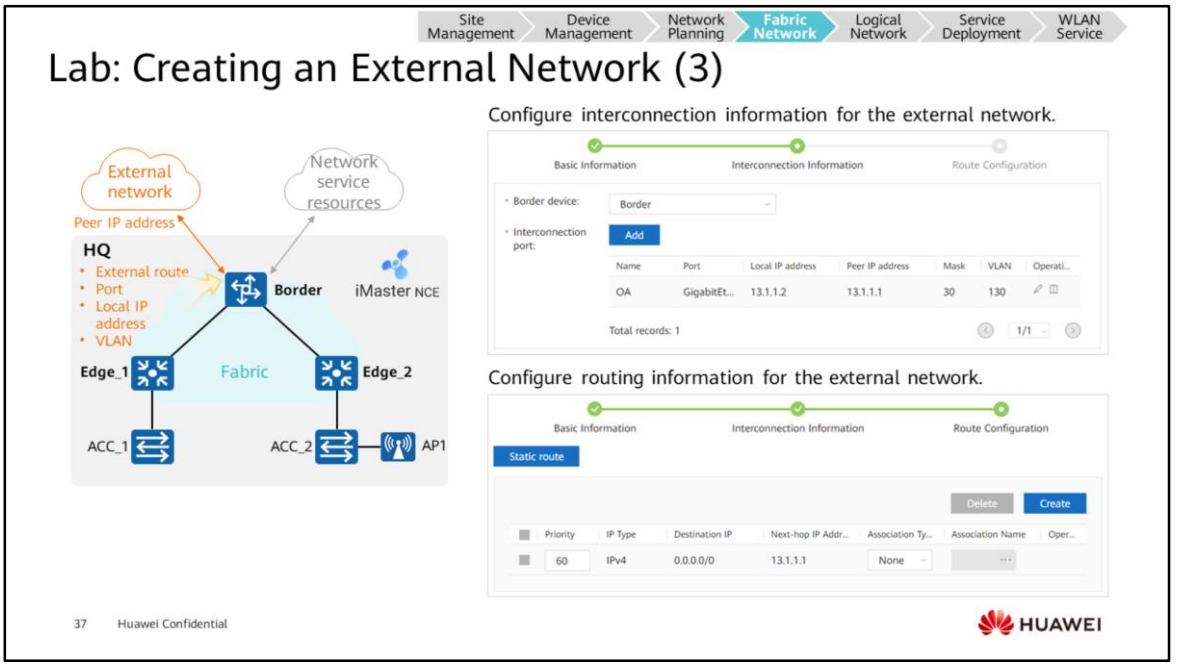
Create an external network for the fabric and select the egress connection mode.



HUAWEI



- Basic information:
  - If **Internet connection** is enabled, iMaster NCE-Campus uses a default route to direct traffic to the corresponding external network. If **Internet connection** is disabled, you need to specify a route prefix.



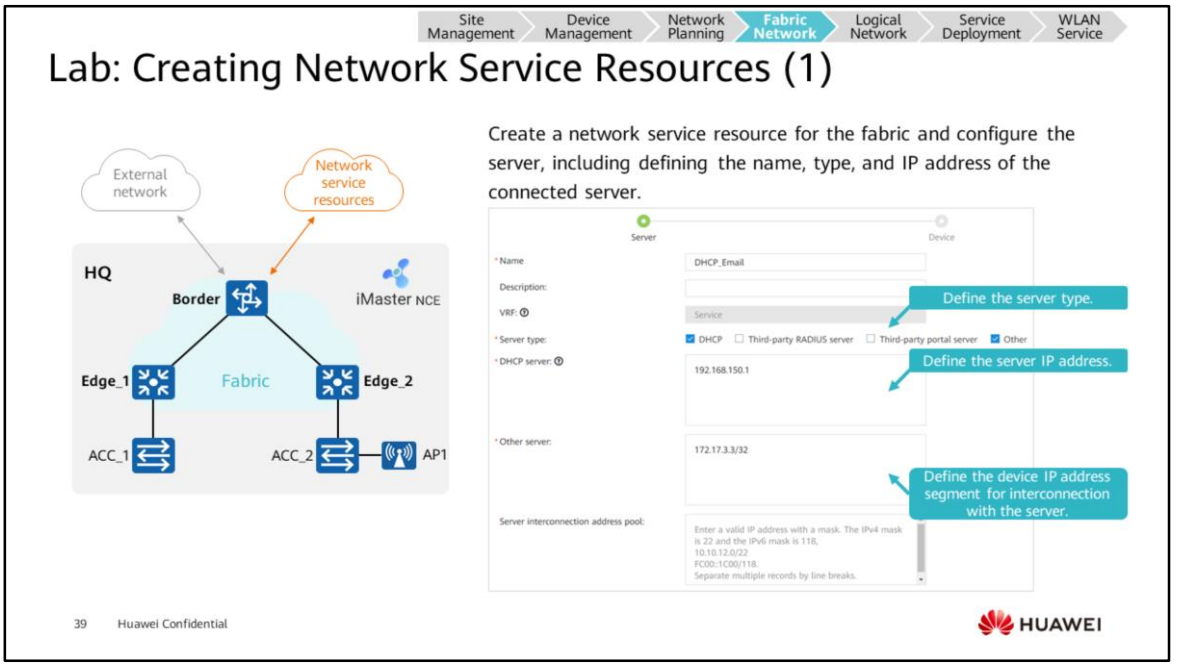
- Interconnection information:
  - Select the border device to be connected to the external network, and select the interconnection interface, interconnection IP address, and interconnection VLAN.
  - Note: The configured interconnection IP address cannot conflict with IP addresses in the underlay automation resource pool. The configured interconnection VLAN belongs to the global resource pool of the fabric.
- Routing information:
  - After you click **Apply**, iMaster NCE-Campus creates a static route to the external network for the border device. (The static route is delivered to the border device only when the external network is invoked by the virtual network.)

Configuring a Fabric: Creating Network Service Resources

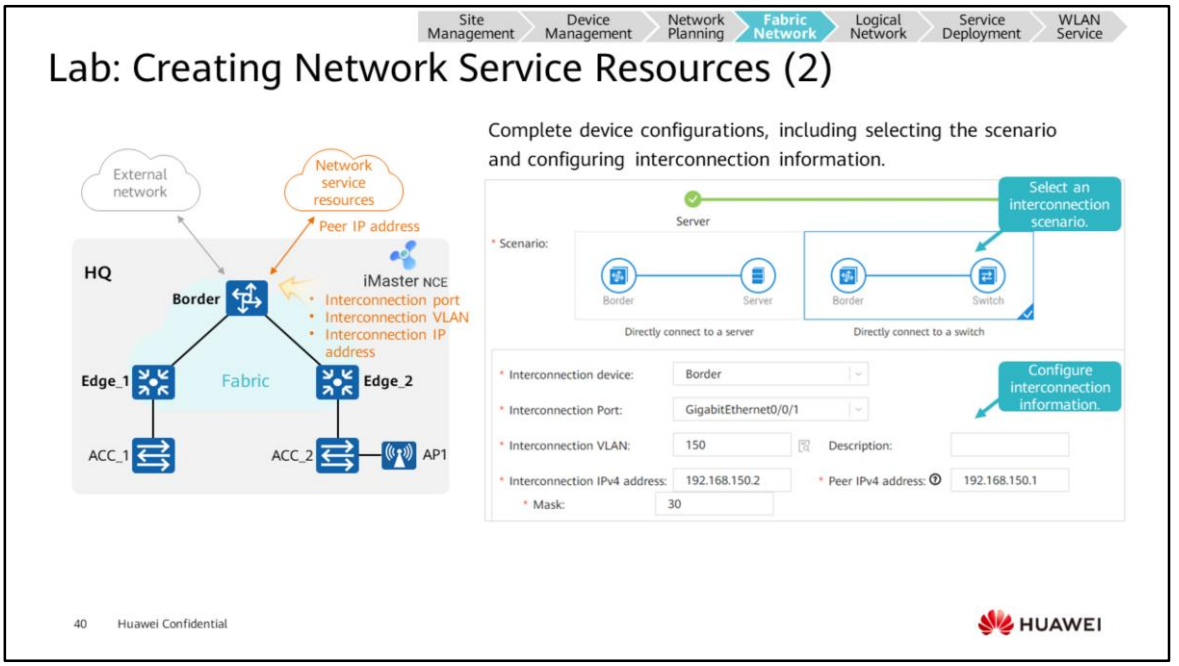
- In the fabric resource model design, network service resources are created on a border node so that service terminals on the campus network can access service resources, such as the DHCP server and NAC server, in the network management zone.
- You can create multiple network service resources or add IP addresses for accessing multiple network service resources to a network service resource model.
- If only a few service resources in the network management zone need to be accessed, you are advised to plan these resources in the same network service resource model. This saves interconnection VLANs and IP addresses and simplifies routing configuration in the network management zone.

Task	Procedure
Create network service resources.	1. Select a server type, which can be DHCP server, third-party RADIUS server, and third-party Portal server. 2. Configure the IP addresses for accessing network service resources, such as the DHCP service address and iMaster NCE-Campus southbound IP address. 3. Select an interconnection scenario, which can be directly connected to a server or directly connected to a switch. 4. Configure interconnection physical interfaces, VLANs, and IP addresses.

HUAWEI



- Server configuration:
  - **Server type:** server resource specified by the network service resource. In this example, the defined network service resource are a DHCP server and a E-mail server.
  - **DHCP server:** IPv4 and/or IPv6 address of a DHCP server.
  - **Server interconnection address pool:** If an IPv6 address pool is configured for the DHCP server, this parameter is mandatory. If only an IPv4 address pool is configured for the DHCP server, this parameter is optional.

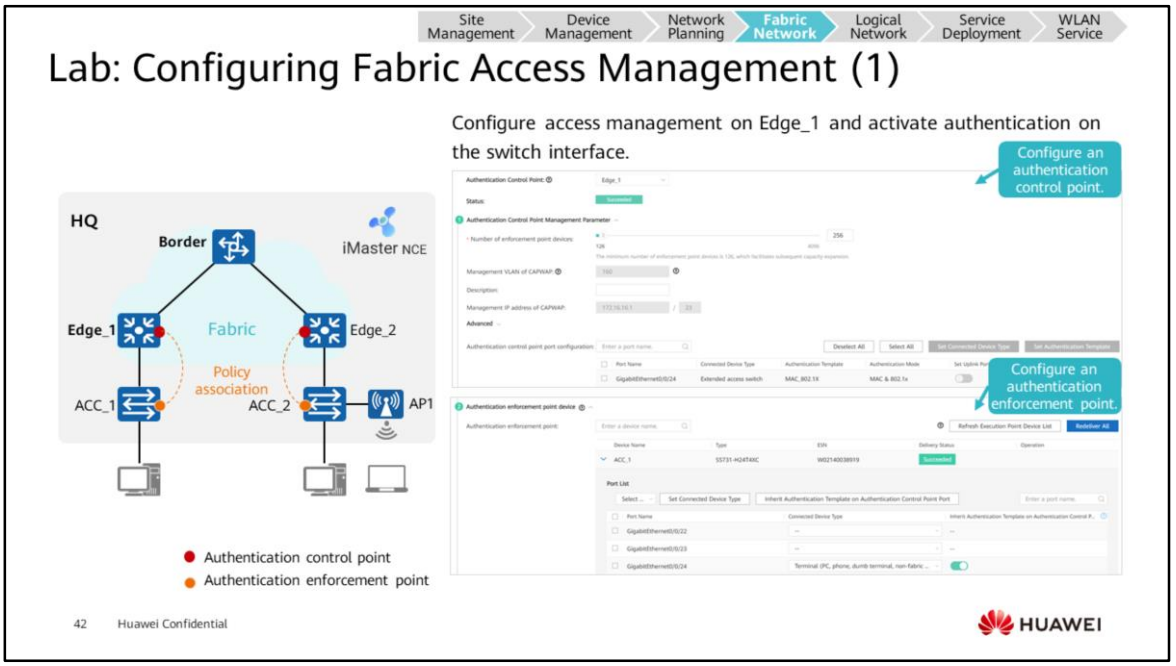


- Device configuration:
  - If **Directly connected to a switch** is selected, the border node adds the interconnection port to the interconnection VLAN in tagged mode to connect to network resources. If **Directly connected to a server** is selected, the switch adds the interconnection port to the interconnection VLAN in untagged mode.
  - **Interconnection device:** Select the border node (Border). The device functions as a border node that connects the fabric to the external network service resource.
  - **Interconnection port:** Select the port used by the border node to connect to the network service resource.
  - **Interconnection VLAN:** Select the VLAN used by the border node to connect to the network service resource.
  - **Interconnection IPv4 address:** Select the IP address of the port used by the border node to connect to the network service resource.
  - **Peer IPv4 address:** Select the IP address of the peer device.
  - **Mask:** Select the network mask.
- Note: The configured interconnection IP address cannot conflict with the IP address in the underlay automation resource pool. The configured interconnection VLAN belongs to the fabric global resource pool.

Configuring a Fabric: Configuring Access Management

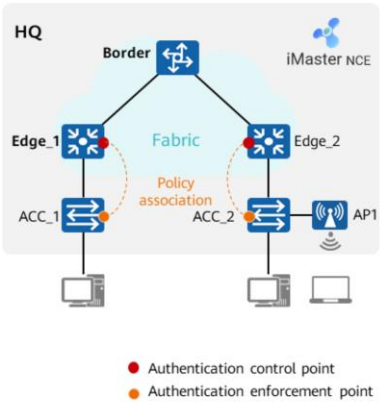
- Access control needs to be deployed on the campus network to authenticate access users, including 802.1X authentication, MAC address authentication, and Portal authentication. The implementation of 802.1X authentication and MAC address authentication requires the use of the RADIUS server, whereas the implementation of Portal authentication requires the use of the RADIUS server and Portal server.
- Configuring access management for the fabric is to configure authentication control points and plan access point resources, which will be used during virtual network creation.
- Wired access point resources refer to switch interfaces connected to terminals, and wireless access point resources refer to SSIDs connected to terminals.
  - During access management configuration for the fabric, three connection types are defined for access ports of switches:
    - **Extended AP**: Huawei Fit APs, which can be managed by iMaster NCE-Campus.
    - **Extended access switch**: Huawei switches, which can be managed by iMaster NCE-Campus.
    - **Terminal (PCs, phones, dumb terminals, and non-fabric extended access switches or APs):** user terminals, switches and APs that cannot be managed by iMaster NCE-Campus, and switches that can be managed by iMaster NCE-Campus but do not support fabric extension.
- Access management configuration for the fabric varies depending on the gateway solution.

HUAWEI

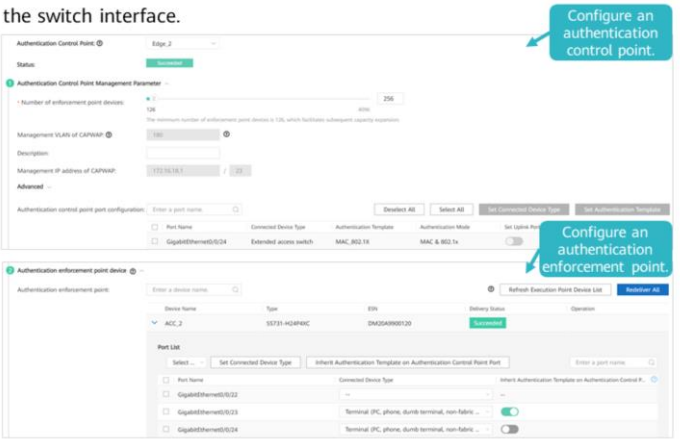


- Policy association:
  - Configure the management VLAN and management IP address for policy association.
  - The configured management IP address cannot conflict with the IP addresses in the underlay automation resource pool. The configured management VLAN belongs to the global resource pool of the fabric.

Lab: Configuring Fabric Access Management (2)



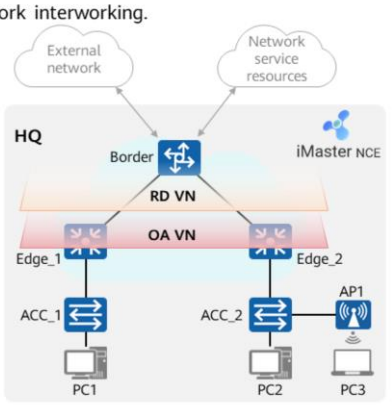
Configure access management on Edge_2 and activate authentication on the switch interface.



HUAWEI

Logical Network

- After a fabric is created, you can select network resources in the fabric to create virtual network instances.
- To enable access users in different virtual networks to communicate with each other, you need to configure virtual network interworking.



Task	Procedure
Logical network	1. Create virtual networks. Different virtual networks represent services of different users and can be used to isolate services of these users. 2. Configure virtual network interworking to ensure that there are reachable underlay network routes between users in different virtual networks.

VN	Access User
OA	Sales_Wired: 172.17.10.0/24 Sales_Wireless: 172.17.11.0/24 Market_Wired: 172.17.20.0/24 Market_Wireless: 172.17.21.0/24
RD	RD: 172.17.30.0/24

HUAWEI

Fabric Resource Pools and Resource Invoking During Virtual Network Creation

Fabric Resource Pool	How to Invoke Resources in the Resource Pool During Virtual Network Creation
VLAN resource pool, which is used in scenarios where terminals are connected to virtual networks and virtual networks communicate with external networks. The VLAN resource pool is planned during the configuration of the fabric global resource pool.	When creating a user gateway in a virtual network, select a resource from the fabric global resource pool to configure a user VLAN.
BD and VNI resource pools, which are used to divide Layer 2 broadcast domains in a virtual network and configure VBDIF interfaces as the gateway interfaces of user subnets. The BD and VNI resource pools are planned during the configuration of the fabric global resource pool.	When a user gateway is created in a virtual network, resources in the BD and VNI resource pools are automatically invoked to create a BD and the corresponding VBDIF interface.
User access point resource pool, which is planned during fabric access management configuration. This resource pool includes the authentication modes that can be configured on access points.	When configuring user access in a virtual network, you can select planned access point resources.
Egress pool, which contains the external resources that can be used by virtual networks. Two types of external resources are created during fabric configuration: • External networks: used for virtual networks to communicate externally • Network service resources: used for virtual networks to communicate with the authentication server and DHCP server, etc.	When creating a virtual network, you can select external networks and network service resources.

HUAWEI

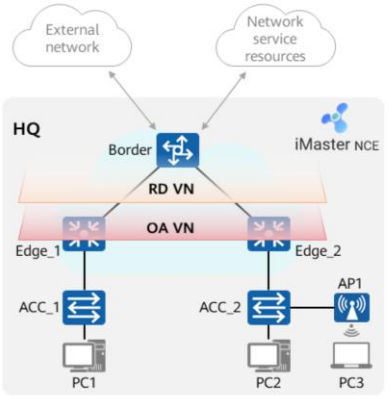
Creating a Virtual Network

- Creating a virtual network is to create a VPN instance to isolate different services.

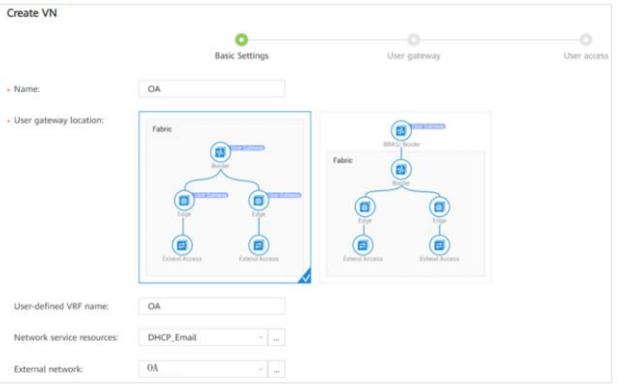
Task	Procedure
Create a virtual network.	1. (Optional) Select network service resources for communication between the virtual network and network service resources (this step is typically performed). 2. (Optional) Select an external network for communication between the virtual network and external network (this step is typically performed). 3. Configure a user gateway, which can be manually specified or automatically allocated. 4. Configure wired access: Select the access ports configured during fabric access management configuration, and add the ports to the service VLAN configured on the user gateway. 5. Configure wireless access: Select the AC that connects to the wireless user subnet. Then the service VLAN configured on the user gateway will be delivered to the AC.

HUAWEI

Lab: Creating a Virtual Network (1)

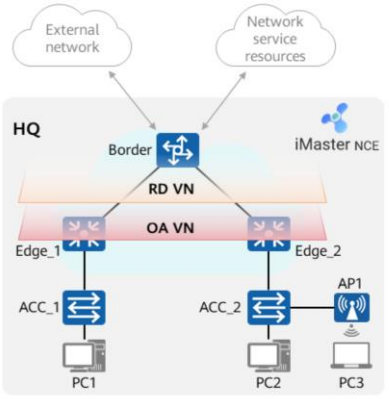


Create a virtual network, define the virtual network name and user gateway location, and configure the external network and network service resources that the virtual network can access.

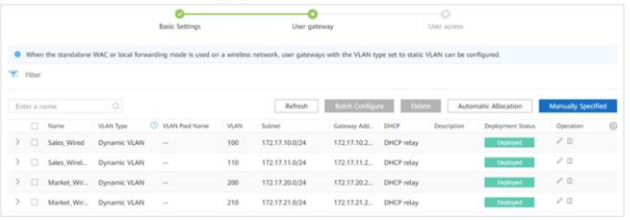


HUAWEI

Lab: Creating a Virtual Network (2)

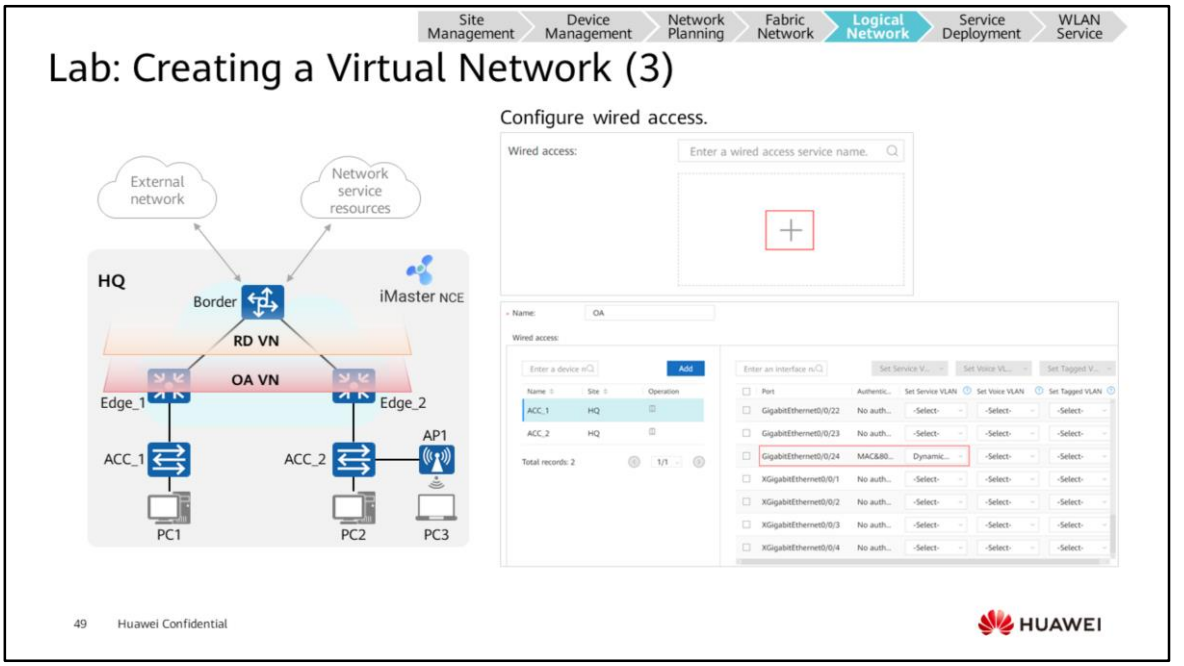


Define user gateways for the virtual network. User gateways can be manually created one by one by the network administrator or created in a batch through automatic allocation.

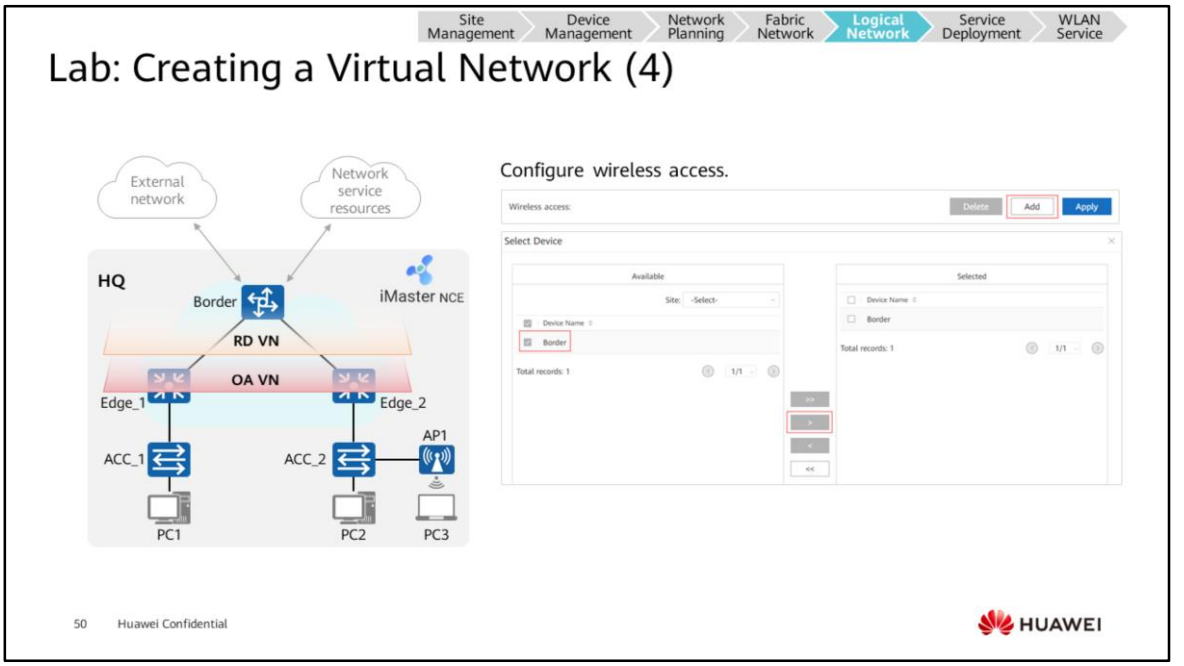


VN	Access User
OA	Sales_Wired: 172.17.10.0/24 Sales_Wireless: 172.17.11.0/24 Market_Wired: 172.17.20.0/24 Market_Wireless: 172.17.21.0/24
RD	RD: 172.17.30.0/24

HUAWEI



- User access is configured on the access points of the current virtual network.
- Wired access: Users on the OA virtual network need to access the network through ACC_1 and ACC_2 and they all need to be authenticated. Therefore, in the wired access configuration, you need to select the interfaces on which authentication has been enabled on the two switches.



- Wireless access: Select the border device in the wireless access configuration. The border device is a switch that provides the native AC, through which it manages APs and provides the wireless access service.

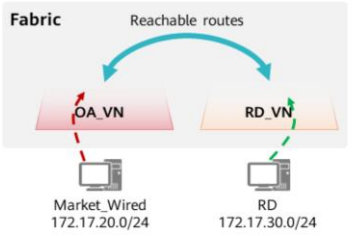
Communication Between Virtual Networks

- By default, virtual networks are isolated from each other. However, in some scenarios, they may need to communicate with each other, so reachable routes must be configured between them.
- Configuring communication between virtual networks is to create static routes for VPN instances so that access users in different virtual networks can communicate with each other.

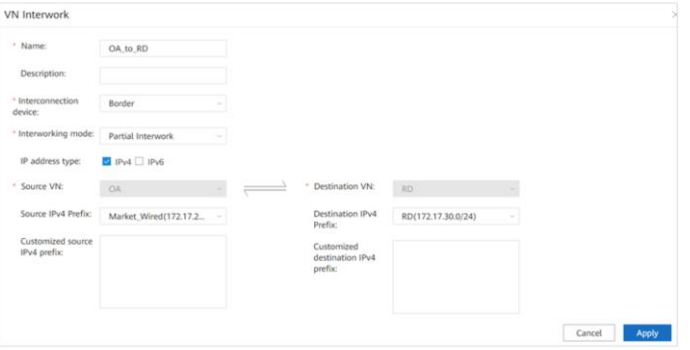
Task	Procedure
Configure communication between virtual networks.	1. Select interconnection devices. 2. Select the interworking mode, including full interworking and partial interworking. 3. Configure the source virtual network, source IPv4 prefix, destination virtual network, and destination IPv4 prefix.

HUAWEI

Lab: Configuring Virtual Network Interworking



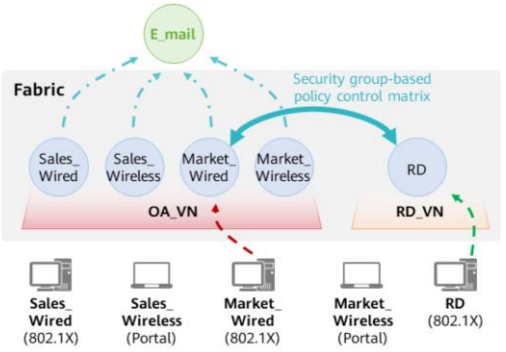
Access the **VN Interwork** page, and click **Add** to configure interworking between virtual networks.



HUAWEI

Service Deployment

- On large campus networks, users are usually allowed access from any location, any VLAN, and any IP network segment with controlled network access rights.

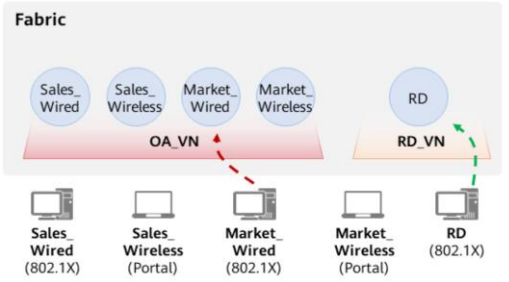


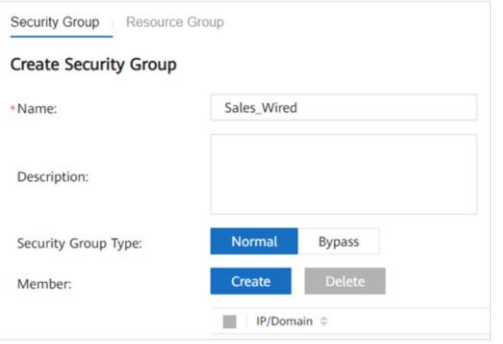
Task	Procedure
Service deployment	1. Configure free mobility, including creating security groups, resource groups, and a policy control matrix. 2. Configure access authentication, including creating user authentication accounts, authentication rules, authorization results, and authorization rules, as well as page management.

HUAWEI

Lab: Creating a Security Group

- A security group is an entity unit for permission control. Users are assigned to different security groups, and access permissions are configured between security groups to implement user permission management on the network.

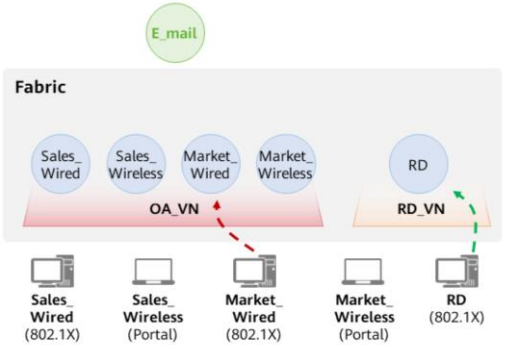


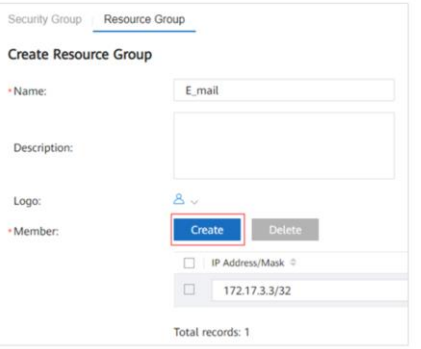


HUAWEI

Lab: Creating a Resource Group

- Different network service resources can be allocated to different resource groups. You can configure access permissions from security groups (source) to resource groups (destination) to manage the access to network service resources.

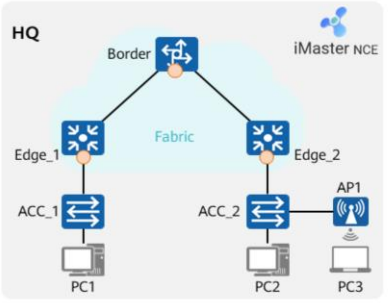


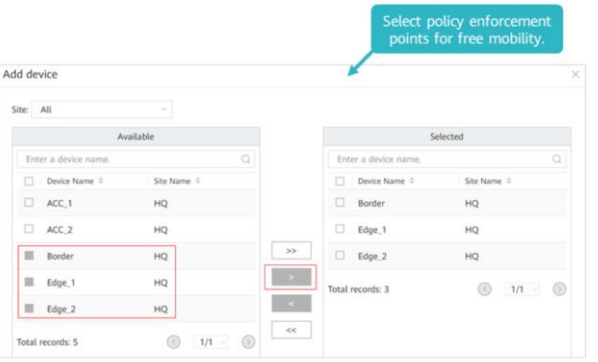


HUAWEI

Lab: Creating a Policy Control Matrix (1)

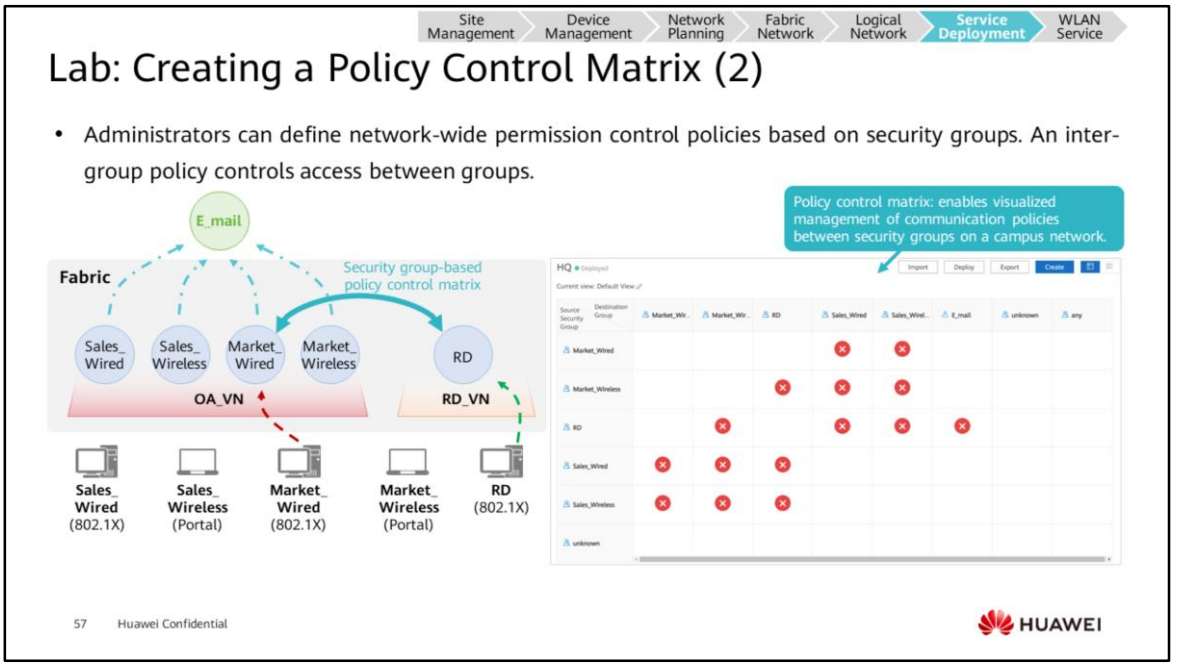
- Administrators can define network-wide permission control policies based on security groups. An inter-group policy controls access between groups.





Policy enforcement point

HUAWEI



- When multiple policies are configured to control access from a source security group to multiple destination groups, the sequence in which these policies are matched can be determined based on the policy priority. For example, if the destination groups are resource groups, in which case the destination IP addresses may be the same, you need to manually adjust the policy priorities to ensure that a specific policy is matched first.

Access Authentication

802.1X authentication

- 802.1X is a port-based network access control protocol. It verifies user identities and controls access permissions of users on ports of LAN access devices.
- When iMaster NCE-Campus functions as a RADIUS server, 802.1X authentication configuration on the server is illustrated in the following table:

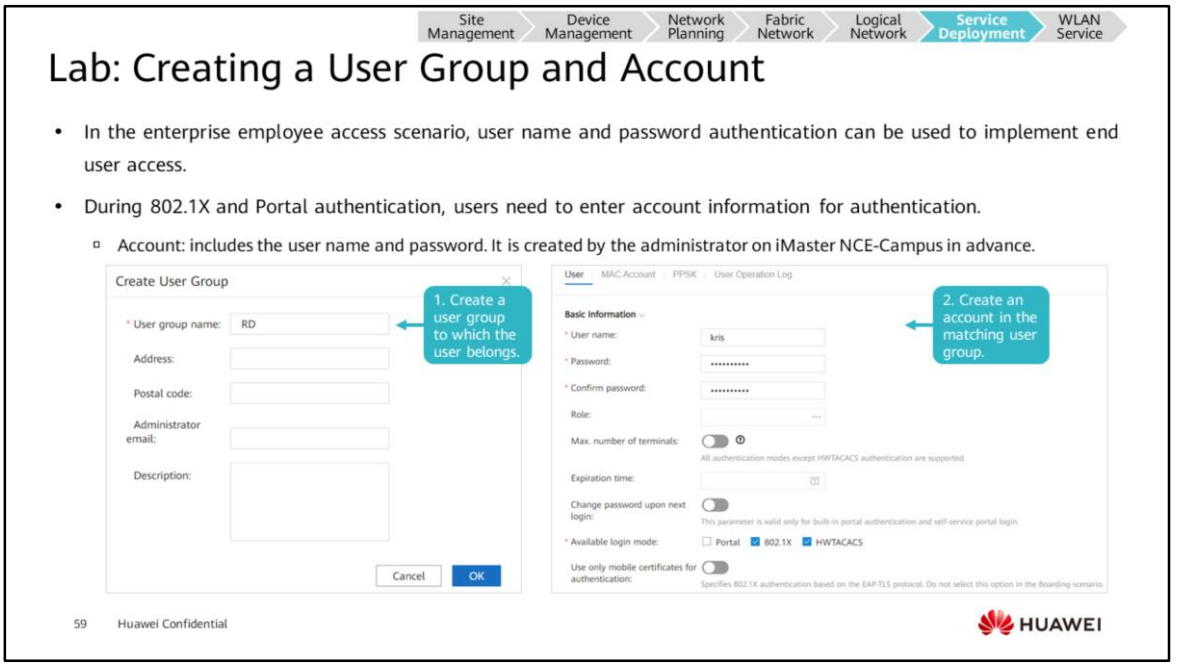
Task	Procedure
Configure 802.1X authentication.	1. Create user authentication accounts. 2. Configure authentication rules. 3. Configure authorization results and authorization rules.

Portal authentication

- Portal authentication is also known as web authentication, and websites for Portal authentication are referred to as web portals. When a user accesses the network, the user must be authenticated on the web portal. If the user fails the authentication, the user can access only specified network resources. The user can access other network resources only after being authenticated successfully.
- When iMaster NCE-Campus functions as the Portal server and RADIUS server, Portal authentication configuration on the server is illustrated in the following table:

Task	Procedure
Configure Portal authentication.	1. Create user authentication accounts. 2. Configure authentication rules. 3. Configure authorization results and authorization rules.
(Optional) Configure a MAC authentication exemption policy.	To enable a user client to be exempt from authentication based on its MAC address within a period after the client passes Portal authentication for the first time, configure a MAC authentication exemption policy in the user control policy.

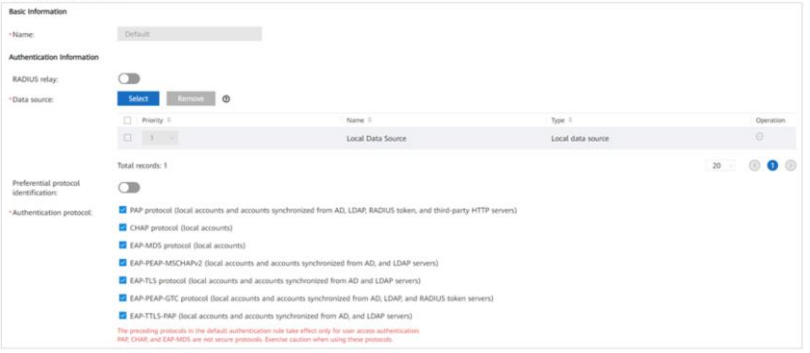
HUAWEI



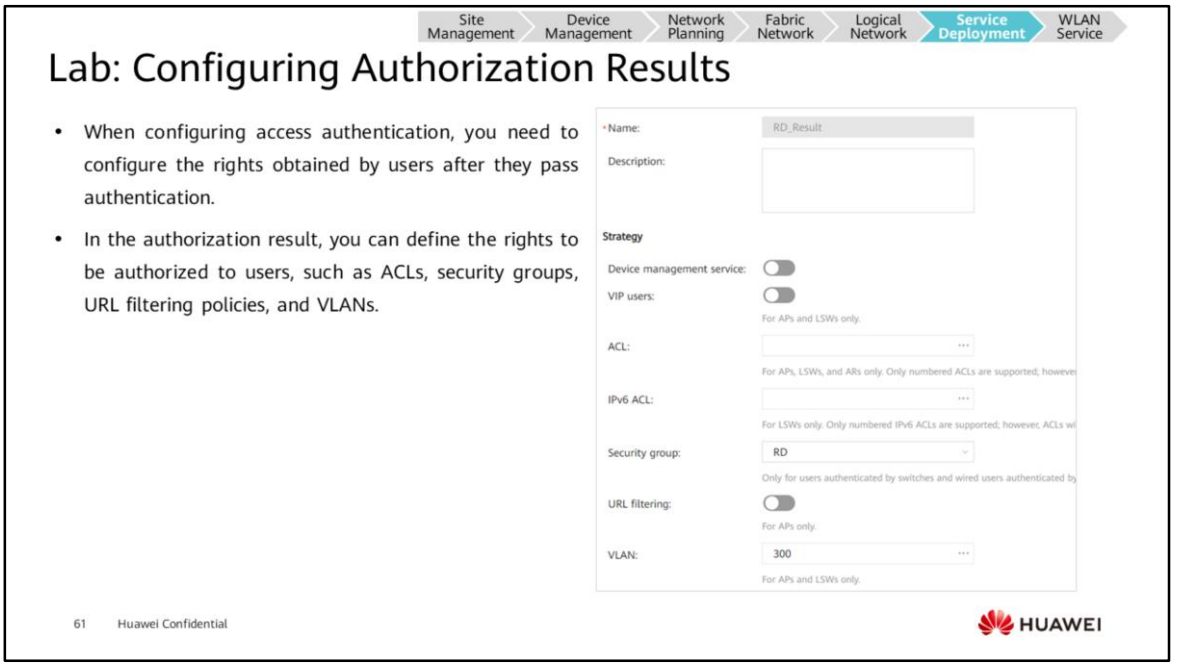
- For example, when creating an account named **kris** (RD user), deselect **Change password upon next login**. As this user belongs to the RD user group, which does not require **Portal** authentication, deselect Portal in the **Available login mode** area.

Lab: Configuring Authentication Rules

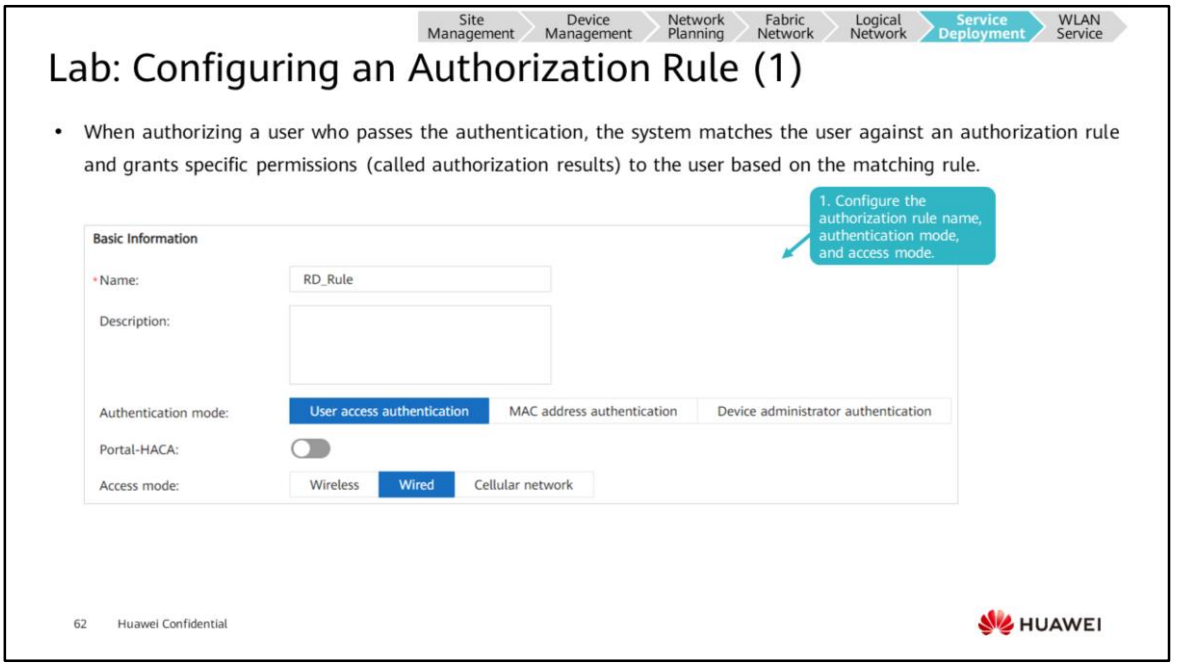
- You can configure authentication rules to authenticate clients and users who access the network, ensuring network security.
- iMaster NCE-Campus has a default authentication rule named **Default**. According to the rule, the local data source is used for authentication.



HUAWEI

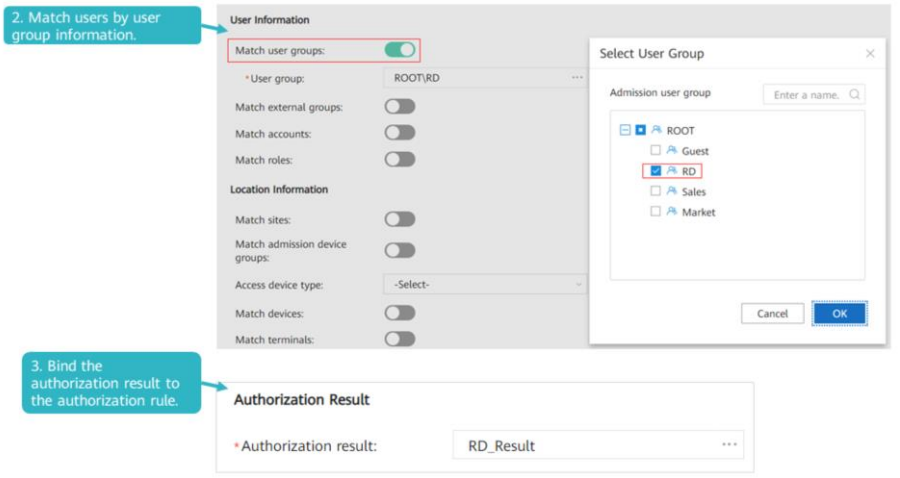


- After configuring authorization results, you need to bind the results to created sites.
- iMaster NCE-Campus provides two default authorization results: permit access and deny access. Once selected, the default authorization result takes effect for all sites and cannot be modified or deleted.



- An authorization rule defines authorization conditions. A user that matches the conditions can obtain the corresponding authorization result. That is, an authorization result specifies the permissions that a user can be granted after the user is authenticated successfully.
- iMaster NCE-Campus provides a default authorization rule named **default**, in which the authorization result is deny access. The default authorization result can be modified to allow the use of another authorization result.

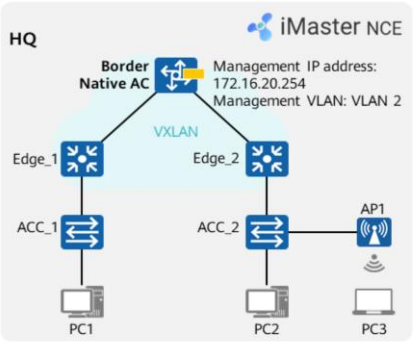
Lab: Configuring an Authorization Rule (2)



HUAWEI

WLAN Service

- In the distributed gateway solution, if the fabric uses a networking model where VXLAN is deployed across the core and aggregation layers and the border device provides the native AC function, APs need to go onboarded and get managed by the border device. APs on the campus need to broadcast an SSID for wireless access, and STAs that have associated with this SSID need to undergo Portal authentication.

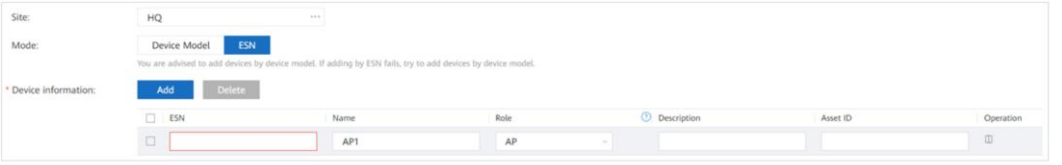


Task	Procedure
AP onboarding	1. Add APs to the HQ site and associate the APs with the border device. 2. Create a management network segment and a management VLAN for AP onboarding. 3. Configure the PnP management VLAN for wireless devices. 4. Configure the source interface for the CAPWAP tunnel on the border device.
Wireless service configuration on the controller	1. Create wireless authentication and configure Portal authentication information.
Wireless service delivery (Web UI of the border device)	1. Configure an SSID profile. 2. Configure a VAP profile, and bind the SSID profile and authentication profile to it. 3. Bind the VAP profile to an AP group.

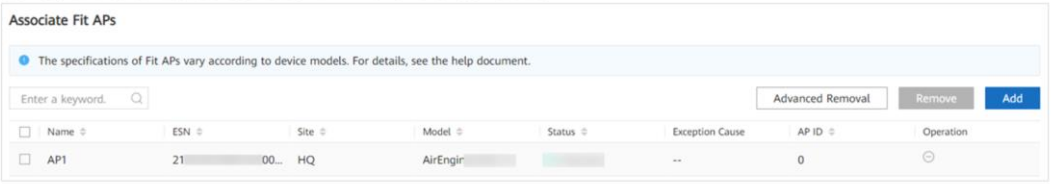
HUAWEI

Lab: AP Onboarding (1)

- Choose **Design** > **Device Management**, and manually add a device to the HQ site. Set the protocol type to NETCONF, and add an AP by entering its ESN. As the AP is not onboarded, it is in the unregistered state.



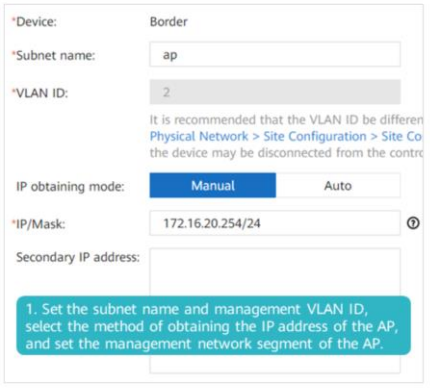
- Choose **Provision** > **Site Configuration**, select the **Fit AP Management** function of the switch, click the border device, and click **Add** to associate AP1 with the border device.

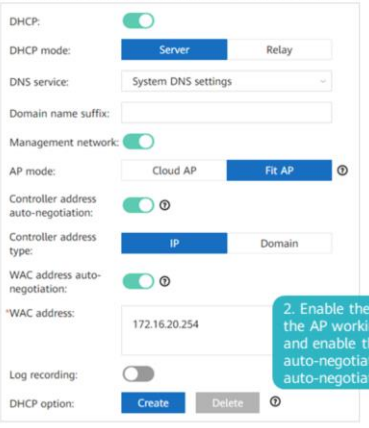


HUAWEI

Lab: AP Onboarding (2)

- Choose **Provision** > **Site Configuration**, select the **Subnet** function of the switch, and create a subnet for AP onboarding. Through this subnet, the AP can obtain the IP addresses of the AC and controller.

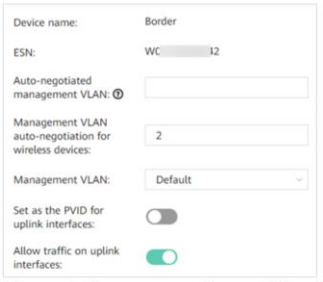




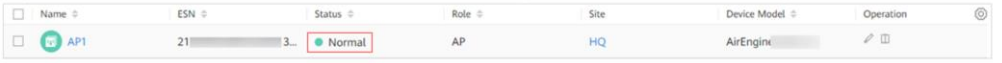
HUAWEI

Lab: AP Onboarding (3)

- Choose **Provision** > **Site Configuration**, select the **Management VLAN** function of the site, and set the wireless PnP VLAN ID of the border device.



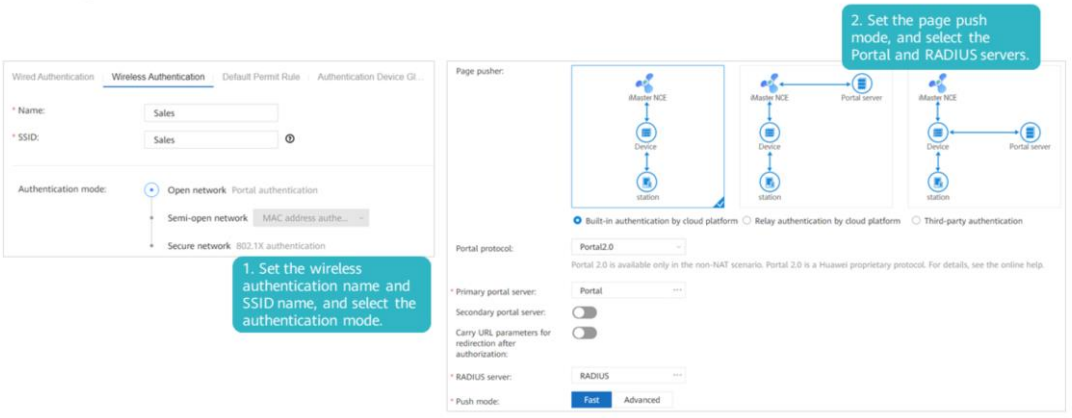
- Then set the source address of the CAPWAP tunnel on the web UI of the border node. The AP can then obtain its own IP address, AC address, and controller address from the address pool of the VLANIF2 interface and successfully register with the border device and controller. After the registration, the AP status changes to normal.



HUAWEI

Lab: Wireless Service Configuration on the Controller (1)

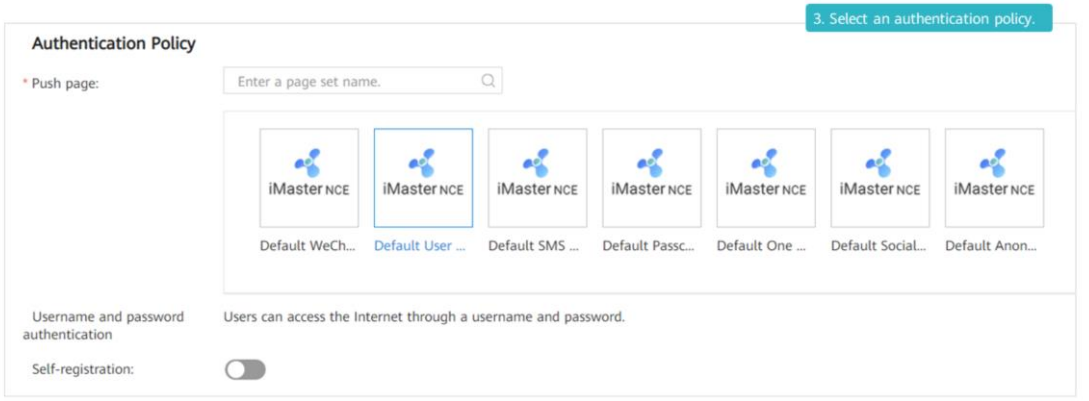
- Choose **Provision > Site Configuration**, select the **Wireless Authentication** function of the switch, and configure wireless authentication.



HUAWEI

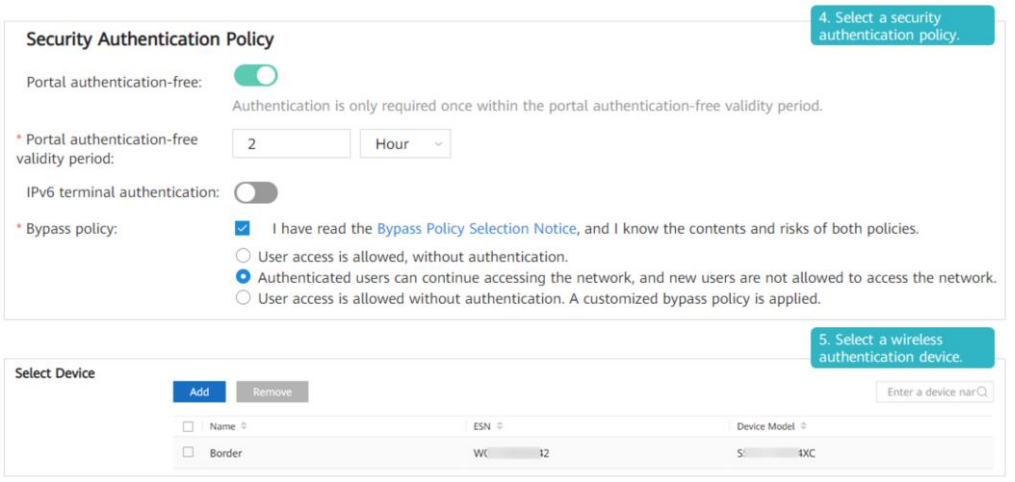
- This example shows the wireless service configuration of sales personnel.

Lab: Wireless Service Configuration on the Controller (2)



HUAWEI

Lab: Wireless Service Configuration on the Controller (3)



HUAWEI

Lab: Wireless Service Delivery

- Choose **Design** > **Device Management**, select the HQ site, and click the border device to access the border device management page.

- Click **Device Configuration** to access the web UI of the border device.

The WLAN service configuration on the web UI of the border device is not provided here.

HUAWEI

Quiz

1. (Short-answer question) What site creation methods are supported by iMaster NCE-Campus?
2. (Multiple-answer question) Which of the following parameters can be entered in the fabric global resource pool? ( )
   A. VLAN
   B. Loopback interface IP address
   C. Bridge domain (BD)
   D. VXLAN network identifier (VNI)

HUAWEI

1. Create sites one by one or in a batch.
   - Create sites one by one: This mode applies when a small number of sites need to be added.
   - Create sites in a batch: This mode applies when a large number of sites need to be added.
2. ACD

Summary

- This course introduces the process of deploying a virtualized campus network, and provides hands-on labs to help trainees understand the functions of the CloudCampus Solution and develop skills required for deploying a virtualized campus network.
- Upon completion of this course, you will be able to deploy the virtualized campus network solution on your own and perform network O&M and management.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Small- and Medium-Sized Cloud-Managed Campus Network Design

HUAWEI

Foreword

- With the development of technologies and digitalization of industries, services in small- and medium-scale scenarios, such as chain stores and branch offices, are gradually connected to the enterprise intranet through cloud-managed networks. A report of IDC reveals the radical difference between traditional products and cloud-based products in terms of annual growth rate from 2014 to 2018. For example, the annual growth rate of traditional Wi-Fi was only 5%, while that of cloud Wi-Fi was 38%. And the annual growth rate of traditional switches and access routers was just 2%, whereas that of cloud switches and access routers was as high as 61%. The above-mentioned data shows that cloud-managed networks have become the trend in constructing small- and medium-sized campus networks. Cloud-based solutions help enterprises improve management efficiency and build networks that can quickly support the development of new services. Huawei's CloudCampus Solution uses cloud computing technology to implement automatic and centralized network management, and provides data collection and analysis capabilities that are unavailable on traditional networks, so as to achieve network (LAN/WLAN) as a service (NaaS).
- This document mainly describes the design and planning methods of small- and medium-sized campus networks in terms of the solution architecture, technical solution comparison, engineering design suggestions, and O&M.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the architecture of Huawei's CloudCampus Solution for small- and medium-sized campus networks.
  - Describe typical networking schemes for small- and medium-sized campus networks.
  - Independently design the CloudCampus solution for small- and medium-sized campus networks based on user requirements, including networking design, physical network design, site deployment design, basic network service design, WLAN design, and admission control design.

HUAWEI

Contents

1. **Small- and Medium-Sized Campus Network Trends and Challenges**
2. Overview of Huawei's CloudCampus Cloud-Managed Network Solution for Small- and Medium-Sized Campus Networks
3. Design of Small- and Medium-Sized Campus Networks with Huawei's CloudCampus Cloud-Managed Network Solution

HUAWEI

Challenges Faced by Small- and Medium-Sized Campus Networks

Accelerated industry changes

Network cloudification brings rapid development in new ICT technologies, such as cloud computing, cloud security, big data, and IoT, leading to tremendous changes in all industries.

- Traditional retailers such as shopping malls and supermarkets offer free Wi-Fi as a way to attract and retain customers, and they also use wireless positioning and customer flow analytics to carry out precision marketing.
- In the education sector, e-classrooms are becoming more and more popular, and diversified multimedia teaching methods further stimulate students' interest in learning.
- Smart healthcare enables hospitals and other healthcare institutions to connect to each other through networks. This implements unified management and analysis of medical data and facilitates medical treatment.

As a pipe that carries upper-layer services and data, the network is becoming more and more complex with a growing number of nodes.

Problems in traditional deployment and management solutions

Low deployment efficiency, which slows down service provisioning

- Site survey, planning, deployment, software commissioning, configuration, and optimization must be completed onsite by professional IT personnel.

Complex network management and high OPEX

- Local professional O&M results in low O&M efficiency and high labor costs. The network management system (NMS), policy control server, charging system, and data analysis platform are deployed independently, causing high management and maintenance costs.

Poor network openness

- The open data provided by multiple management systems on the traditional network needs to be further integrated. In addition, due to the incompatibility of API, the network and applications are connected at a slower speed than application development.

HUAWEI

Service Requirement Analysis for Small- and Medium-Sized Campus Networks

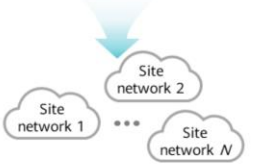
Plug-and-play network devices improve deployment efficiency

Unified management and centralized configuration

- Configurations of multiple sites are centrally delivered, reducing onsite configuration and commissioning workload and improving deployment efficiency.
- Network devices are plug-and-play and able to be expanded on demand, requiring low cost for upgrades.

Cloud-based centralized O&M simplifies multi-site O&M

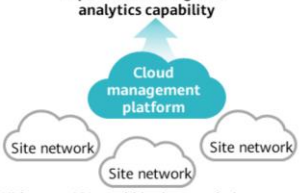
Centralized cloud management of multiple branches and remote automatic O&M



- Geographically dispersed campus branch networks are centrally managed on the cloud through the Internet.
- Troubleshooting and monitoring tools as well as many other automation tools are integrated for remote automatic O&M.

Open APIs accelerate business application integration

Openness and big data analytics capability



- With open APIs and big data analytics capabilities, the cloud management platform can interconnect with multiple management systems to achieve unified network management.
- It is also able to provide diversified value-added applications to lead enterprises into digital transformation.

HUAWEI

Small- and Medium-Sized Campus Network Trends

Network cloudification

- Due to evolution to the cloud architecture, enterprises can focus more on their mission-critical services, without paying too much attention to IT architecture construction.
- To support service cloudification, enterprises need to create a ubiquitous, intelligent, controllable, and on-demand network.
- The network needs to become more a service than a solution.

Cloud security

- Cloud security become more important than ever.
- Facing cloudification, enterprises are vulnerable to attacks that are fundamentally different from those on traditional networks when providing various services.
- Security has shifted from passive defense to proactive defense.
- Detection and response have become as important as defense.

IoT

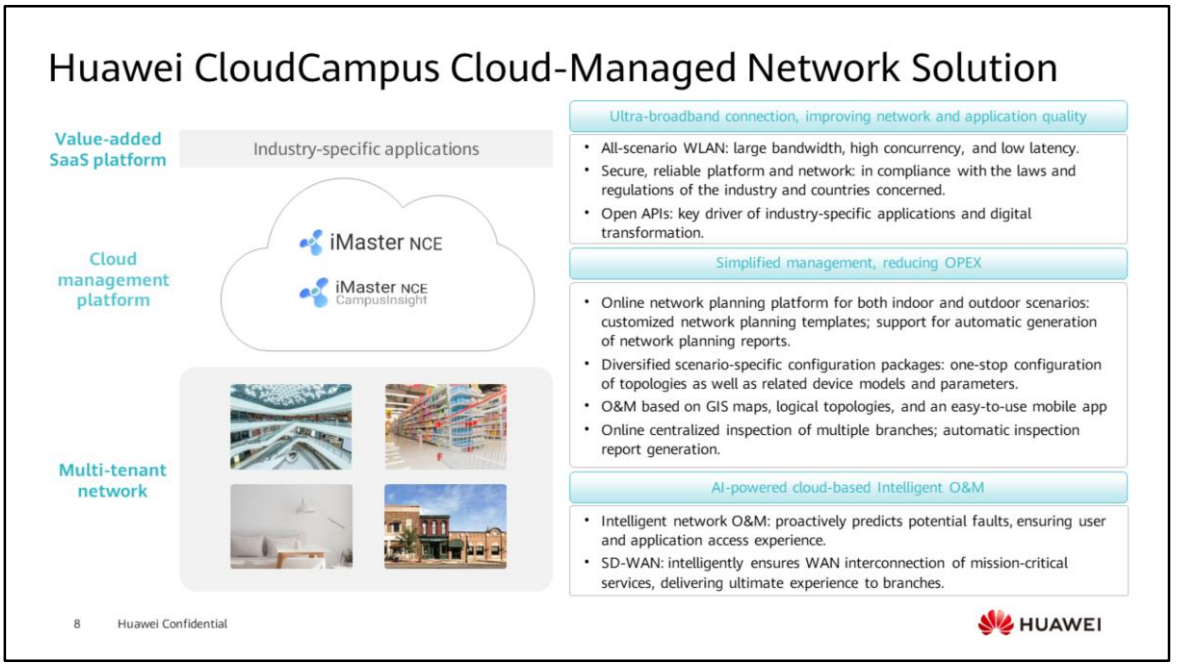
- The rise of IoT leads to a huge increase in the number and types of terminals accessing the network, and these terminals generate a large amount of data.
- Diversified IoT sensing networks need to be smoothly connected to the existing campus network.
- The types of terminals connected to the campus network are more complex than ever. As a result, the campus network becomes a converged network that accommodates multiple types of terminals and media.

HUAWEI

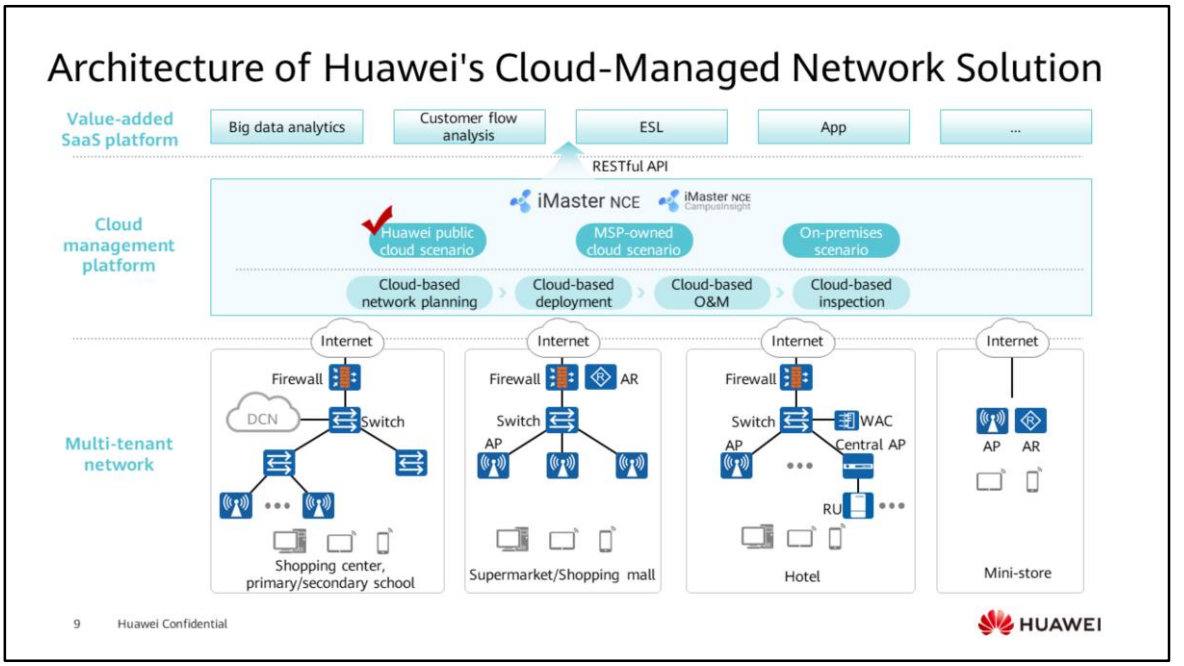
Contents

1. Small- and Medium-Sized Campus Network Trends and Challenges
2. **Overview of Huawei's CloudCampus Cloud-Managed Network Solution for Small- and Medium-Sized Campus Networks**
3. Design of Small- and Medium-Sized Campus Networks with Huawei's CloudCampus Cloud-Managed Network Solution

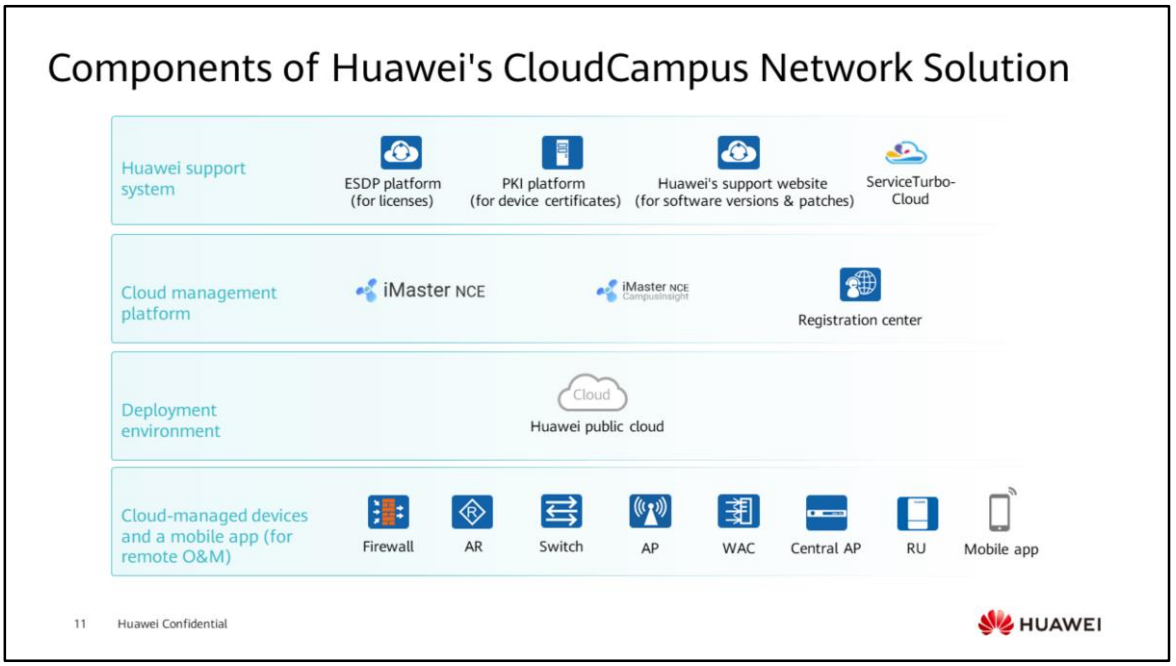
HUAWEI



- In this slide, CloudCampus Network refers to Huawei's CloudCampus Cloud-Managed Network Solution for Small- and Medium-Sized Campus Networks.



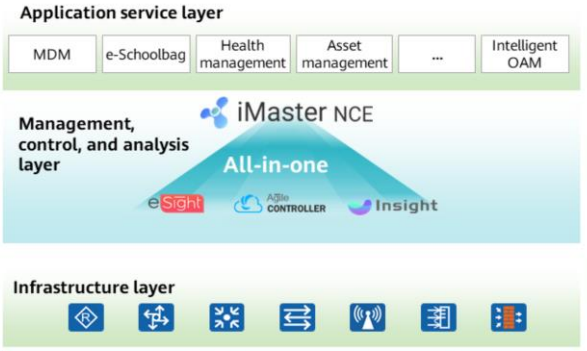
- Huawei's CloudCampus Solution applies to three deployment scenarios: Huawei public cloud, MSP-owned cloud, and on-premises. The on-premises scenario applies to large- and medium-sized campus networks. The Huawei public cloud and MSP-owned cloud scenarios apply to small- and medium-sized campus networks. This document focuses on the Huawei public cloud scenario, which is a cloud management scenario.
- Huawei's CloudCampus Solution for small- and medium-sized campus networks uses cloud computing technology to implement automatic and centralized network management, and provides data collection and analysis capabilities that are unavailable on traditional networks, so as to achieve network (LAN/WLAN) as a service (NaaS).
- There are three layers in the architecture of Huawei CloudCampus Solution for small- and medium-sized campus networks: multi-tenant network, cloud management platform, and value-added SaaS platform.
- Multi-tenant network: It consists of over one hundred network devices, including APs, switches, firewalls, and ARs, and is deployed at the customer side to provide user access.



- ServiceTurbo-Cloud (https://serviceturbo-cloud.huawei.com/) is an enterprise tool service cloud platform. It provides one-stop tool services for Huawei, partners, and customer service engineers in government and enterprise service scenarios.

Solution Component: iMaster NCE-Campus

- iMaster NCE-Campus is a main configuration and management platform in the CloudCampus Solution. It is a main portal for CloudCampus service configuration, O&M, and monitoring.

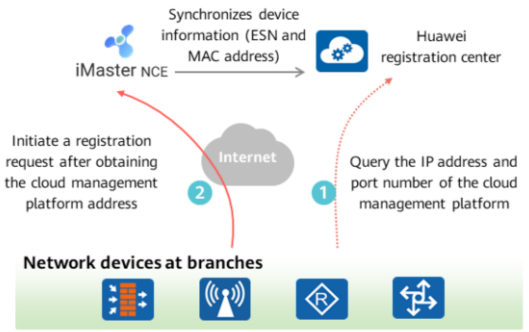


Automated + Intelligent	• SDN-based automatic service configuration/deployment • AI-powered intelligent analysis/prediction/troubleshooting
Management + Control + Analysis	• Unified data base • Centralized detection/locating/processing
Plan + Construct + Maintain + Optimize	• Full-lifecycle management • Simulation/Verification/Monitoring/Optimization

iMaster NCE-Campus, an autonomous driving campus network management and control system.

HUAWEI

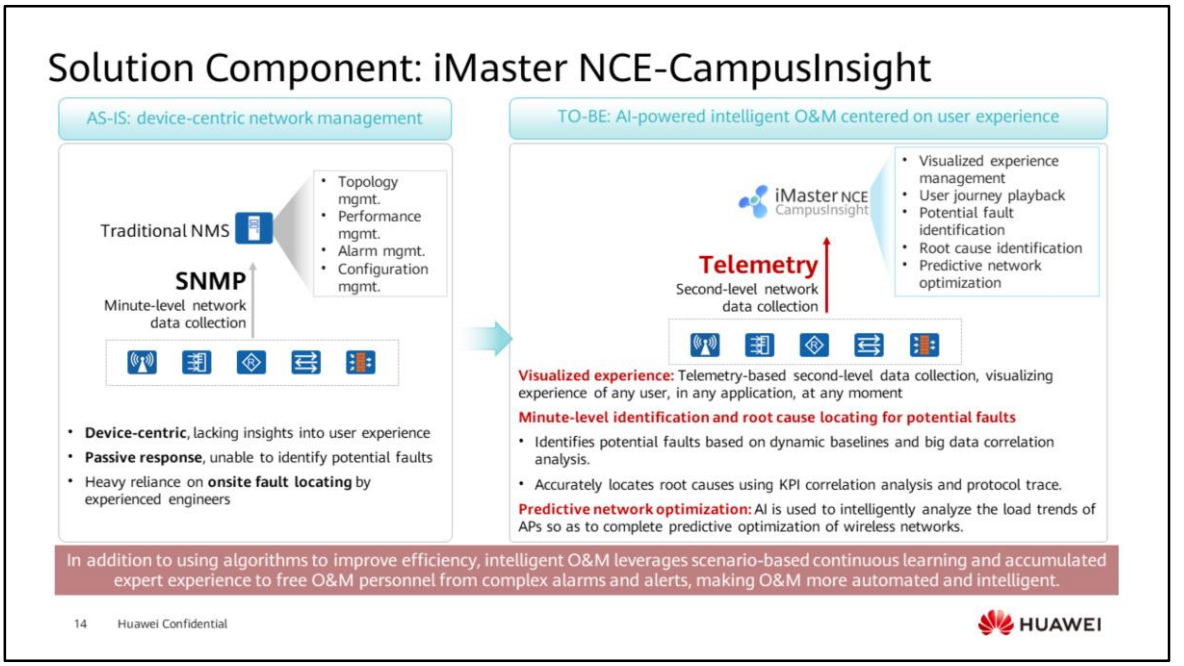
Solution Component: Registration Center



Work out of the box	Devices are already preset with the registration center domain (register.naas.huawei.com) before factory delivery. No additional configuration is required.
Secure transmission	Bidirectional authentication and encryption are performed for query channels based on HTTP/2, ensuring that user device information is not disclosed.
Wide coverage	Multiple registration centers are deployed around the world.
Quick response	The smart DNS returns the address of the registration center server nearest to devices, ensuring fast query.

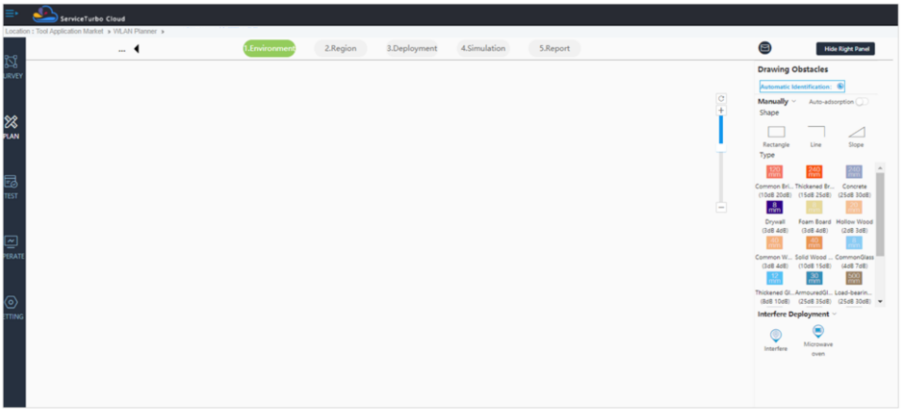
At branches, devices can automatically go online on the cloud management platform without additional configuration.

HUAWEI



- AS-IS: the current status
- TO-BE: future ideal status

Solution Component: WLAN Planner

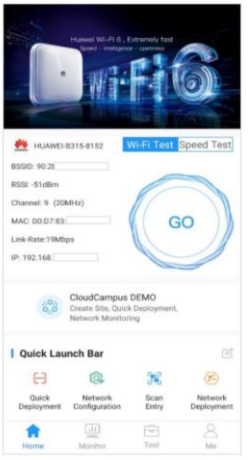


The WLAN Planner is an efficient WLAN network planning tool. It enables signal simulation, helping determine AP deployment locations and signal coverage results.

HUAWEI

- Efficiently supports WLAN network planning:
  - No installation: Cloud-based tool, without the need of software installation.
  - All-scenario: All scenarios supported, including indoor, outdoor, agile distributed, and high-density scenarios.
  - Efficient: Improved simulation efficiency when compared with the traditional standalone deployment.
  - High quality: Supports tens of thousands of WLAN projects.

Solution Component: CloudCampus APP



Mobile APP that covers full lifecycle management for campus networks

Site survey	Connects to the cloud-based WLAN Planner to record photos and texts based on drawings.
Network planning	Connects to the cloud-based WLAN Planner to display network planning results, heat maps, and AP attributes anytime anywhere.
Deployment	Deployment by scanning barcode
Acceptance	One-click test, single service test, project test Deployment position acceptance based on the cloud-based network planning project
O&M	Mobile O&M, device and application monitoring AP access through Bluetooth/management VAP, AP offline diagnosis

E2E simplicity from planning and deployment, all the way to O&M.

HUAWEI

Solution Component: Hardware Products

CloudEngine S series switches

High-performance campus switches

NetEngine ARs

Campus egress routers
(integrating routing, switching, Wi-Fi, 5G, and security functions)

AirEngine Wi-Fi 6 APs

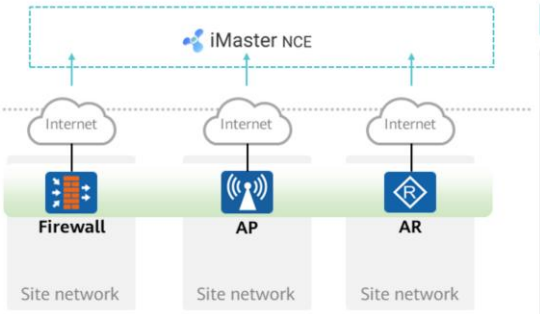
Campus APs
(e.g., indoor settled APs, indoor wall plate APs, agile distributed APs, and outdoor APs)

HiSecEngine AI firewalls

Firewalls
(providing comprehensive network security protection capabilities)

HUAWEI

Solution Technology: Plug-and-Play of Devices



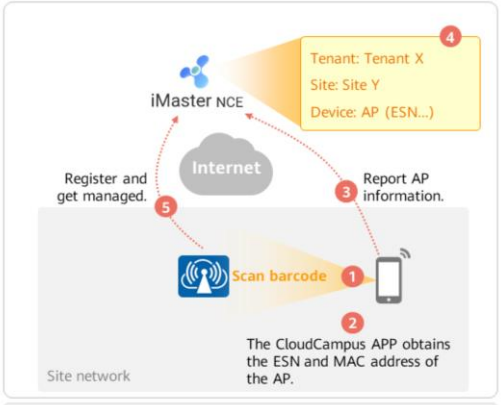
Simplified network deployment

- The time and labor costs for initial device installation and configuration as well as upgrade are reduced.
- After devices on the campus are powered on and connected to the Internet, they can obtain the IP address of the controller iMaster NCE and register with it through multiple methods. All subsequent operations can be performed on the controller, without the need of onsite visits.
- Services are deployed on the iMaster NCE in advance, greatly shortening the deployment time.
- The configuration error rate is reduced due to GUI-based operations.

HUAWEI

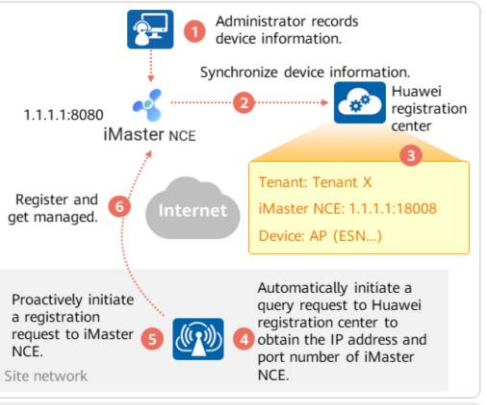
Plug-and-Play of Devices (1)

Through CloudCampus APP (barcode scanning)



Devices supported: AP

Through registration center

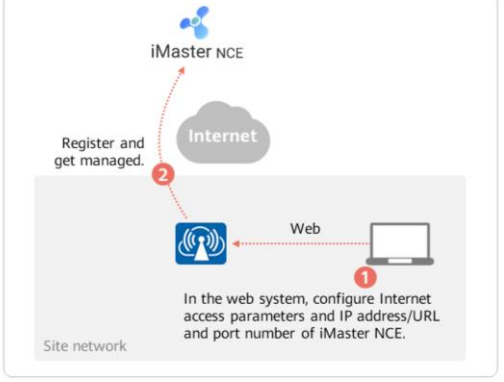


Devices supported: AR, firewall, switch, AP

HUAWEI

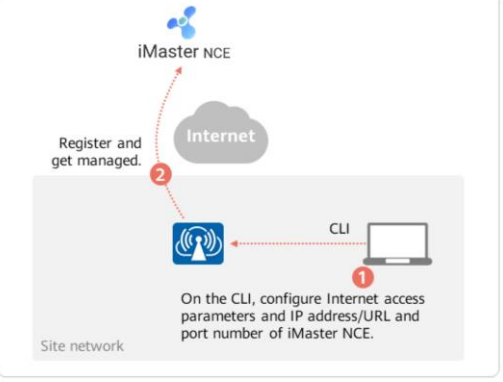
Plug-and-Play of Devices (2)

Through web system



Devices supported: AR, firewall, switch, AP

Through CLI

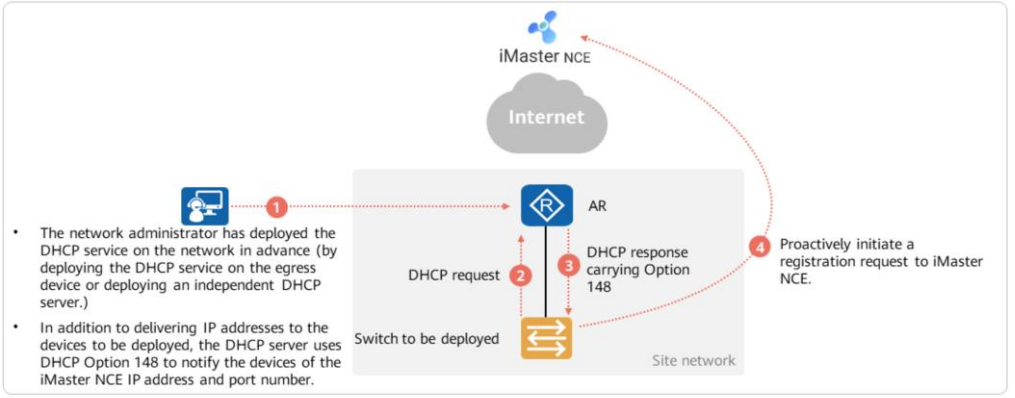


Devices supported: AR, firewall, switch, AP

HUAWEI

Plug-and-Play of Devices (3)

Through DHCP Option 148



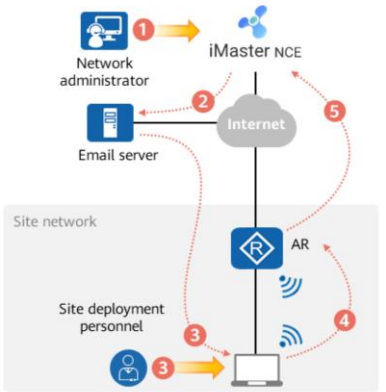
- The network administrator has deployed the DHCP service on the network in advance (by deploying the DHCP service on the egress device or deploying an independent DHCP server.)
- In addition to delivering IP addresses to the devices to be deployed, the DHCP server uses DHCP Option 148 to notify the devices of the iMaster NCE IP address and port number.

Devices supported: AR, switch, AP

HUAWEI

Plug-and-Play of Devices (4)

Email-based deployment

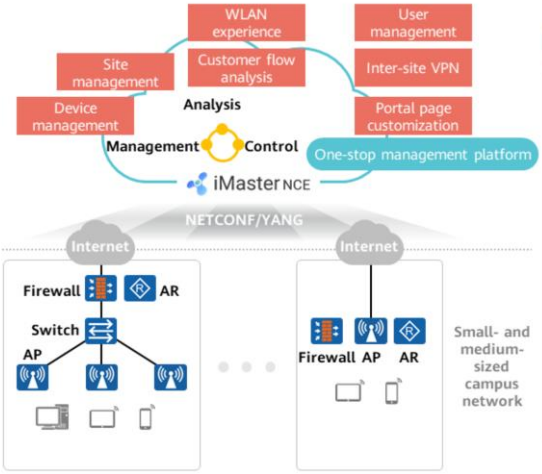


1. The network administrator configures ZTP and selects the email-based deployment mode on iMaster NCE.
2. iMaster NCE packs the deployment configuration into a series of character strings and sends them to the email address specified by the administrator.
3. The site deployment personnel log in to the email box on a PC at the site and receive the deployment email.
4. During site deployment, the site deployment personnel connect the PC to the AR in wired or wireless mode and click the hyperlink in the email body. The PC then automatically logs in to the AR, parses the parameters in the hyperlink to obtain the deployment configuration of the AR, and writes the configuration into the AR.
5. The AR connects to the Internet and automatically sends a registration request to iMaster NCE.

Devices supported: AR

HUAWEI

Solution Technology: Cloud Management



Higher management and O&M efficiency

- After devices go online, iMaster NCE provisions all required configurations and services. iMaster NCE delivers configurations to devices at the branch sites through NETCONF, without the need of CLI-based operations. This simplifies network O&M.
- iMaster NCE performs intelligent analysis and display of the operating status of branch sites, presenting intuitive insights into site health.
- iMaster NCE provides operating status monitoring, intelligent analysis, and remote O&M, so that O&M personnel do not need to visit sites. This improves O&M efficiency and reduces O&M costs.

HUAWEI

Highlights of Huawei CloudCampus Solution for Small- and Medium-Sized Campus Networks

- **Automatic deployment:** Devices can be easily and quickly deployed.
- **Cloud-based network planning and mobile O&M:** WLAN design and device O&M are simplified.
- **Diversified product portfolios:** Huawei provides different product portfolios, including full series of network devices (switches, firewalls, ARs, and APs), meeting diversified network requirements of tenants.
- **Dual-working-mode:** All network devices used in this solution can work in either cloud-based or traditional management mode. Tenants can implement cloud management of traditional devices after devices are upgraded.
- **VASs:** Terminal behavior analysis is a value-added application of iMaster NCE-Campus. More VASs can be developed based on terminal behavior analysis.

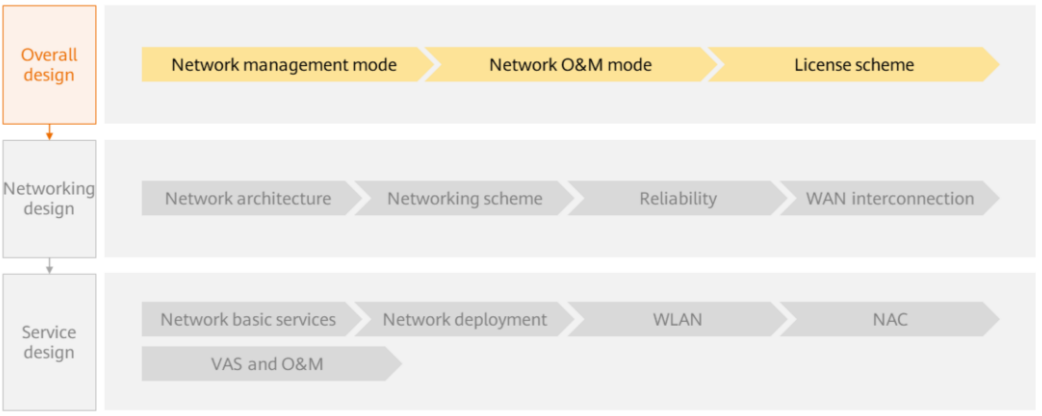
HUAWEI

Contents

1. Small- and Medium-Sized Campus Network Trends and Challenges
2. Overview of Huawei's CloudCampus Cloud-Managed Network Solution for Small- and Medium-Sized Campus Networks
3. **Design of Small- and Medium-Sized Campus Networks with Huawei's CloudCampus Cloud-Managed Network Solution**

HUAWEI

Overall Design Process



HUAWEI

Choose the Network Management Mode

- Which management mode to use: cloud management or on-premises?

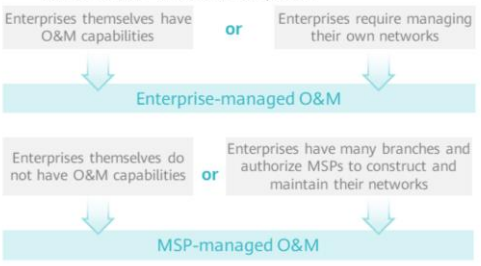
Small- and medium-sized campus networks are small in scale and are sensitive to CAPEX and OPEX. Therefore, the public cloud management mode is recommended for such networks. In this mode, Huawei or MSPs provide SaaS services to manage the networks.

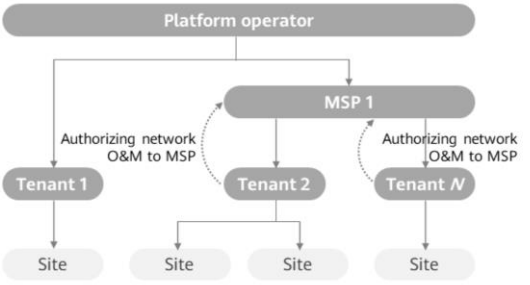
HUAWEI

- The public cloud management mode can be Huawei public cloud management mode or MSP-owned cloud management mode. The two modes are essentially the same. The only difference lies in the operational entity and the provider that offers cloud management services.
- Unless otherwise specified, the Huawei public cloud management mode is used as an example in this document.

Choose the Network O&M Mode

- Which O&M mode to use: enterprise-managed O&M (enterprises perform O&M by themselves) or MSP-managed O&M (enterprises authorize network O&M to MSPs)?
  - The network O&M mode is closely related to the network management mode. If an enterprise manages the network by itself, O&M is also performed by the enterprise. This is called enterprise-managed O&M. If an enterprise authorizes network O&M to an MSP, this is called MSP-managed O&M.
  - Huawei's CloudCampus Solution supports both enterprise-managed O&M and MSP-managed O&M. Enterprises can flexibly choose either of them as required.





HUAWEI

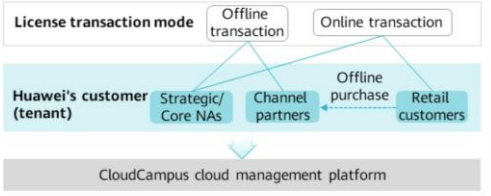
- With certain IT capabilities, a tenant administrator can deploy and maintain a campus network. This scenario is called tenant-managed construction and maintenance. The tenant administrator is the main implementer, and the MSP administrator only provides simple deployment assistance. The tenant administrator can apply to the MSP for the managed construction and maintenance services. After being authorized, the MSP constructs and maintains the campus network for the tenant. This scenario is called MSP-managed construction and maintenance, in which the MSP administrator is the main implementer.
- The following lists the differences between the deployment processes in the tenant-managed construction and maintenance and MSP-managed construction and maintenance scenarios:
- Tenant-managed construction and maintenance:
  - The tenant administrator installs cloud managed devices onsite and registers the cloud managed devices.
  - The tenant administrator logs in to iMaster NCE using their own account and deploys services.

Choose the License Transaction/Purchase Mode

- Select the cloud management deployment scenario (Huawei public cloud or MSP-owned cloud), then select the license transaction or purchase mode accordingly.

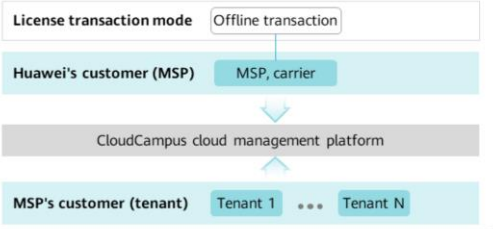
Scenario 1: Huawei public cloud

- Licenses control the available resources of tenants.
- Online transaction mode is supported. That is, licenses can be purchased online from HUAWEI CLOUD. License pooling is not supported in this mode.
- Offline transaction mode is also supported. License activation codes need to be loaded on the CloudCampus cloud management platform.



Scenario 2: MSP-owned cloud

- Licenses control the available resources of the CloudCampus cloud management platform, but not the available resources of tenants.
- Only offline transaction mode is supported. The license file needs to be loaded on the CloudCampus cloud management platform. License pooling and co-termination are supported in this mode.



HUAWEI

Choose the License Consumption and Termination Modes

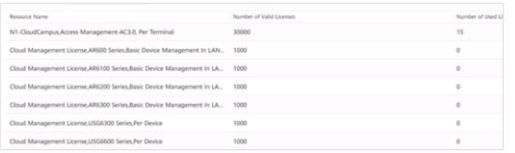
- In the Huawei public cloud scenario, both the co-termination and non-co-termination licensing models are supported.

Co-termination licensing mode

- Licenses are pooled and shared by all devices, and co-terminate for all devices.
- Co-termination of licenses cannot be undone. Therefore, the license termination time must be properly designed.

Non-co-termination licensing mode

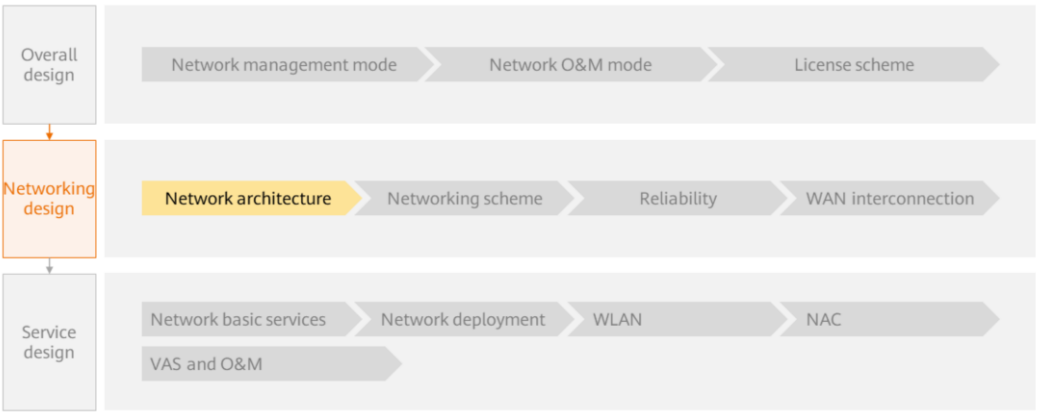
- Licenses are pooled and shared by device type. License resources are exclusive to devices of one type and not shared by devices of other types.
- Licenses for different types of devices terminate separately, while licenses for devices of the same type co-terminate.



HUAWEI

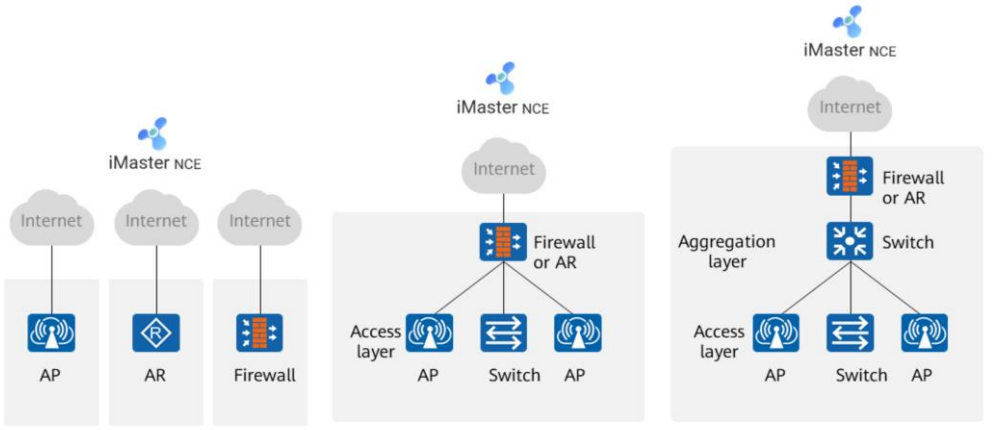
- For small- and medium-sized campus networks, the co-termination licensing model is recommended, facilitating device management and operations. If the license validity periods of devices of different types need to be precisely controlled, use the non-co-termination licensing model.

Overall Design Process



HUAWEI

Network Architecture Design: Intranet Architecture (1)



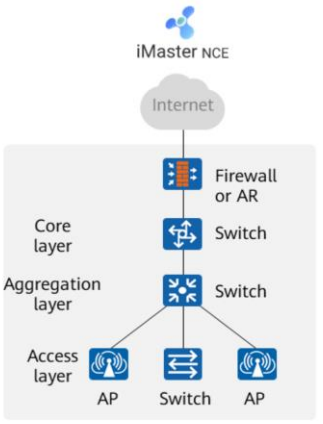
Single-device networking architecture

Single-layer networking architecture

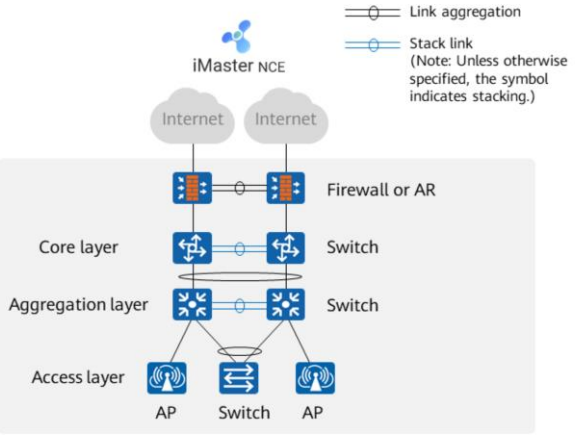
Two-layer networking architecture

HUAWEI

Network Architecture Design: Intranet Architecture (2)

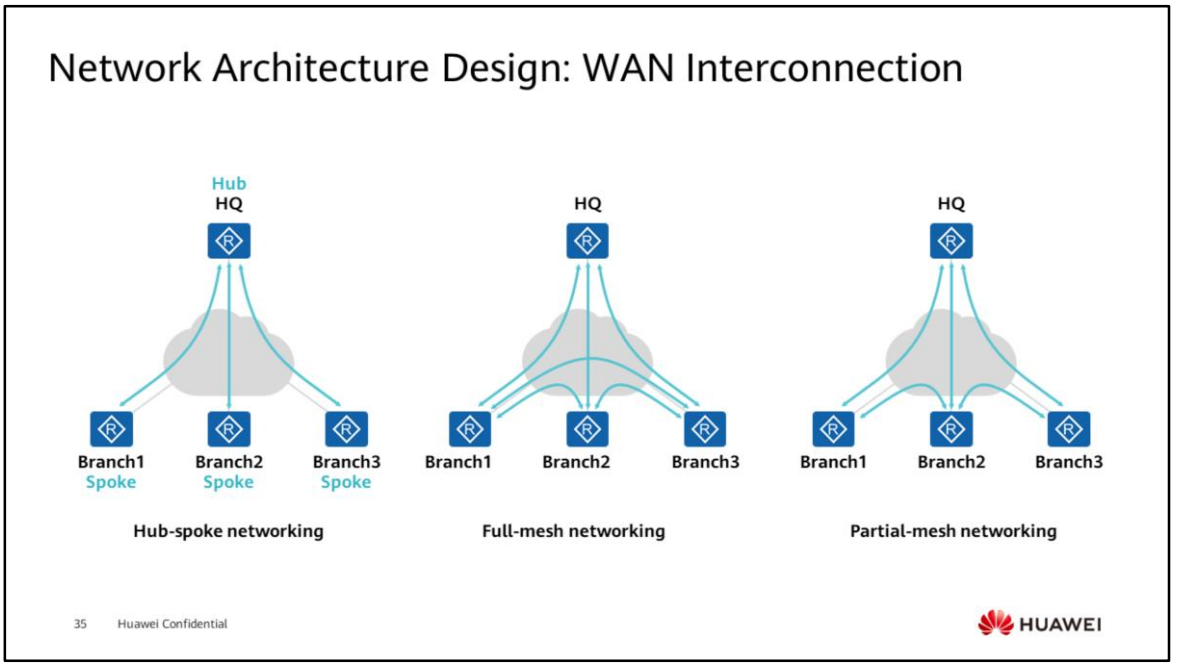


Three-layer networking architecture



Highly reliable networking architecture

HUAWEI



- Hub-spoke: Generally, the enterprise headquarters or DC functions as a hub site. Each branch site of an enterprise can communicate with the hub site and can communicate with other branch sites through the hub site. This model is applicable to scenarios where traffic between all branch sites of an enterprise must pass through the headquarters for centralized security monitoring.
- Full-mesh: All sites of an enterprise can communicate with each other. If traffic needs to be transmitted between the headquarters and branches or between branches, data is directly exchanged without traversing an intermediate node. This model is applicable to scenarios where all sites of an enterprise need to directly access each other. This model eliminates the delay caused by traffic transmission through the headquarters.
- Partial-mesh: Most sites of an enterprise can directly communicate with each other. However, the underlay WAN networks of a small number of sites cannot directly communicate with each other. For example, in the figure, branch 1 and branch 3 need to communicate with each other through the headquarters.

Network Architecture Design Considerations

1 Determine the intranet architecture

Determine the network architecture layers based on the number of access terminals or device capabilities. Determine the device model at each layer and the mapping between upper-layer and lower-layer devices.

- Single-device networking architecture
- Single-layer networking architecture
- Two-layer networking architecture
- Three-layer networking architecture

2 (Optional) Determine the WAN interconnection model

Determine the WAN interconnection model based on the communication requirements between sites.

- Hub-spoke networking
- Full-mesh networking
- Partial-mesh networking

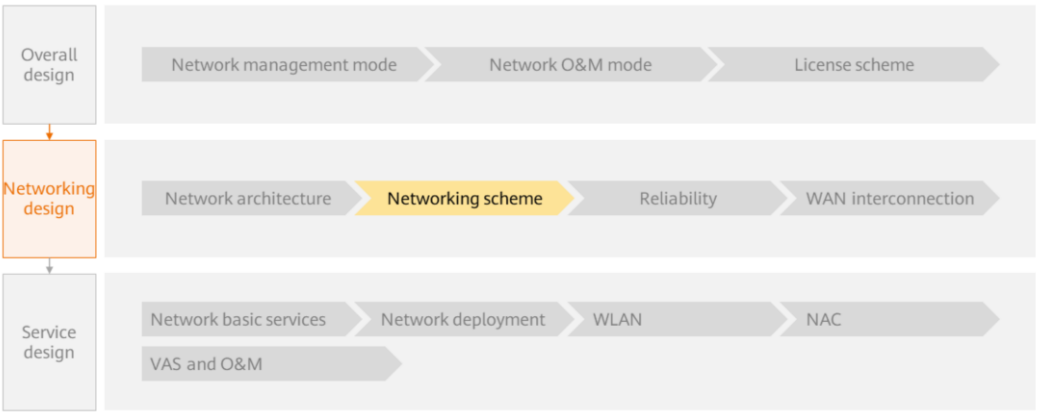
3 (Optional) Determine reliability requirements

Determine the highly reliable networking model based on the device type and capability

- Firewalls are deployed in hot standby mode
- Switch stacking
- Eth-Trunk

HUAWEI

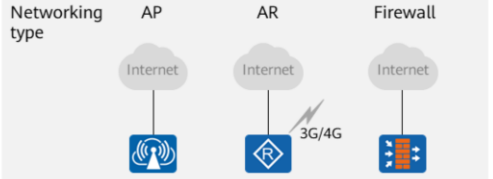
Overall Design Process



HUAWEI

Single-Device Networking

Network scale	< 50 terminals; area < 50 m^2



Overview

- Applicable to single stores and small stores (such as agent stores and gas stations)
- A single device provides gateway features, such as PPPoE dialup, DHCP, and NAT.
- An AP is deployed if only wireless access is required and only one wired Internet egress is available.
- An AR is deployed if wired access is needed and multiple uplinks, especially a 3G/4G backup link, are required.
- A firewall is deployed if security-sensitive stores, such as small logistics, office, and finance organizations, require advanced security features, including URL filtering, security protection, and antivirus.

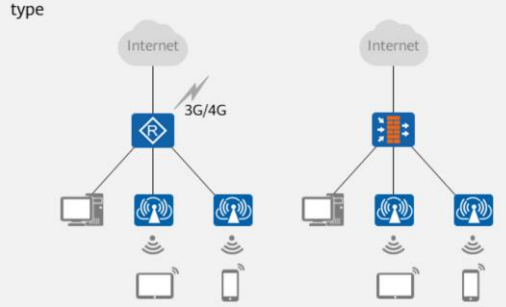
Constraints

- Wireless features such as roaming and load balancing are not supported.
- The AP supports only one Internet link.
- The AP and AR support wired user access, but do not support wired user authentication.
- The AR supports wireless user authentication in only Open, PSK, and Portal authentication modes, but not 802.1X or MAC address authentication mode. The firewall provides wired user authentication and supports only Portal authentication.
- Currently, no firewall model supports Wi-Fi and LTE, and therefore it is not recommended in actual deployments.

HUAWEI

Egress Gateway + AP

Network scale	< 200 terminals; area < 300 m²
Networking type	AR + AP / Firewall + AP



Overview

- Applies to small- and medium-sized clothing stores, supermarkets, shopping malls, etc.
- The egress gateway provides features such as PPPoE, DHCP, NAT, and LTE.
- This networking mode supports continuous coverage of multiple APs, as well as multiple uplinks, including a 3G/4G backup uplink.
- APs support mesh networking.
- A firewall is deployed at the egress for scenarios where advanced security features, including URL filtering, security protection, and antivirus, need to be met.

Constraints

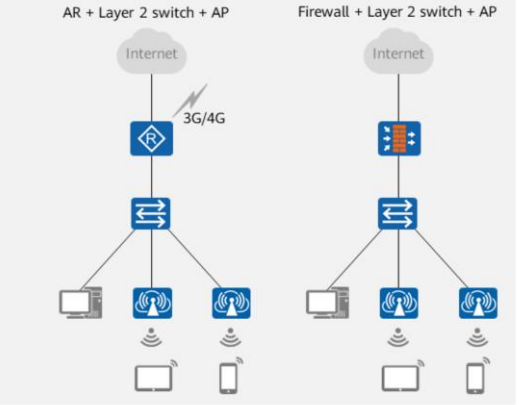
- The AP and AR support wired user access, but do not support wired user authentication.
- The firewall provides wired user authentication and supports only Portal authentication.
- If the AR or firewall does not support PoE, the AP can use an external PoE power adapter.

HUAWEI

Egress Gateway + Layer 2 Switch + AP

Network scale	< 2000 terminals; area < 3000 m²

Networking type



Overview

- Applies to small- and medium-sized clothing stores, retail stores, etc.
- The egress gateway provides features, such as WAN access, DHCP, and NAT.
- The Layer 2 switch provides PoE extended access and wired terminal access functions.
- APs provide access services for wireless terminals at the site.
- APs support mesh networking.
- A firewall is deployed at the egress for scenarios where advanced security features, including URL filtering, security protection, and antivirus, need to be met.

Constraints

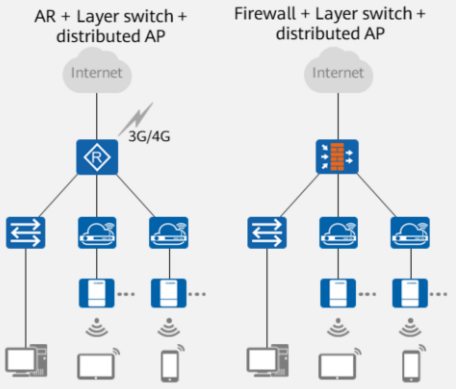
- The AR and firewall in cloud management mode do not support Eth-Trunks and cannot connect to the switch through Eth-Trunks.
- If switches need to be stacked, it is recommended to deploy multiple layers of switches.
- The AR can be deployed only in a single-node system.

HUAWEI

Egress Gateway + Layer 2 Switch + Distributed AP

Network scale	< 2000 terminals; area < 3000 m²

Networking type



Overview

- Applies to dense-room building scenarios, such as dormitories and hotels, where network planning is not required and each room transmits signals separately.
- The AR functions as the egress gateway and provides features, such as WAN access, DHCP, and NAT.
- The Layer 2 switch provides PoE extended access and wired terminal access functions.
- APs provide access services for wireless terminals at the site.
- APs support mesh networking.
- A firewall is deployed at the egress for scenarios where advanced security features, including URL filtering, security protection, and antivirus, need to be met.

Constraints

- The AR and firewall in cloud management mode do not support Eth-Trunks and cannot connect to the switch through Eth-Trunks.
- If switches need to be stacked, it is recommended to deploy multiple layers of switches.
- The AR can be deployed only in a single-node system.

HUAWEI

Summary of Typical Networking Schemes (1)

Deployment Scenario	Network Scale	Key Networking Requirements	Networking Model		Remarks
Many branches, where they are no demands for branch interconnection	< 50 terminals; area < 50 m²	Single store; wireless access only; single Internet egress	Singe-device networking	AP	/
		Single store; mainly wireless access; Ethernet or LTE uplink		AR	/
		Single store; wired and wireless authentication and access; multiple Internet uplinks and LTE uplinks; high security requirements (URL filtering/IPS/security protection/antivirus)		Firewall	Currently, firewalls have no Wi-Fi/LTE models.
	< 200 terminals; area < 300 m²	Wireless access only; multiple Internet uplinks and LTE uplinks	Egress gateway + AP	AR + AP	Currently, ARs and firewalls do not have PoE models. APs need to connect to external power supplies.
		Wired and wireless authentication and access; multiple Internet uplinks; high security requirements (URL filtering/IPS/security protection/antivirus)		Firewall + AP	
	< 2000 terminals; area < 3000 m²	Wired and wireless access; multiple Internet uplinks and LTE uplinks	Egress gateway + Layer 2 switch + AP	AR + Layer 2 switch + AP	/
		Wired and wireless authentication and access; multiple Internet uplinks; high security requirements (URL filtering/IPS/security protection/antivirus)		Firewall + Layer 2 switch + AP	/

HUAWEI

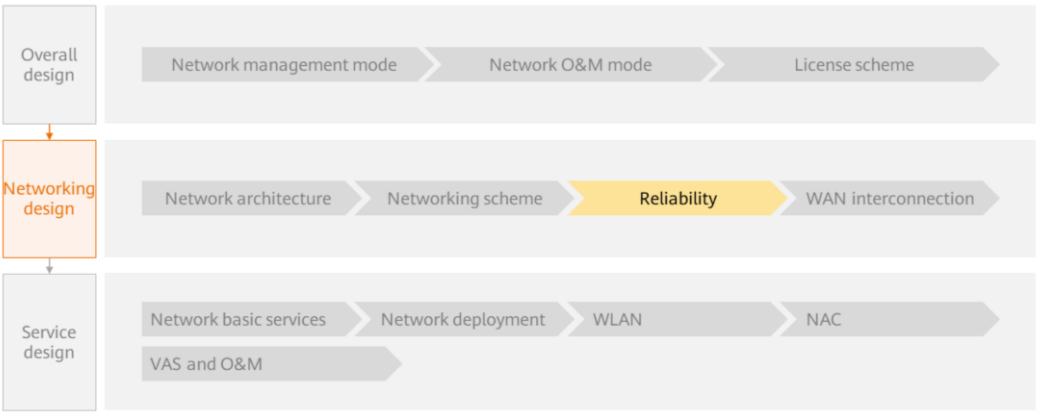
- Cloud management for small- and medium-sized campus networks applies to small- and medium-sized enterprises (SMEs) and branches, such as shopping malls/supermarkets, retail stores, hotels, educational institutions, and financial services organizations. Different networking models are used based on different network scales.

Summary of Typical Networking Schemes (2)

Deployment Scenario	Network Scale	Key Networking Requirements	Networking Model		Remarks
Many branches, where they are no demands for branch interconnection	< 2000 terminals; area < 3000 m²	Dense-room scenarios such as hotels and dormitories; wired and wireless access; multiple Internet uplinks and LTE uplinks	Egress gateway + Layer 2 switch + distributed AP	AR + Layer 2 switch + central AP + RU	/
		Dense-room scenarios such as hotels and dormitories; wired and wireless access; multiple Internet uplinks; high security requirements (URL filtering/IPS/security defense/antivirus)		Firewall + Layer 2 switch + central AP + RU	/
Multiple branches, where there are demands for communication with the headquarters	Same as that in the "many branches, branches, where they are no demands for branch interconnection" scenario	Only basic communication between branches and the headquarters is required. Third-party VPN gateways may be deployed at the headquarters.	IPsec VPN interconnection	AP/AR/firewall as the egress gateway	/
		Multiple links are deployed between the headquarters and branches, and intelligent traffic steering is required.	SD-WAN Interconnection	AR as the egress gateway	/

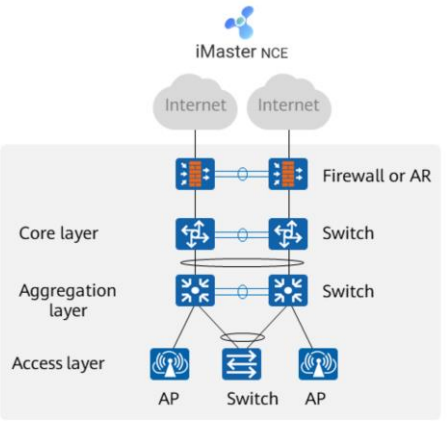
HUAWEI

Overall Design Process



HUAWEI

Network Reliability Design



Highly reliable networking architecture

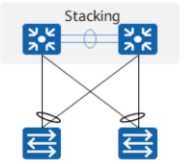
- Networking reliability refers to the reliability design of the tenant network. You can add devices or links in any typical networking scheme to enhance network reliability. Network reliability design can be classified into three levels based on the involved scope:
  - Link-level reliability design
  - Device-level reliability design
  - Network-level reliability design (which combines link and device reliability)

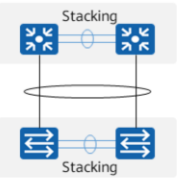
HUAWEI

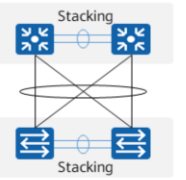
Link-Level Reliability Design (1)

Intranet link reliability

- Multiple links can be deployed between switches and bonded using Eth-Trunks to improve link reliability.
- When switches are stacked, it is recommended that Eth-Trunks be used across switches to ensure link reliability.







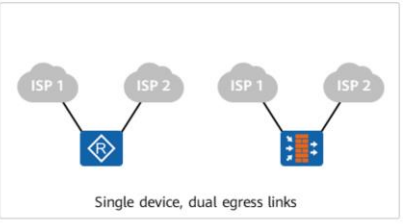
HUAWEI

Link-Level Reliability Design (2)

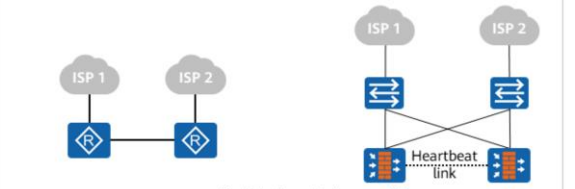
Egress link reliability

Multiple links are deployed at the campus egress and work in active/standby mode. The networking scheme is as follows:

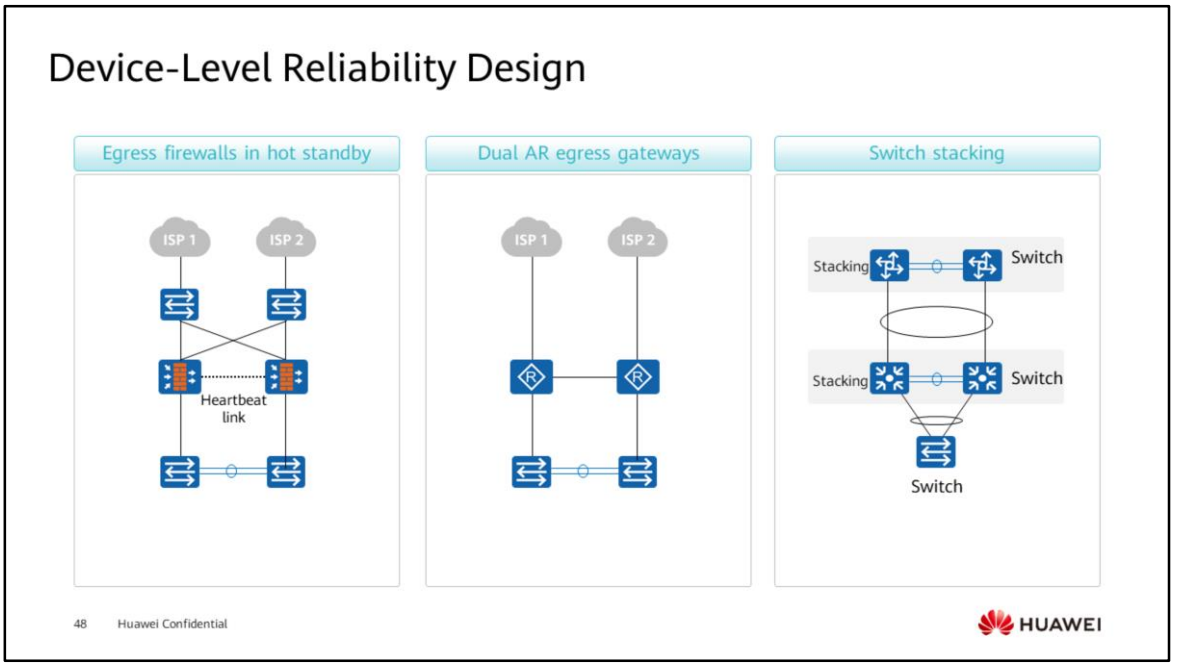
- The "single device, multiple egress links" scheme or the "multiple devices, at least one egress link for each device" scheme can be used. The latter scheme is recommended.
- To further improve reliability, it is recommended that different egress links be connected to different carrier networks.



Single device, dual egress links

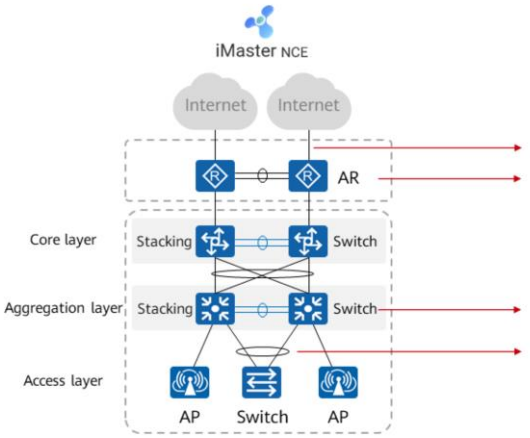


Dual devices, dual egress links



- Device reliability: Fixed-configuration devices are generally used on small- and medium-sized campus networks.
- Device reliability design includes the following two aspects:
- Redundant components are used to improve device reliability.
    - In addition to selecting devices with high reliability, you can further improve the reliability of the devices by using dual power supplies or redundant components (e.g. boards).
- Joint networking is used to provide device reliability.
    - Multiple devices are deployed together to improve device reliability, for example, active/standby or stacking.
    - Two devices are deployed in the egress zone of the campus network to work in active/standby mode. Currently, in a two-node system, firewalls support only the hot standby mirroring mode. ARs support dual-CPE networking, and the two devices and two egress links can work concurrently.
    - Intranet network devices, such as LAN switches at the core or aggregation layer, are stacked to implement device-level backup. Currently, only switches can be stacked.

Network-Level Reliability Design



Typically, Eth-Trunk and stacking technologies are used in dual-device and multi-link scenarios to improve network-wide reliability.

Egress reliability design:

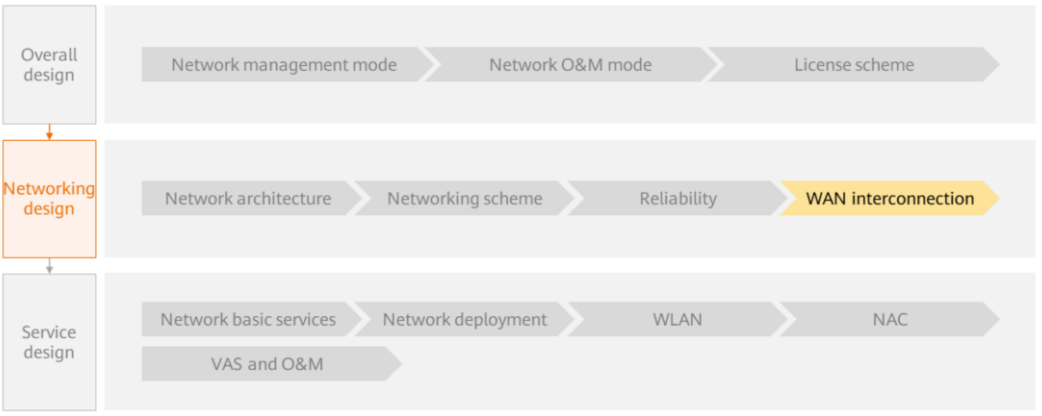
- Dual-device: active/standby, load balancing
- Multi-link, multi-carrier network

Intranet reliability design: switch networking

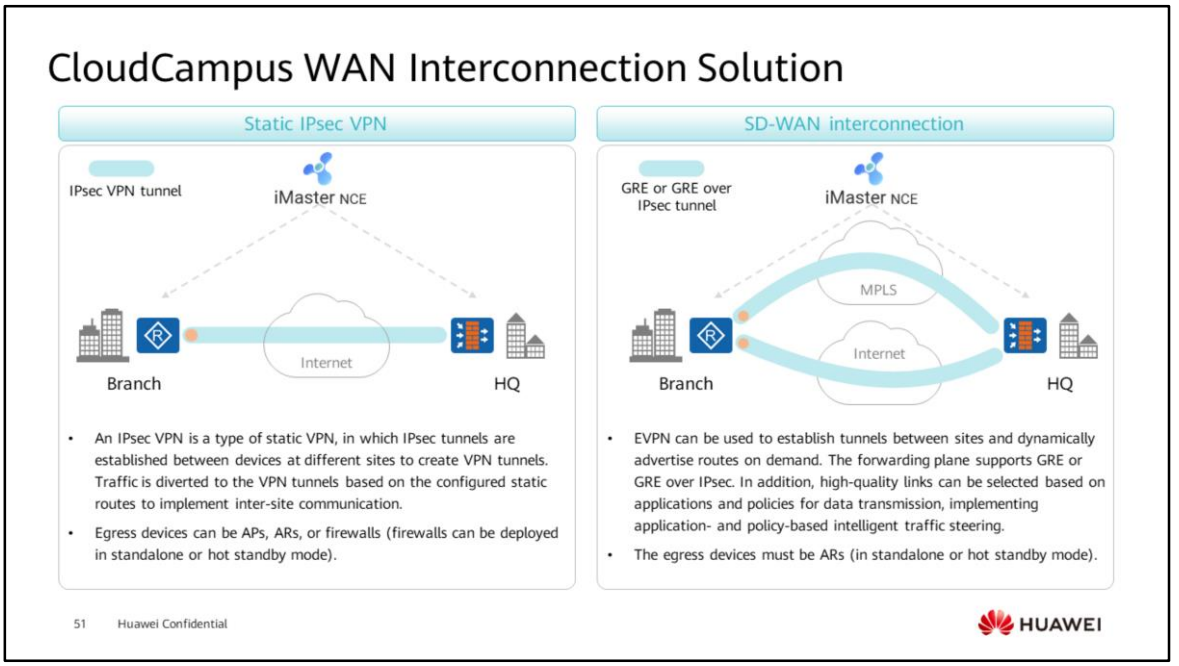
- Dual-/multi-device: stacking
- Dual-/multi-link: Eth-Trunk

HUAWEI

Overall Design Process

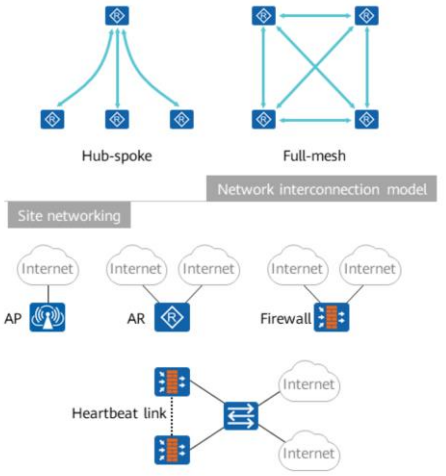


HUAWEI



- The multi-campus network interconnection solution is a sub-solution provided in the CloudCampus Solution for the interconnection between branch campuses and between branches and the HQ or DCs. With the SD-WAN solution integrated, the multi-campus network interconnection solution provides two models for WAN interconnection: static IPsec VPN and EVPN.
- An IPsec VPN is a type of static VPN, in which IPsec tunnels are established between devices at different sites to create VPN tunnels. Traffic is diverted to the VPN tunnels based on the configured static routes so that services between sites can be accessed through the VPN tunnels.
- An EVPN is a type of dynamic VPN that can establish tunnels between sites and dynamically advertise routes on demand. EVPN establishes GRE tunnels between sites to establish VPN tunnels and supports IPsec encryption on GRE tunnels to ensure tunnel encryption security. In addition, the EVPN solution offers application- and policy-based intelligent traffic steering, allowing high-quality links to be selected based on applications and policies for data transmission.

IPsec VPN Interconnection Networking Model



Network interconnection model

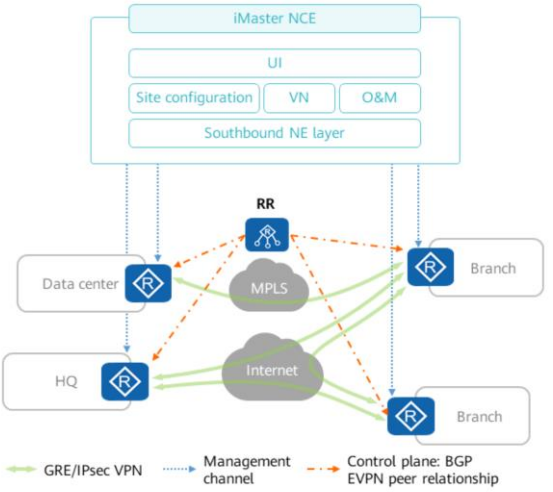
- The hub-spoke and full-mesh models are supported.
- The hub-spoke model is applicable when data traffic is mainly transmitted between branches and the HQ. In this model, branches can also communicate with each other through the HQ.
- If branches are of similar scales and a large amount of traffic is transmitted between the branches, the full-mesh model can be used. In this case, all egress devices must use public IP addresses, and only firewalls support this model. In the full-mesh model, the number of sites is limited (32 at most). Therefore, the hub-spoke model is recommended for IPsec interconnection.

Site networking

- APs, ARs, or firewalls can be used as egress devices based on site requirements.
- In some scenarios requiring high reliability, two egress links can be deployed. In scenarios requiring wireless uplinks, the egress gateway must be an LTE-capable device.
- For large campus networks, the egress gateways can be deployed in hot standby mode to ensure high reliability. Currently, only firewalls support hot standby.
- In the hub-spoke model, the hub site is usually the HQ or DC site, which requires high-performance devices, rich policies, and security configurations. Currently, the cloud management mode supports only a few security and policy features. Therefore, it is recommended that the traditional management mode be configured for devices at the hub site, and the devices work in hot standby mode.

HUAWEI

SD-WAN Solution Overview (1)



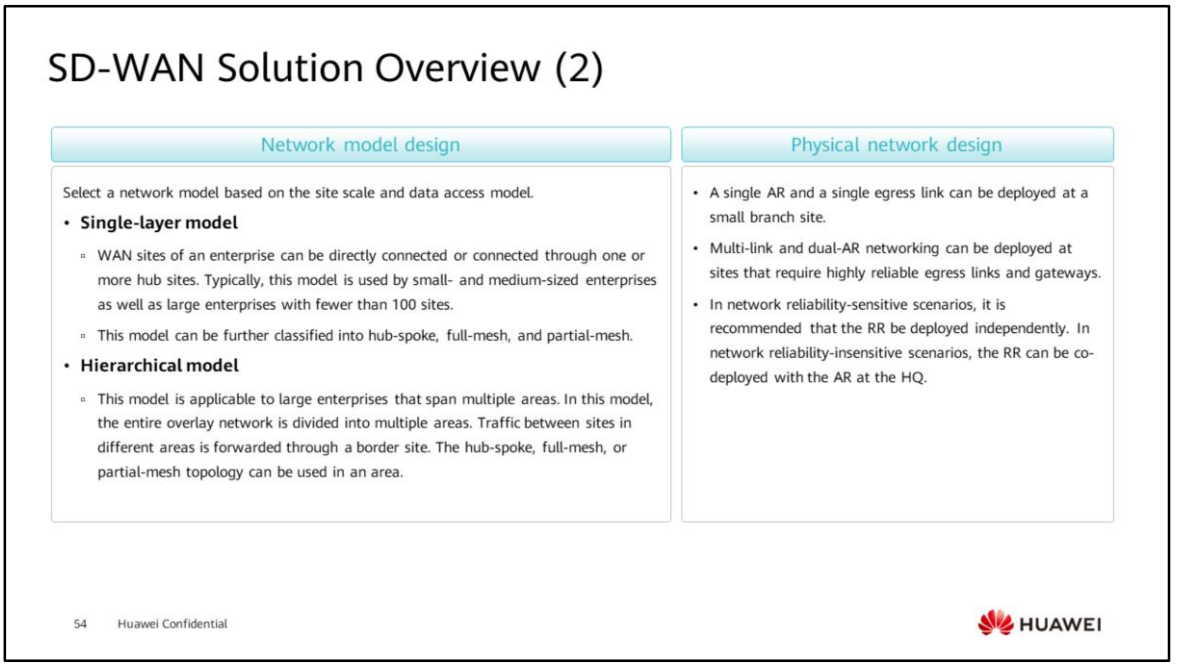
Solution overview

- The EVPN interconnection mode of the SD-WAN solution is applicable to the scenarios where a large number of branches communicate with the HQ/DC (hub site). This mode can meet the networking requirements of multiple HQs/regional centers. In this mode, the branches can communicate with the HQ through multiple links. In addition, this mode supports application- and quality-based intelligent traffic steering, QoS scheduling, and link quality display.
- The hub-spoke, full-mesh, and partial-mesh networking models are supported.
- Hub sites can work in active/standby mode to improve network access reliability at the HQ.
- Multi-VPN (VRF) and WAN-side multi-VPN interconnection are supported.
- In scenarios requiring high security, such as the finance industry, both firewalls and ARs can be deployed at sites.

Solution constraints

- ARs deployed at the HQ and branches must support SD-WAN.

HUAWEI



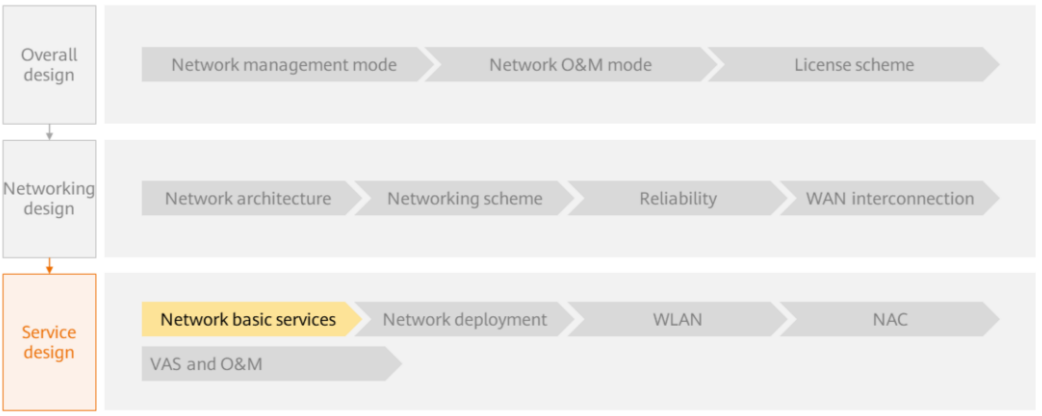
- The single-layer network model is also called the flat network model. In this model, WAN sites of an enterprise can be directly connected or connected through one or more hub sites. Typically, this model is used by small- and medium-sized enterprises as well as large enterprises with fewer than 100 sites. The single-layer network model can be further classified into hub-spoke, full-mesh, and partial-mesh.
- The hierarchical network model is applicable to large enterprises that span multiple areas. In this model, the entire overlay network is divided into multiple areas. Traffic between sites in different areas is forwarded through a border site. The hub-spoke, full-mesh, or partial-mesh topology can be used in an area.

SD-WAN Solution: Service Design

- **VPN planning:** A network can be divided into multiple VPNs based on services or departments. Policies can be independently configured for services of each VPN, and VPNs can be isolated from each other.
- **Intelligent traffic steering:** Traffic policies can be defined (based on network segments or applications) to design intelligent traffic steering, so that high-reliability links (such as MPLS links) are preferentially selected for high-priority applications (such as audio and video services) and low-priority links are selected to transmit common data.
- **QoS application:** Multi-level QoS scheduling, such as three-level QoS scheduling for ports, VPNs, and applications.
- **Security:** ACL-based traffic filtering, URL filtering, firewall, and IPS are configured to ensure traffic security.
- **Site-to-Internet access:** The local Internet access, centralized Internet access, and hybrid Internet access modes are supported. If the Internet egress link is available, local Internet access can be configured. If centralized traffic policing is required, centralized Internet access will suffice.
- **Interworking with traditional networks:** The solution supports interworking with traditional networks.

HUAWEI

Overall Design Process



HUAWEI

VLAN Planning

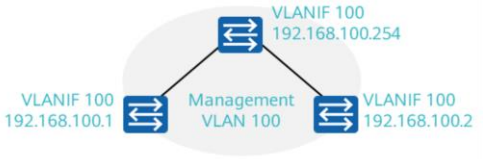
- Classify VLANs into service VLANs, management VLANs, and interconnection VLANs.
- Allocate consecutive VLAN IDs to ensure proper use of VLAN resources.
- Reserve a specific number of VLANs for future use.
- Typically, VLANs are divided based on interfaces. According to different design principles, interfaces of access switches are added to different VLANs so that users of different service types can be isolated.

VLAN Assignment	Example
VLAN assignment by logical area	• Core network area: VLANs 100–199 • Server area: VLANs 200–999 (VLANs 1000–1999 are reserved.) • Access network area: VLANs 2000–3499 • Service network area: VLANs 3500–3999
VLAN assignment by geographic area	• Area A: VLANs 2000–2199 • Area B: VLANs 2200–2399
VLAN assignment by personnel structure	• Department A in area A: VLANs 2000–2009 • Department B in area A: VLANs 2010–2019
VLAN assignment by service type	• Web server area: VLANs 200–299 • Application server area: VLANs 300–399 • Database server area: VLANs 400–499

HUAWEI

IP Address Planning (1)

Management IP address

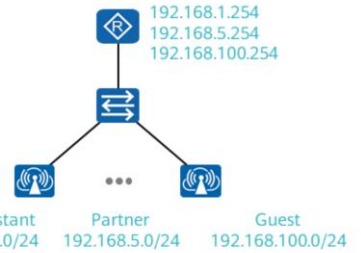


A Layer 2 device uses the VLANIF interface's IP address as the management IP address. It is recommended that all Layer 2 switches connected to a gateway be on the same network segment.

Interconnection IP address

It is recommended that the interconnection IP address use a 30-bit mask. The core device uses a smaller host IP address.

Service IP address



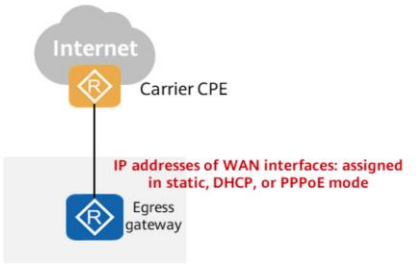
The service IP address is the IP address of a server, host, or gateway.

- It is recommended that the gateway addresses use the same rightmost three digits, such as .254.
- The IP address range of each service must be clearly distinguished. The IP addresses of each type of service terminals must be contiguous and can be aggregated.
- You are advised to use an IP address segment with a 24-bit mask.

HUAWEI

IP Address Planning (2)

Egress gateway

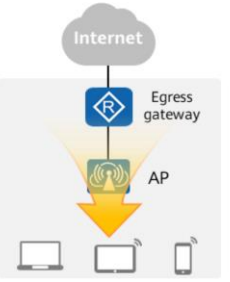


IP addresses of WAN interfaces on egress gateways are assigned by the carrier in static, DHCP, or PPPoE mode. The IP addresses of these interfaces need to be obtained from the carrier in advance.

Devices such as servers and printers

It is recommended that servers and special terminals (such as punch-card machines, printing servers, and IP video surveillance devices) use static IP addresses.

End user

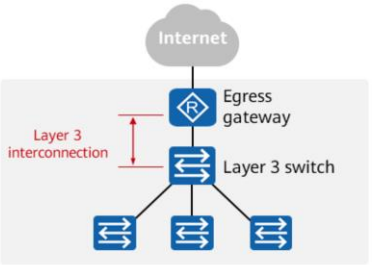


It is recommended that end users be assigned IP addresses in DHCP mode and the gateway provide the DHCP service.

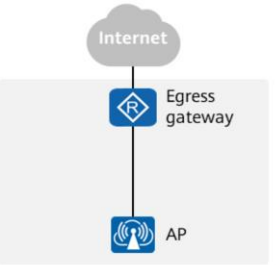
HUAWEI

IP Address Planning (3)

LAN-side devices



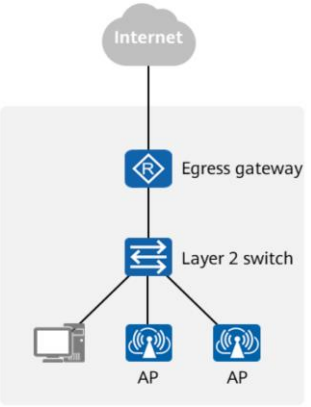
When the egress gateway interconnects with a Layer 3 switch, it is recommended that the interconnection IP addresses be manually configured in static mode.



It is recommended that the DHCP server be deployed on the gateway to dynamically allocate IP addresses to APs.

HUAWEI

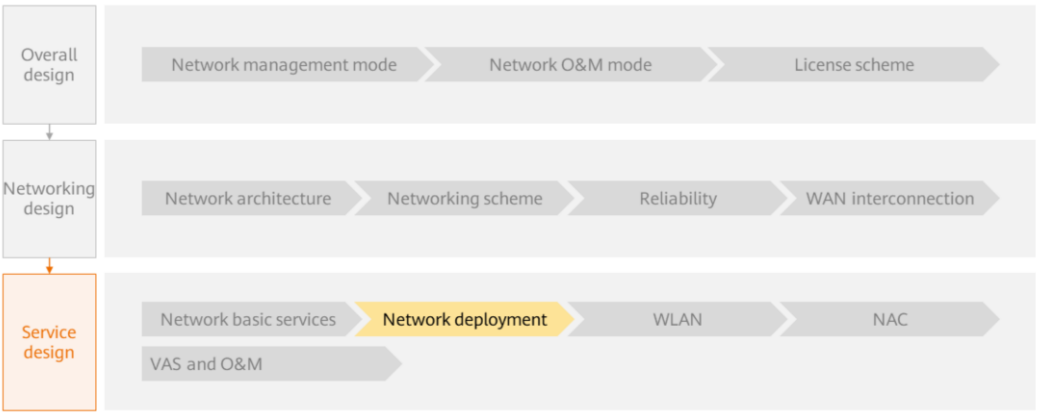
Routing Design



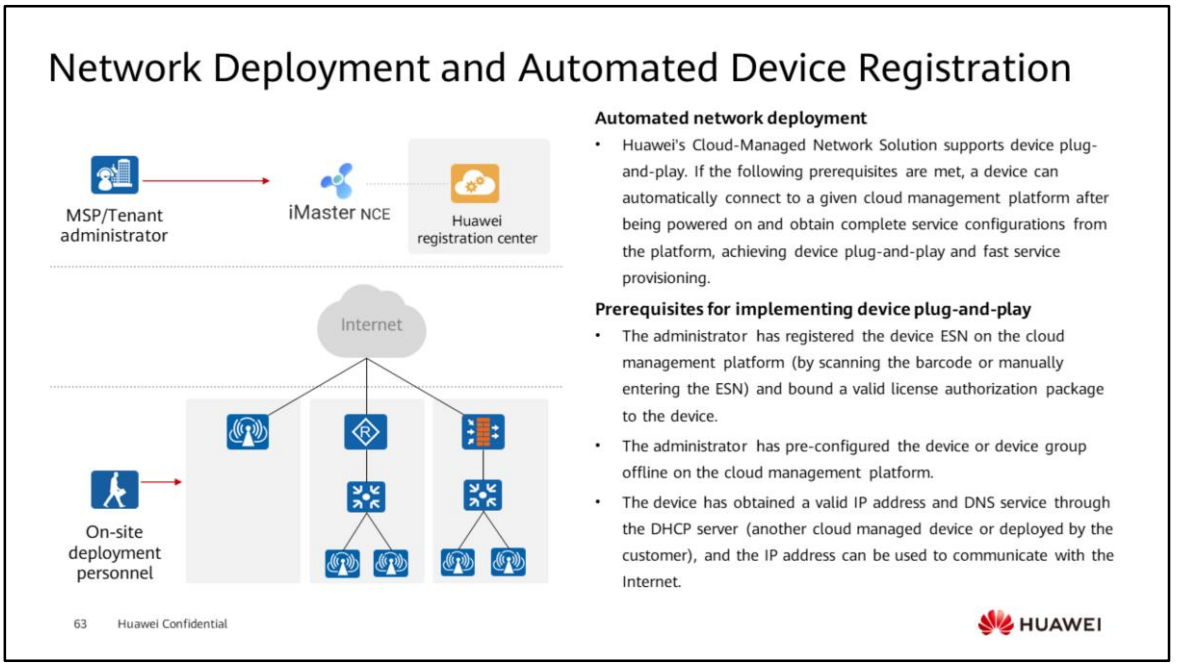
- **Static routes are recommended for small- and medium-sized campus networks.**
- **Internal routing design of the campus network:**
  - APs: After an AP obtains an IP address through DHCP, a default route is generated by default.
  - Switches and gateways: Static routes can be used to meet requirements. No complex routing protocol needs to be deployed.
- **Routing design for the campus egress:**
  - You are advised to configure static routes on the egress device.

HUAWEI

Overall Design Process



HUAWEI



- Traditional network deployment requires the participation of engineering installation personnel and network commissioning personnel. This results in long installation and commissioning periods and slow service provisioning, and consumes a large amount of manpower. The CloudCampus Cloud-Managed Network Solution supports automated network deployment.

Device Registration

- Enterprise network scenarios are complex, and different device registration modes are required in different scenarios. Huawei devices support multiple registration modes to meet complex networking requirements. The registration model also varies with device model. The following table lists the registration modes of different devices:

Device	Registration Mode					
	Registration Center	DHCP Option	USB Flash Drive	Mobile App	Web/CLI	Email
AP	√	√		√	√	
Firewall	√		√		√	
Switch	√	√			√	
AR	√	√	√		√	√

- On the same network, devices at different locations may use different registration modes. Typically, devices to be registered on a network are classified into:
  - Egress gateway
  - Intranet (LAN-side) device

HUAWEI

Egress Gateway Registration

- Egress line configuration of the egress gateway usually depends on the carrier network. Therefore, the registration mode is complex.
- In the CloudCampus Solution, ARs, firewalls, and APs can function as egress gateways.
- The registration modes of egress gateways vary with scenarios. For details, see the following table.
- If a DHCP server exists on the live network, the egress gateway can be registered using the DHCP option only when the DHCP server supports DHCP Option 148.
- In LAN-WAN convergence scenarios, ARs support email-based and DHCP option-based registration modes. However, in the DHCP option-based registration mode, the DHCP server must support DHCP Option 148.

Networking Scenario		Recommended Egress Gateway Registration Mode		
		AR	Firewall	AP
Huawei public cloud scenario	The egress gateway cannot automatically obtain an IP address.	Web system	Web system	CloudCampus APP
	The egress gateway can automatically obtain an IP address.	Registration center		
MSP-owned cloud scenario	The egress gateway cannot automatically obtain an IP address.	Web system	Web system	CloudCampus APP
	The egress gateway can automatically obtain an IP address.	DHCP option	Web system	DHCP option
LAN-WAN convergence scenario (where the SD-WAN function needs to be configured on the egress gateway)		Email	---	---

HUAWEI

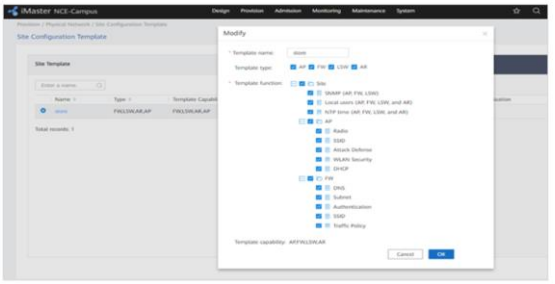
Intranet (LAN-Side) Device Registration

- A LAN-side network refers to the network under the egress gateway, which is mainly built on LAN switches and APs.
- In Huawei public cloud scenarios, the Huawei registration center-based registration mode is recommended.
- If an enterprise does not want to synchronize device information to the registration center, the DHCP option-based registration mode can be used.

Networking Scenario		LAN-Side Device Registration Mode	
		LAN Switch	AP
Huawei public cloud scenario	DHCP options cannot be configured on the network.	Registration center	
	DHCP options can be configured on the network.	Registration center or DHCP option	
MSP-owned cloud scenario	DHCP options cannot be configured on the network.	Web system	CloudCampus APP
	DHCP options can be configured on the network.	Registration center or DHCP option	

HUAWEI

Scenario-Specific Automated Deployment



Solution overview

- The scenario-specific automated deployment function is provided to further simplify network deployment.
- Network configurations, such as SSIDs, can be automatically generated based on the industry scenario selected by a customer. This means that users do not need to plan or configure the network in advance, eliminating the need for complex configuration design and planning. In addition, users can customize a scenario template to plan the site network so that the site network configuration can be quickly generated for deployment.
- Currently, templates for small- and medium-sized office scenarios are preset in the solution. Templates for other scenarios will be added in the future.

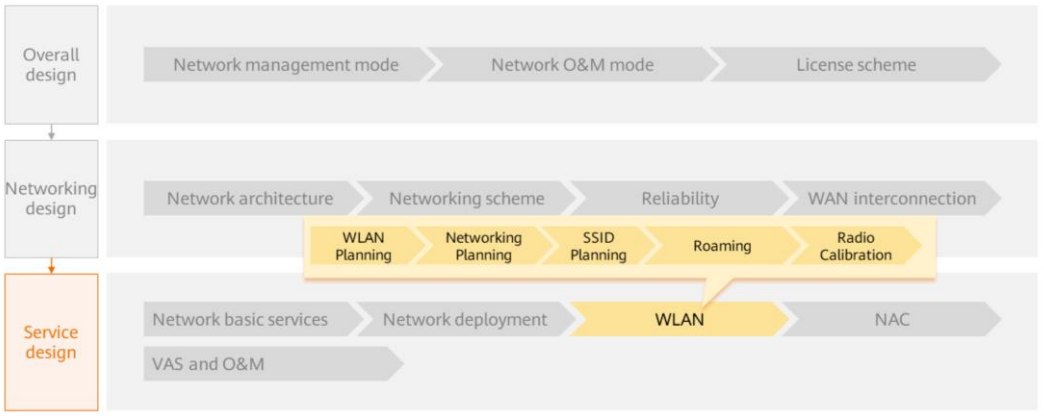
Key design points:

- Select a service scenario, that is, select a scenario template.
- Select related devices and network topologies.
- Confirm the related configuration or adjust the preset configuration.

HUAWEI

- A scenario template includes the following contents and functions related to network deployment:
  - Various networking modes: typical networking models of small- and medium-sized campus networks.
  - Automated DHCP server configuration.
  - Dual uplinks (active/standby mode by default) on the firewall egress for higher reliability.
  - Automated Eth-Trunk configuration.
  - Automated configuration of network services, including SSID, PSK password, QoS, access authentication, traffic monitoring, and application monitoring.

Overall Design Process



HUAWEI

WLAN Planning: Wireless Signal Coverage (1)

Field strength reference

Coverage Area	Field Strength Requirements	Typical Scenario
Key coverage area	–40 to –65 dBm	Dormitory room, library, hotel room, lobby, conference room, office, and exhibition hall
Common coverage area	> –75 dBm	Corridor, kitchen, storeroom, and dressing room
Special coverage area	N/A	Areas that have limitations on or do not allow coverage or installation due to considerations such as service security or property management

Physical coverage reference

Scenario	Coverage Recommendation
Indoor scenario	Plan the coverage radius of 15–20 m for each AP.
Outdoor scenario	Plan the coverage radius of 50–80 m for each AP.
Indoor high-density scenario	Use small-angle directional antennas. During network planning, select AP positions and spacing based on the antenna angle.

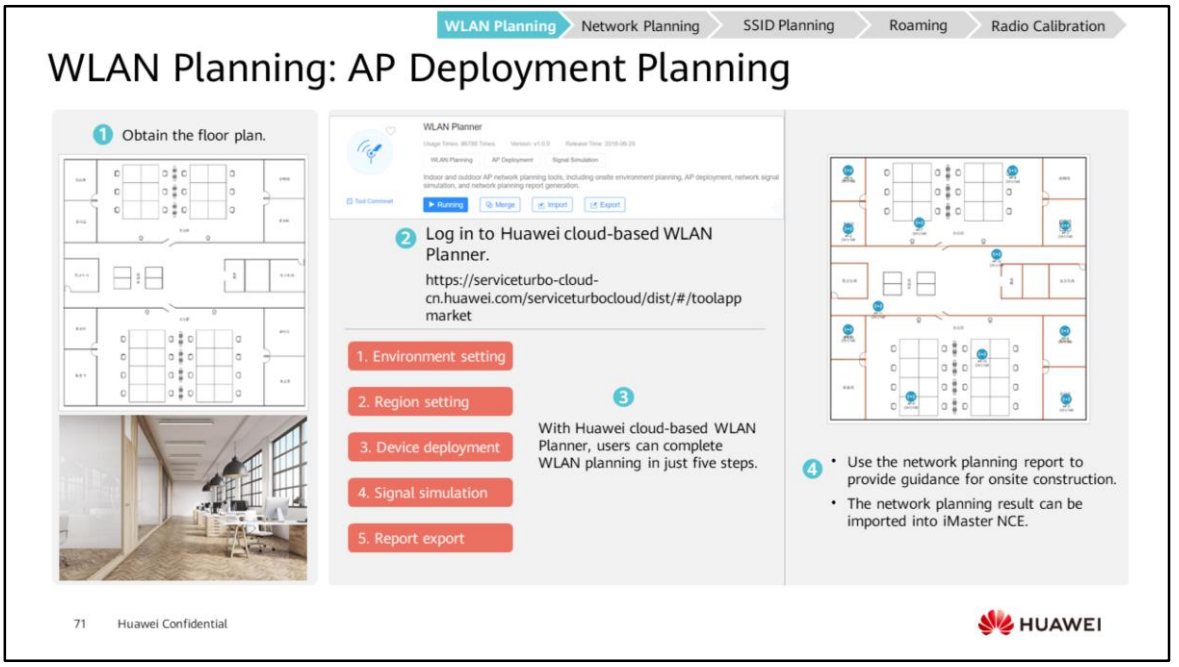
To guarantee a good access and roaming experience, you need to plan proper signal coverage depending on different application scenarios; select proper AP models based on signal coverage demands and AP capabilities; and design a proper WLAN using the WLAN Planner.

HUAWEI

WLAN Planning: Wireless Signal Coverage (2)

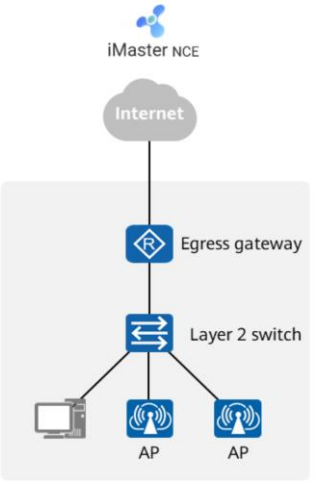
- Observe the following rules when deploying APs:
  - When installing an AP, try to reduce the number of obstacles that signals pass through.
  - Ensure that the front side of an AP faces the target coverage area for good coverage.
  - Place APs away from interference sources, such as electronic devices. Do not deploy microwave ovens, wireless cameras, wireless phones, or other electronic devices in the coverage area.
  - For areas with roaming requirements, keep a 10% to 15% signal overlap between the coverage areas of neighboring APs to ensure smooth STA roaming between APs.

HUAWEI



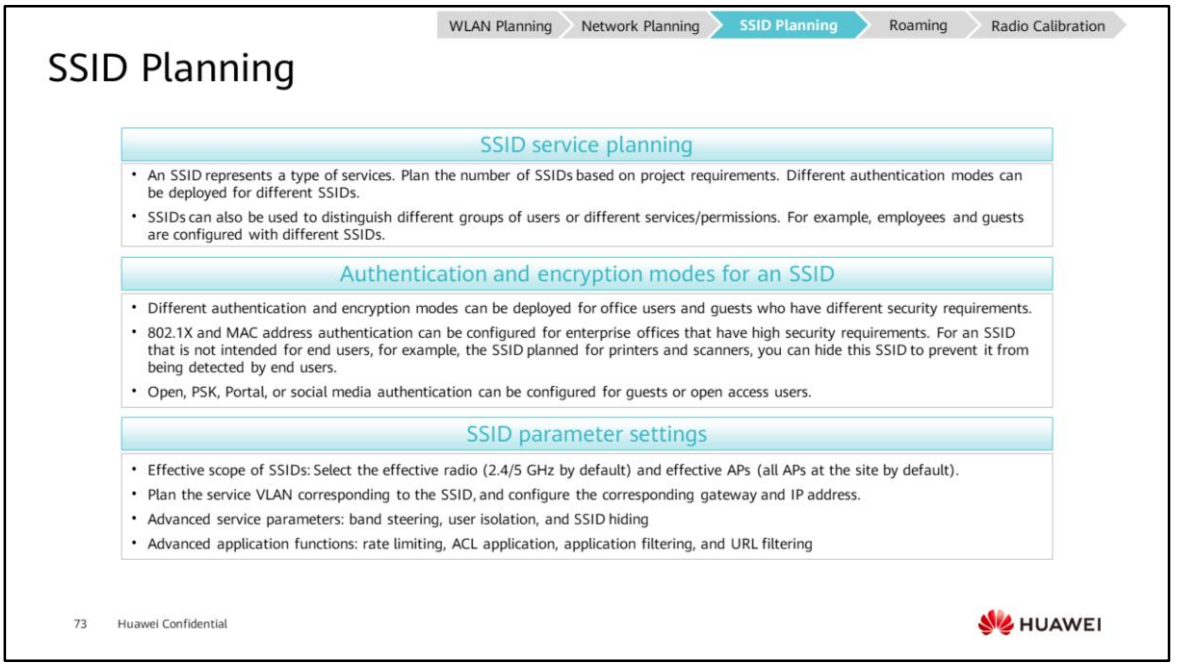
- After WLAN signal coverage requirements are determined based on scenarios, AP deployment can be planned using the WLAN Planner.
  - This tool can be used to create network planning projects based on drawings or Amap.
  - This tool supports region settings, obstacle generation, automated or manual AP deployment with one click, and WLAN signal coverage simulation.
  - This tool can generate network planning files and allows users to export the files. Users can import network planning results on the tenant management page of iMaster NCE to display AP locations and help tenants install APs.
  - CampusInsight allows users to import network planning files. After network deployment, users can view the actual signal heatmap of the network.

Network Planning

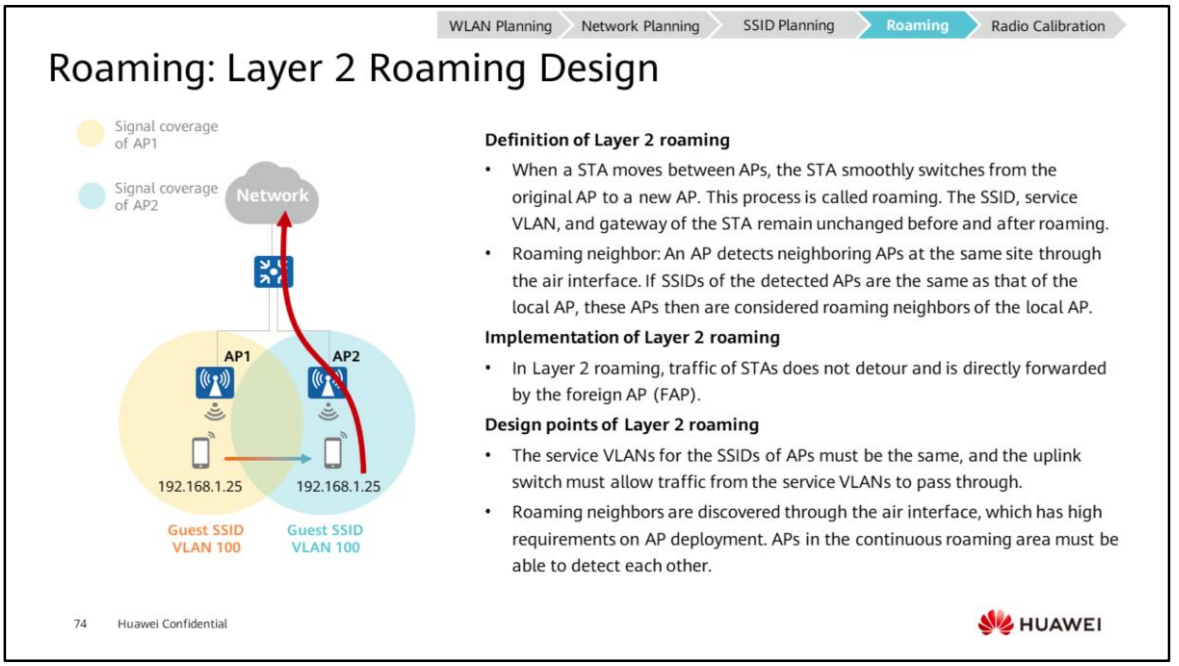


- **Number of APs**: Determine the number of APs based on the WLAN planning result.
- **Access switch selection**: PoE switches are recommended. Select switches based on the number of APs and PoE power supply requirements of APs.
- **Number of access switches**: Determine the number of access switches based on the number of APs, access relationships, and the number of ports on the selected switches.

HUAWEI

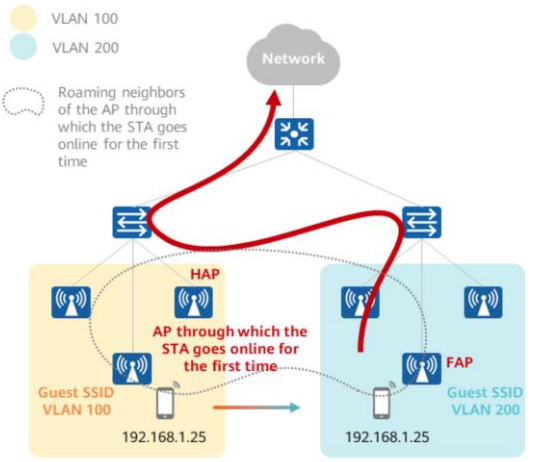


- SSIDs are used to represent WLANs. They are WLAN names displayed on a STA such as a mobile phone when you search for WLANs available for access on the STA. You can select an SSID to access the corresponding WLAN.
- Generally, SSIDs are planned based on user roles or services. Different authentication modes, network access rights, and control policies can be deployed for different SSIDs.
- Portal authentication can be used on open networks (for example, networks for external user access). PSK or PPSK authentication can be used on semi-open networks (such as guest room networks of hotels). 802.1X authentication can be used on secure networks (such as office networks). MAC address authentication can be used for dumb terminals such as printers.
- For an SSID that is not intended for end users, for example, the SSID planned for printers and scanners, you can hide this SSID to prevent it from being detected by end users.



- Definition of Layer 2 roaming: When a STA moves between APs, the STA smoothly switches from the original AP to a new AP. This process is called roaming. The SSID, service VLAN, and gateway of the STA remain unchanged before and after roaming.
- Implementation of Layer 2 roaming:
    - Roaming neighbor: An AP detects neighboring APs at the same site through the air interface. If SSIDs of the APs are the same as that of the local AP, and they can detect each other, these APs are considered roaming neighbors of the local AP. (An AP can establish a maximum of 64 roaming neighbors.).
    - Each AP establishes a wired CAPWAP tunnel (control plane tunnel) with a roaming neighbor to transmit roaming information.
    - The establishment of roaming neighbors is irrelevant to the calibration group. The calibration group is divided based on the management VLAN. Layer 2 roaming neighbors are established based on whether service VLANs are the same.

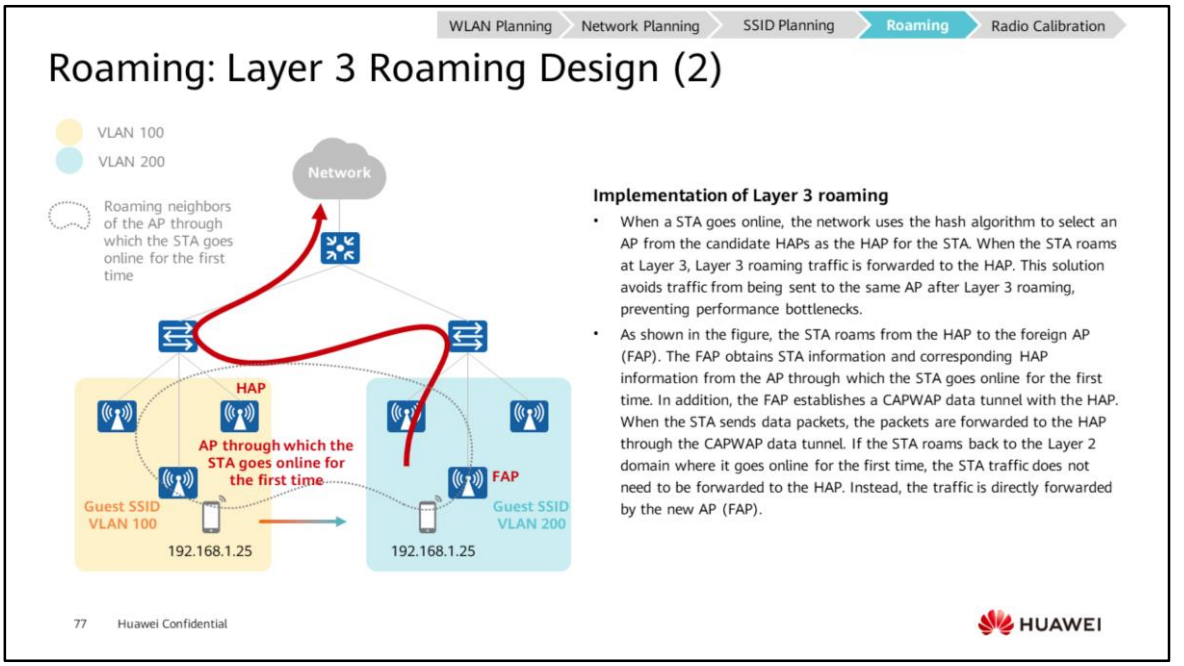
Roaming: Layer 3 Roaming Design (1)



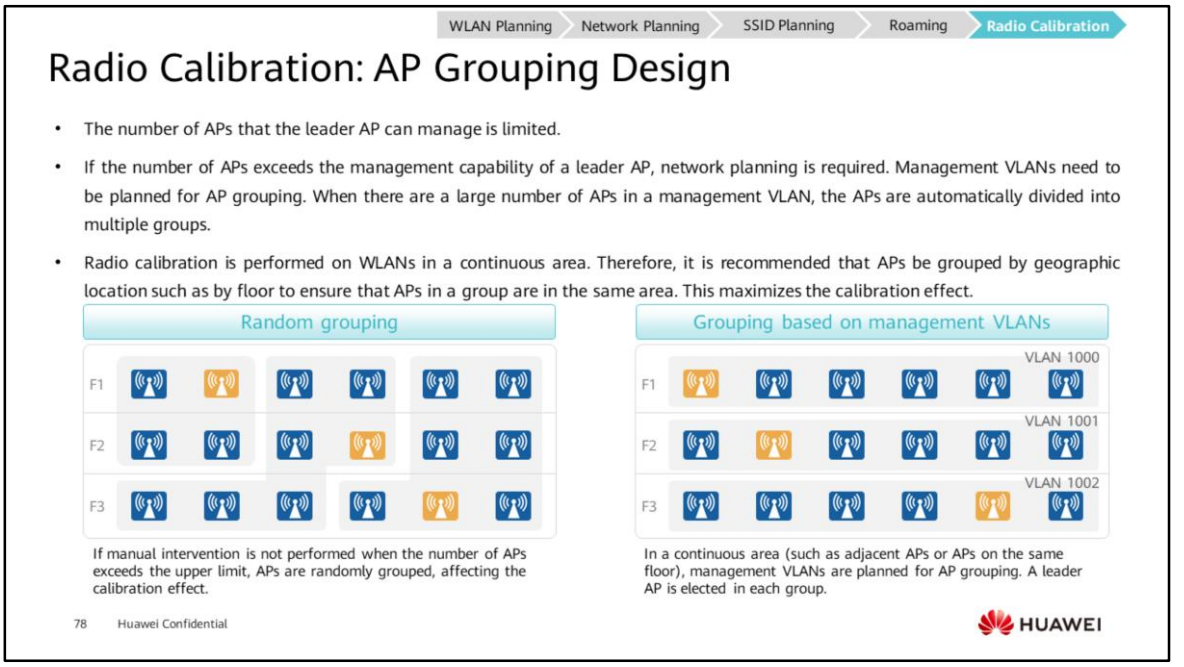
Definition of Layer 3 roaming

- **Layer 3 roaming**: The service VLANs for the SSIDs before and after STA roaming are different and correspond to different gateways. In this case, to keep the IP address of the roaming STA unchanged, the STA traffic needs to be detoured to an AP that resides on the same network segment as the AP through which the STA goes online for the first time.
- **AP through which the STA goes online for the first time**: AP that the STA first associates with in a mobility group
- **Home AP (HAP)**: is selected using the hash algorithm from the neighbors that belong to the same Layer 2 roaming domain as the AP through which the STA goes online for the first time. Once selected, the HAP forwards Layer 3 roaming traffic of the STA.

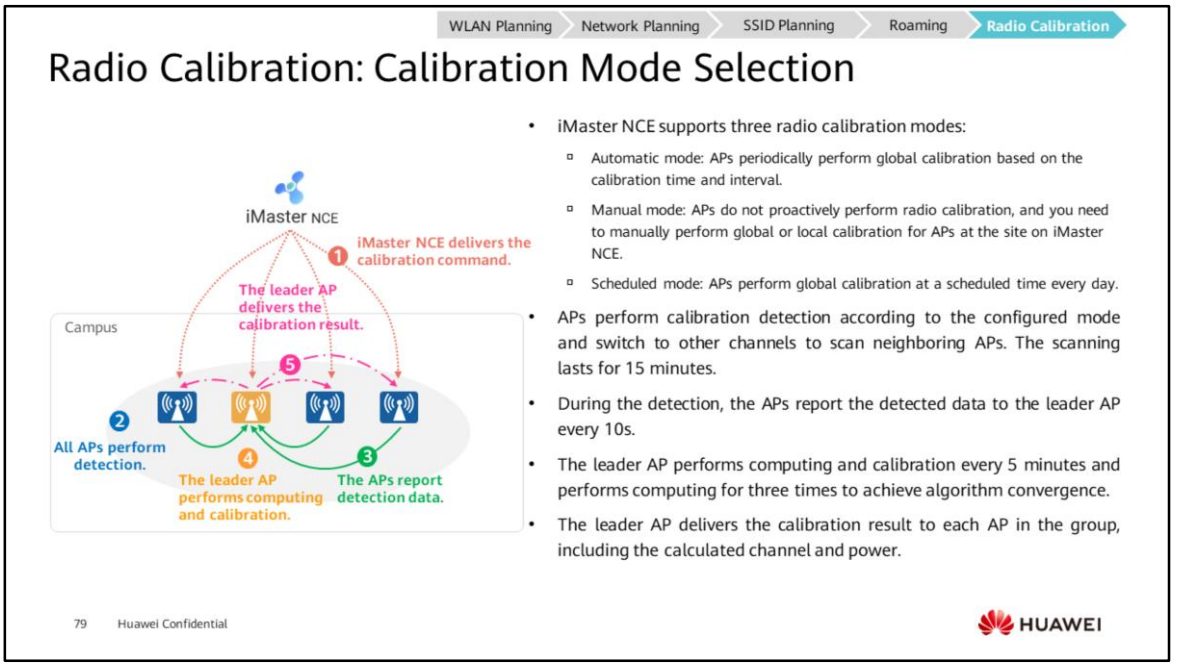
HUAWEI



- Design points of Layer 3 roaming
  - Layer 3 roaming traffic will be detoured to the HAP. Therefore, a large Layer 2 domain needs to be planned at the entrance of the entrance hall so that the detoured traffic can be distributed to different APs during roaming.
  - Each AP supports up to 64 Layer 3 roaming STAs. When there are a large number of Layer 3 roaming STAs, the roaming fails and the STAs need to go offline and go online again. (In actual deployment scenarios, you are advised to deploy Layer 2 roaming on the same floor or in the same area based on the service VLAN planning.)



- A cloud AP is essentially a Fat AP and is independent of each other. Some WLAN services, such as radio calibration, need to be processed centrally. However, no WAC is deployed. To ensure high network reliability and performance and meet local computing requirements, a global control role similar to a WAC is required.
- A leader AP is an AP with strong capabilities in an AP group. It is responsible for global calibration of the entire group (automatically elected in a Layer 2 domain).
- The calibration region is automatically divided based on the management VLAN. On a small network, a leader AP can manage all APs (no more than 128 high-performance APs and no more than 50 low-performance APs). In this case, only one management VLAN needs to be planned for APs.
- When the number of APs exceeds the management specifications of the leader AP, you can divide multiple management VLANs to manage the network. You are advised to plan management VLANs based on floors or physical continuous coverage areas to ensure the continuity of the calibration regions.
- Port isolation cannot be configured on the uplink access switch of the cloud AP. (The calibration group is negotiated through wired-side broadcasting.)



- During scheduled radio calibration, you can enable intelligent radio calibration and use the analyzer to analyze historical data of the WLAN and predict interference sources on the network. During network optimization, APs can avoid possible interference sources on the network in advance to improve the quality of the entire WLAN.
- During deployment, you are advised to perform manual calibration to automatically plan the channels and power of APs after APs are deployed and go online.

Radio Calibration: Channel and Frequency Bandwidth Selection

AP channel selection

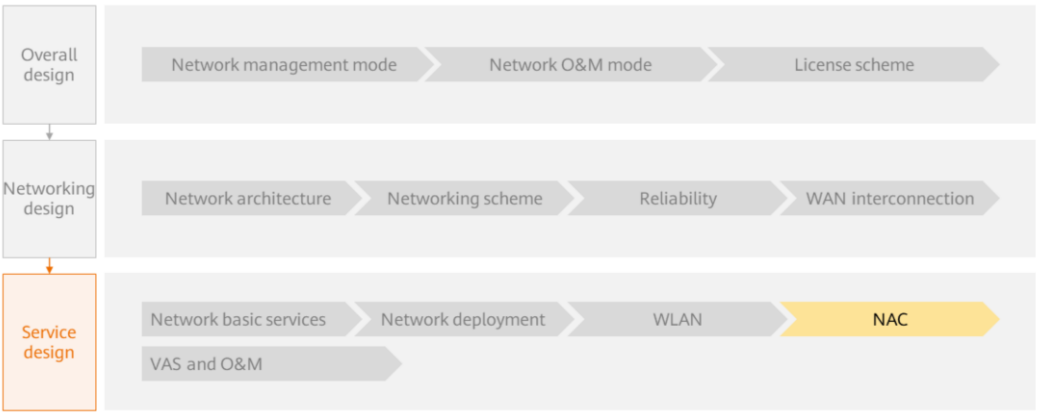
- **2.4 GHz frequency band:** Channel sets 1, 6, and 11 are recommended. If APs are densely deployed, channel sets 1, 5, 9, and 13 are recommended.
- **5 GHz frequency band:** When an AP uses a single 5 GHz radio, it is recommended that high and low frequency channels of neighboring APs be staggered. When an AP uses dual 5 GHz radios, it is recommended that two 5 GHz radios be planned at low and high frequencies respectively.

Frequency bandwidth selection

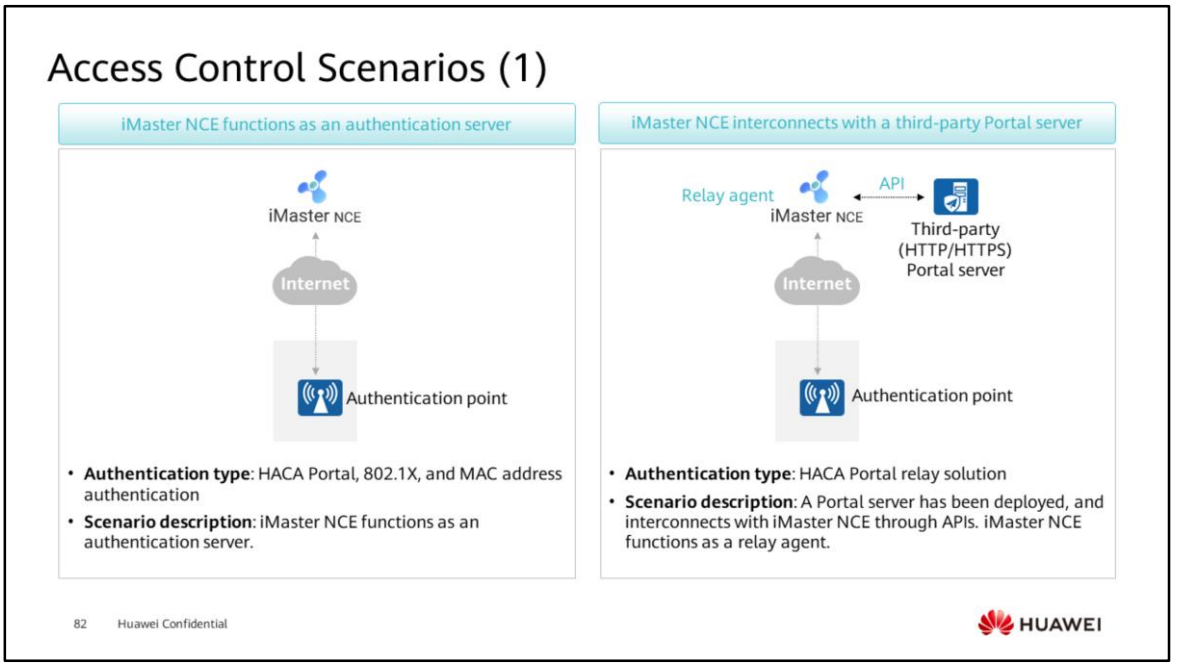
- **2.4 GHz frequency band:** Only the 20 MHz frequency bandwidth can be selected.
- **5 GHz frequency band:** The 40 MHz frequency bandwidth is recommended. The 80 or 160 MHz frequency bandwidth can be used in bandwidth-hungry scenarios.

HUAWEI

Overall Design Process



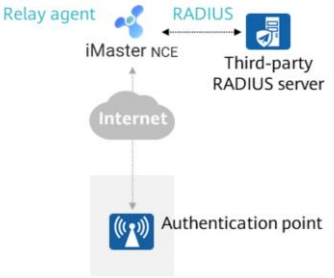
HUAWEI



- Deployment mode recommendation:
  - It is recommended that iMaster NCE functions as an authentication server to implement authentication and policy control.
  - If an enterprise has its own authentication server, the authentication server can interconnect with a third-party server in proxy mode. In addition, iMaster NCE can also implement user management and some policy management functions.
  - If an enterprise uses a third-party authentication server to provide the access authentication function, the device that functions as an authentication point can directly interconnect with the third-party authentication server.

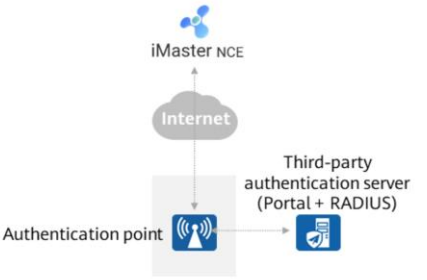
Access Control Scenarios (2)

iMaster NCE interconnects with a third-party RADIUS server



- **Authentication type**: HACA Portal-to-RADIUS solution
- **Scenario description**: A RADIUS server has been deployed, and interconnects with iMaster NCE through RADIUS. iMaster NCE functions as a RADIUS relay agent.

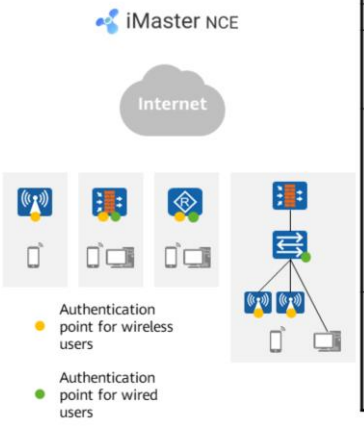
The authentication point directly interconnects with a third-party authentication server



- **Authentication type**: Portal, HTTPS Portal, 802.1X, and MAC address authentication
- **Scenario description**: The network device that functions as an authentication point directly interconnects with a third-party server. (Note: HTTPS Portal is configured using CLIs.)

HUAWEI

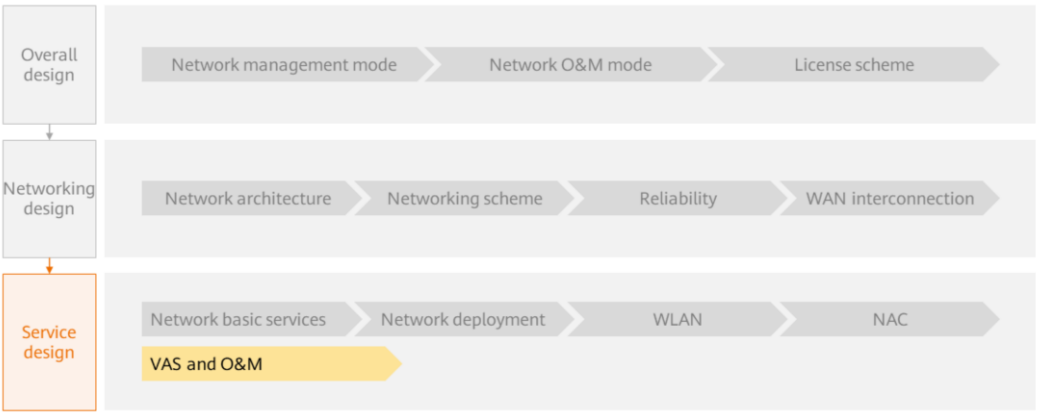
Authentication Solution Deployment



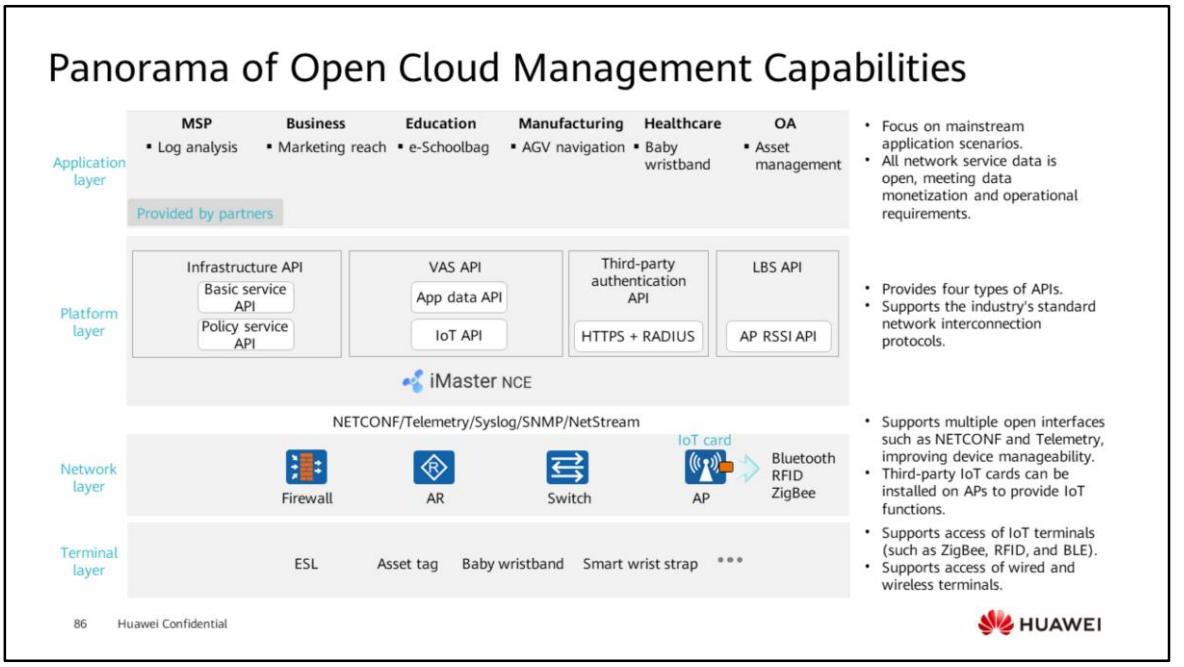
Terminal Role	Guest	Enterprise employee	Dumb terminal
Access Mode	Wireless	Wireless/Wired	Wired
Authentication Type	Portal authentication	• Portal authentication (recommended for wireless users) • 802.1X authentication (optional for both wireless and wired users)	MAC address authentication
Authentication Point	Access devices are recommended as authentication points. This has the following advantages: • Multiple access devices perform user authentication separately, reducing the authentication load. • Authentication points are closer to terminals, improving authentication security. • The configuration planning is simple. If authentication points are deployed at the upper layer, the following factors must be considered: performance specifications of the devices acting as authentication points, Layer 2 isolation at the access layer, and configuration for transparent transmission of 802.1X protocol packets at the access layer.		
Networking Scenario	Single-device networking: The authentication point is the local device, and Portal authentication is recommended. (In the cloud management scenario, ARs and firewalls do not support 802.1X or MAC address authentication.) Other networking scenarios: • APs are used as authentication points for wireless users. • Access switches are used as authentication points for wired users.		

HUAWEI

Overall Design Process

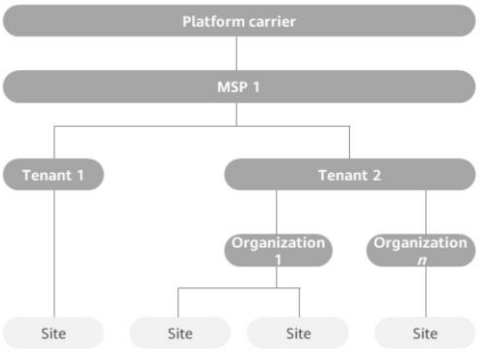


HUAWEI



- Open capabilities at all layers, for all services, and in all scenarios help MSPs, VAS application partners, and customers quickly implement system interconnection, service convergence, and data monetization.

Network Management Design: Management Level



Key points of network management design:

- Management level design is performed to determine whether multiple or multi-level organizations are needed to manage different sites. Huawei's Cloud-Managed Network Solution supports the design of multiple or multi-level organizations for a tenant, which meets the requirements of managing large branches. Each organization can manage multiple sites or sub-organizations.
- **(Optional) Organization planning and design**: A large organization with multiple branches needs to be managed by area. That is, the organization is divided into multiple areas to manage the branches on demand.
- **Site planning**: A network with independent network management services is managed as a site. It can be an independent campus/branch network, or a relatively independent network on a campus/branch network, for example, a network of a building or even a floor. Sites can be flexibly planned based on actual management requirements.
- Rights- and domain-based management of network administrators' rights is adopted.

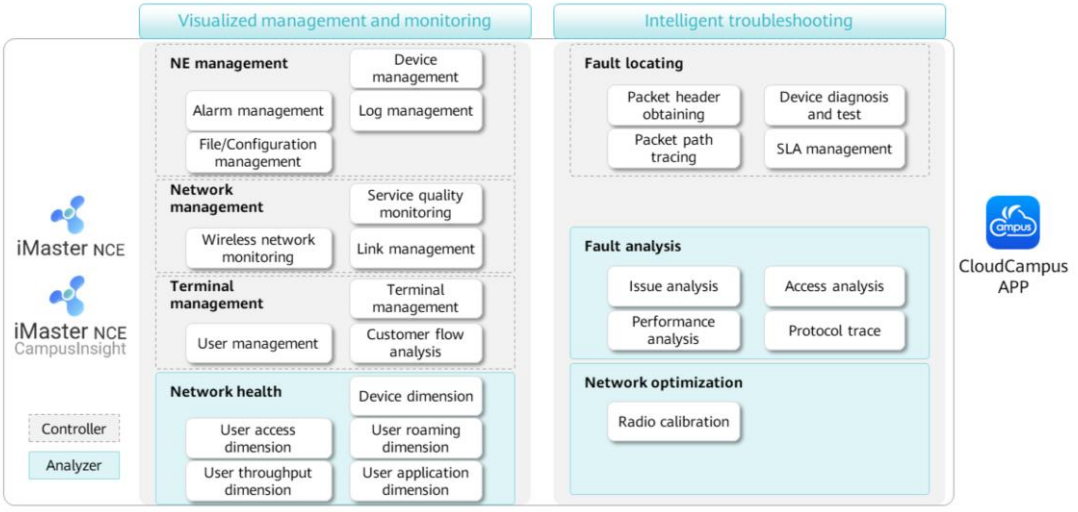
HUAWEI

Network Management Design: Rights- and Domain-Based Management

Authorization Scope	Tenant Level	MSP Level
Rights-based: controls the functions that can be performed	• The super administrator can grant different rights to tenant administrators. • To facilitate rights assignment and management, the system presets three roles with different rights: Monitor, Open API Operator, and Tenant Administrator. • Tenant administrators authorize MSP administrators based on the preceding three roles.	• The super administrator can assign different rights to MSP administrators. • Similarly, there are three types of roles: Monitor, Open API Operator, and MSP Administrator.
Domain-based: controls the range of devices that an administrator can manage	• The minimum granularity of domain-based management is site. That is, tenant administrators perform authorization by site. Different tenant administrators can manage devices at different sites. • When a tenant administrator authorizes an MSP administrator to manage its network, domain-based management is not supported. Instead, the MSP administrator can manage all sites of the tenant.	Domain-based management cannot be configured for MSPs at different levels. Once a tenant authorizes an MSP administrator, the MSP administrator can manage all devices of the tenant whatever level the MSP administrator is at.

HUAWEI

O&M Function Panorama of CloudCampus



HUAWEI

Quiz

1. (Multiple-answer question) Which of the following device deployment solutions can be used on small- and medium-sized campus networks to ensure device reliability? ( )
   A. Deploy firewalls in hot standby mode.
   B. Deploy two ARs as egress gateways.
   C. Deploy switches in stacking mode.
   D. Deploy multi-chassis link aggregation.
2. (True or false) 802.1X authentication is recommended for Wi-Fi networks deployed in stores. ( )
   A. True
   B. False

HUAWEI

1. ABC
2. B

Summary

- This course describes how to use Huawei's CloudCampus Cloud-Managed Network Solution to design small- and medium-sized campus networks. Key design points include the overall design (network management mode design, network O&M mode design, and license scheme design), networking design (network architecture design, networking solution design, WAN interconnection design, and reliability design), and service design (network basic service design, network deployment design, WLAN design, NAC design, security design, QoS design, VAS design, as well as network O&M design).

HUAWEI

Recommendations

- Huawei's CloudCampus Solution V100R020C10 Product Documentation
  - https://support.huawei.com/hedex/hdx.do?docid=DOC1100797422&lang=en

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

CampusInsight Intelligent O&M

HUAWEI

Foreword

- In the Big Data era, traditional O&M based on specified rules cannot meet network O&M requirements, and the lack of automatic O&M becomes increasingly prominent. It is becoming an emergency to use a large amount of data generated on the network to perform intelligent O&M and improve the O&M efficiency.
- Huawei iMaster NCE-CampusInsight, an intelligent network O&M platform, overrides traditional resource monitoring methods and applies AI to the O&M domain. Based on existing O&M data (such as device metrics and client logs), big data, AI algorithms, and more advanced analysis technologies, CampusInsight digitizes user experience on the network to help customers detect network access issues in a timely manner, improving user experience.
- iMaster NCE-CampusInsight uses the Telemetry technology to collect metrics and logs of network devices and detects network exceptions based on real service traffic. The big data platform supports centralized data collection, storage, and analysis to process big data efficiently.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the pain points and requirements of intelligent O&M of campus networks.
  - Describe the logical architecture and external interfaces of CampusInsight.
  - Describe the application scenarios and deployment modes of CampusInsight.
  - Describe the main functions and applications of CampusInsight.
  - Complete major operations on CampusInsight.

HUAWEI

Contents

1. **CampusInsight Overview**
2. CampusInsight Functions and Demonstration
   - Telemetry
   - Visibility
   - Analysis
   - Optimization
   - Assurance

HUAWEI

Challenges Facing Campus Network O&M

Precise detection

Traditional O&M is based on SNMP and data is collected in minutes. Once a fault occurs, data at the fault occurrence time cannot be obtained in real time and no convenient backtracking method is available.

Difficult fault locating and analysis

Experience detection

During traditional O&M, only device metrics are monitored. However, user experience may be poor when the metrics are normal. Traditional O&M lacks means of correlatively analyzing the client and network.

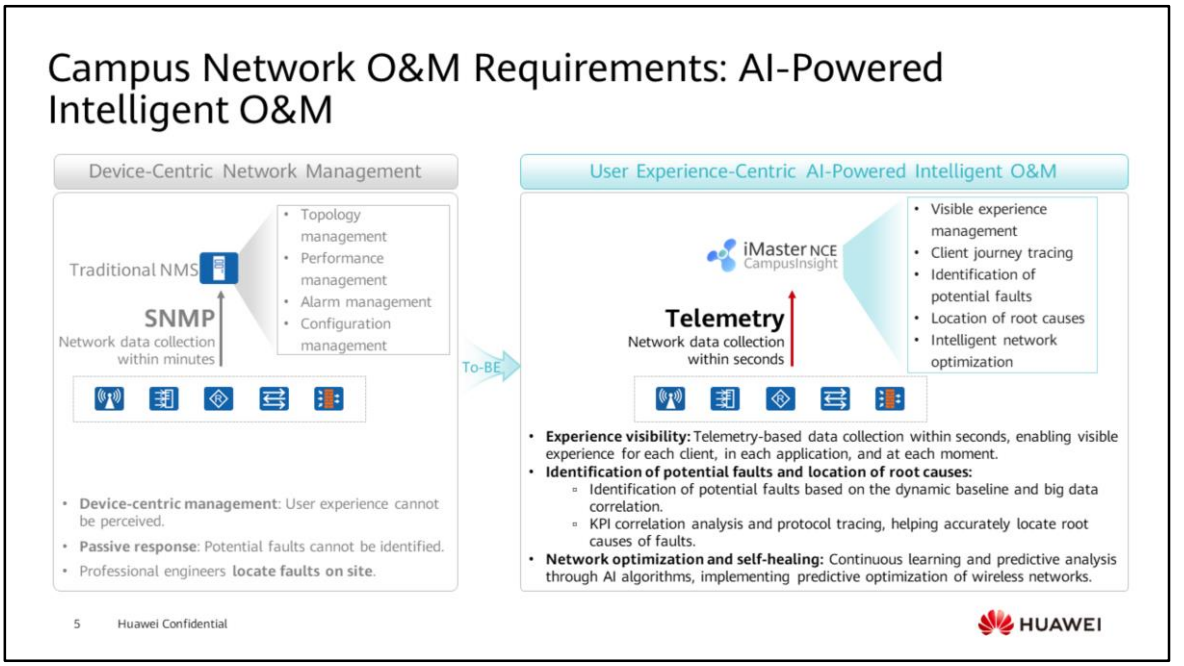
Difficult user experience measurement

Issue identification

During traditional O&M, network faults can be detected only after receiving clients' complaints. As a result, faults cannot be effectively and proactively identified and analyzed.

Difficult to proactively identify problems

HUAWEI



- Customer benefits: The transformation is to improve efficiency by using algorithms. With scenario-based continuous learning and expert experience, intelligent O&M frees O&M personnel from complex alarms and noises, making O&M more automatic and intelligent.

CampusInsight: Improving User and Service Experience Based on Prediction and AI

Real-time Experience Visibility

- **Each region:** Intuitively display the network status and user experience on the entire network or in each region through multi-dimensional evaluation of wired and wireless network health.
- **Each client:** Display network experience (who, when, which AP to connect, experience, and issue) of all clients in real time throughout the journey.
- **Each application:** Perceive experience of audio and video applications in real time, demarcate faulty devices quickly and intelligently, and analyze the root cause of poor quality.

Fault Locating Within Minutes

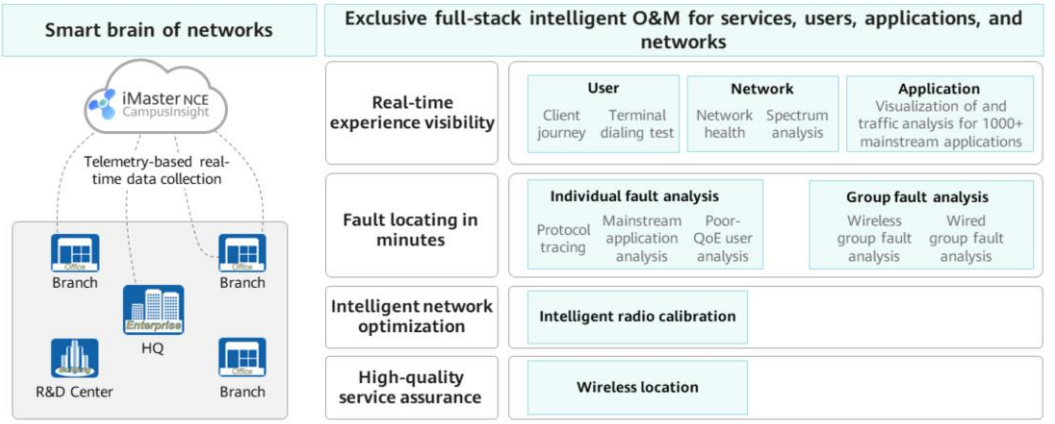
- **Proactive issue identification:** Proactively identify 85% of potential network issues through the AI algorithms that are continuously trained via Huawei's 200,000+ terminals.
- **Fault locating within minutes:** Locate issues within minutes, identify the root causes of issues, and provide effective fault rectification suggestions based on the fault reasoning engine.
- **Intelligent fault prediction:** Learn historical data through AI to dynamically generate a baseline, and compare and analyze the baseline with real-time data to predict possible faults.

Intelligent Network Optimization

- **Real-time simulation feedback:** Evaluate channel conflicts on wireless networks in real time and provide optimization suggestions based on the neighbor relationship and radio information of devices on each floor.
- **Predictive optimization:** Identify edge APs, predict the load trend of APs, perform predictive optimization on wireless networks, and compare the gains before and after the optimization based on historical data analysis. This practice improves the network-wide performance by 50%+ (Tolly-verified).

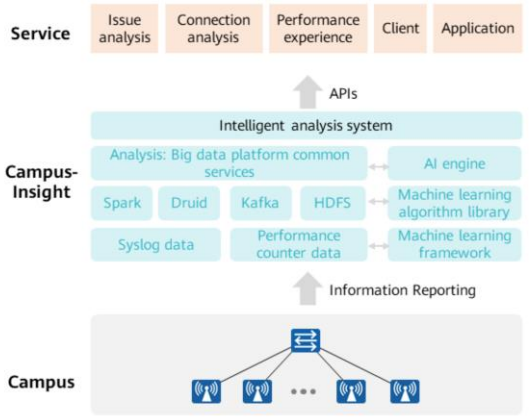
HUAWEI

CampusInsight: Enabling Exclusive Full-Stack Intelligent O&M



HUAWEI

CampusInsight: Logical Architecture



Business Services: Data analysis service based on scenarios

- **Issue identification:** intelligent identification of connection, air interface performance, roaming, and device issues.
- **Access analysis & performance experience:** analysis of connection and performance experience issues.
- **User and network profiles:** client journey retrospection and AP details analysis.
- **Application analysis:** audio and video quality detection.

Data Analysis: Big data platform and data analysis service

- **Data storage:** real-time preprocessing flow, offline distributed processing flow, and data storage service.
- **Data analysis:** mode identification, AI engine, and data aggregation and query.

Data Collection: Second-level performance counter and log collection

- **Data collection:** multi-dimensional data related to client, radios, APs, and client logs.

HUAWEI

CampusInsight: External Interfaces

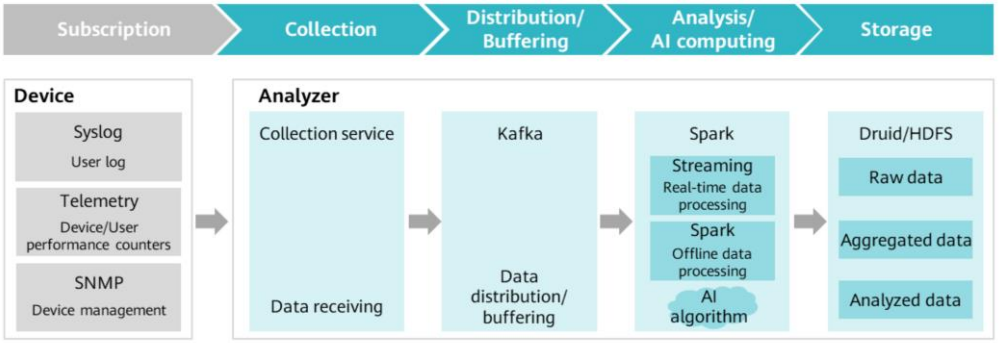
- The CampusInsight southbound interface implements interconnection between CampusInsight and devices, enabling CampusInsight to manage devices. CampusInsight supports the following southbound interface types: SNMP, HTTP2+ProtoBuf, Syslog.

SNMP	HTTP2+ProtoBuf	Syslog
• Supports standard SNMPv2c and SNMPv3. • SNMP can be used to connect CampusInsight to network devices. • SNMP is an application-layer network management protocol based on TCP/IP. It uses UDP as the transport-layer protocol to manage network devices that support proxy processes.	• CampusInsight uses HTTP2+ProtoBuf interfaces to collect device metric packets. • The security layer of the HTTP2 protocol uses SSL and TLS to authenticate and encrypt communication channels. • ProtoBuf is a data serialization protocol developed by Google (similar to XML, JSON, and hessian). ProtoBuf can serialize data and is widely used in data storage and communication protocols.	• The Syslog protocol is a standard for forwarding system logs on an IP network. • Industry standard protocol used to record device logs. • CampusInsight receives log data reported by devices through the Syslog protocol.

HUAWEI

- ProtoBuf: Protocol Buffers

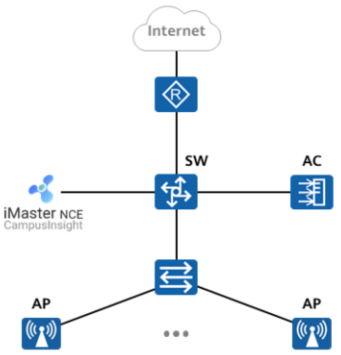
CampusInsight: Data Processing Flowchart



After data subscription, the collection service collects data in seconds. After the data is buffered and distributed by a high-throughput distributed message system, each application service completes data analysis and computing based on AI algorithms and expert experience, and saves processed data to a fast and column-based distributed data storage system. Then pages can access the data to display related functions.

HUAWEI

CampusInsight Deployment Scenario: Independent Deployment (Local Deployment)



Scenario Description

- CampusInsight is deployed independently on the network.
- Supports intelligent analysis of wireless and wired network devices in campus enterprises.

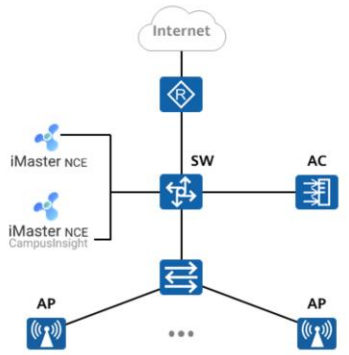
Network Description

The following networking modes are supported:

- All ACs (including standalone ACs, native ACs, and ACUs) + Fit AP
- All ACs (including standalone ACs, native ACs, and ACUs) + central AP + RU
- Hybrid networking of switches and WLAN devices

HUAWEI

CampusInsight Deployment Scenario: CloudCampus Deployment (Local Deployment)



Scenario Description

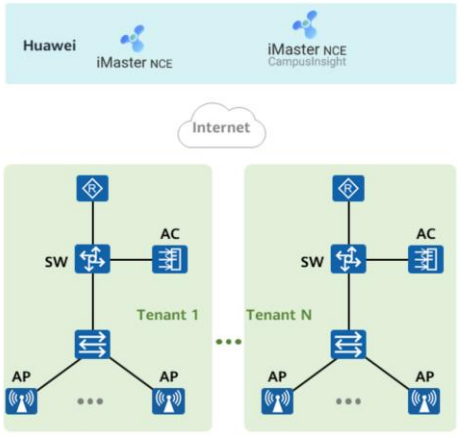
- CampusInsight is co-deployed with iMaster NCE-Campus.
- An enterprise purchases the Huawei cloud management platform (iMaster NCE-Campus and CampusInsight) and deploys the platform in the enterprise data center. O&M personnel of the enterprise maintains the cloud management platform and enterprise network. The platform is used within the enterprise. The enterprise purchases related licenses from the Huawei service team.

Network Description

- Management and intelligent analysis are supported for Huawei cloud switches, cloud ACs, and cloud APs. For details about supported device models, see the iMaster NCE-CampusInsight Specifications List.
- An enterprise deploys CampusInsight and iMaster NCE-Campus in the enterprise data center, and manages devices through iMaster NCE-Campus. CampusInsight synchronizes device management information from iMaster NCE-Campus.

HUAWEI

CampusInsight Deployment Scenario: CloudCampus Deployment (Huawei Public Cloud Scenario)



Scenario Description

- In the Huawei public cloud scenario, the cloud management platform (iMaster NCE-Campus and CampusInsight) is uniformly managed by the cloud management and operations team of Huawei and provides the SaaS service for end users.
- Tenant network devices are connected to the Huawei cloud management platform through the carrier network.

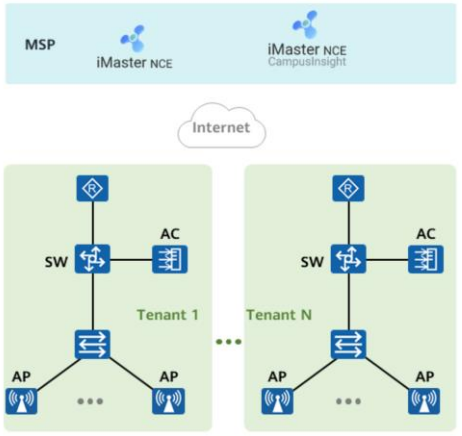
Network Description

- Management and intelligent analysis are supported for Huawei cloud switches, cloud ACs, and cloud APs. For details about supported device models, see the iMaster NCE-CampusInsight Specifications List (CloudCampus).
- Huawei cloud management and operations team deploys CampusInsight and iMaster NCE-Campus on Huawei data public cloud, and manages devices through the iMaster NCE-Campus. CampusInsight synchronizes device management information from iMaster NCE-Campus.

HUAWEI

- SaaS: software as a service

CampusInsight Deployment Scenario: CloudCampus Deployment (MSP Self-built Cloud Scenario)



Scenario Description

- MSPs purchase the controller (iMaster NCE-Campus) and analyzer (iMaster NCE-CampusInsight) for operational purposes. Software can be deployed in their data centers or on the public cloud IaaS.
- MSPs develop their tenants and provide SaaS services for tenants.
- Tenant network devices connect to the MSP data center or public cloud IaaS through the carrier network.

Network Description

- Management and intelligent analysis are supported for Huawei cloud switches, cloud ACs, and cloud APs. For details about supported device models, see the iMaster NCE-CampusInsight Specifications List (CloudCampus).

HUAWEI

- IaaS: infrastructure as a service

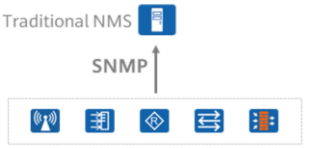
Contents

1. CampusInsight Overview
2. **CampusInsight Functions and Demonstration**
   - Telemetry
   - Visibility
   - Analysis
   - Optimization
   - Assurance

HUAWEI

Meeting Real-Time Analysis Requirements Based on the Telemetry Technology

SNMP



- Protocol development stagnation --- SNMP is designed for limited processing capabilities.
- Polling technology --- The minute-level polling cycle cannot meet the service requirements of real-time management.
- Rigid data structure --- Fixed data structures are defined, and multiple data requests are required to complete each effective data collection.

Telemetry

- Based on HTTP2 and ProtoBuf.
- Subscription-based release and on-demand use.
- Efficient encoding and decoding technology to obtain multiple data records at a time, implementing second-level data acquisition.

The quasi-real-time data acquisition capability is the key dependency for the analyzer to mine data.

HUAWEI

Monitoring Telemetry Metrics on Wireless Networks

- Monitor key metrics on wireless networks based on the telemetry technology, display the wireless network quality from the AP, radio, and client dimensions, and proactively identify air interface performance issues, such as weak-signal coverage, high interference, and high channel utilization.

Real-time data display	Automatic issue identification
Display key metrics on wireless networks and single out abnormal metric status.	Automatically identify air interface performance issues based on AI algorithms, correlative analysis, and exception modes.

Telemetry metric collection on wireless networks

Measured Object	Measurement Metric	Supported Device Type	Default Collection Period
AP	CPU usage, memory usage, and number of online clients	AP	1 minute
Radio	Number of online clients, channel utilization, noise, traffic, backpressure queue, interference rate, and power	AP	1 minute
Client	RSSI, negotiated rate, packet loss rate, and latency	AP	1 minute

HUAWEI

- RSSI: Received Signal Strength Indicator

Monitoring Telemetry Metrics on Wired Networks

- Analyze telemetry metric data of devices, interfaces, and optical links collected through Telemetry on wired networks, and proactively monitor and predict network issues.

Real-time display of key metrics

Display key metrics on wired networks in real time, including top N devices and historical trend.

Exception detection based on dynamic baselines

Use AI algorithms to predict the baselines of key metrics, such as the CPU/memory usage. Identify network metric deterioration before service interruptions through comparison with dynamic baselines.

Telemetry metric collection on wired networks

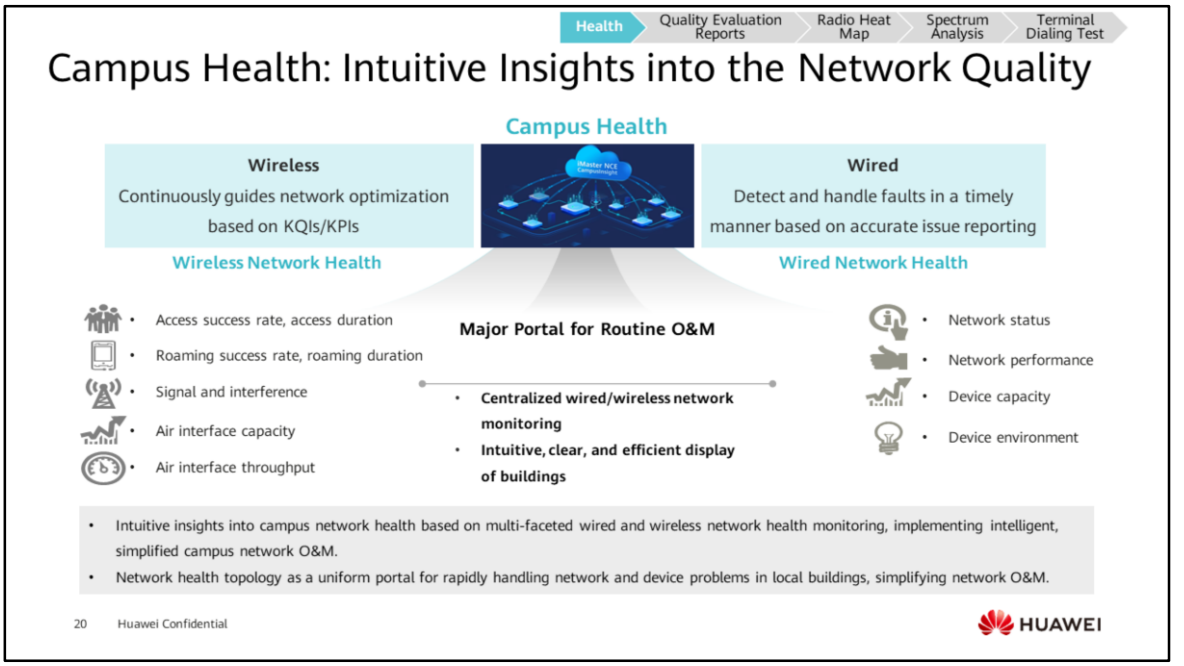
Measured Object	Measurement Metric	Supported Device Type	Default Collection Period
Device/ Card	CPU usage	Switch and AC	1 minute
	Memory usage	Switch and AC	1 minute
Interface	Number of received/sent packets, number of received/sent broadcast packets, number of received/sent multicast packets, number of received/sent unicast packets, number of received/sent packets that are discarded, and number of received/sent error packets	Switch and AC	1 minute

HUAWEI

Contents

1. CampusInsight Overview
2. **CampusInsight Functions and Demonstration**
   - Telemetry
   - Visibility
   - Analysis
   - Optimization
   - Assurance

HUAWEI



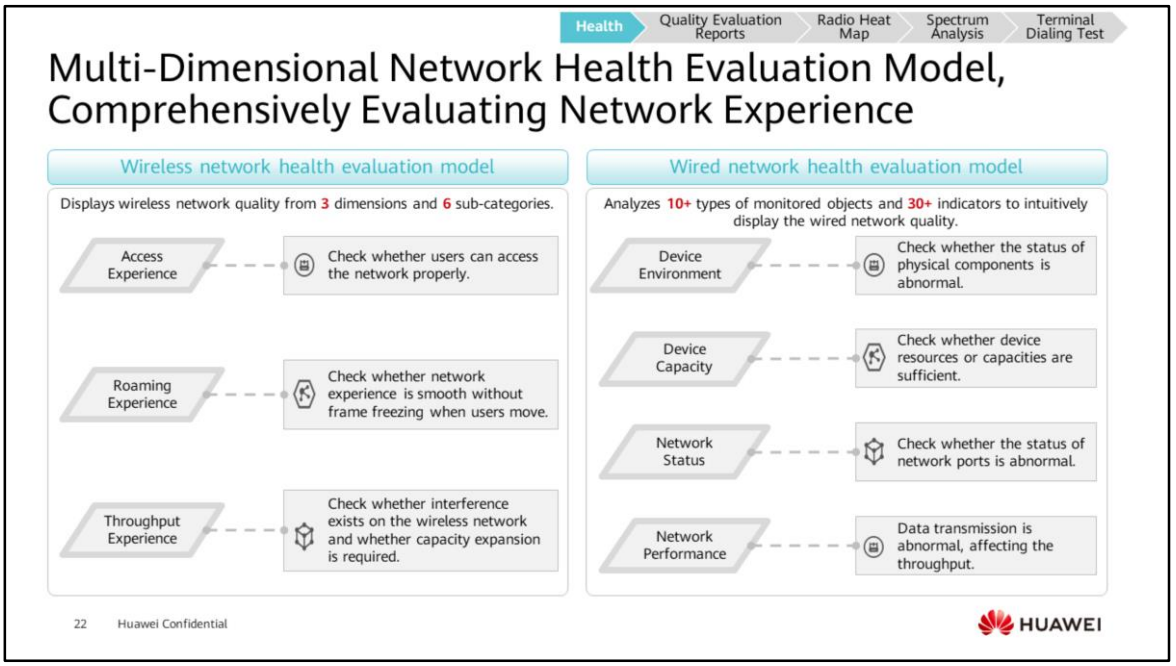
- KQI: key quality indicator

Building Topology: Displaying Key KPI and Network Issues Based on Buildings

Intuitive, clear problem display from the building perspective

- Your network is abnormal.
  Device off-boarding, intermittent port disconnection, AP off-boarding.
- An Access fault occurs on your network.
  Users fail to be authenticated in a batch.
- Traffic congestion occurs on your network.
  Port or queue congestion.
- Packet error occurs on your network.
  The number of error packets on a network port exceeds the threshold, or the number of error packets keeps increasing.

HUAWEI



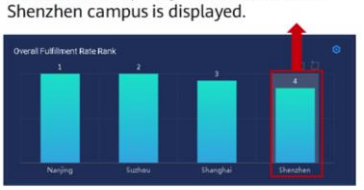
- Wireless Network Health Evaluation Model:
  - Access Experience
    - Access success rate: Association/Authentication/DHCP success rate.
    - Access duration: Association/Authentication/DHCP duration.
  - Roaming Experience
    - Roaming fulfillment rate: Roaming success rate/Roaming duration.
  - Throughput Experience
    - Signal and interference: STA signal strength and interference rate.
    - Capacity fulfillment rate: Channel utilization/Number of users.
    - Throughput fulfillment rate: Interference rate/Non-5G-prior access/Air interface congestion fulfillment rate.
- Wired Network Health Evaluation Model:
  - Device Environment: Fault of a device, board, fan, power supply, or file system.
  - Device Capacity: ARP/MAC/FIB entry capacity, ACL resources, storage capacity.
  - Network status: Intermittent port disconnection, port suspension, optical module exception.
  - Network performance: Port congestion, queue congestion, port error packets.

Case 1: Wireless Network Health (1)

Case: Using the wireless health of CampusInsight, O&M personnel at a site find that the network quality of Shenzhen campus is much lower than that of other campuses. The access success rate, coverage, and throughput fulfillment rate of Shenzhen campus are below the industry benchmark. By drilling down data, O&M personnel successfully locate the cause and rectify the fault.

1. Choose Health > Wireless Health.

2. In fulfillment rate rankings, the fulfillment rate of Shenzhen campus is lower than that of other campuses. Click Shenzhen. The quality evaluation data of Shenzhen campus is displayed.



3. Check data of the multiple dimensions. It is found that the quality evaluation result of the coverage is Good, which is below the industry benchmark. Click the RSSI fulfillment rate metric to drill down data for analysis.

HUAWEI

Case 1: Wireless Network Health (2)

Case: Using the wireless health of CampusInsight, O&M personnel at a site find that the network quality of Shenzhen campus is much lower than that of other campuses. The access success rate, coverage, and throughput fulfillment rate of Shenzhen campus are below the industry benchmark. By drilling down data, O&M personnel successfully locate the cause and rectify the fault.

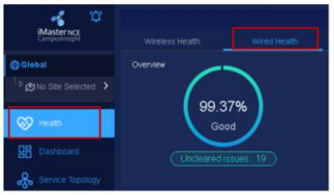
4. On the RSSI fulfillment rate analysis page, two weak-signal coverage issues are detected. Click the link in the Issue Name column to go to the issue details page and check the specific issue.

HUAWEI

Case 2: Wired Network Health

Case: The O&M personnel at a site use CampusInsight to check the wired network health and find that the network performance is abnormal. The O&M personnel view the anomaly details, find the abnormal monitoring item, go to the corresponding issue analysis page to view the root cause, and rectify the anomaly.

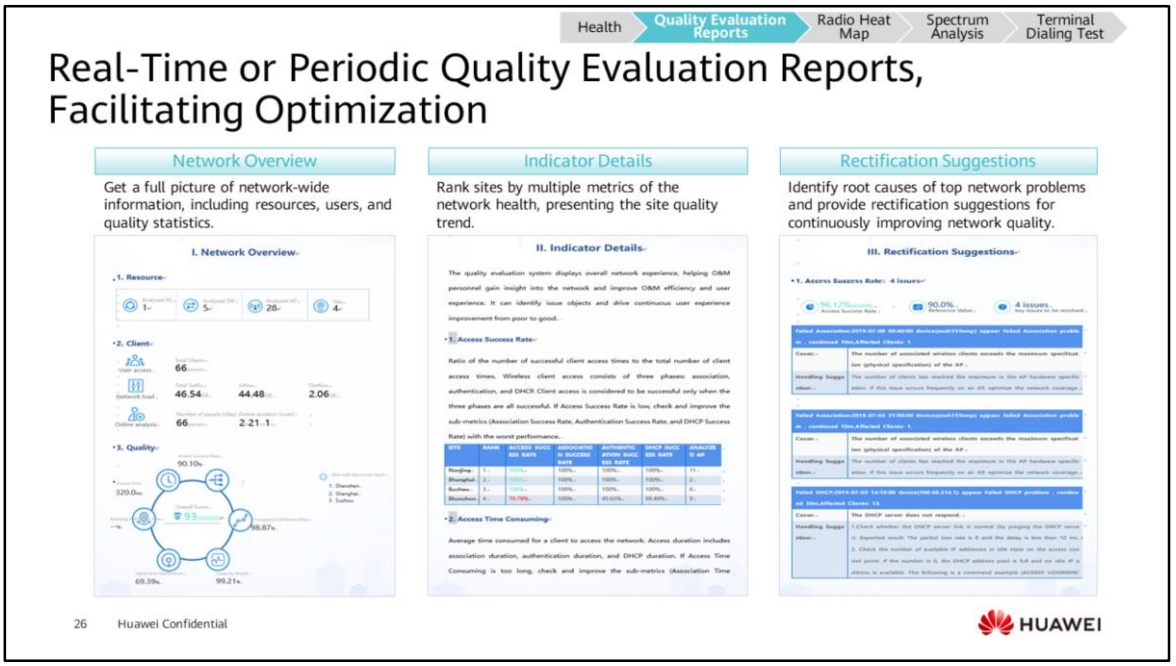
1. Choose Health > Wired Health.



2. If Network Performance is abnormal, click View Details.

3. View the abnormal monitoring item. Click Issues to go to the issue analysis page, view the root cause and rectification suggestions, and resolve the anomaly.

HUAWEI

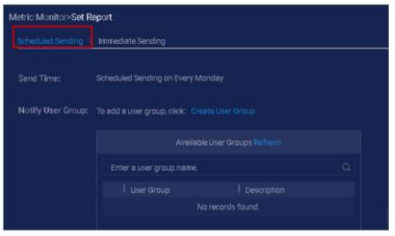


- CampusInsight provides professional evaluation report services. Specifically, network quality evaluation reports on the network overview, indicator details, and rectification suggestions are generated in real time or periodically. These deliver measurable network experience.

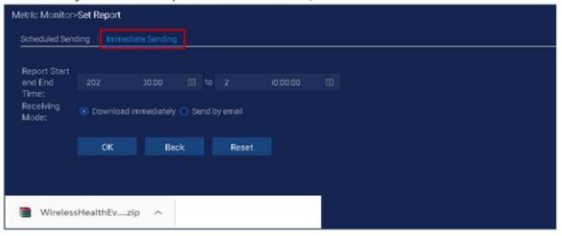
Case 3: Quality Evaluation Reports

1. Choose Health. On the Wireless Health page that is displayed, click Report Export in the upper right corner.

2. Scheduled Sending: Set Send Time and Notify User Group for the report. The report then is sent to the specified email boxes at the specified interval.



3. Immediate Sending: Set Report Start and End Time and Receiving Mode for the report. The report then is generated immediately. (Download immediately: The report can be directly downloaded from the browser. Send by email: The report is immediately sent to the specified email boxes.)



HUAWEI

Health | Quality Evaluation Reports | Radio Heat Map | Spectrum Analysis | Terminal Dialing Test

Radio Heat Map: Visualizing Network Coverage Simulation

Comparison with planning: Is the coverage of the deployed wireless network consistent with the planning?

Simulate the coverage based on the actual radio power and power after the network is deployed.

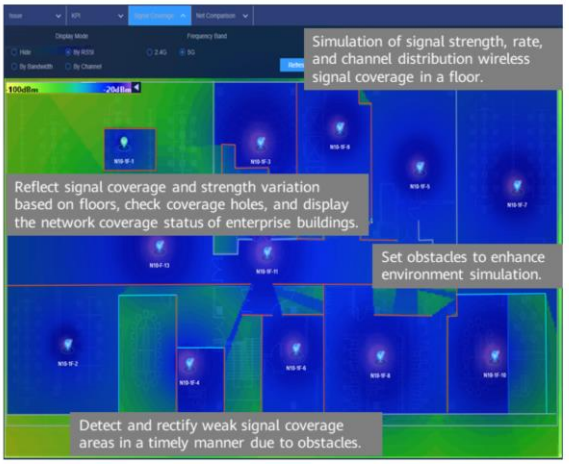
Impact: Is there any impact on the network if an isolation wall needs to be added?

Simulate the impact on wireless coverage by adding faulty objects.

Less interference: Do device configurations need to be modified to interfere with other devices?

After device configurations such as power and channel are modified, simulate whether interference occurs between devices in real time.

Ability	Description
By RSSI	Wi-Fi signal strength at each location of the simulation floor.
By rate	Indicates the attainable Wi-Fi access rate at each location of the simulation floor.
By Channel	Simulate the signal strength and conflict of each floor on the specified channel.



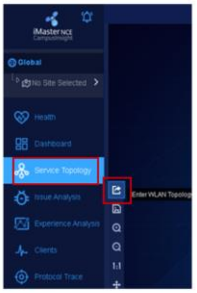
HUAWEI

- The service topology collects statistics on the status, access, congestion, and error packet issues, displays the number of clients and traffic volume based on sites, regions, buildings, and floors. This allows administrators to quickly search for and view the buildings that users pass by, helping administrators quickly identify campus network issues.
- In the Service Topology of CampusInsight, you can access WLAN Topology to view the radio heat map of the network.

Case 4: Radio Heat Map (1)

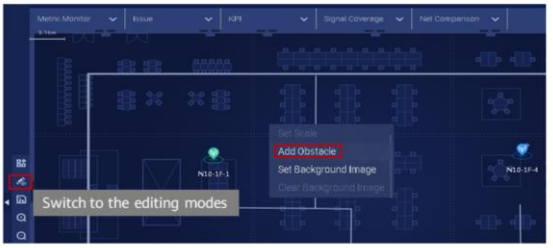
Case: A company has insufficient conference room resources and need to partition the office area to form some conference rooms. The construction design personnel consult the company's IT department about whether this practice affects the office Wi-Fi. The IT personnel edit the coverage obstacles such as conference room walls and glass doors based on the radio heat map, and fill coverage holes or adjust APs based on the simulation result.

1. Choose Service Topology > WLAN Topology.



2. Select Shenzhen > N10 > N10-1F from the left navigation tree.

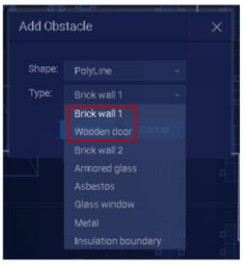
3. Switch to editing state, right-click and select Add Obstacle.



HUAWEI

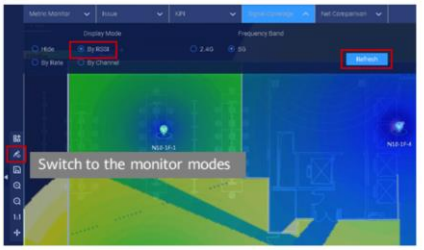
Case 4: Radio Heat Map (2)

4. Set Shape to PolyLine and Type to Brick wall 1 or Wooden door.



5. Draw a conference room on the topology (as shown by the quadrate in the figure), double-click the left button of the mouse to exit the obstacle drawing view, and click Save.

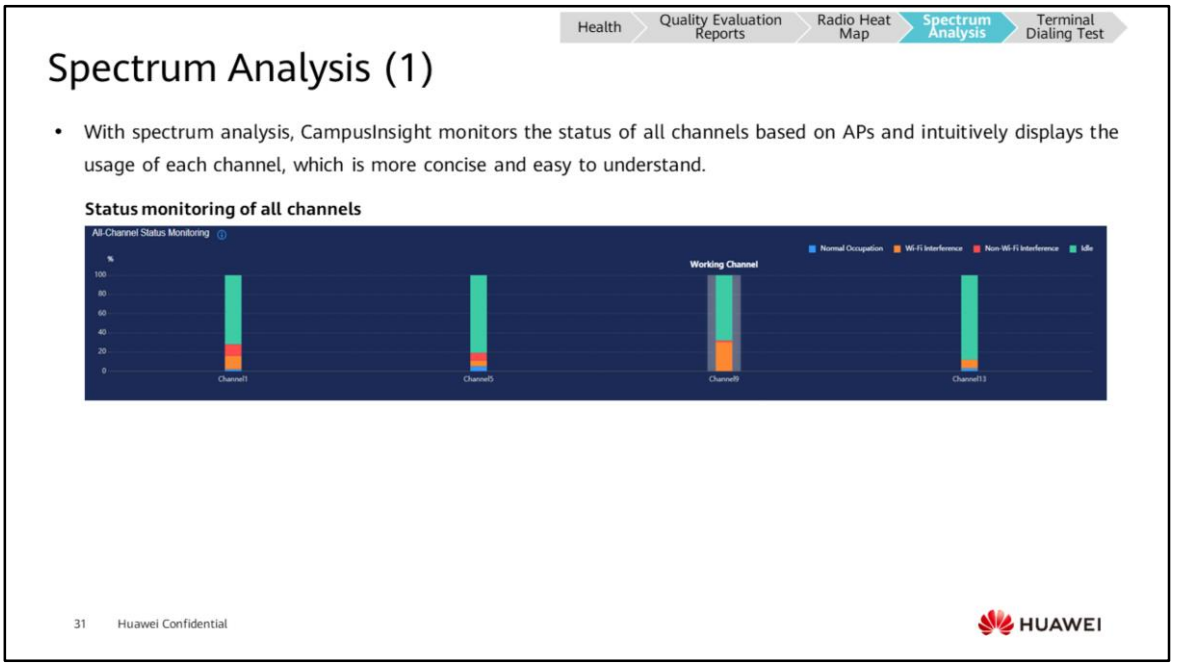
6. Enter the monitoring state, select By RSSI, and click Refresh to check whether there are coverage holes. Perform operations according to the simulation result.



Use Case Benefits

Using radio heat map, CampusInsight can simulate Wi-Fi coverage based on the radio signal transmission model upon changes of the radio environment or signal configuration. In this manner, CampusInsight can proactively identify coverage holes and areas affected by conflicts, and provide references for filling coverage holes and adjusting network configurations

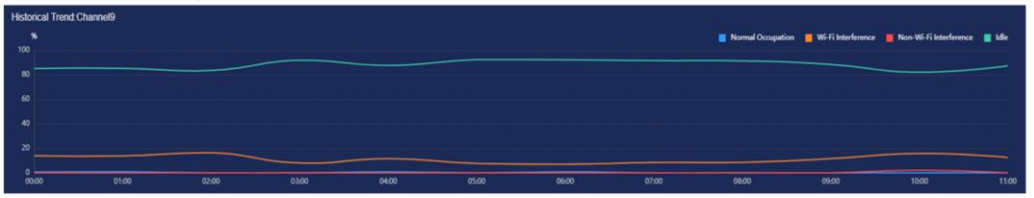
HUAWEI



- The **Spectrum Analysis** tab page displays the channel status and surrounding interference of APs, including the channel utilization, current working channel, historical trend of channel status, non-Wi-Fi interference sources, Wi-Fi interference sources, and Wi-Fi interference source distribution, facilitating device status analysis.
- The spectrum analysis process is as follows:
    1. APs scan all channels in real time, including co-channel interference, non-Wi-Fi interference, and normal usage ratio of channels.
    2. The APs report channel scanning data to CampusInsight.
    3. CampusInsight monitors the status of all channels by AP in real time and allows administrators to view the historical trend chart, non-Wi-Fi interference source types, and RSSI.

Spectrum Analysis (2)

Historical trend chart by channel



List of detected Wi-Fi/non-Wi-Fi interference sources

Non-Wi-Fi Interference Source | Wi-Fi Interference Source

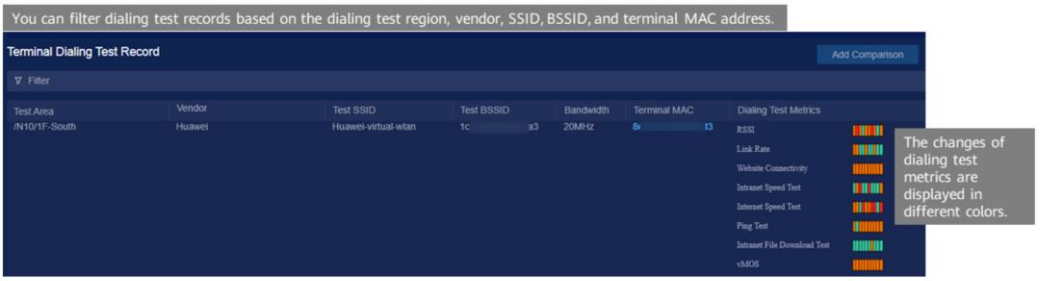
Device Type	RF Type	Channel	Last Detection Time	RSSI(dBm)
BLUETOOTH	Broadband	7,8,9,10	20: :20:04	-57
2G WIRELESS TRANSMITTER	Broadband	3,5	20: :53:30	-30
BABY MONITOR	Broadband	5	20: :46:21	-59
FIXFREQ DEVICE	Broadband	1	20: :46:24	-42

HUAWEI

Health > Quality Evaluation Reports > Radio Heat Map > Spectrum Analysis > Terminal Dialing Test

Terminal Dialing Test (1)

- With terminal dialing test, CampusInsight displays the overall WLAN experience quality from the perspective of users based on multiple metrics collected from the CloudCampus APP.



Dialing Test Type	Content
Wi-Fi network metric	Signal strength, bandwidth, ping delay, ping packet loss rate, and AP association performance
Wi-Fi service metric	Website connectivity, intranet speed test, intranet file download speed, Internet performance test, and video experience

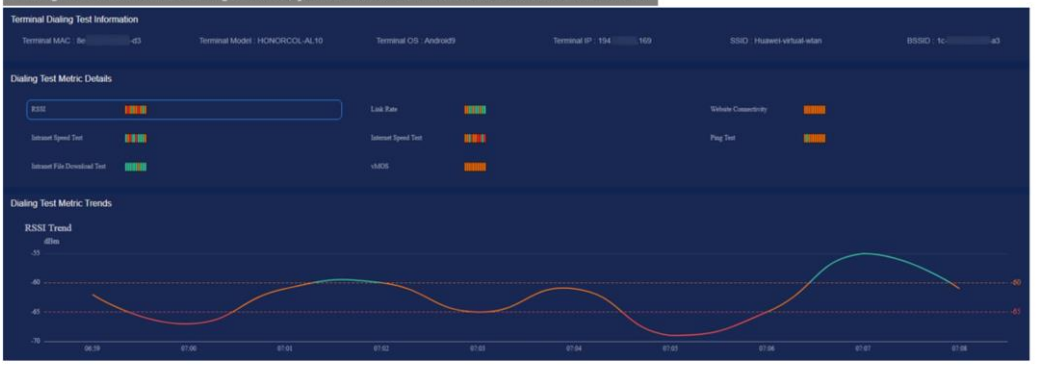
CloudCampus APP

HUAWEI

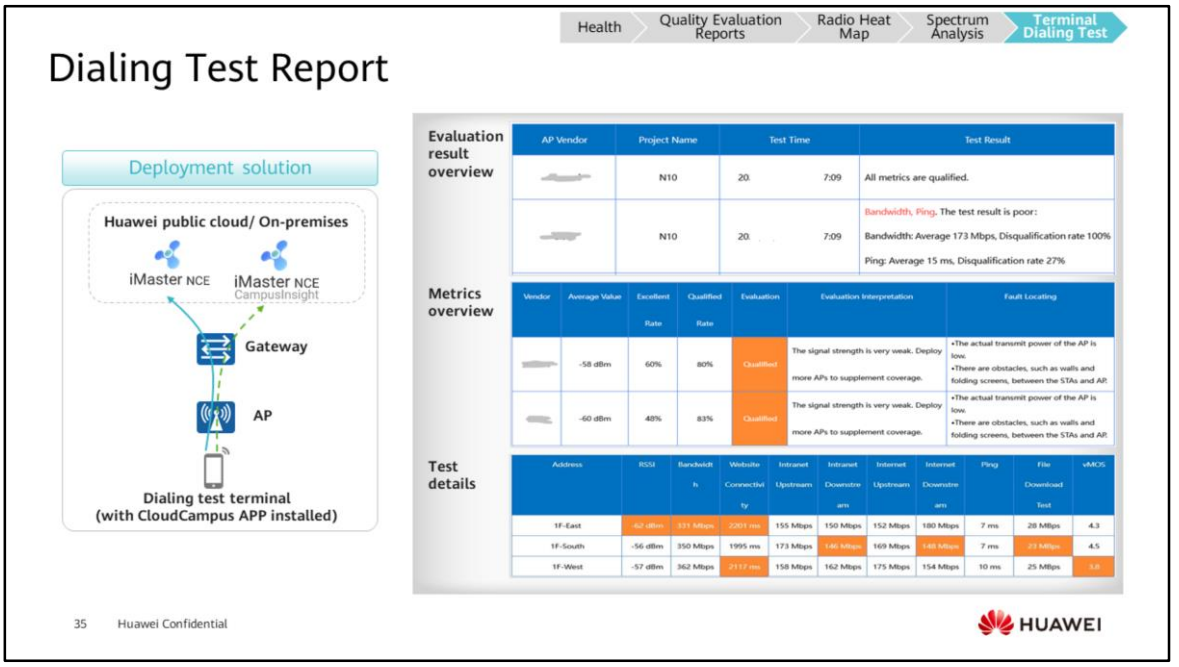
- Note: Intranet servers are required for testing the intranet speed, intranet file download rate, and video experience.

Terminal Dialing Test (1)

Dialing test details: After clicking a metric, you can view the historical trend chart of the metric.



HUAWEI



- CampusInsight provides comparison reports of networks built by different vendors, comprehensively evaluate the overall user experience of Wi-Fi networks, analyze network issues, and provide network optimization guidance.
- Zero-cost rapid deployment:
  - Network reachability can be tested.
  - Analysis reports are immediately generated.

Contents

1. CampusInsight Overview
2. **CampusInsight Functions and Demonstration**
   - Telemetry
   - Visibility
   - Analysis
   - Optimization
   - Assurance

HUAWEI

CampusInsight: Individual and Group Issue Analysis

During campus network O&M, administrators encounter the following types of issues:

1. Individual issues: for example, access failures caused by incorrect client configurations.

2. Group issues: for example, group authentication failures caused by authentication server faults and group weak-signal coverage issues caused by insufficient AP coverage.

Solution to Individual Issues

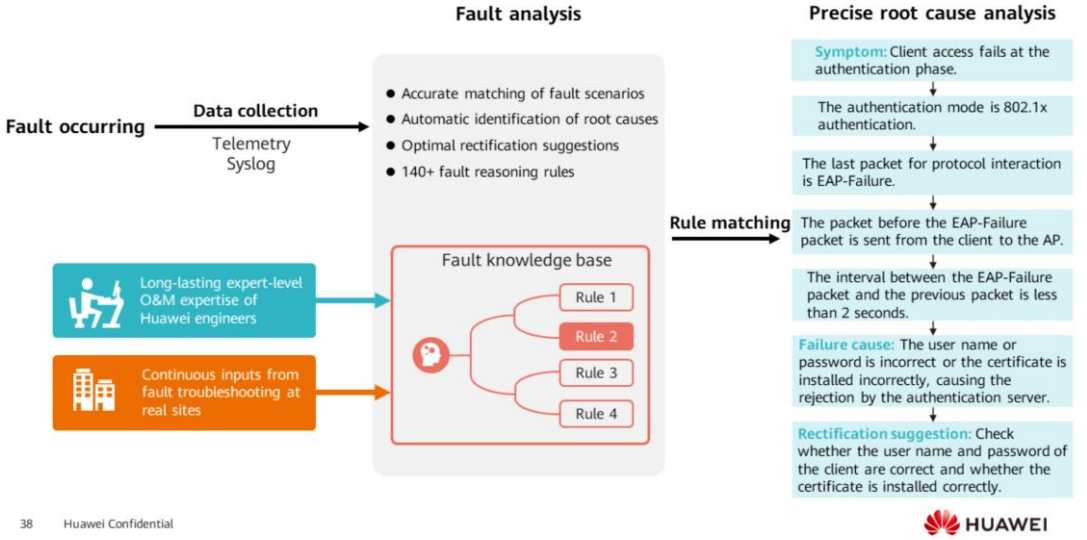
1. Each metric: Collect data within seconds and detect issues.
2. Each client: Analyze the protocol-level access process, trace the client journey and analyze experience, and detect audio and video application quality.

Solution to Group Issues

Issue identification: Intelligently identify connection, air interface performance, roaming issues, and device issues.

HUAWEI

CampusInsight: Fault Reasoning



HUAWEI

Individual Issue | Group Issue

CampusInsight: Intelligently Analyze Individual Issues from Four Aspects

Challenges	Requirements
1. Unable to analyze the phase where access failures occur 2. Unable to analyze client locations and experience 3. Unable to effectively detect application quality	Identify network experience issues of individual clients in E2E mode and analyze the root causes of the issues.

Solution and Customer Benefits

Journey Analysis Client journey (wireless + wired networks)	**Access Analysis** Protocol tracing (wireless + wired networks)	CampusInsight analyzes individual issues encountered during network O&M from the access network, client journey, experience, and application detection dimensions, and implements protocol process analysis, visible client journey tracing, correlative analysis of poor experience, and audio and video application quality detection, helping administrators maintain networks and ensure the experience of VIP clients.
Experience Analysis Correlative analysis of poor-experience clients (wireless networks)	**Application Analysis** Application experience awareness and assurance (wireless + wired networks)	

HUAWEI

Client Journey: Real-Time Experience Visibility for Each Client at Each Moment (1)

Step1 Experience Overview

Check the overall experience metrics.

HUAWEI

Client Journey: Real-Time Experience Visibility for Each Client at Each Moment (2)

Step2 Experience Trend

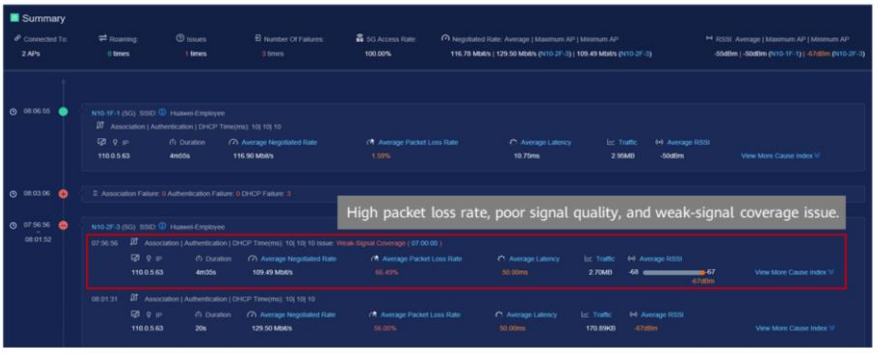
Check for experience trend changes or deterioration.

HUAWEI

Client Journey: Real-Time Experience Visibility for Each Client at Each Moment (3)

Step3 Client Journey

Check APs that connect to the network at each moment, their experience, and issues that occur.



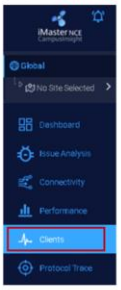
HUAWEI

Individual Issue > Journey Analysis > Access Analysis > Experience Analysis > Application Analysis

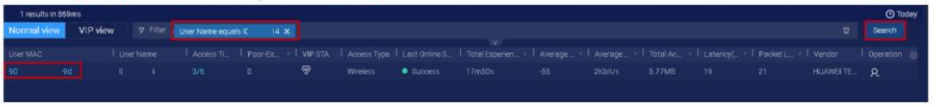
Case 5: Wi-Fi User Experience Visibility (1)

Case: A VIP user reports that the Wi-Fi experience is poor. O&M personnel use the client journey function and find that the packet loss rate is high and the signal quality is poor in the cafe. In addition, the weak-signal coverage issue occurs.

1. Click Clients.



2. Enter l0***14 in the search box, and click Search.



3. Click the MAC address to go to the client journey page. (For detailed operations, see the next slide.)

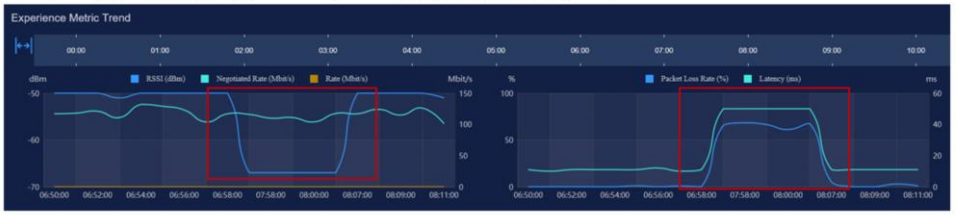
HUAWEI

- Note: The process of viewing wired user experience visibility is the same as that of viewing wireless user experience visibility.

Case 5: Wi-Fi User Experience Visibility (2)

1. Check the experience overview and find that the average packet loss rate is high (15%) among Wi-Fi experience metrics.

2. Check the experience metric trend. The client experience deteriorates significantly after a period of connecting to the Wi-Fi, with poor signal quality (< -65 dBm) and a high packet loss rate (> 5%).



HUAWEI

Case 5: Wi-Fi User Experience Visibility (3)

3. When the client accesses the AP N10-2F-3, the weak-signal coverage issue occurs. In this case, the signal quality is poor and the packet loss rate is high.

Use Case Benefits

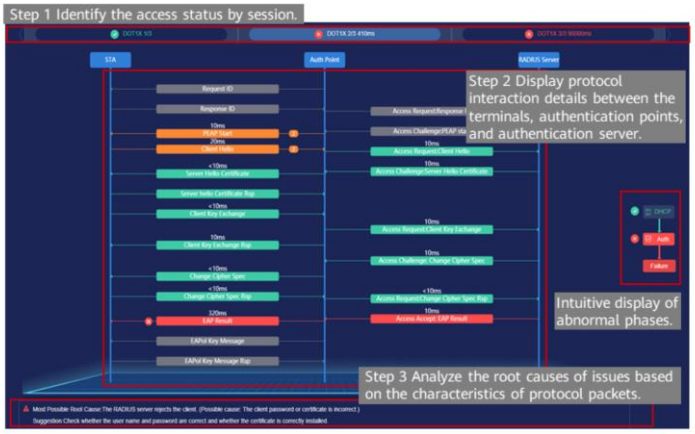
Using client journey, CampusInsight focuses on the real Wi-Fi experience of clients and accurately traces the entire Wi-Fi access process of each client (regarded as a sensor). The traced information includes: client, time, location, connected AP, experience, and issue.

HUAWEI

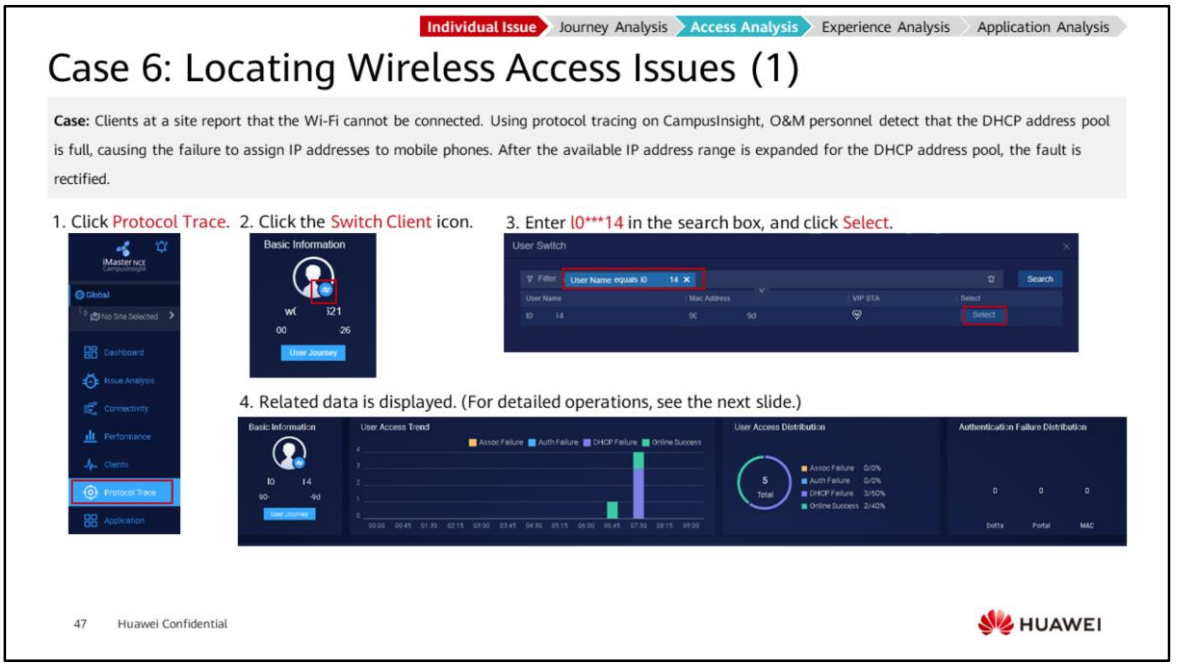
Protocol Tracing: Locating the Root Cause of Access Faults Within Minutes

3 simple steps to resolve connection issues

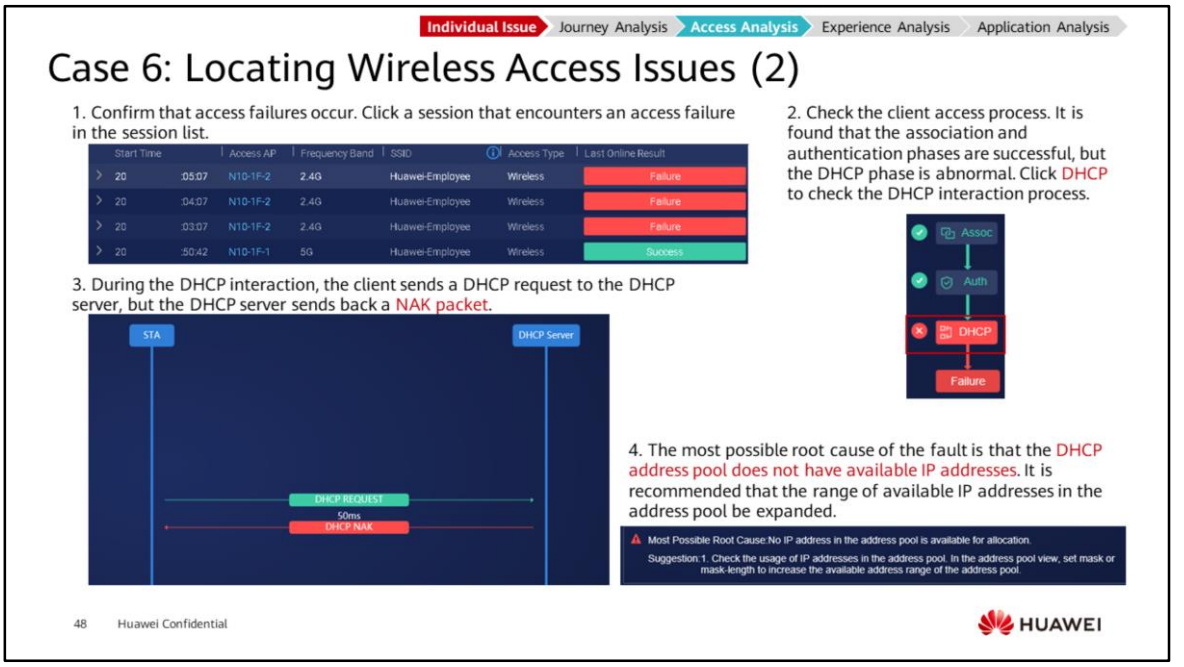
1. **Status: Check terminal connections**
   Check the session access result to determine whether access issues occur.
2. **Interaction: Check protocol interaction**
   Check the association, authentication, and DHCP phases to determine the abnormal phase.
3. **Root Cause: Check root causes**
   Check possible causes and rectification suggestions.



HUAWEI



- Note: The procedure for locating wired access problems is the same as that for locating wireless access problems.



- Use Case Benefits:
  - Using protocol tracing, CampusInsight can analyze the protocol interaction details in the three phases (association, authentication, and DHCP) for a client to access the wireless network in a fine-grained manner, and provide root causes and rectification suggestions. This facilitates the resolution of connection issues.

Correlation Analysis for Wi-Fi Experience Deterioration (1)

Step1 AI Identification

Use the AI algorithm to detect outliers and identify clients with deteriorated Wi-Fi experience.

Step2 AI Analysis

Use the correlation analysis algorithm to analyze the network metric with the highest relevance and locate the root cause.

HUAWEI

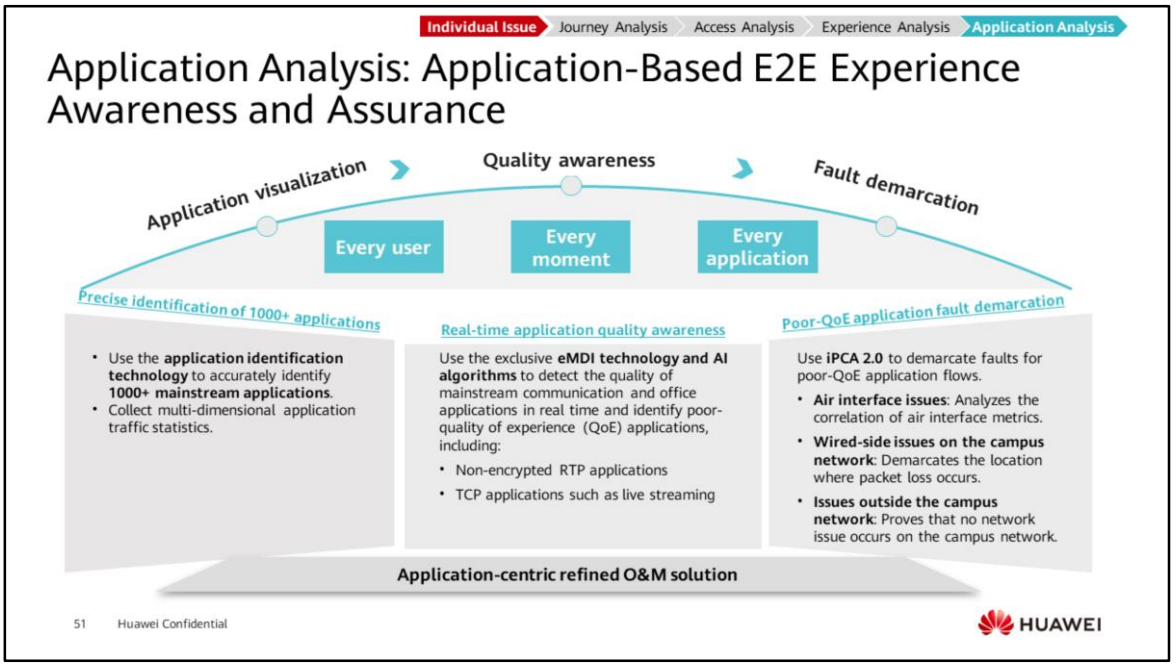
Correlation Analysis for Wi-Fi Experience Deterioration (2)

Step3 AI Closed-loop Management

Provide the most reasonable rectification suggestions based on Huawei's accumulated O&M expertise.

Possible Cause	Troubleshooting Object	Suggestions
Interference exists between co-channel neighboring APs.	Neighboring AP	Suggestion 1: Manually trigger RF calibration. Take AC6605 V200R009C00 as an example. Run the calibrate manual startup command in the WLAN view to manually trigger RF calibration. In the cloud scenario, choose Provision > Site Configuration > AP > Radio and trigger radio calibration. For other versions or device types, see the corresponding configuration manual. Suggestion 2: Reduce the number of broadcast frames sent by neighboring APs on the air interface. The following uses AC6605 V200R009C00 as an example: Step 1: Run the beacon-5g-rate command in the SSID profile view to set the transmission rate of 5GHz Beacon frames to 9 or 12 Mbit/s. Step 2: Run the beacon-interval command in the RF profile view to set the interval for the AP to send Beacon frames to 200 TUs.In the cloud scenario, choose Provision > Site Configuration > AP > SSID and modify the rate set.

HUAWEI



- Application analysis:
  - Application analysis monitors network-wide applications and supports NetStream and service awareness (SA) data sources. This module displays overview data such as the number of applications and traffic, and sorts applications by traffic in ascending or descending order. In wired scenarios, SA can be configured only on access switches, and the incoming and outgoing traffic displayed on the page is the incoming and outgoing traffic over the data reporting devices.
- eMDI: Enhanced Media Delivery Index
- RTP: Real-time Transport Protocol
- iPCA: Packet Conservation Algorithm for Internet

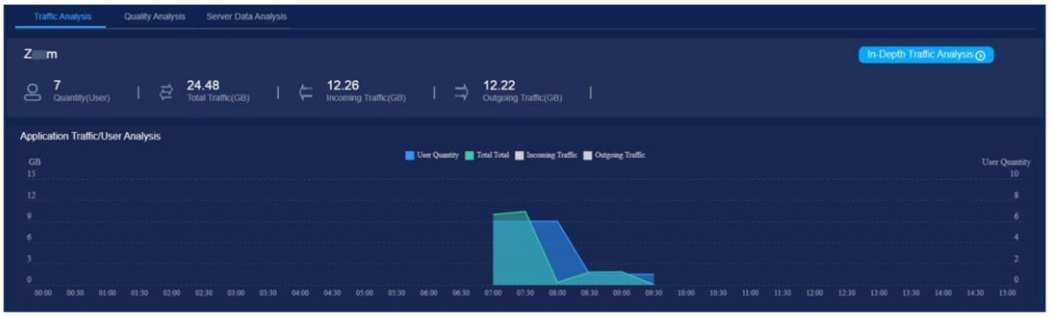
Application Identification and Traffic Analysis (1)

- Analyze the network-wide application traffic and the number of users based on applications.

HUAWEI

Application Identification and Traffic Analysis (2)

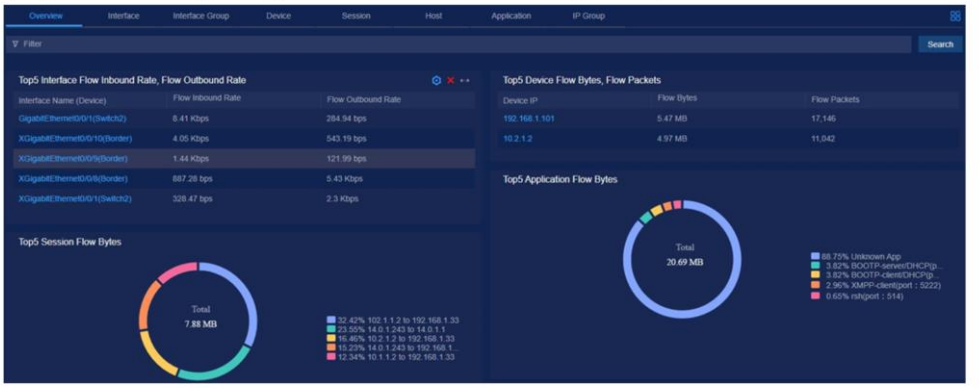
- Analyze the network-wide application traffic and the number of users based on applications.



HUAWEI

Application Identification and Traffic Analysis (3)

- Display the traffic usage of applications based on interfaces, devices, and hosts.



- Retrospect the application usage based on client journey. (For details, see "Journey Analysis".)

HUAWEI

- The application details page displays the application experience awareness and poor-QoE analysis.
    - The **Metric Overview** area displays the average packet loss rate, average jitter, disorder rate, packet rate, and byte rate of the session.
    - If the device role is correctly set on the resource side and LLDP is enabled, the **Analysis and Demarcation** area displays the full link topology of the session from the initiator to the responder. You can view devices such as APs, switches, and ACs that the session passes through. When a device is faulty, the device is marked in red and displayed as a poor-quality device.
    - **Application quality and air interface:** You can click a device to view the performance metrics of the device and its interface or air interface in the session.
        - The device metrics include the MOS value, packet loss rate, maximum number of consecutively lost packets, jitter, disorder rate, and deterioration time ratio.
        - The metrics of an interface or air interface include the signal strength, channel utilization, interference rate, latency, and retransmission rate.
- MOS: Mean opinion score. It is used to measure quality of voices, especially voices over the Internet, at the line termination.

Case 7: Application Experience Awareness and Fault Demarcation (1)

Case: O&M personnel at a site are responsible for guaranteeing important video conferences attended by Huawei executives and proactively inspect the conference application quality of the executives. If a conferencing quality issue is found, the application troubleshooting function can be used to quickly demarcate the packet loss location of application flows and rectify the issue.

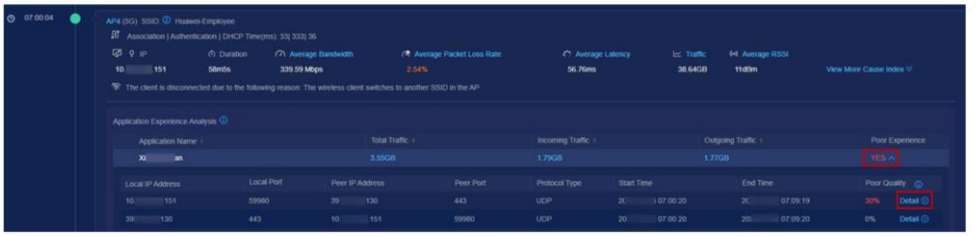
1. Choose Clients.

2. In the VIP view, click the MAC address hyperlink of user d0***06.

HUAWEI

Case 7: Application Experience Awareness and Fault Demarcation (2)

3. On the client journey page, check the experience analysis result of the Xi***an application at 7:00 AM, and find that a receive poor-QoE issues occurs in the application flows of the client at that time point. Click YES and click Detail to view details about the issue.



HUAWEI

Case 7: Application Experience Awareness and Fault Demarcation (3)

- Basic application information and quality overview: View basic information about application flows, such as the 5-tuple information, and obtain the network quality overview, such as packet loss, disorder, and jitter.

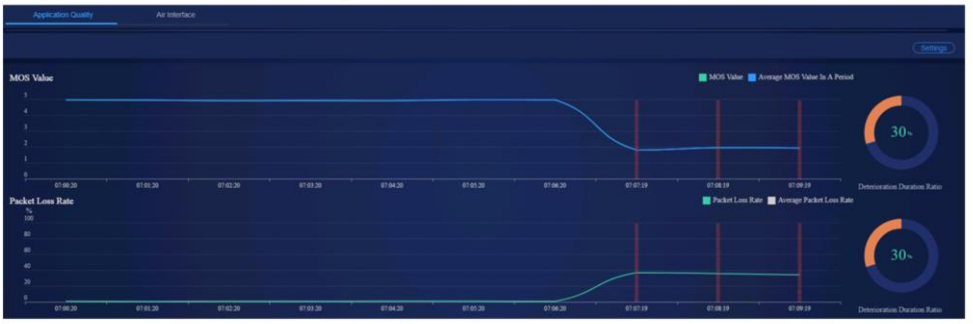
- Analysis and demarcation: Check the physical topology paths of application flows and actual application paths. The location where air interface quality is poor is marked using a yellow line. The issue is suspected to be on AP4.

HUAWEI

Individual Issue | Journey Analysis | Access Analysis | Experience Analysis | Application Analysis

Case 7: Application Experience Awareness and Fault Demarcation (4)

- Indicator details: Click the device to view the packet loss trend of application flows on the port as well as KPIs of the port, in order to facilitate fault locating.



HUAWEI

- Use Case Benefits:
  - Application quality awareness and fault demarcation allow administrators to view the quality of specific application flows. If an application encounters a poor-QoE issue, the system can directly demarcate the issue. If the issue occurs inside the campus network, and we could troubleshoot it based on the packet loss location. If the issue occurs outside the campus network, who is responsible can be clarified.

Individual Issue | Group Issue

CampusInsight: Intelligent Identification of Four Types of Group Issues

Scenario and Requirements

1. In the digital era, passive O&M performed after events occur cannot ensure service experience.
2. More intelligent methods are required to automatically identify potential faults and accurately locate root causes to reduce the fault response time.

Identify potential faults: The fault response time is reduced from hours to minutes.

Connection

- Association failure
- Slow association
- Authentication failure
- Authentication timeout
- Slow authentication
- DHCP connection failure
- Slow DHCP connection
- Client gateway unreachable

Air Interface Performance

- Weak-signal coverage
- High channel utilization
- High interference
- Dual band capable clients prefer 2.4 GHz
- Client capacity
- Air interface congestion

Roaming

- Repeated roaming
- Roaming exception

Device

- Network status: Switch port error-down, Intermittent port disconnection, etc.
- Network performance: Layer 2 loop, Port congestion, etc.
- Device environment: Device offline, Switch LPU fault, etc.
- Device capacity: Abnormal switch ARP entry increasing, Switch CPU threshold exceeded, etc.

Wireless

Wired

HUAWEI

CampusInsight: Intelligent Identification of Four Typical Issues (1)

Type	Description	Issue
Connection Issues	CampusInsight quickly identifies a variety of network access issues, such as group authentication failure and slow authentication, that occur at the association, authentication, and DHCP phases. CampusInsight quickly and accurately identifies the root causes of each issue based on the fault knowledge base and provides troubleshooting suggestions accordingly.	Authentication failure (wireless + wired networks)
		Authentication timeout (wireless + wired networks)
		Slow authentication (wireless + wired networks)
		Association failure
		Slow association
		DHCP failure (wireless + wired networks)
		Slow DHCP (wireless + wired networks)
		Unreachable user gateway
		User Offline
Performance Issues	CampusInsight monitors air interface performance data in real time. On the basis of real-time performance monitoring and Huawei's expertise in WLAN field, CampusInsight intelligently identifies six classes of air interface issues that affect network access experience after users get connected to the wireless network and provides troubleshooting suggestions accordingly.	Weak-signal coverage
		High interference
		High channel utilization
		Air interface congestion
		Dual band capable client prefers 2.4G
		Client capacity

HUAWEI

CampusInsight: Intelligent Identification of Four Typical Issues (2)

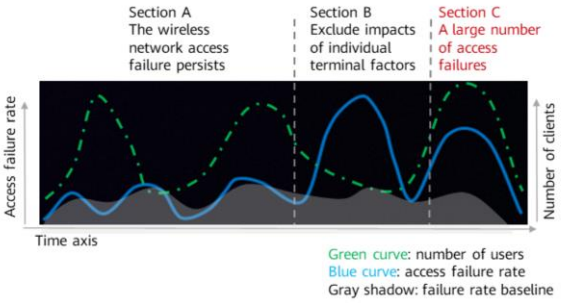
Type	Description	Issue
Roaming Issues	CampusInsight analyzes the process when a user roams between APs to intelligently identify network access experience issues when the user moves and provides troubleshooting suggestions accordingly.	Repeated roaming
		Roaming exception
Device Issues	Intelligently identifies network status, network performance, device environment, and device capacity issues and provides rectification suggestions.	Network status: Switch interconnection port's protocol status was down, Intermittent port disconnection, Switch port error-down, Switch physical port suspended, Suspected optical link fault
		Network performance: Layer 2 loop, Port congestion, Queue congestion, Abnormal high number of error packets on the port, CPCAR packet loss
		Device environment: Management channel interruption, Device offline, Repeated device restarts, Switch LPU fault, Switch SFU fault, Switch MPU fault, etc.
		Device capacity: Abnormal switch ARP entry increasing, Abnormal switch MAC entry increasing, Abnormal switch FIB4 entry increasing, Switch CPU threshold exceeded, Switch CPU threshold exceeded, etc.

HUAWEI

Connection Issue Analysis (1)

Key Technologies

- **Exception detection:** Detection of network access exceptions.
  - Normal connection failure (no fault, as shown by section A in the right figure): The wireless network access failure persists, but the failure is not caused by a network fault.
  - Noise reduction for abnormal terminals (as shown by section B in the right figure): Exclude impacts of individual terminal factors. The access failure rate increases due to faulty terminals. Although the access failure rate already exceeds the threshold, but it does not indicate that a network fault occurs.
  - Intelligent fault identification (as shown by section C in the right figure): Intelligently identify group issues with large impact scopes (issues with a large number of failed clients and a large failure rate).



HUAWEI

Connection Issue Analysis (2)

Key Technologies

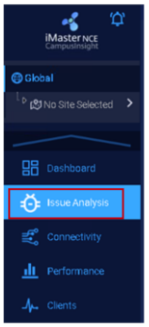
- **Pattern identification:** Causes may be different for issues with the same symptom. Through identifying patterns, CampusInsight can locate possible causes. In addition, CampusInsight abstracts features of clients with access failures and performs group analysis using aggregation algorithms.
- **Root cause analysis:** Analyze possible root causes based on client online logs and provide rectification suggestions, helping O&M personnel resolve issues.

HUAWEI

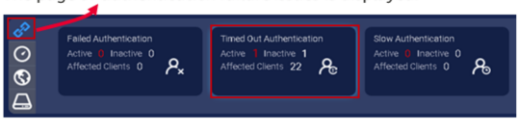
Case 8: Connection Issue Analysis (1)

Case: At noon of a day, O&M personnel at a site received a complaint from the customer that mobile phones failed to access the network during working hours in the morning. The IT personnel found that the RADIUS server had poor performance and could not process a large number of authentication requests in peak hours. As a result, the authentication failed. The fault was rectified after the legacy RADIUS server was replaced with a server with higher performance.

1. Click Issue Analysis.

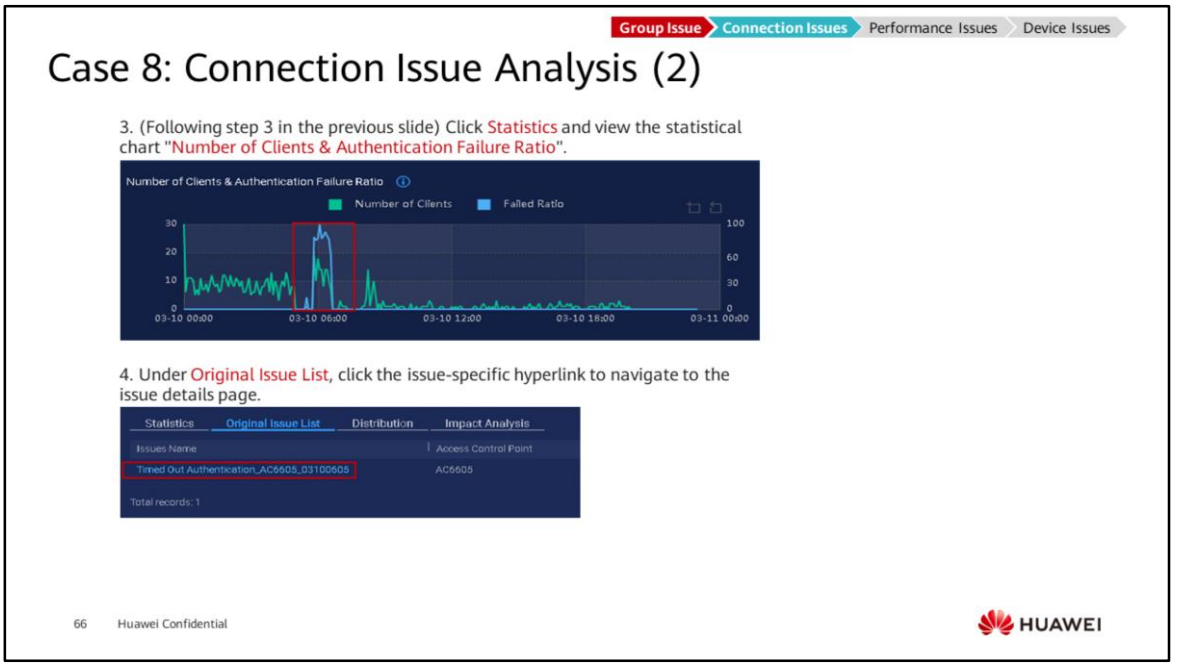


2. Click the connection issue tab and click Timed Out Authentication.
The page of authentication failure issues is displayed.

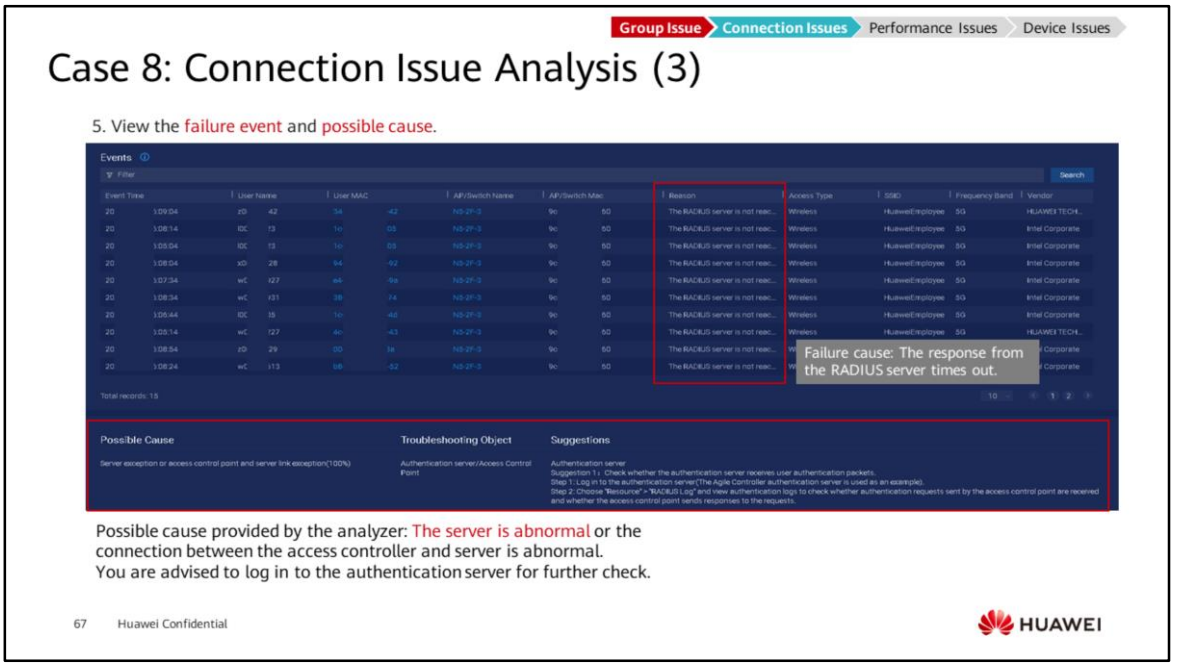


3. Click Statistics and view the statistical analysis chart "Number of Clients & Authentication Failure Ratio". (For the analysis details, see the next slide.)

HUAWEI



- The statistics of the "Number of Clients & Authentication Failure Ratio”:
  - Noise reduction for abnormal terminals: The authentication failure rate increases sharply in some time periods. The analysis result shows that the abnormal terminals initiate a large number of authentication requests, and the failure rate is 100%. This issue is not caused by network faults, and noise data needs to be removed. (The client revisit shows that these clients are new employees and the terminals are not installed with Huawei Wi-Fi certificate. In this manner, frequent re-authentication causes a large number of failure events.)
  - Intelligent identification of issues with a large number of failed clients and a large failure rate: When the number of wireless access users increases sharply at 6:00 a.m., the authentication failure rate reaches 30%. (The RADIUS server cannot respond to authentication requests in a timely manner due to its limited performance.) This is a typical issue with a large number of failed clients and a large failure rate, indicating that a group fault occurs.
  - Connection failure but not a fault: Due to the instability of wireless client access (for example, when a client moves or passes through a coverage hole), the user authentication failure persists in each time segment, but does not affect user experience. The fault is rectified after the user automatically accesses the network again.



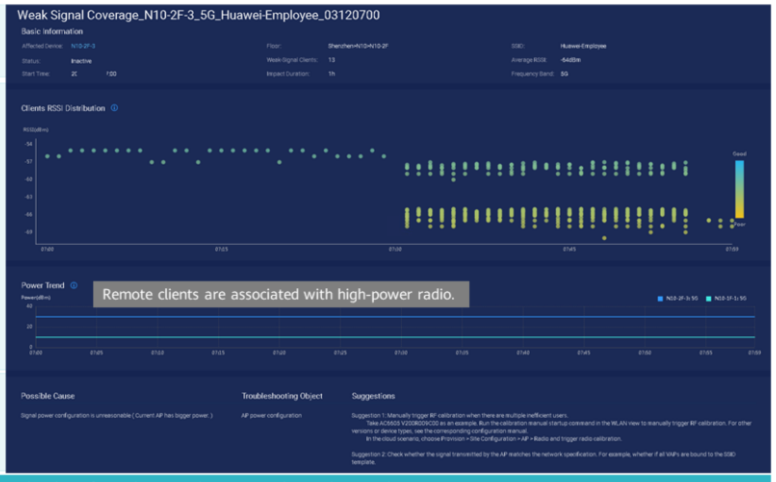
- Use Case Benefits:
  - Due to factors such as signal coverage, obstacles in physical environments, and radio interference, user access failures persist on the wireless network. CampusInsight analyzes client access data to identify abnormal clients and eliminate invalid failure events. In addition, CampusInsight uses the machine learning algorithm to intelligently identify issues with a large number of failed clients and a large failure rate, accurately identify group connection faults on the network in a timely manner, and provide proper rectification suggestions based on expert experience.

Weak-Signal Coverage Issue Analysis

Step1 Issue Identification
Intelligent identification of weak-signal coverage issues

Step2 Issue Analysis
Extensive information for issue analysis

Step3 Cause Locating
Troubleshooting based on suggestions



Intelligently identify and remotely cope with weak-signal coverage issues

HUAWEI

Case 9: Weak-Signal Coverage Issue Analysis (1)

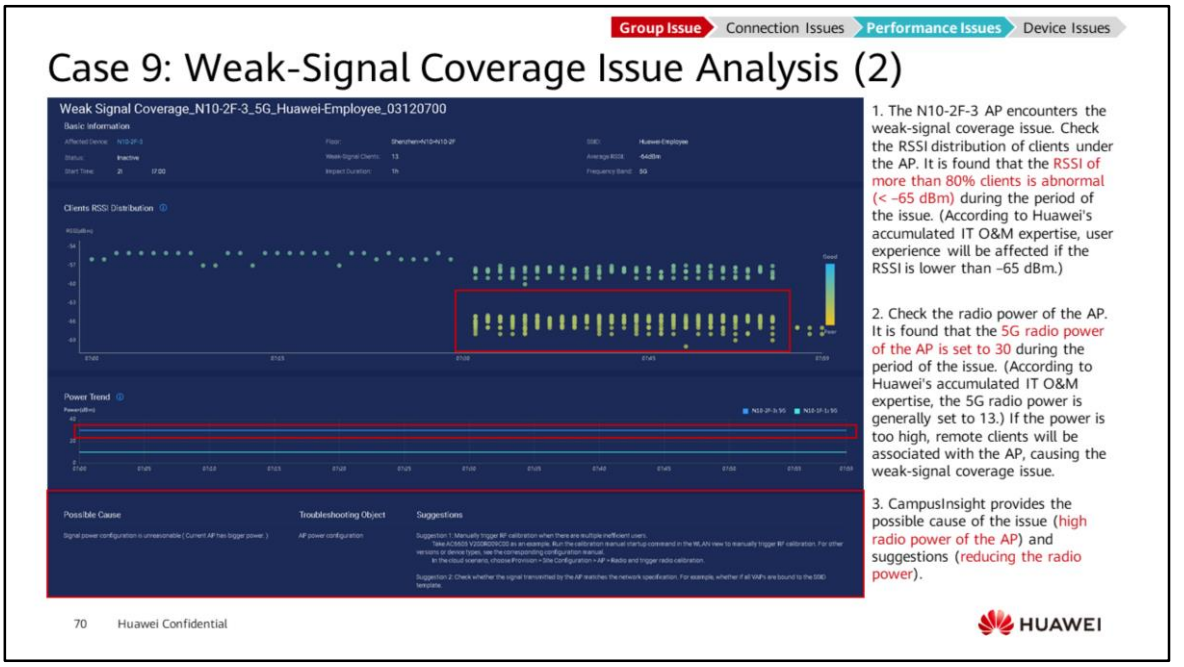
Case: On CampusInsight, O&M personnel at a site detect multiple weak-signal coverage issues. The issue details show that many remote clients are associated with this high-power radio and therefore most clients have a low RSSI. O&M personnel perform troubleshooting based on the rectification suggestions. The fault is rectified.

1. Click Issue Analysis.

2. Click the air interface issue tab and click Weak-Signal Coverage. The page of weak-signal coverage issues is displayed.

3. Click Original Issue List. In the issue list, click the first issue. The issue details page is displayed. (For detailed operations, see the next slide.)

HUAWEI



- Use Case Benefits:
  - Weak-signal coverage issues easily occur on Wi-Fi networks due to improper network planning and configuration. Such issues lead to poor RSSI of clients and deteriorate Wi-Fi experience. Using weak-signal coverage issue analysis, CampusInsight can automatically identify weak-signal coverage issues, analyze data such as the RSSI and radio power, and provide possible causes and rectification suggestions.

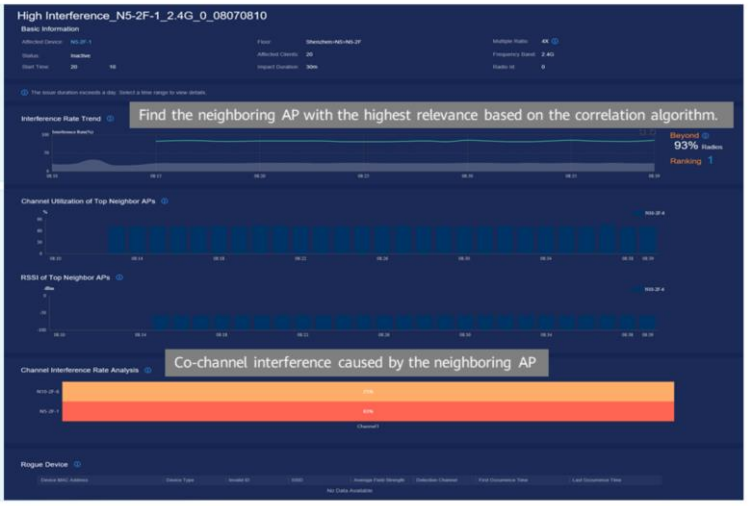
High Interference Issue Analysis (1)

Step1 Intelligent Identification

Proactively identify high interference issues on the network.

Step2 Correlation Analysis

Correlation analysis on metrics related to high interference issues.



HUAWEI

High Interference Issue Analysis (2)

Step3 Fault Diagnosis

Locate the root cause and rectify the fault based on possible causes and rectification suggestions.

Possible Cause	Troubleshooting Object	Suggestions
Interference exists between(N10-2F-6)co-channel neighboring APs.	Neighboring AP	Suggestion 1: Manually trigger RF calibration. Take AC6605 V200R009C00 as an example. Run the calibrate manual startup command in the WLAN view to manually trigger RF calibration. In the cloud scenario, choose Provision > Site Configuration > AP > Radio and trigger radio calibration. For other versions or device types, see the corresponding configuration manual. Suggestion 2: Configure the channel sets for calibration so that the default channel is avoided during calibration and RF calibration is triggered. The following uses AC6605 V200R009C00 as an example: Step 1: Run the dca-channel channel-set command in the regulatory domain profile view and set the calibration channel sets to 1, 5, 9, and 13. Step 2: Run the calibrate manual startup command in the WLAN view to manually trigger RF calibration. In the cloud scenario, choose Provision > Site Configuration > AP > Radio and modify the calibration channel set. For other versions or device types, see the corresponding configuration manual. Suggestion 3: Reduce the number of broadcast frames sent by neighboring APs on the air interface. The following uses AC6605 V200R009C00 as an example: Step 1: Run the beacon-2g-rate command in the SSID profile view to set the transmission rate of 2.4 GHz Beacon frames to 5.5 Mbit/s. Step 2: Run the beacon-interval command in the RF profile view to set the interval for the AP to send Beacon frames to 200 TUs. In the cloud scenario, choose Provision > Site Configuration > AP > SSID and modify the rate set. For other versions or device types, see the corresponding configuration manual.

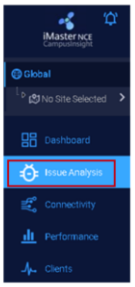
Passive O&M is changed to proactive O&M.

HUAWEI

Case 10: High Interference Issue Analysis (1)

Case: On CampusInsight, O&M personnel at a site detect a co-channel interference issue caused by the neighboring AP through correlation analysis. Then they perform troubleshooting based on the rectification suggestions. The fault is rectified.

1. Click Issue Analysis.



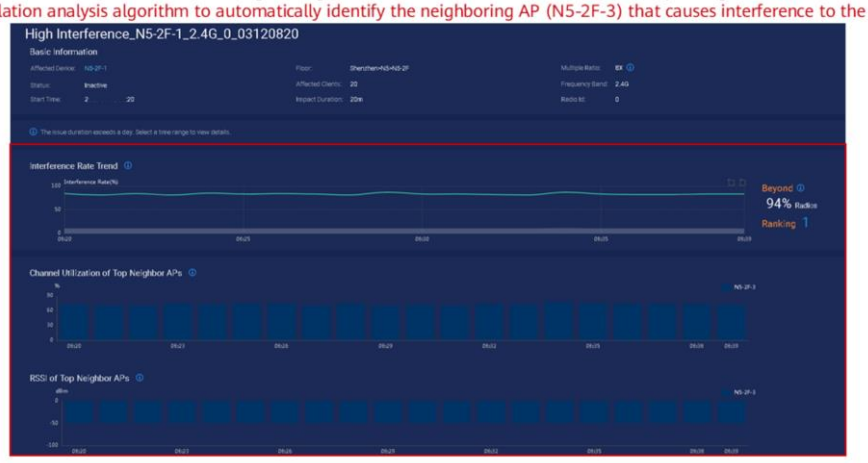
2. Click the air interface issue tab and click High Interference. The page of high interference issues is displayed.

3. Click Original Issue List. In the issue list, click the first issue. The issue details page is displayed. (For detailed operations, see the next slide.)

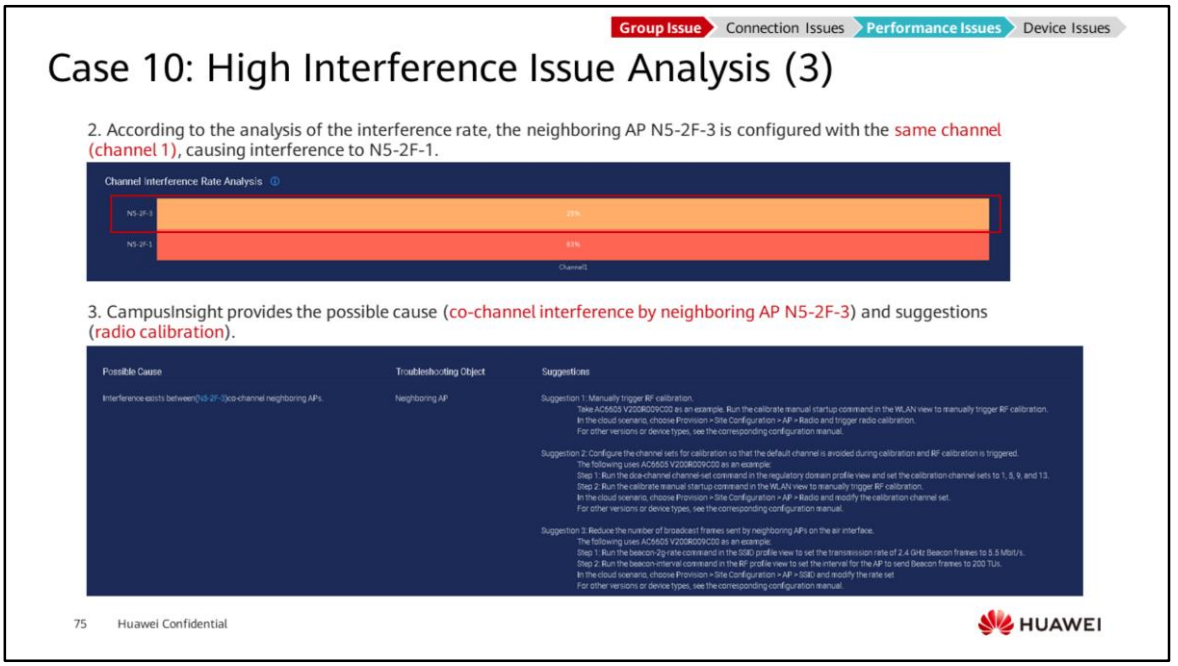
HUAWEI

Case 10: High Interference Issue Analysis (2)

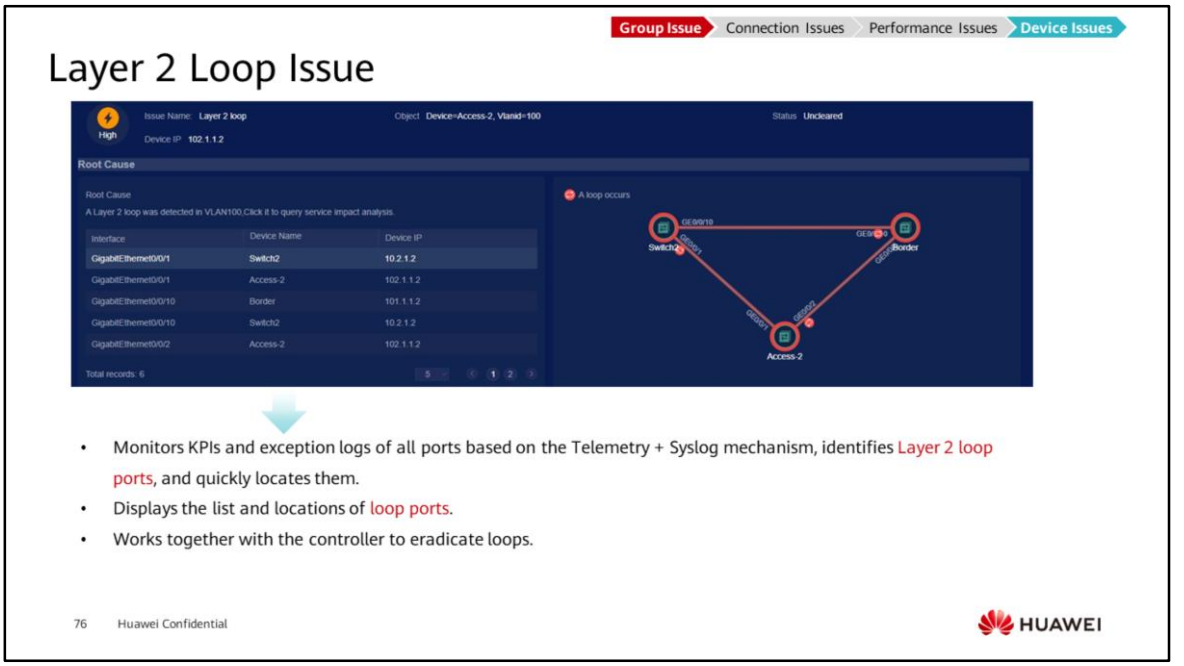
1. The N5-2F-1 AP encounters the high interference issue. The interference rate of an AP is closely related to the channel utilization and RSSI of the neighboring AP that causes interference to the AP. CampusInsight uses the correlation analysis algorithm to automatically identify the neighboring AP (N5-2F-3) that causes interference to the AP.



HUAWEI



- Use Case Benefits:
  - High interference issues easily occur on Wi-Fi networks. These issues affect user experience and cause problems such as long network delay and congested network access. As a result, user experience on the Wi-Fi network deteriorates. Using correlation analysis and big data analytics, CampusInsight can detect high interference issues on the network, find the most possible causes, and provide rectification suggestions.



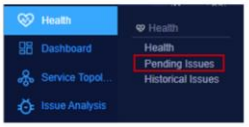
- Key Technologies:
  - To improve network reliability, redundant devices and links are usually used on an Ethernet switching network. However, due to network adjustment, configuration modification, upgrade, and cutover, data or protocol packets are often forwarded in a ring, which inevitably leads to loops.
  - Topology-based loop path display: Restores the loop path of a Layer 2 loop based on the switch neighbor relationship.

Case 11: Layer 2 Loop Issue Analysis (1)

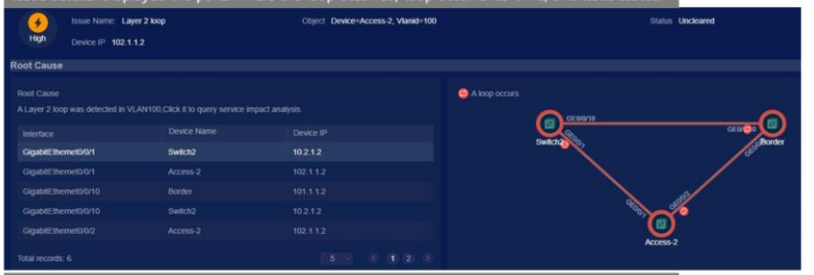
Case: On a campus network, the traffic on a device port surged, causing the port load to increase rapidly. As a result, the services on the port were interrupted, and therefore the fault needed to be located and rectified as soon as possible.

An O&M engineer used CampusInsight to rapidly locate the fault, and found that a Layer 2 loop occurred on the port. Services were quickly restored after the port was isolated.

1. Choose Health > Pending Issues. This displays existing issues. From this list, we can see that a Layer 2 loop occurred. Expand the list to view further details.



Issue details: Displayed the ports where the loop occurred, loop occurrence time, and issue status.

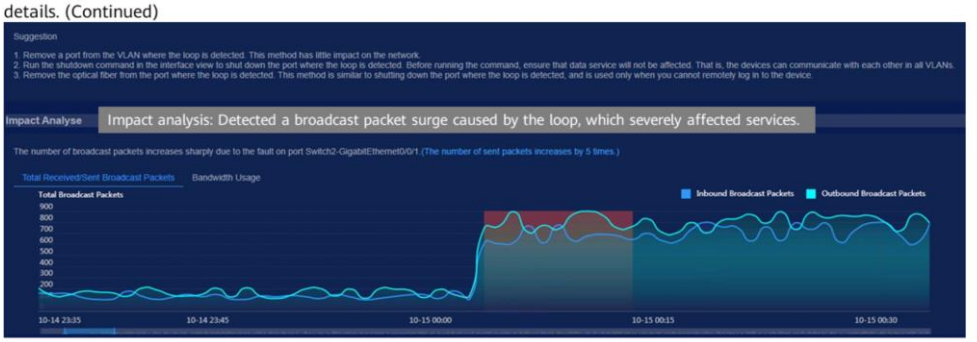


Topology restoration: Displayed information about the ports where the loop occurred, and restored the devices and ports affected by the loop based on the topology.

HUAWEI

Case 11: Layer 2 Loop Issue Analysis (2)

2. This displays existing issues. From this list, we can see that a Layer 2 loop occurred. Expand the list to view further details. (Continued)

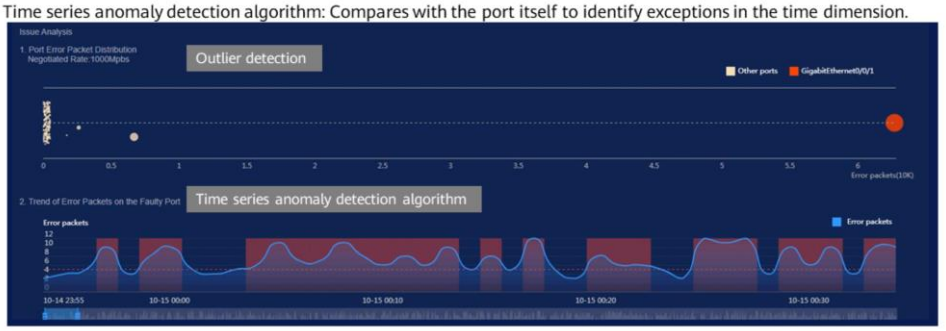


HUAWEI

- Use Case Benefits:
  - The Layer 2 loop issue function of CampusInsight proactively detects Layer 2 loops on the network, restores the loop topology based on switch link information, and locates the loop location.

Port Packet Error Issue

- Service impact: Service packets will be directly discarded on a port with error packets, leading to slow file downloads, multicast artifacts, or voice blurring.
- Issue Identification:
  - Outlier detection algorithm: Compares with other ports to identify exceptions in the space dimension.
  - Time series anomaly detection algorithm: Compares with the port itself to identify exceptions in the time dimension.



HUAWEI

- Issue Identification:
  - Outlier detection algorithm typical scenario: The negotiated rate supported by a terminal is different from that supported by the switch port connected to the terminal. As a result, the bit error rate of the port is significantly higher than that of other switch ports.
  - Time series anomaly detection algorithm typical scenario: Aged physical components such as Ethernet cables and RJ45 connectors can cause the bit error rate to gradually and continuously increase over a long period of time.
- Root cause analysis
  1. Check whether error packets are caused by inconsistent configurations (such as the negotiated rate) at both ends of a link.
  2. Guide customers to test port cables in order to figure out whether error packets are caused by cable aging or internal crosstalk.

Cass 12: Port Packet Error Issue Analysis (1)

Case: Jack is an O&M engineer in a company. Just this morning he viewed the health topology for building of Shenzhen Area on CampusInsight and discovered that error packets were detected on the network. Jack visited his colleague Sam who accessed the network through the faulty port. According to Sam's feedback, frame freezing occurred during video playback, and this affected routine office work.

By using CampusInsight, Jack was able to rapidly locate the fault and pinpoint that continuous error packets occurred on the faulty port. This was because the Ethernet cable connected to the port aged out. The fault was rectified and services were restored after the Ethernet cable was replaced.

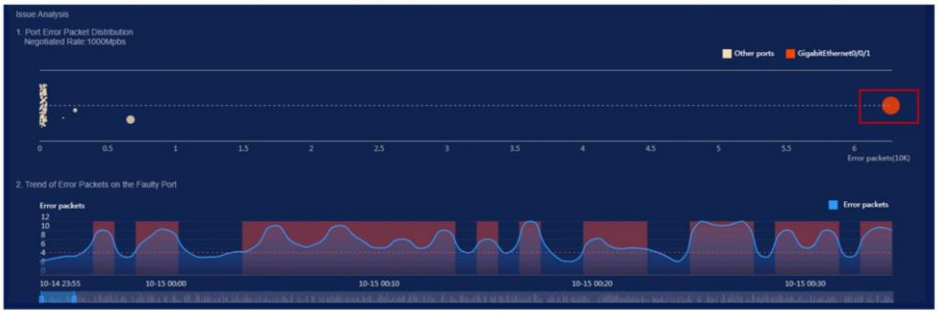
1. Choose Service Topology.

2. Error packets are detected on the network of *Shenzhen*.
Directly click the issue to bring up further details.

HUAWEI

Cass 12: Port Packet Error Issue Analysis (2)

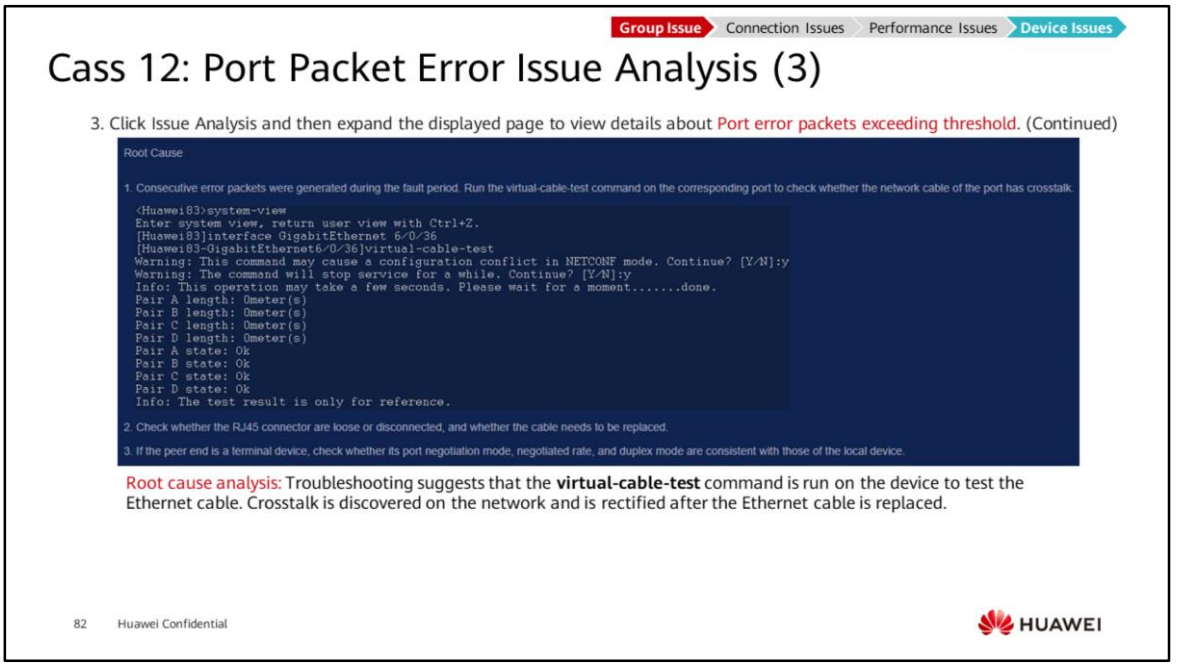
3. Click Issue Analysis and then expand the displayed page to view details about Port error packets exceeding threshold.



Skewness analysis: According to the historical big data analysis and the skewness algorithm, the number of error packets on the faulty port is significantly greater than that on other ports. This indicates that the port is faulty.

Exception identification: Check the number of error packets on the port and discover that error packets on the port are persistent in each period.

HUAWEI



- Use Case Benefits:
  - CampusInsight continuously monitors the number of error packets on ports, detects any substantial increase in error packets based on the dynamic baseline, and automatically generates corresponding issues. This approach accurately identifies error packet issues on the network in a timely manner, and provides expert suggestions.

Frequent Port Up/Down Issue

- Service impact: A device port frequently alternates between Up and Down states. As a result, the port is physically disconnected, interrupting services.
- Issue identification: Monitor the Up and Down events of each port. If a port frequently alternates between Up and Down states in a short period of time, an exception occurs.
- Root cause analysis:
  - Correlation analysis: Check whether the port status changes after/upon port negotiation to determine whether the port frequently alternates between Up and Down states due to inconsistent duplex modes and speeds negotiated between the local and peer ports.
  - Port cable test: Guide customers to test port cables in order to figure out whether the port repeatedly negotiates with the peer port due to damaged internal wires of the cable's RJ45 connector.

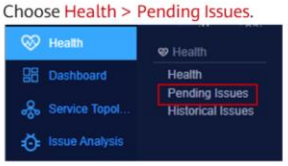
HUAWEI

Case 13: Frequent Port Up/Down Issue (1)

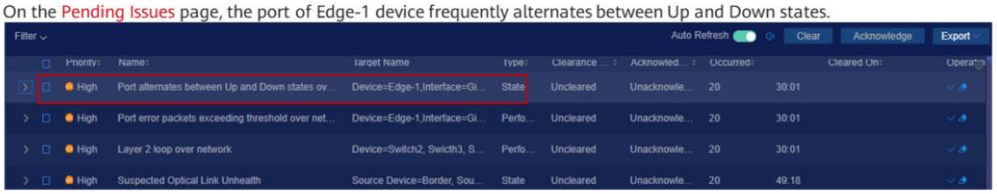
Case: George is a company employee. He turned on his laptop to work and found that the laptop frequently disconnected from the network.

He contacted an O&M engineer who used CampusInsight to rapidly locate the fault, and discovered that the wired port that was connecting George to the network frequently alternated between Up and Down states. A further check indicated that the internal wires of the Ethernet cable's RJ45 connector were damaged, and this was rectified after the Ethernet cable was replaced.

1. Choose Health > Pending Issues.



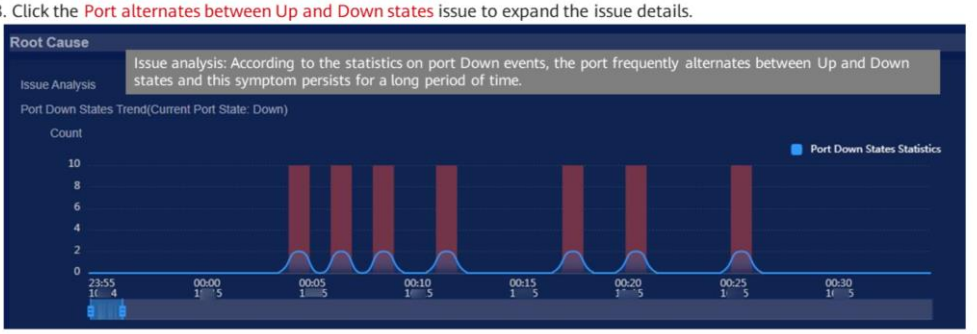
2. On the Pending Issues page, the port of Edge-1 device frequently alternates between Up and Down states.



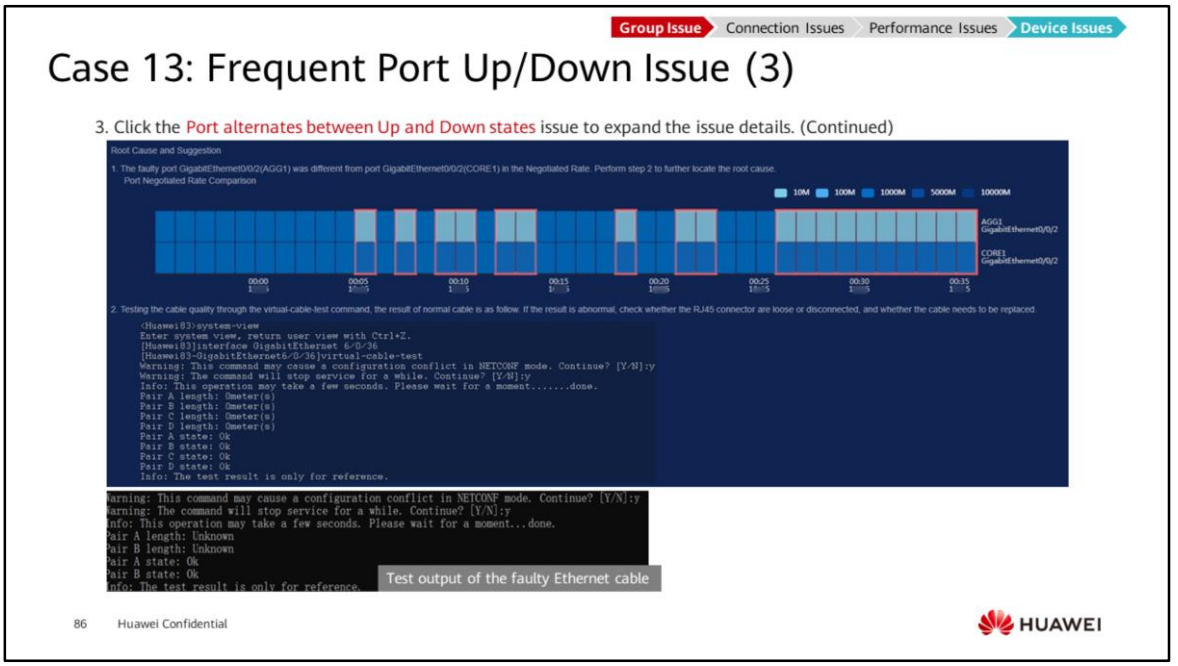
HUAWEI

Case 13: Frequent Port Up/Down Issue (2)

3. Click the Port alternates between Up and Down states issue to expand the issue details.



HUAWEI

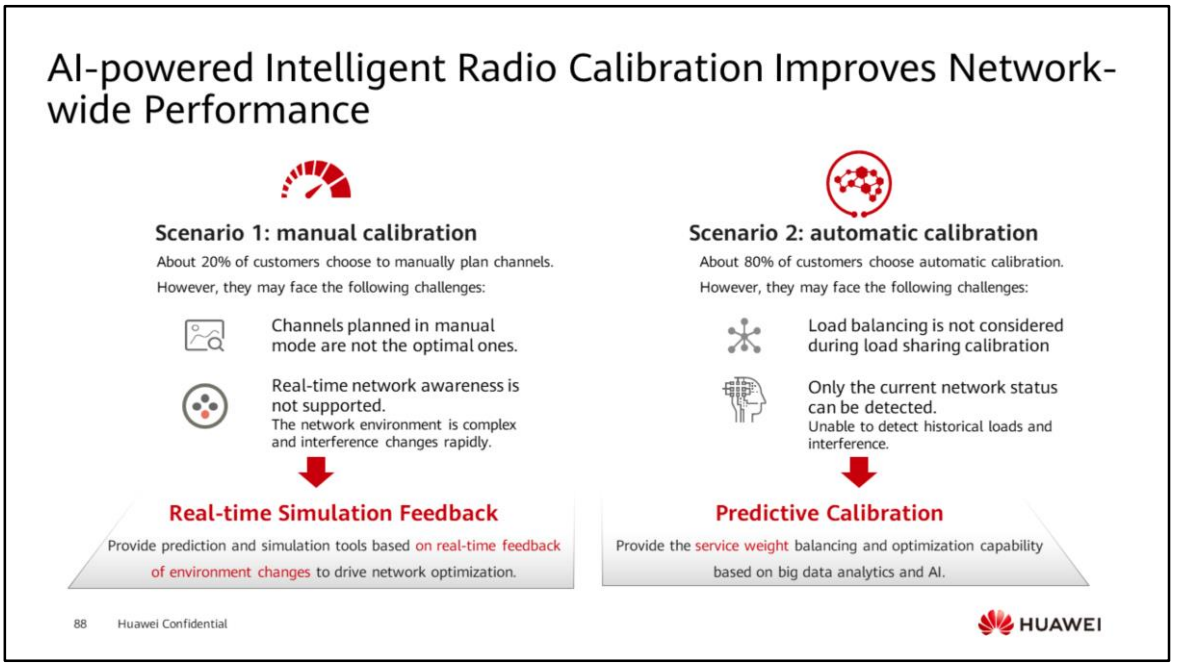


- Root cause analysis:
  - When the port alternates between Up and Down states, the negotiated rate of the port changes multiple times. As a result, the port repeatedly performs rate negotiation, causing intermittent disconnection.
  - Troubleshooting suggests that the **virtual-cable-test** command is run to check the Ethernet cable connected to the port. The command output indicates that the Ethernet cable has only 4 wires in 2 pairs. (A typical Ethernet cable has 8 wires in 4 pairs.) As a result, the port rate changes, causing intermittent disconnection. A further check indicates that the RJ45 connector of the Ethernet cable is faulty. The fault is rectified after the RJ45 connector is remade.
- Use Case Benefits:
  - CampusInsight continuously monitors the Down event of each port and can accurately detect any intermittent port disconnection on the network in a timely manner, providing expert suggestions.

Contents

1. CampusInsight Overview
2. **CampusInsight Functions and Demonstration**
   - Telemetry
   - Visibility
   - Analysis
   - Optimization
   - Assurance

HUAWEI



- Simulation feedback: CampusInsight evaluates the radio score and the number of APs waiting for calibration based on the radio and neighbor information of APs, displays the calibration simulation effect through the AI algorithm, and provides channel adjustment suggestions. The APs with this function must be deployed on the floor.
- Intelligent radio calibration: Historical big data is analyzed using the AI algorithm. Network devices periodically request big data and the analytics results based on the calibration policy to implement intelligent radio calibration.

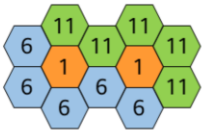
Scenario 1: Providing Optimal Channel Planning Suggestions Based on Neural Network Simulation Feedback (1)

Problem	Challenge
An IT engineer optimizes the network in the area where a fault is reported. However, faults are reported in surrounding areas after the optimization. The optimization is performed multiple times in several days, but the optimization result is not satisfactory, and the network stability becomes worse.	Expert experience-based optimization, high professional requirements; heavy analysis workload; wireless network not planned from the entire network perspective in manual mode.

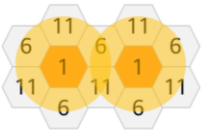
HUAWEI

Scenario 1: Providing Optimal Channel Planning Suggestions Based on Neural Network Simulation Feedback (2)

Technical Root Cause



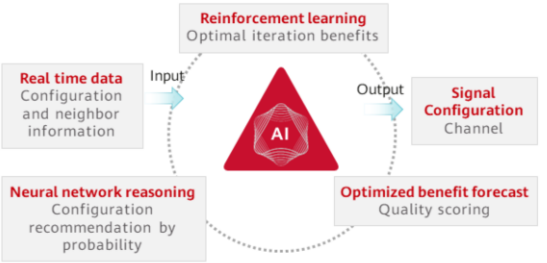
Interference between APs caused by improper channel planning



Two APs using channel 1 are too close to each other. As a result, the two APs interfere with each other.

Technical root cause: The impact of inter-AP interference, surrounding interference, and distance must be fully considered during AP configuration. In most cases, optimization can only ensure that the local network is at its optimal state, and comprehensive network-wide evaluation cannot be provided.

Wireless Network Simulation Feedback Solution



Expected result: network-wide optimal reasoning to properly allocate air interface resources; simulation capability (customers evaluate the simulation result based on network scores and determine whether to deliver the simulation result).

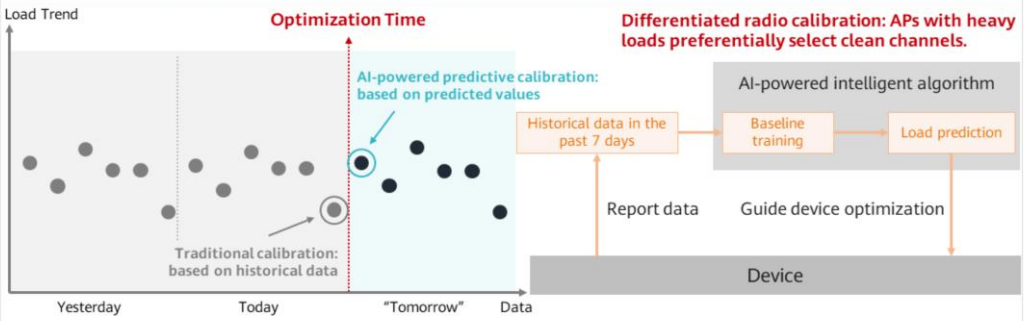
HUAWEI

Scenario 2: AI-Powered Predictive Calibration

Challenge: How to efficiently use spectrum resources based on AP load data?

On a wireless network, APs are busy to different extents. How to accurately predict the load trend of each AP and perform differentiated predictive calibration on the entire network?

Solution: Integrates multiple prediction models to accurately predict AP loads and implement differentiated radio calibration.



HUAWEI

Case 14: AI-Powered Predictive Calibration (1)

Case: The wireless network office area of a company was upgraded and reconstructed. Employees were temporarily moved to building C4 for centralized office. As a result, the number of employees in building C4 increased, and the network load also increased. Employees complained that the wireless network response was getting slower. Using the intelligent radio calibration, CampusInsight automatically identified high-load areas in building C4 and accordingly adjusted APs' frequency bandwidth, thereby improving client bandwidth and network experience.

1. Choose Intelligent Radio Calibration and Big Data Calibration.

2. Enable intelligent radio calibration. You are advised to enable this function in advance to improve the data training accuracy.

HUAWEI

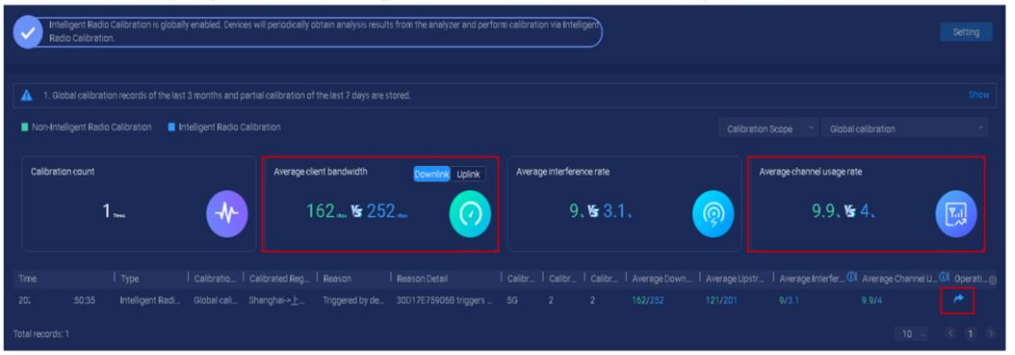
Case 14: AI-Powered Predictive Calibration (2)

3. Click Next. On the Load Optimization page, many high-load APs are identified by 4C-3F.

HUAWEI

Case 14: AI-Powered Predictive Calibration (3)

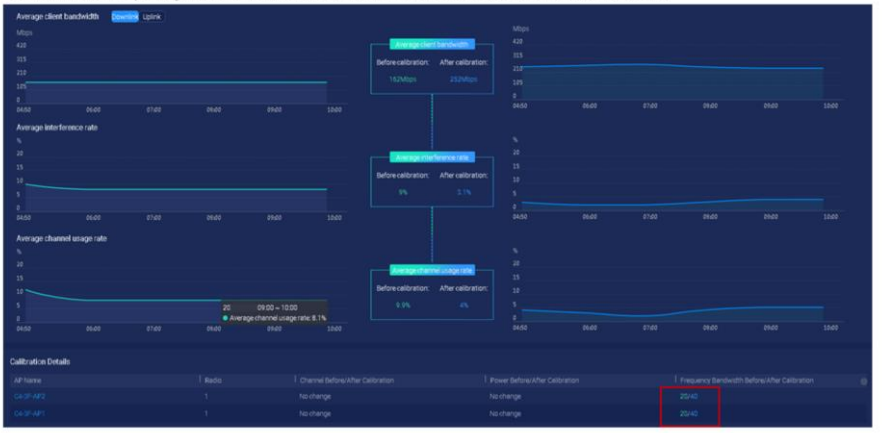
4. On the second day after big data calibration is enabled, the bandwidth of the APs on the third floor of building C4 is increased to 252 Mbps (by 50%) and the average channel utilization is reduced to 4% (by 50%).



HUAWEI

Case 14: AI-Powered Predictive Calibration (4)

5. Check the calibration details. The 5 GHz frequency band of the high-load APs on the third floor of building 4C is changed from 20 MHz to 40 MHz. The Internet access experience is improved and no frame freezing occurs.
Note: If the frequency band of APs increases, the client bandwidth will also increase.



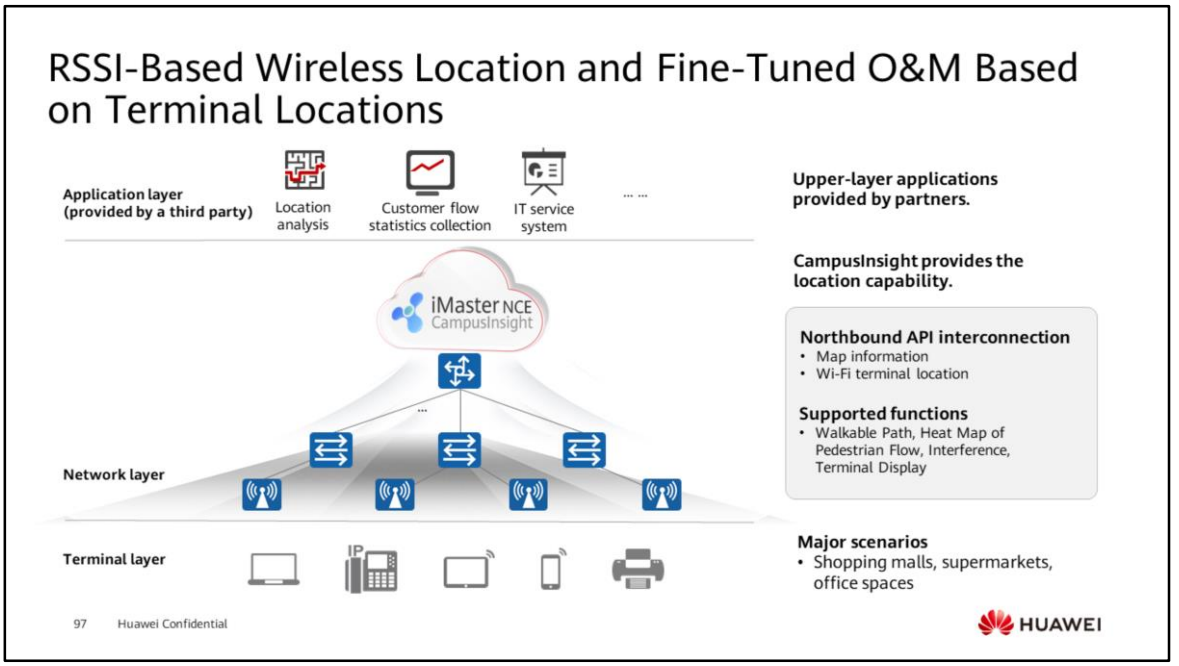
HUAWEI

- Use Case Benefits:
  - Using intelligent radio calibration, CampusInsight continuously collects massive data of real clients, identifies high-load APs and edge APs through AI algorithms, and provides decision-making data for differentiated system optimization, enabling the network to follow clients.

Contents

1. CampusInsight Overview
2. **CampusInsight Functions and Demonstration**
   - Telemetry
   - Visibility
   - Analysis
   - Optimization
   - Assurance

HUAWEI



- Note:
  - This version supports only RSSI-based wireless location, and other location methods are not supported.
  - Wireless location is applicable only to indoor scenarios.
  - Location accuracy: within 10 m, 60% of accuracy (independent RF scanning), 50% of accuracy (non-independent RF scanning); location delay: within 20s.
  - Wireless location data can be stored for up to 7 days.

Interference Source Locating

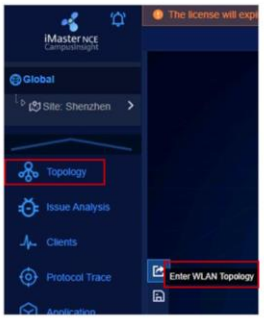
- CampusInsight supports location of Wi-Fi interference sources and non-Wi-Fi interference sources (such as Bluetooth, microwave, and audio/video devices).

HUAWEI

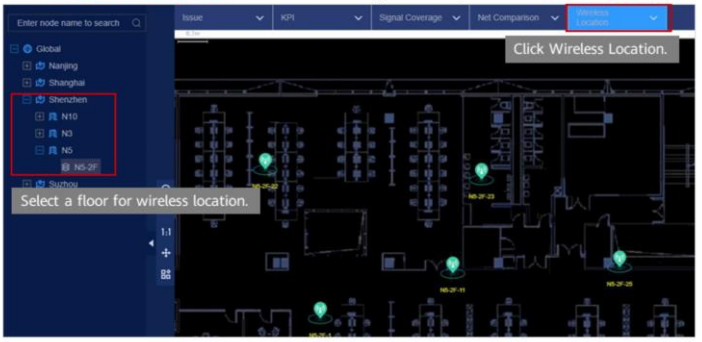
Case 15: Wireless Location (1)

Case: During the pandemic, a scenic spot wishes to strictly control personnel contact to reduce the pandemic spreading risks. The wireless location function enables administrators to check the client heat map in a specified period of time and identify crowded areas. In this manner, the scenic spot managers can perform crowd dispersing in advance and reduce contact risks.

1. Choose Service Topology > Enter WLAN Topology.



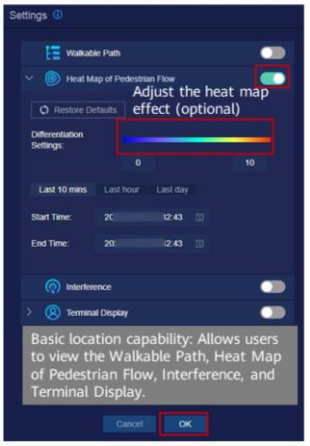
2. Select Shenzhen > N5 > N5-2F from the left navigation tree, and click Wireless Location.



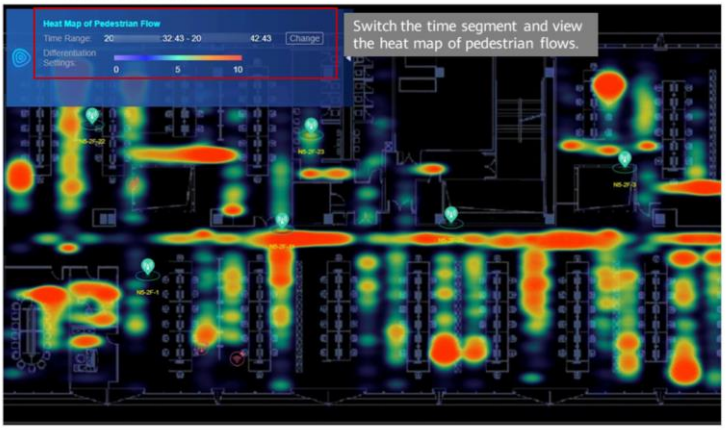
HUAWEI

Case 15: Wireless Location (2)

3. On the Settings page, enable Heat Map of Pedestrian Flow, and click OK.



4. Check the Heat Map of Pedestrian Flow. The crowded location can be identified by switching the time segment.



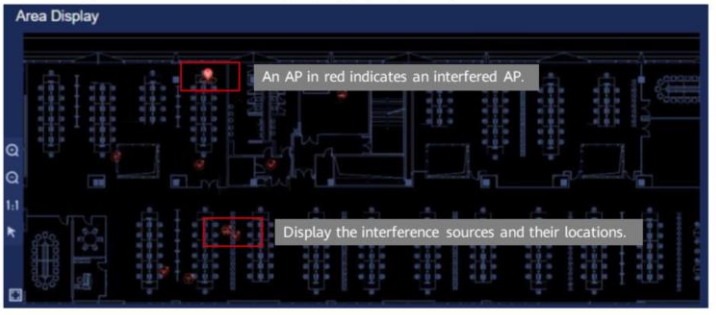
HUAWEI

- CampusInsight supports the display of heat maps, terminal locations, and walking paths.
  - Displays the people distribution heat map based on the specified time period.
  - Displays the locations of all Wi-Fi-enabled terminals, the location of a single user, and the walking path within a specified time segment.
  - Supports terminal MAC address anonymization.

Case 16: Interference Source Locating

Case: Employees of a company department report that the network quality is poor near their office seats, and file downloading and email exchange are very slow, deteriorating their work efficiency. The issue analysis result shows that the APs to which employees get connected are experiencing high interference, and the network quality becomes poor due to external interference sources. After interference source locating, it is found that there are rogue APs near the employees' office seats. The fault is rectified after the interference sources are eliminated.

1. Choose Issue Analysis > High Interference, click Original Issue List and click the issue.
2. On the high-interference issue analysis page, query the specific interference source information.



HUAWEI

Quiz

1. (Multiple-answer question) Which of the following functions can CampusInsight provide?( )
   A. Network Health Evaluation
   B. Wired-user Experience Visualization
   C. Protocol Tracing for Locating Wireless Access Issues
   D. Minute-level Fault Locating

HUAWEI

1. ABCD

Summary

- CampusInsight is a campus network analyzer launched by Huawei. It uses big data analytics technologies and machine learning algorithms to provide excellent network service assurance experience by analyzing data of each user at any time.
- CampusInsight can implement real-time experience visualization, minute-level fault locating, and intelligent network optimization. This course describes the main functions and application cases of CampusInsight, including wired and wireless network health, individual and massive fault analysis, intelligent optimization, Wireless Location, and Interference Source Locating.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

WAN Interconnection Solution and Technology Overview

HUAWEI

Foreword

- Wide area networks (WANs) have undergone a long history, and the technologies used on WANs have been updated many times. Enterprises have more diversified requirements over WANs than campus networks. In addition, the requirements for network connection mode, reliability, and security vary with enterprises.
- To meet the interconnection requirements of different enterprises, a variety of WAN interconnection technologies are developed.
- This course will focus on the WAN interconnection technologies and their application scenarios.

HUAWEI

Objectives

- Upon completion of this course, you will be able to:
  - Describe the basic architecture of a typical WAN interconnection solution.
  - Describe WAN interconnection networking technologies and their application scenarios.
  - Describe WAN interconnection reliability technologies and their application scenarios.
  - Describe WAN interconnection optimization technologies and their application scenarios.

HUAWEI

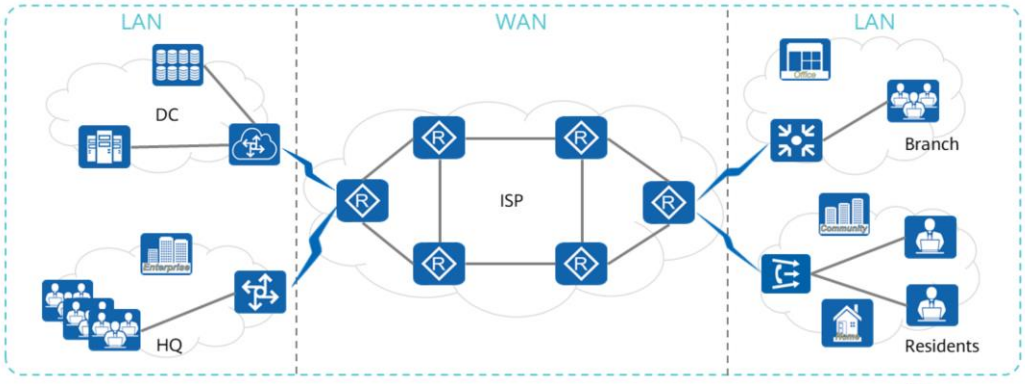
Contents

1. **Basic Architecture of a Typical WAN Interconnection Solution**
2. Networking Technologies and Their Applications of WAN Interconnection
3. Reliability Technologies and Their Applications of WAN Interconnection
4. Optimization Technologies and Their Applications of WAN Interconnection

HUAWEI

What Is a WAN?

- WANs provide interconnection services across regions, cities, and countries. A WAN usually spans a long distance (dozens or even thousands of kilometers). To meet long-distance transmission requirements on a WAN, optical fibers are often used as the interconnection media.

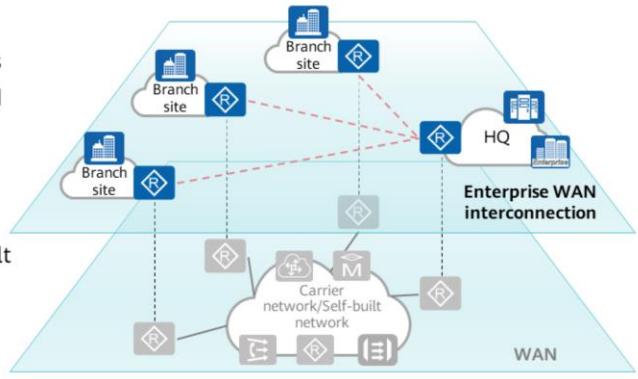


HUAWEI

- LAN
  - A local area network (LAN) is a computer network that connects computers, peripheral devices, databases, and other devices in a limited geographical area (such as a campus, factory, or organization) within thousands of meters.
- WAN
  - WANs provide wider coverage than LANs and metropolitan area networks (MANs). The communication subnet of a WAN mainly uses the packet switching technology. The communication subnet of a WAN can use the public packet switching network, satellite communication network, and wireless packet switching network to interconnect the LANs or computer systems in different areas for resource sharing.
  - The Internet is the largest WAN in the world.
- Relationship between the LAN and WAN:
  - A LAN is located in an area, whereas a WAN spans a larger area. For example, the headquarters of a large company is located in Beijing, and its branches are distributed all over the country. If the company connects all its branches through a network, a branch is a LAN, and the company network is a WAN.

WAN and Enterprise WAN Interconnection

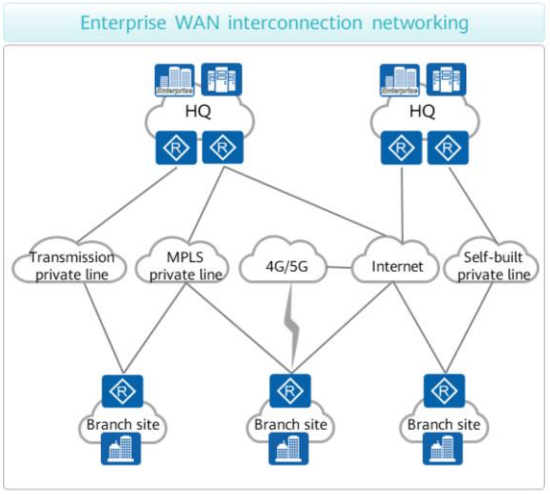
- Enterprise WAN interconnection refers to the interconnection between nodes at different levels, such as the headquarters (HQ), data centers (DCs), branches, fixed offices, and mobile offices within enterprises.
- Generally, enterprise WAN interconnection depends on a carrier-built or self-built WAN.



HUAWEI

Major Enterprise WAN Interconnection Modes

- Generally, enterprise WANs can be interconnected in the following modes:
  - Carriers' transmission or MPLS private lines are used to connect regional networks. This mode is expensive and is applicable to enterprises with high SLA requirements.
  - The carrier Internet + VPN technology is used for connection. This mode is applicable to small- and medium-sized branches that do not have high SLA requirements.
  - Carriers' point-to-point (P2P) private lines are used to implement cross-city or cross-border connections. This mode is mainly used for connections between DCs, headquarters, or important branches, and is expensive.
  - Industries such as electric power and transportation build network connections through self-built private lines.
- Usually, enterprise WANs are interconnected using a combination of the preceding modes.

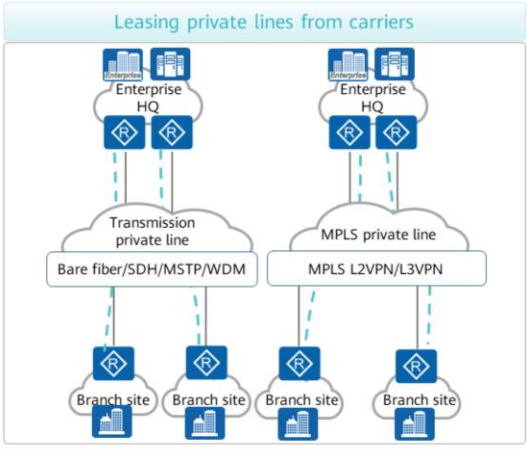


HUAWEI

- SLA is an agreement between the network service provider and the customer. It defines terms such as service type, service quality, and customer payment.
- According to the SLA requirements, the service provider uses multiple technologies and solutions to monitor and manage the network performance and traffic so as to meet the requirements defined in the SLA and generate reports about customers' network performance.

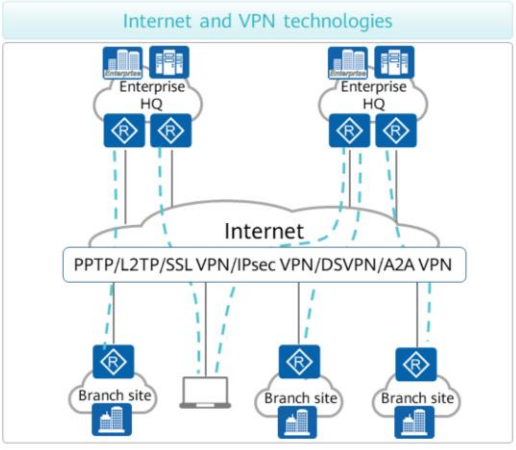
Enterprise WAN Interconnection Technologies: Carriers' Private Lines

- To ensure network reliability and security, enterprises lease transmission or MPLS private lines from carriers when constructing enterprise WANs.
  - Transmission private lines are expensive, but data is carried on dedicated lines, ensuring service quality and security.
  - MPLS private lines are cheaper than transmission private lines and can ensure service security. However, service reliability is not as good as that of transmission private lines.
  - A small number of enterprises (such as transportation and electric power enterprises) can deploy optical fibers and build their own backbone networks.



Enterprise WAN Interconnection Technologies: Internet and VPN

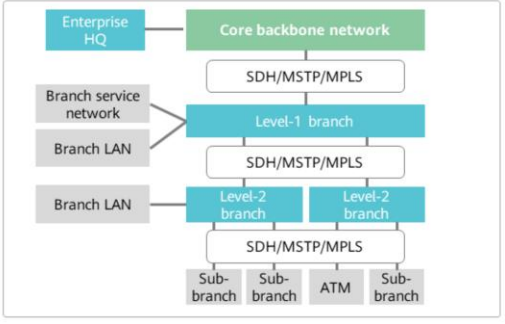
- With the development of the Internet, some enterprise services can be carried over the Internet.
- The Internet has security risks, so VPN technology is used to provide secure and reliable connections.
- Virtual Private Dial-up Network (VPDN) technologies, such as Point-to-Point Tunneling Protocol (PPTP), Layer 2 Tunneling Protocol (L2TP), and Point-to-Point Protocol over Ethernet (PPPoE), allow terminal users or branches to dial up to the carrier network or HQ network.
- Typically, Internet Protocol security virtual private network (IPsec VPN) technology is used to build network connections between enterprise branches or between enterprise branches and the HQ.
- To simplify IPsec configuration on large-scale networks, technologies such as Dynamic Smart VPN (DSVPN) and Any to Any (A2A) VPN have been developed and widely used.



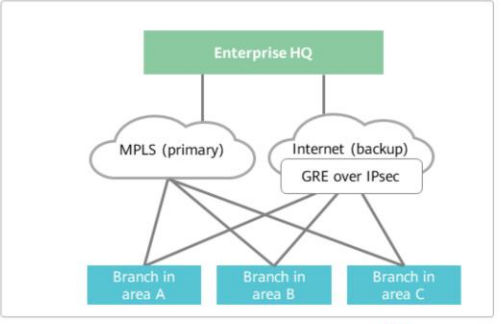
Common Application Scenarios of Enterprise WAN Interconnection Networking

- Enterprise WAN interconnection needs to be deployed based on enterprise requirements. For example, in the financial service industry, most enterprises lease transmission or MPLS private lines to guarantee network reliability and security. Considering network costs, other enterprises usually lease MPLS private lines as primary lines and Internet + VPN lines as backup lines.

WAN interconnection in the financial service industry



WAN interconnection for a wine enterprise



HUAWEI

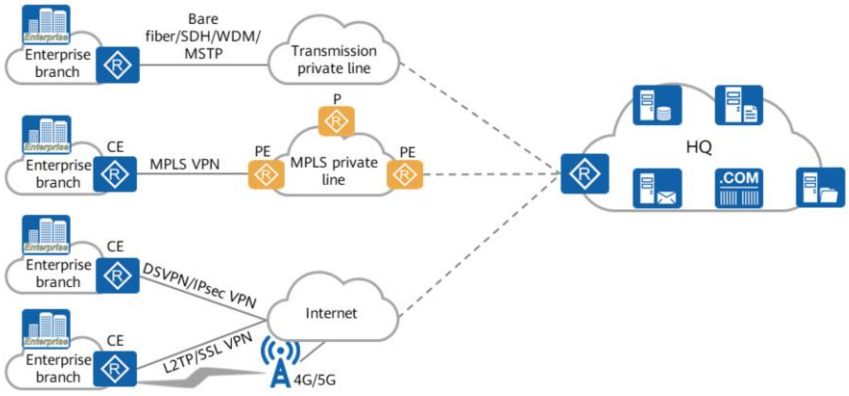
Contents

1. Basic Architecture of a Typical WAN Interconnection Solution
2. **Networking Technologies and Their Applications of WAN Interconnection**
   - WAN Interconnection Networking Solution
   - Private Line Technologies and Their Applications
   - VPN Technologies and Their Applications
3. Reliability Technologies and Their Applications of WAN Interconnection
4. Optimization Technologies and Their Applications of WAN Interconnection

HUAWEI

Enterprise WAN Interconnection Networking Solution

- There are many interconnection modes for enterprise WANs. Generally, one or more interconnection modes are used based on different enterprise requirements.



HUAWEI

Comparison Between Private Line and VPN Technologies

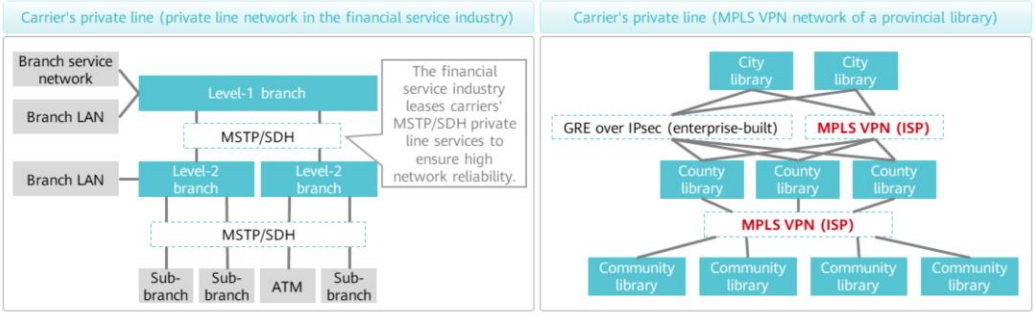
- Private lines were introduced very early. They can meet interconnection requirements of enterprises and ensure high network reliability and security. However, private lines are expensive.
- As networks develop, VPN technologies start to occupy more market shares. However, some industries demanding high security and reliability, such as the financial service industry, still prefer private line technologies.
- Selecting private line or VPN technologies depend on enterprises' services. The following table compares the two technologies.

Item	Private Line Technology	VPN Technology
Security	Relatively high: depending on ISPs	Very high: Data is encrypted before being transmitted and security control is in the hands of users.
Reliability	High: depending on ISP network reliability	Comparatively high: depending on reliability of Internet lines
Scalability	Medium: depending on ISPs	Based on TCP/IP technology, the access mode is flexible. A network can be scaled out as long as it is reachable.
Investment cost	The private line expense is very high and needs to be paid every month. In addition, the device expense needs to be invested in the initial period of network construction.	One-off device expense is invested, so there is no need to pay monthly operating expenses.
Support for mobile users	Mobile users can only connect to the network to which private lines are reachable, and internal mobile users leaving a LAN cannot access a private network.	Internal mobile users can use the Internet for secure access, eliminating geographical differences.
Transmission bandwidth	Leased bandwidth is low because of the high price.	The Internet is cheap, and the leased bandwidth is high.

HUAWEI

Overview of Carriers' Private Lines

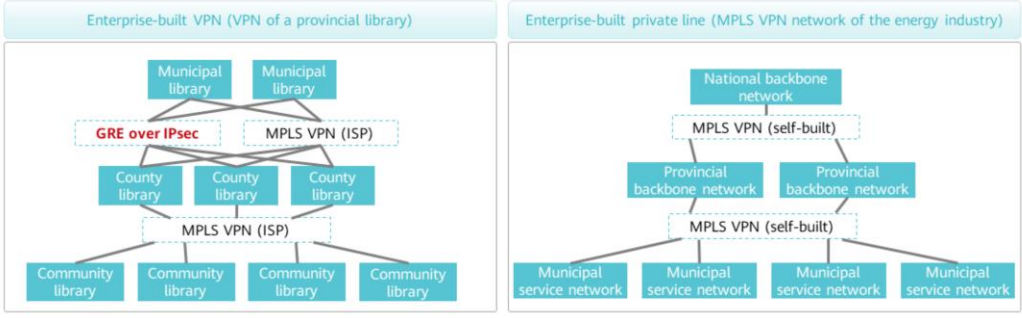
- Carriers have a large number of line resources and launch different private line services based on different industries and scenarios.
- Carriers' high-quality transmission private line services mainly include SDH, MSTP, and bare fiber services, which are expensive but deliver excellent performance.
- MPLS VPN is another type of private line services provided by carriers. MPLS VPN private lines provide slightly lower performance than transmission private lines but are less expensive.



HUAWEI

Overview of Enterprise-Built Private Line and VPN

- Enterprises can establish VPNs, such as SSL VPN, DSVPN, and IPsec VPN, through carriers' networks.
- Some large enterprises can lay out optical fibers by themselves and set up private lines. However, only few enterprises can lay out optical fibers by themselves.
- VPNs built by enterprises are more and more widely used because they are cost-effective, easy to expand, and controllable.



HUAWEI

Contents

1. Basic Architecture of a Typical WAN Interconnection Solution
2. **Networking Technologies and Their Applications of WAN Interconnection**
   - WAN Interconnection Networking Solution
   - Private Line Technologies and Their Applications
   - VPN Technologies and Their Applications
3. Reliability Technologies and Their Applications of WAN Interconnection
4. Optimization Technologies and Their Applications of WAN Interconnection

HUAWEI

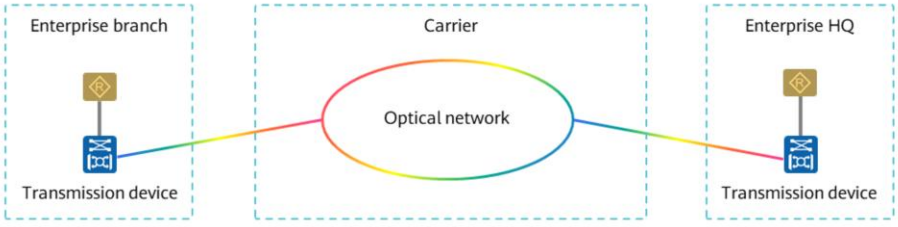
Overview of Private Line Technologies

- Private line technologies were introduced very early. As networks develop, many private line technologies, such as frame relay (FR) and ATM, are no longer used. Currently, the following private line technologies are widely used:
  - Bare fiber: Carriers provide bare fibers along which no intermediate device is deployed. Therefore, bare fibers are expensive.
  - SDH/MSTP/WDM: Transmission private lines use transmission devices to build hard pipes over optical fibers, ensuring good performance. The price of such private lines is lower than that of bare fibers.
  - MPLS VPN: MPLS private lines use Ethernet for network access and do not have hard pipes. The performance of MPLS VPN is poorer than that of transmission private lines, but MPLS VPN is the cheapest among all types of private lines.

HUAWEI

Overview of Bare Fibers

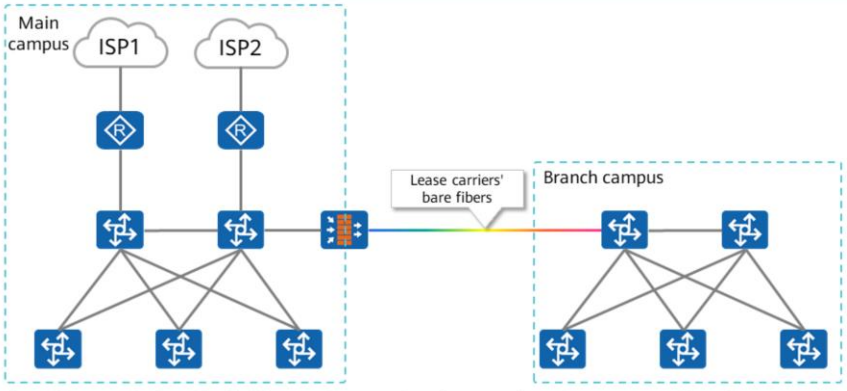
- A carrier provides a bare fiber line along with no intermediate device is deployed. The network capacity depends on the enterprise devices at both ends of the bare fiber.
- Bare fibers are charged based on the distance. A longer distance indicates a higher cost. Generally, the maximum transmission distance of a hop of an optical fiber is 300 km. If the distance between two sites exceeds 300 km, a regeneration device needs to be deployed.



HUAWEI

Exemplary Application Scenario of Bare Fibers

Application scenario of bare fibers

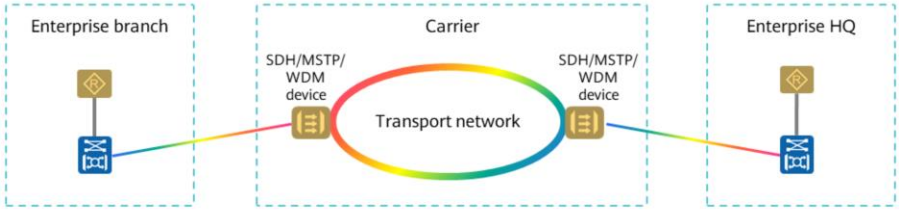


- Carriers' bare fibers can be leased to build network connections between the main campus and branch campus in the same city. This practice simplifies network management and access authentication management.

HUAWEI

Overview of SDH/MSTP/WDM Private Lines

- Enterprises that require long-distance transmission and high network reliability and security can lease SDH/MSTP/WDM private lines.
- This type of private line is a transmission private line. Tenants exclusively occupy part of the bandwidth of the transmission private line. Because multiple users share the transmission private line, its price is lower than that of bare fibers. Although transmission private lines are shared by tenants, they exclusively occupy bandwidth and use hard pipes. Therefore, they deliver high network reliability and security.
- MSTP and WDM private lines are widely used on the live network, and SDH private lines are still used in a few areas.

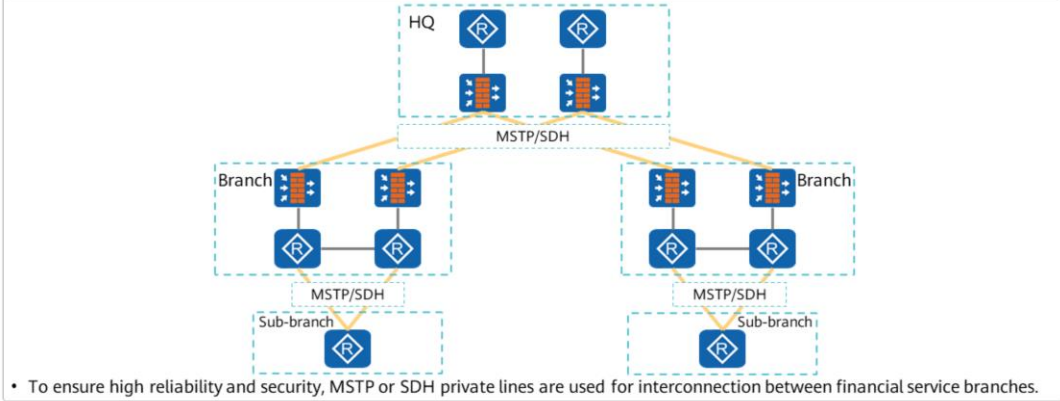


HUAWEI

- SDH is a TDM system and a traditional circuit scheduling mode.
- The multi-service transport platform (MSTP) receives, processes, and transmits TDM, ATM, and Ethernet services.
- WDM uses multiple lasers to transmit multiple beams of lasers with different wavelengths over a single optical fiber. The transmission bandwidth of WDM devices is high, and the live-network bandwidth can reach up to 8 Tbit/s.

Exemplary Application Scenario of SDH/MSTP/WDM Private Lines

Application scenario of SDH/MSTP/WDM private lines

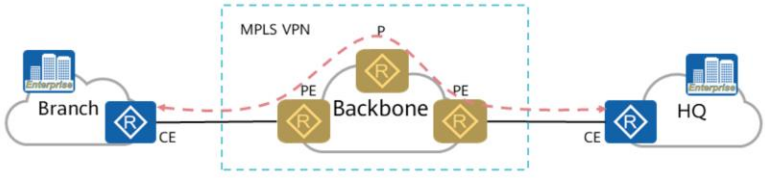


- To ensure high reliability and security, MSTP or SDH private lines are used for interconnection between financial service branches.

HUAWEI

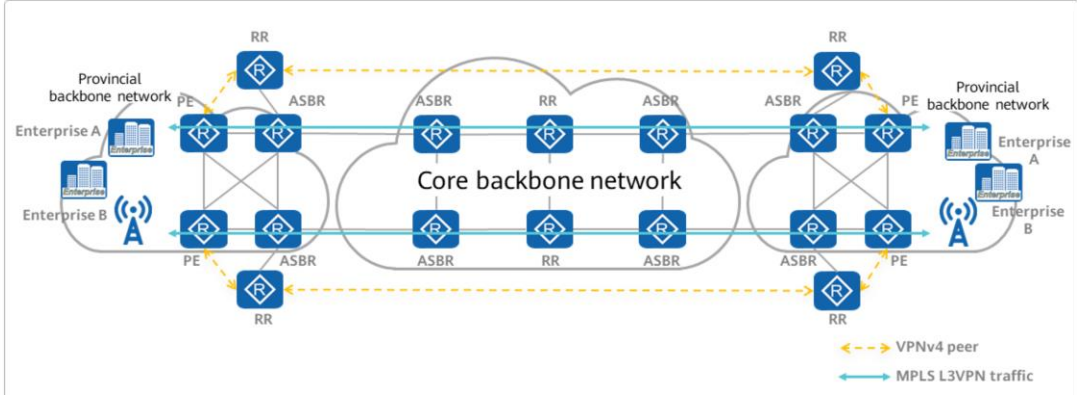
MPLS VPN Private Line

- MPLS VPN technology is widely used in enterprise interconnection scenarios. MPLS L2VPN or MPLS L3VPN can be deployed based on enterprise requirements. MPLS VPN makes a compromise between the cost and performance, so it is very popular.
- For enterprises that can build their own WANs, such as railways and electric power companies, MPLS VPN is an easy-to-manage and low-cost VPN technology. For enterprises that cannot build their own WANs, MPLS VPN is expensive.
- The enterprises that require high network reliability and security can use the MPLS VPN private line as the primary link and the GRE over IPsec line as the backup link.



Exemplary Application Scenario of MPLS VPN

Application scenario of MPLS VPN



- The MPLS L3VPN solution is widely used on the live network. Some large networks may use inter-AS MPLS L3VPN.

HUAWEI

- There are three types of inter-AS MPLS L3VPN solutions: Option A, Option B, and Option C.
- Option A applies to small inter-AS MPLS L3VPNs. Option B applies to midsize and large inter-AS MPLS L3VPNs. Option C applies to large or super-large inter-AS MPLS L3VPNs.

Contents

1. Basic Architecture of a Typical WAN Interconnection Solution
2. **Networking Technologies and Their Applications of WAN Interconnection**
   - WAN Interconnection Networking Solution
   - Private Line Technologies and Their Applications
   - VPN Technologies and Their Applications
3. Reliability Technologies and Their Applications of WAN Interconnection
4. Optimization Technologies and Their Applications of WAN Interconnection

HUAWEI

Overview of VPN Technologies

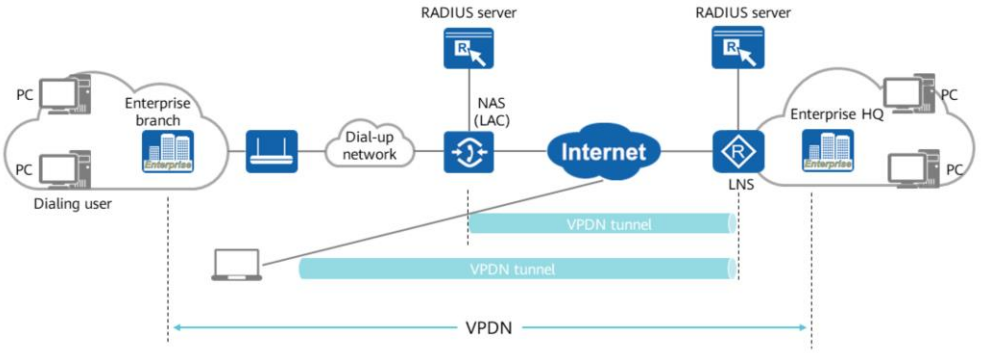
- VPN technologies are widely used in scenarios where enterprises build their own Internet.
- VPN technologies can be classified into the following three types based on the service usage:
  - Access VPN (virtual private network for remote access): also called dial-up VPN or VPDN. Generally, L2TP VPN technology is used.
  - Intranet VPN (internal virtual private network of an enterprise): connects gateways and connects resources of the same company through the company's network architecture. Generally, GRE or DSVPN technology is used.
  - Extranet VPN (extended internal virtual private network of an enterprise): is used to build an extranet with the enterprise network of a partner. Generally, SSL VPN technology is used.

HUAWEI

- By service usage:
  - Access VPN: enables mobile employees, remote office employees, and remote small offices to establish private network connections with enterprise intranet and extranet through a public network. There are two types of access VPN connections: client-initiated and NAS-initiated.
  - Intranet VPN: Intranet VPN is an extension or replacement of traditional private lines or other enterprise networks to connect distribution points within an enterprise through a public network.
  - Extranet VPN: extends enterprise networks to suppliers, partners, and even clients over a public network.
- According to the layers of tunnels in the OSI model:
  - Layer 2 tunneling protocol: encapsulates PPP frames into a tunnel. Layer 2 tunneling protocols include the Point-to-Point Tunneling Protocol (PPTP), Layer 2 Forwarding (L2F), and Layer 2 Tunneling Protocol (L2TP).
  - Layer 3 tunneling protocol: Only Layer 3 packets are carried in a tunnel. Existing Layer 3 tunneling protocols include Generic Routing Encapsulation (GRE) and IPsec. IPsec includes the Authentication Header (AH) protocol and Encapsulating Security Payload (ESP) protocol.

Access VPN Overview

- Access VPN uses virtual private dial-up network (VPDN) technology, and is a type of VPN service based on dial-ups. It can be used for enterprise interconnection or remote access to enterprise networks.



- VPDN is implemented by using a tunneling technology. That is, data of an enterprise network is encapsulated in a tunnel for transmission. On an interface between the source LAN and public network, the tunneling technology encapsulates data as a payload in a data format that can be transmitted on a public network. On an interface between the destination LAN and the public network, it decapsulates data to extract the payload. The logical path through which encapsulated data packets are transmitted on the Internet is called a tunnel. To ensure that data is encapsulated, transmitted, and decapsulated smoothly, the communication protocol is the core.
- VPDN provides three common tunneling technologies:
  - Point-to-Point Tunneling Protocol (PPTP)
  - Layer 2 Forwarding (L2F)
  - Layer 2 Tunneling Protocol (L2TP)

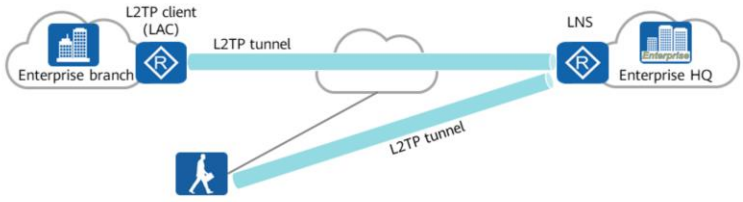
PPTP Overview

- PPTP is a type of VPN protocol and has a history of more than 20 years. This protocol relies on encryption, authentication, and Point-to-Point Protocol (PPP) for negotiation. It requires only the user name, password, and server address for connection setup.
- PPTP is fast, but has weak encryption. Among all VPN protocols, PPTP has the lowest encryption level and must be based on IP networks.

HUAWEI

L2TP Overview

- L2TP, an open standard of IETF, combines advantages of PPTP. L2TP is especially suitable for setting up an access VPN and has become a de facto industry standard.
- L2TP is only a tunneling protocol and does not provide encryption or privacy protection. Therefore, L2TP is usually used together with IPsec.
- L2TP is one of commonly used enterprise interconnection technologies. When L2TP is used, an AAA server is required. L2TP is a good choice for constructing an L2VPN.

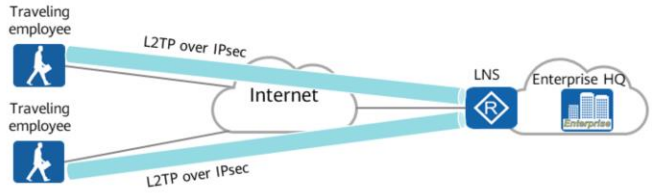


HUAWEI

Live-Network Application of Access VPN

- Access VPN is mainly used for remote access of intranet users, and L2TP over IPsec is most widely used.
- PPTP requires the support of the Windows operating system. In addition, the IP address of the Windows server on the intranet needs to be mapped through NAT for extranet access. Therefore, PPTP is difficult to deploy and is seldom used.

Live-network application of L2TP over IPsec



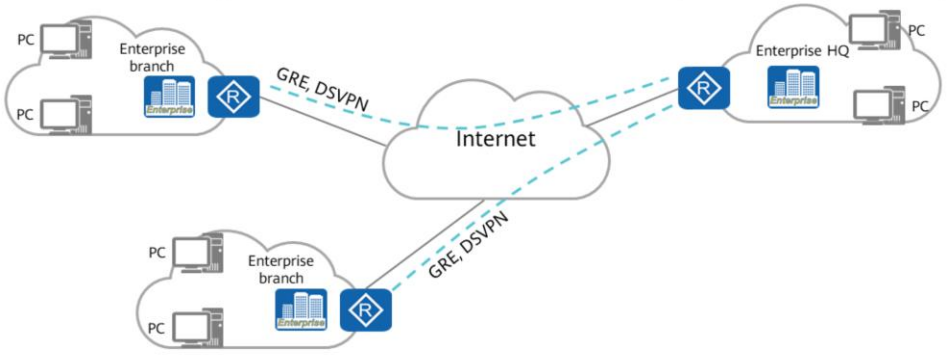
- L2TP connections can also be used between branch sites. However, L2TP cannot transmit multicast data or advertise routes between the HQ and branches. Therefore, L2TP is mainly used for remote user access on the live network.

HUAWEI

Access VPN Overview | Intranet VPN Overview | Extranet VPN Overview

Intranet VPN Overview

- Intranet VPN technology is used to construct a VPN between gateways based on the Internet. GRE and DSVPN technologies are mainly used.
- GRE and DSVPN technologies are used to establish VPNs between enterprise branches and the HQ.

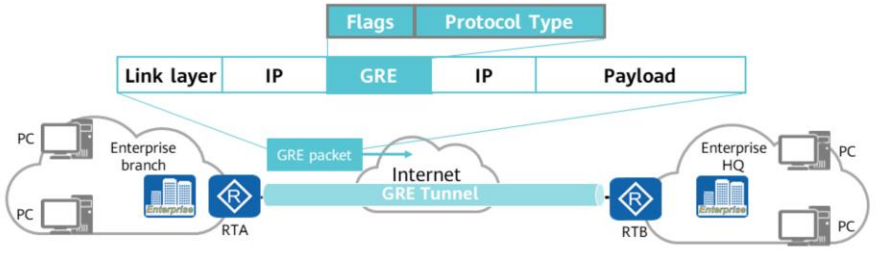


HUAWEI

- Intranet VPN mainly uses the following technologies:
  - GRE
  - GRE over IPsec
  - DSVPN
  - DSVPN IPsec

GRE Overview

- Generic Routing Encapsulation (GRE) is used to encapsulate packets of some network-layer protocols (such as IP, IPX, and AppleTalk) so that the encapsulated packets can be transmitted over the network on which another network-layer protocol is applied.
- GRE is typically used on networks with a few branch sites.



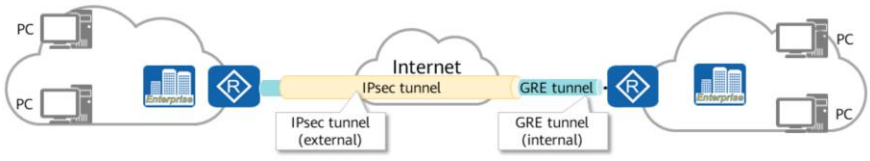
HUAWEI

- GRE is a Layer 3 tunneling technology. A GRE tunnel is a virtual P2P connection that transmits encapsulated data packets.
- The two ends of a GRE tunnel are tunnel interfaces which encapsulate and decapsulate data packets. The tunnel interface that sends encapsulated packets is called the tunnel source interface, and the one that receives these packets on the peer end is called the tunnel destination interface.
- The packet encapsulation process in the figure is as follows:
  - After receiving an IP packet, RTA's interface that connects to the enterprise branch sends a packet to the IP protocol module.
  - The IP protocol module checks the destination address in the packet header to determine how to forward this packet. If the packet is destined for the other end of the GRE tunnel, the IP protocol module sends the packet to the tunnel interface.
  - After receiving the packet, the tunnel interface encapsulates the packet using GRE and delivers the packet to the IP protocol module.

Access VPN Overview | Intranet VPN Overview | Extranet VPN Overview

GRE over IPsec Overview

- GRE does not support encryption. IPsec supports only the IP protocol and does not support multicast.
- GRE over IPsec combines advantages of both GRE and IPsec, and offsets their disadvantages.
- GRE over IPsec is a point-to-point VPN technology commonly used by enterprises.



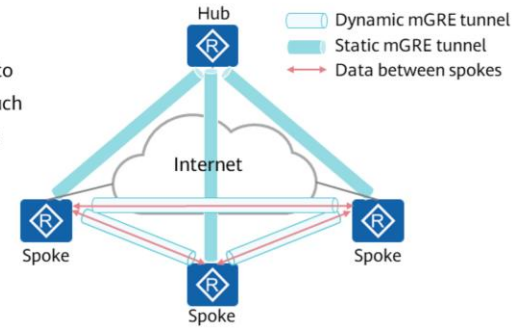
HUAWEI

- GRE encapsulates multicast data to allow data to be transmitted through GRE tunnels. Currently, IPsec can encrypt only unicast data. If multicast data, such as routing protocol, voice, and video data, needs to be transmitted over IPsec tunnels, a GRE tunnel can be established to encapsulate multicast data, and then IPsec encrypts the encapsulated packets. In this way, multicast data is encrypted and transmitted in the IPsec tunnel.
- GRE over IPsec combines advantages of both GRE and IPsec. It enables a network to support multiple upper-layer protocols and multicast packets, as well as packet encryption, identity authentication, and data integrity check.
- GRE over IPsec encapsulates packets using GRE, and then IPsec.
- GRE over IPsec supports the following encapsulation modes:
  - Tunnel mode
  - Transport mode

Access VPN Overview | Intranet VPN Overview | Extranet VPN Overview

DSVPN Overview

- DSVPN overcomes defects of GRE over IPsec and enables enterprises with a large number of branches to easily build VPNs.
- DSVPN is a technology that dynamically establishes GRE tunnels. It uses the Next Hop Resolution Protocol (NHRP) to dynamically collect, maintain, and advertise information such as the public IP address of each spoke, allowing the source branch to obtain the public IP address of the destination branch.
- DSVPN uses mGRE technology to enable VPN tunnels to transmit multicast and broadcast packets, and a tunnel interface can establish VPN tunnels with multiple peers.
- The GRE tunnel established by DSVPN can still use IPsec technology to ensure tunnel security.



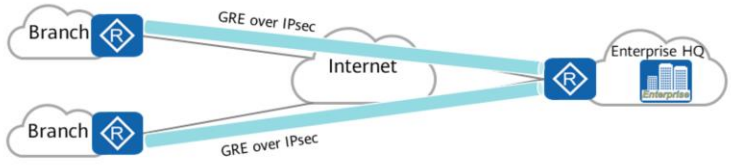
HUAWEI

- DSVPN resolves the following defects of GRE over IPsec:
  - All traffic must pass through the hub.
  - The hub configuration needs to be modified when a site is added.
  - If spokes use dynamic addresses, problems may occur when P2P GRE is deployed.

Live-Network Application of Intranet VPN

- On the live network, intranet VPN is mainly used for interconnection between enterprise branches and the HQ or between branches.
- GRE over IPsec is widely used on the live network. For enterprises with many branches, Efficient VPN can be used to simplify the branch configuration. IPsec link redundancy can be deployed to ensure GRE reliability.

Live-network application of GRE over IPsec

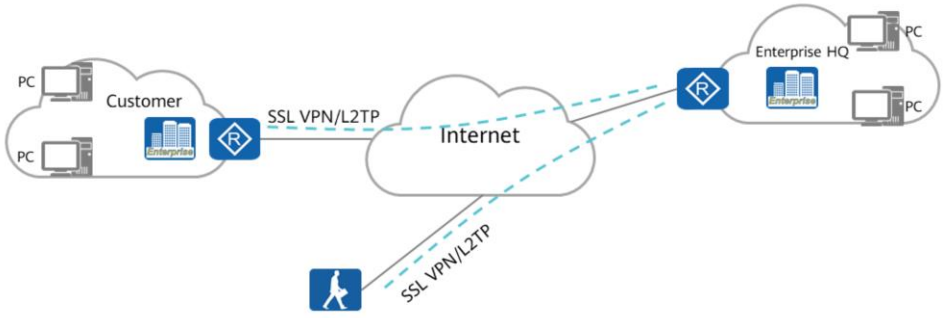


- DSVPN is seldom used on the live network due to the following reasons: 1. Many enterprises expect that traffic between branches can pass through the HQ to facilitate management. 2. Enterprise O&M personnel just have a basic understanding of DSVPN, which is inconvenient for O&M.

HUAWEI

Extranet VPN Overview

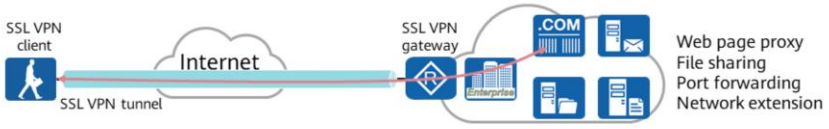
- Extranet VPN is mainly used to deliver secure network access between customers or suppliers. Traveling employees can also use extranet VPN to access the enterprise network.
- Extranet VPN mainly uses SSL VPN and L2TP.



HUAWEI

SSL VPN Overview

- SSL VPN is mainly used to enable traveling employees to remotely access the enterprise intranet, which is an extension of the enterprise intranet on the WAN.
- SSL VPN authenticates and controls users based on HTTP. Users do not need to configure SSL VPN, which is easy to use.



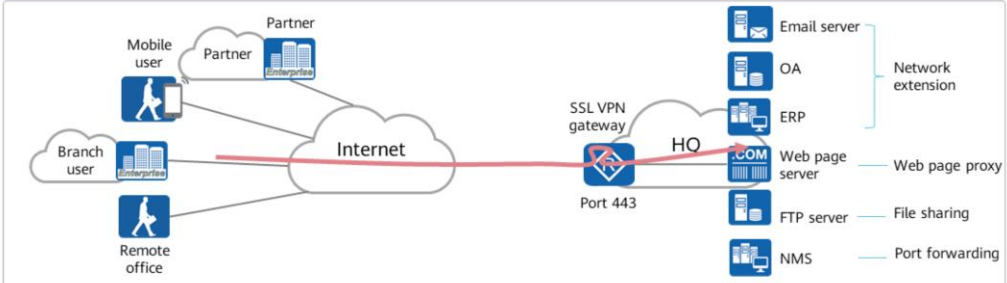
HUAWEI

- SSL VPN is a VPN remote access technology based on SSL. Mobile users (referred to as remote users in SSL VPN) can use SSL VPN to securely and conveniently access enterprise intranets and intranet resources, improving work efficiency.
- Before SSL VPN is developed, VPN technologies such as IPsec and L2TP are used to enable remote user access. However, these VPN technologies have the following disadvantages:
  - Remote users need to install specific client software on their terminals, leading to difficult network deployment and maintenance.
  - The IPsec or L2TP VPN configuration is complex.
  - Network management personnel cannot perform fine-tuned control over the remote users' access permission to enterprise intranet resources.

Exemplary Application Scenario of Extranet VPN

- It is convenient for external users to access the intranet through SSL VPN. They can directly access the intranet after passing web page authentication, without having to install a client. In addition, there are many SSL VPN service options, such as web page proxy, file sharing, port forwarding, and network extension.

Application scenario of SSL VPN



- SSL VPN is easy to use. You can use a browser to access the enterprise intranet. However, not all routers support SSL VPN. For routers that do not support SSL VPN, you can use L2TP to establish connections between the extranet and intranet.

HUAWEI

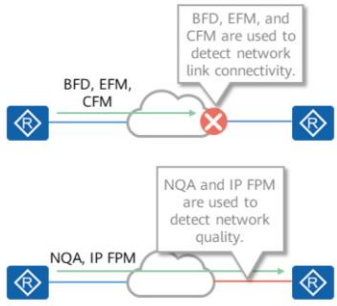
Contents

1. Basic Architecture of a Typical WAN Interconnection Solution
2. Networking Technologies and Their Applications of WAN Interconnection
3. **Reliability Technologies and Their Applications of WAN Interconnection**
   - Link Detection Technologies and Their Applications
   - Network and Service Reliability Technologies and Their Applications
4. Optimization Technologies and Their Applications of WAN Interconnection

HUAWEI

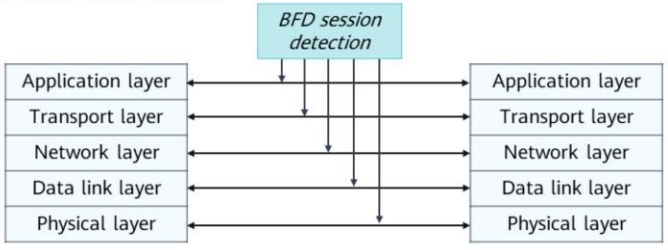
Overview of Link Detection

- Fluctuation of network link quality affects service quality. How to quickly detect the link quality fluctuation is the first step in improving link quality.
- There are many protocols and technologies for detecting link quality, which are classified into two types:
  - Link connectivity detection technologies:
    - Bidirectional Forwarding Detection (BFD)
    - Ethernet in the First Mile (EFM)
    - Connectivity Fault Management (CFM)
  - Link quality detection technologies:
    - Network Quality Analysis (NQA)
    - IP Flow Performance Measurement (IP FPM)
- On the live network, BFD is typically used to detect link connectivity, and NQA or IP FPM is typically used to detect link quality.



Overview of BFD

- BFD provides a universal, standardized, media-independent, and protocol-independent fast failure detection mechanism. It has the following advantages:
  - Provides low-overhead and fast failure detection for channels between adjacent forwarding engines.
  - Performs uniform detection for all media and protocol layers in real time.
- BFD is a simple Hello protocol. Two systems establish a BFD session channel and periodically send BFD packets to each other. If one system does not receive BFD packets from the other system within a certain period, the system considers that a fault occurs on the channel.

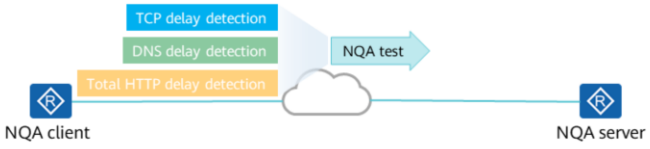


HUAWEI

- Only one BFD session can be established in a data path. If different applications need to use different BFD parameters on the same data path, use the BFD parameters that can meet the requirements of all applications to configure a unique BFD session and enable the status changes of the BFD session to be reported to all the applications bound with the BFD session.

Overview of NQA

- To visualize the quality of network services and allow users to check whether the quality of network services meets requirements, the following measures must be taken:
  - Enable devices to provide network service quality information.
  - Deploy probe devices to monitor network service quality.
- The preceding measures require devices to provide statistical parameters such as the delay, jitter, and packet loss rate and require dedicated probe devices. These requirements increase investments on devices.
- When NQA is deployed on devices, dedicated probe devices do not need to be deployed, effectively reducing costs. NQA can accurately test the network running status and output statistics.
- NQA measures network performance and collects statistics about the response time, network jitter, and packet loss rate in real time.

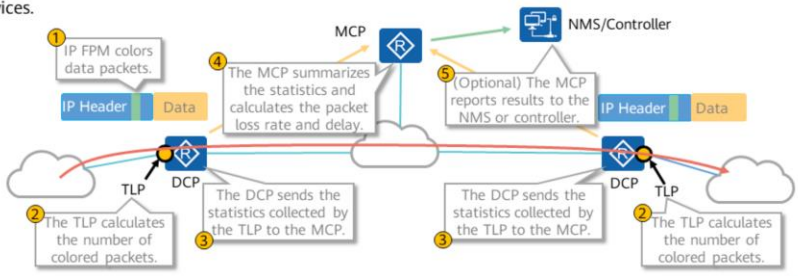


HUAWEI

- Additionally, NQA measures the performance of different protocols running on the network. This facilitates real-time collection of network performance counters, such as the total HTTP connection delay, TCP connection delay, DNS resolution delay, file transfer rate, FTP connection delay, and DNS resolution error rate.

Overview of IP FPM

- With the advent of the cloud computing era, end-to-end service performance measurement becomes essential. However, the commonly used NQA technology has the following defects in end-to-end network performance measurement scenarios:
  - NQA simulates service packet forwarding on the network by constructing service packets. Therefore, the collected performance statistics are not accurate.
  - NQA does not support end-to-end performance measurement across network layers, and cannot monitor or measure network performance in a multipath scenario of IP networks.
- IP flow performance measurement (IP FPM) can effectively solve these problems. It is a general IP network performance measurement solution. IP FPM can directly measure service packets, and the measurement data can reflect the performance of IP networks. In addition, IP FPM can monitor the changes of services carried by IP networks online and accurately reflect the running status of services.

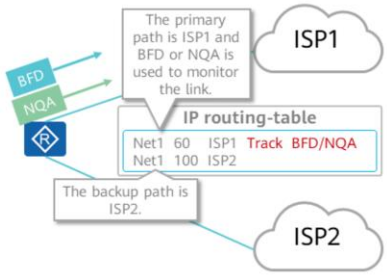


HUAWEI

- TLP is short for Target Logical Port.
  - TLPs are interfaces on the edge nodes of the network and provide the following functions:
    - Collect statistics about the packet loss rate and delay.
    - Generate statistics, such as the number of packets sent and received, volume of traffic sent and received, and timestamp.
  - An In-Point-TLP collects statistics about service flows it receives. An Out-Point-TLP collects statistics about service flows it sends.
- DCP is short for Data Collecting Point.
  - DCPs are edge nodes on the network and provide the following functions:
    - Manage and control TLPs.
    - Collect the statistics generated by TLPs.
    - Report the statistics to an MCP.

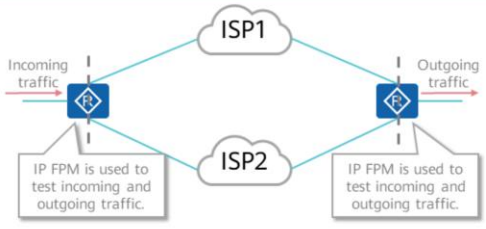
Exemplary Application Scenarios of Link Detection

Common link quality detection scenario



- NQA and BFD are usually associated with floating routes.
- When NQA or BFD detects a link fault, the high-priority route becomes invalid and the backup route is used.

E2E link quality detection



- IP FPM colors and marks data packets to help ingress and egress devices measure the packet loss or jitter of specific traffic.
- IP FPM can measure end-to-end network quality. Therefore, it can measure the actual network quality of different links in the multi-link egress scenario.

HUAWEI

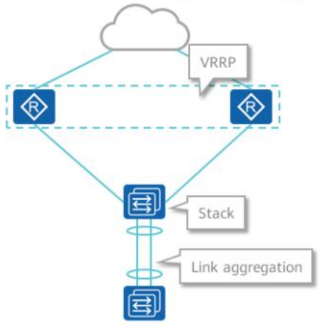
Contents

1. Basic Architecture of a Typical WAN Interconnection Solution
2. Networking Technologies and Their Applications of WAN Interconnection
3. **Reliability Technologies and Their Applications of WAN Interconnection**
   - Link Detection Technologies and Their Applications
   - Network and Service Reliability Technologies and Their Applications
4. Optimization Technologies and Their Applications of WAN Interconnection

HUAWEI

Overview of Network Reliability

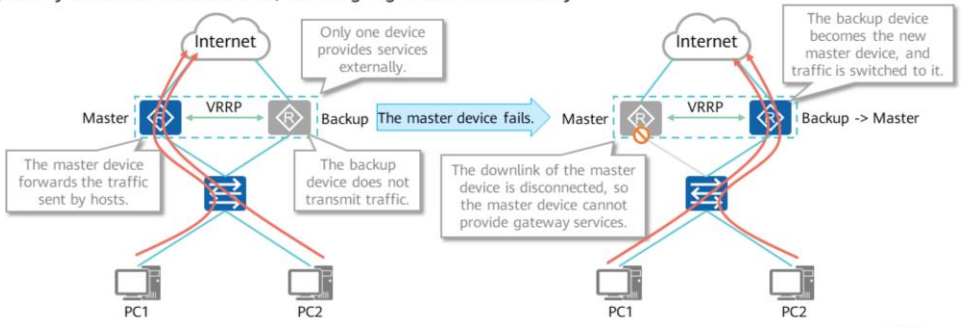
- If a fault occurs on the network, the fault may not be detected or rectified in a timely manner. To resolve this issue, redundancy technologies are required.
- Common redundancy technologies include stack, link aggregation, and VRRP.
- VRRP is the most widely used network redundancy technology on egress devices or gateways.



HUAWEI

Overview of VRRP

- Hosts are connected to external networks through gateways. If a single gateway fails, services will be interrupted for a long time. Adding egress gateways is a common method to improve system reliability. In this case, route selection among multiple egresses becomes essential.
- VRRP groups multiple routing devices into a single virtual routing device. If a gateway fails, VRRP selects a new gateway to transmit data traffic, ensuring high network reliability.



HUAWEI

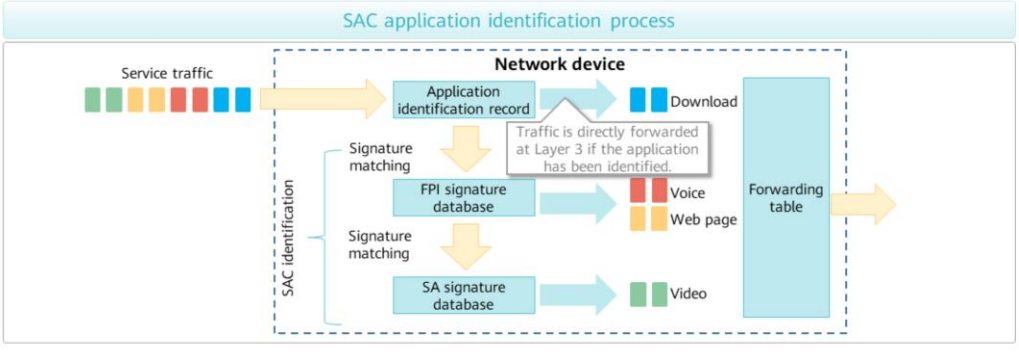
Service Reliability

- In the cloud computing era, network reliability cannot meet user requirements. Users want to understand the live network status based on applications and adjust the network based on the application status.
- Such requirements pose the following challenges to traditional networks:
  - Traditional networks cannot accurately identify applications.
  - Traditional networks cannot be adjusted based on applications.
- To cope with the challenges, two technologies are developed:
  - Smart Application Control (SAC): This technology can flexibly identify applications.
  - Smart Policy Routing (SPR): This technology can switch forwarding paths based on the network or application status.

HUAWEI

Overview of SAC

- Typically, routing and switching devices cannot identify application-layer information. Therefore, it is difficult to manage networks based on applications. SAC technology helps routing and switching devices identify classified applications.
- SAC uses service awareness (SA) and first packet identification (FPI) technologies to detect and identify Layer 4 to Layer 7 information (such as HTTP and RTP) in packets.

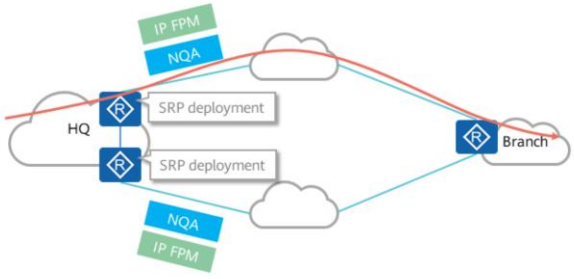


HUAWEI

- After a packet enters an SAC-enabled device, the device determines whether the corresponding application has been identified based on the 5-tuple information carried in the packet. If the application has been identified, the device forwards the packet at Layer 3 without identifying the application again. If the application has not been identified, the device performs SAC application identification. The device then processes the packet based on the SAC identification result and forwards the packet at Layer 3. The SAC application identification process is as follows: The device identifies an application based on the ACL rules defined in FPI. If the application cannot be identified, the device identifies the application based on the DNS entries defined in FPI. If the application still cannot be identified, the device identifies the application based on the protocol and port mapping table defined in FPI. If the application still cannot be identified, the device starts the SA identification process.

Overview of SPR

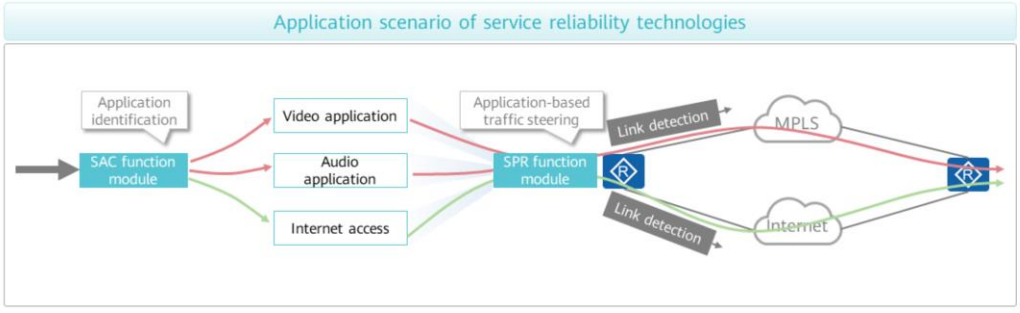
- In the cloud computing era, more users shift their attention from network connectivity to service availability, such as service response speed and service quality. However, traditional networks cannot detect link quality and service requirements, resulting in poor user experience.
- SPR addresses this problem. It actively detects the link quality and matches service requirements to select an optimal link to forward service data. SPR prevents network blackholes and flappings.



HUAWEI

Exemplary Application Scenario of Service Reliability Technologies

- Deploying both SAC and SPR can ensure the reliability of specific services on the network.
  - SAC uses the SA signature database and FPI signature database to identify applications and group traffic.
  - SPR determines the link quality and forwarding path based on the probe packets.



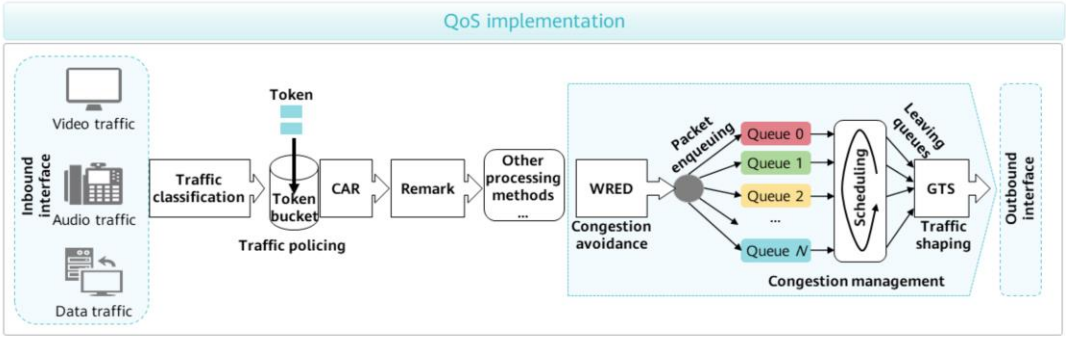
Contents

1. Basic Architecture of a Typical WAN Interconnection Solution
2. Networking Technologies and Their Applications of WAN Interconnection
3. Reliability Technologies and Their Applications of WAN Interconnection
4. **Optimization Technologies and Their Applications of WAN Interconnection**
   - QoS Technology and Its Application
   - FEC Technology and Its Application

HUAWEI

QoS Overview

- Quality of Service (QoS) defines a service provider's ability to guarantee a certain level of performance required by customers. The QoS-enabled device controls enterprise network traffic, implements congestion management and congestion avoidance, reduces the packet loss rate, and provides dedicated bandwidth for enterprise users or differentiated services (such as audio, video, and data services).



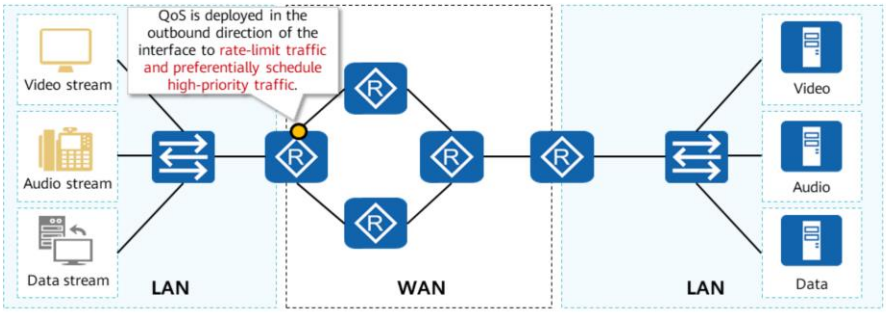
HUAWEI

- To meet SLA requirements of different services (such as audio, video, and data services), the network is required to distinguish different communication modes before providing corresponding QoS guarantee.
    - For example, real-time services such as voice over IP (VoIP) demand a shorter delay. A long delay for packet transmission is unacceptable. Email and File Transfer Protocol (FTP) services are comparatively insensitive to the delay.
    - Conventional IP networks provide the best-effect mode, which cannot help identify or differentiate the communication types on the network. The capability of differentiating communication types is the prerequisite for providing differentiated services. Therefore, the best-effect mode cannot meet the requirements of the emerging applications. This is where QoS comes in.
- QoS is designed to provide differentiated services based on application requirements. For example:
    - The bandwidth used by the FTP service on the backbone network can be limited, and a higher priority can be assigned to database access.

QoS Application Example

- In the enterprise WAN interconnection scenario, QoS is typically deployed on the egress link.
- QoS technology is used to ensure high network bandwidth and short delays for high-value services.

Application scenario of QoS

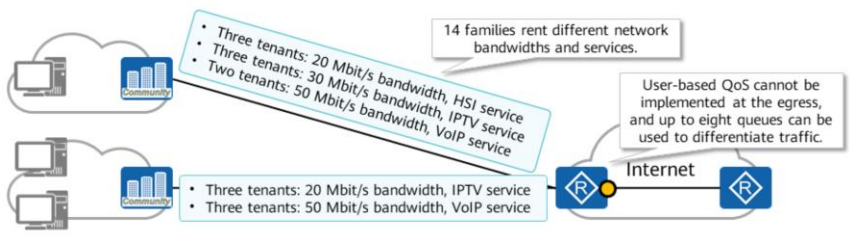


HUAWEI

Overview of HQoS

- Conventional QoS schedules traffic based on interfaces. Interfaces themselves can only differentiate service priorities. Traffic of the same priority uses the same interface queue and compete for the same queue resources. Therefore, conventional QoS is unable to provide differentiated services based on the types of traffic and users.
- HQoS implements hierarchical scheduling based on multiple levels of queues, differentiating both services and users to provide fine-tuned QoS guarantee.

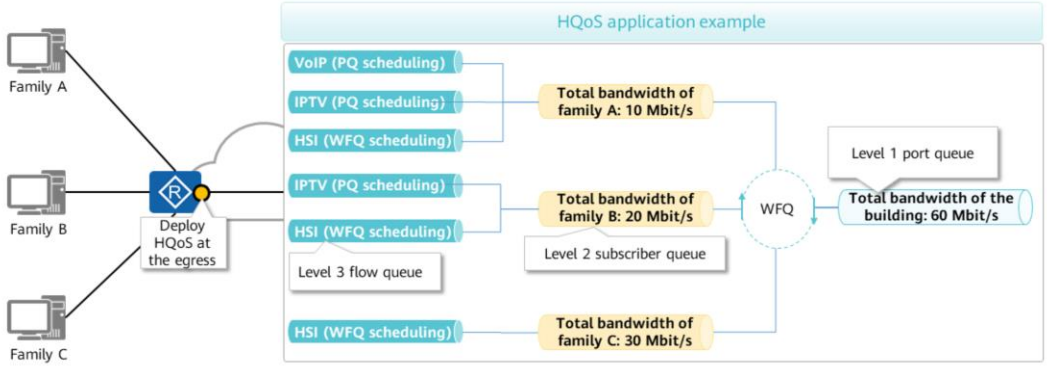
QoS limitations in multi-tenant scenarios



- In home broadband scenarios, different families may lease different network bandwidths and network services. QoS cannot be used to implement fine-tuned management of these families.

HQoS Application Example

- HQoS is mainly used in multi-tenant scenarios. For example, there are three families in a building. Family A has purchased 10 Mbit/s bandwidth and subscribed to VoIP, IPTV, and HSI services. Family B has purchased 20 Mbit/s bandwidth and subscribed to IPTV and HSI services. Family C has purchased 30 Mbit/s bandwidth and subscribed to the HSI service. These three families have different requirements. HQoS is the best choice for this scenario.



HUAWEI

- VoIP: voice over IP service
- IPTV: Internet protocol TV service
- HSI: high-speed Internet service

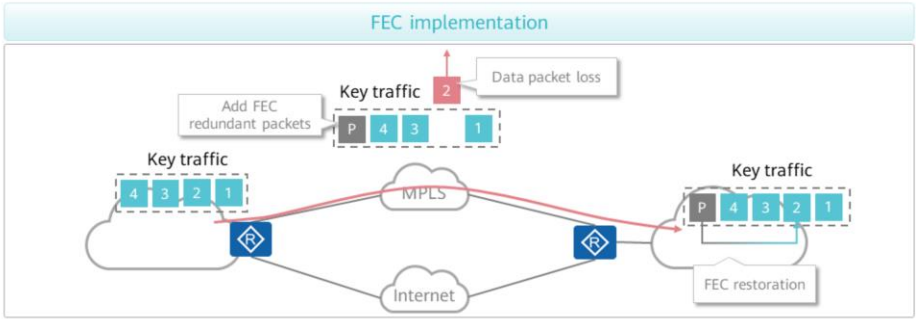
Contents

1. Basic Architecture of a Typical WAN Interconnection Solution
2. Networking Technologies and Their Applications of WAN Interconnection
3. Reliability Technologies and Their Applications of WAN Interconnection
4. **Optimization Technologies and Their Applications of WAN Interconnection**
   - QoS Technology and Its Application
   - FEC Technology and Its Application

HUAWEI

Overview of FEC

- An IP video call often encounters image distortion or audio interruption due to packet loss. To prevent this, error control technologies are required.
- Forward error correction (FEC) is such a technology. The sender attaches FEC redundant packets to the data to be transmitted. If an error is detected, the receiver can correct the data based on the redundant packets.
- FEC is applicable to networks where random packet loss occurs or the RTT is large.

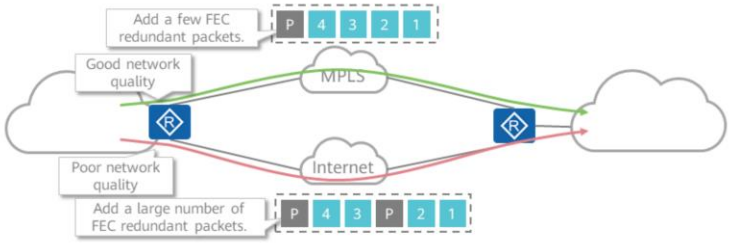


HUAWEI

- Error control technologies are classified into forward error correction (FEC) and backward error correction (BEC).
  - Automatic repeat request (ARQ) is an on-demand retransmission mechanism. The sender uses a "sending-acknowledgement" mechanism to detect whether the receiver receives data packets. If the sender does not receive any acknowledgment packet from the receiver, the sender retransmits the corresponding packet. In this error correction mode, packets need to be sent repeatedly. As a result, an extra delay is introduced. In this mode, the sender retransmits packets to implement error correction. This is why this mode is called BEC.
  - In FEC, both FEC redundant packets and data packets are sent to the receiver. If an error is found, the receiver directly restores the lost data packets by using the FEC redundant packets. In this mode, error correction is performed on the receiver. This is why this mode is called FEC.
    - The error correction capability of FEC is limited. For example, if three out of four packets are lost, FEC cannot restore these three packets based on the remaining packet.

Overview of A-FEC

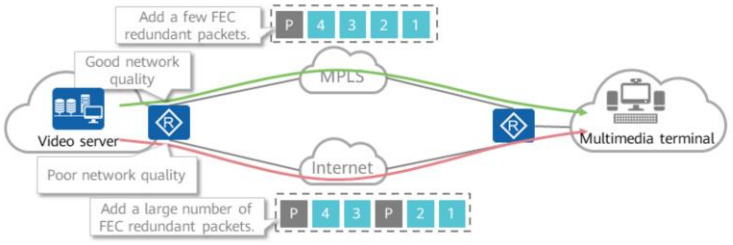
- In terms of FEC, the ratio of the original data to FEC redundant packets is fixed. If burst packet loss occurs, the original data may fail to be restored based on the redundant packets. Considering this, FEC is not applicable to the network where burst packet loss may occur. In addition, since FEC will add redundant packets, the bandwidth utilization will decrease. This is where adaptive-FEC (A-FEC) technology comes in.
- A-FEC technology can flexibly add redundant packets based on the live network quality.
  - When the network quality is good, A-FEC adds a few redundant packets or even does not add any redundant packet to improve bandwidth utilization.
  - When the network quality is poor, A-FEC adds more redundant packets to prevent data restoration failures caused by burst packet loss.



Exemplary Application Scenario of FEC/A-FEC

- FEC or A-FEC can be used to guarantee mission-critical video applications and reduce frame freezing and artifacts.

Application scenario of FEC or A-FEC



- The A-FEC algorithm is used to flexibly increase redundant packets at the transmit end based on the application and link quality to prevent continuous packet loss on the network. This helps prevent mission-critical video applications from frame freezing or artifacts even when the packet loss rate reaches up to 20%.

Quiz

1. (Multiple-answer question) Which of the following are private line technologies? ( )
   A. SDH
   B. L2TP
   C. MPLS VPN
   D. IPsec VPN
2. (True or false) SRP technology can flexibly select egress links based on the link quality. ( )
   A. True
   B. False

HUAWEI

1. AC
2. A

Summary

- There are two WAN interconnection modes: private line and VPN.
- To ensure WAN reliability, the link quality needs to be detected first, and then specific technologies are used to select a proper egress or egress link based on the link quality.
- There are many WAN optimization methods. Typically, QoS is used to control egress traffic and FEC is used to ensure reliable data forwarding on WANs.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Key Technologies of WAN Interconnection

HUAWEI

Foreword

- Wide area interconnection technologies are diverse, and different technologies have different application scenarios and backgrounds.
- On the live network, GRE and L2TP are used as WAN interconnection technologies. To ensure the security of GRE and L2TP tunnels, IPsec is used as an auxiliary technology.
- On the live network, some WAN interconnection devices are located behind NAT devices. Therefore, NAT traversal can be used in many WAN interconnection scenarios.
- To ensure WAN interconnection reliability, multiple links are generally deployed in the uplink. Multi-link routing is also an important WAN interconnection technology.
- This course describes common technologies used in WAN interconnection scenarios.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the working principles of common WAN interconnection technologies.
  - Describe the working principles of IPsec.
  - Describe the common solutions and basic principles of NAT traversal.
  - Describe how Wide Area Interconnect intelligent traffic steering works

HUAWEI

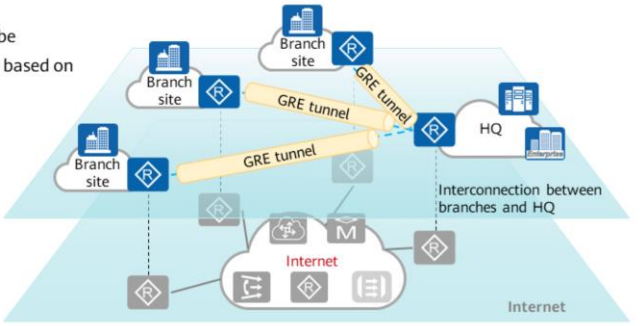
Contents

1. **Common Networking Technologies in WAN Interconnection**
   - GRE Fundamentals and Application Scenarios
   - L2TP Fundamentals and Application Scenarios
2. Security Technologies in WAN Interconnection
3. NAT Traversal Technologies in WAN Interconnection
4. Intelligent Traffic Steering Technologies in WAN Interconnection

HUAWEI

GRE Background

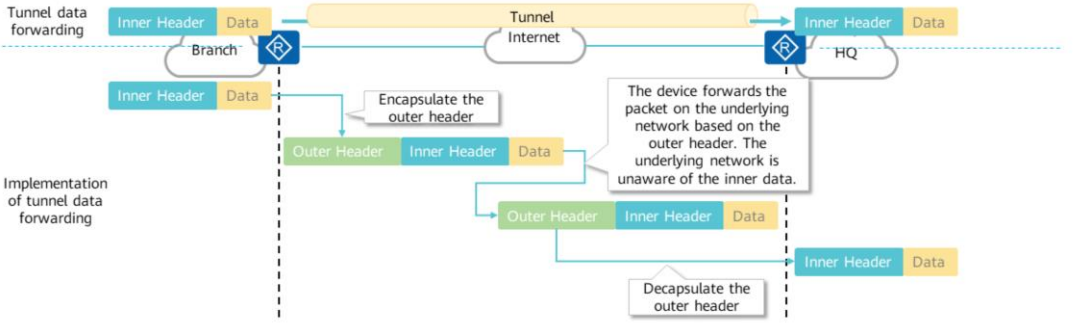
- With the development of enterprises, more and more enterprises need to communicate between branches and headquarters. Private lines (such as MPLS and SDH/MSTP private lines) need to be leased for communication between the headquarters and branches. However, private lines are expensive. For small- and medium-sized enterprises or cross-border companies, the cost is high.
- With the development of the Internet, the Internet has sufficient bandwidth and coverage. Therefore, it is more feasible to implement communication on the intranet between the headquarters and branches through the Internet. GRE is proposed in this background.
- Through GRE tunnels, the enterprise network can be established between the branch and headquarters based on the Internet.



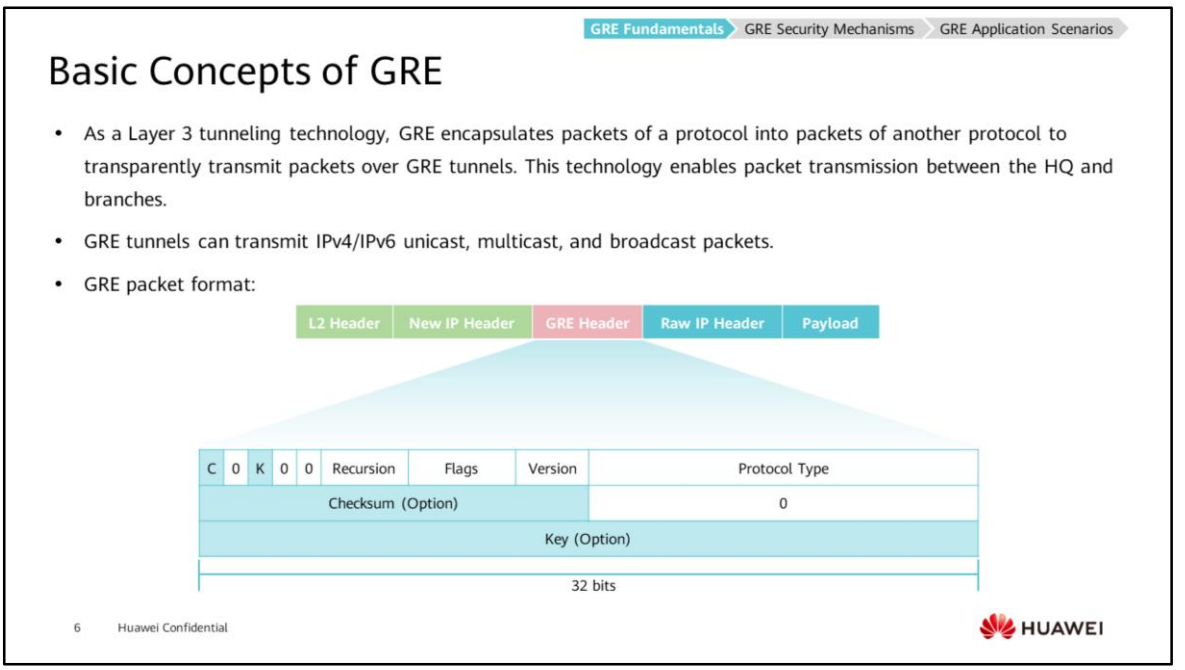
HUAWEI

Introduction to Tunneling Technologies

- GRE is one of tunneling technologies. A tunnel is similar to a bridge. Forwarding channels are established on the underlying network (for example, Internet). Users can establish a tunnel network by themselves without the intervention of the underlying network provider (for example, an ISP).
- There are many tunneling technologies, such as MPLS, GRE, Layer 2 Tunneling Protocol (L2TP), and Virtual Extensible LAN (VXLAN). The following figure shows the implementation of tunnel data forwarding.



HUAWEI

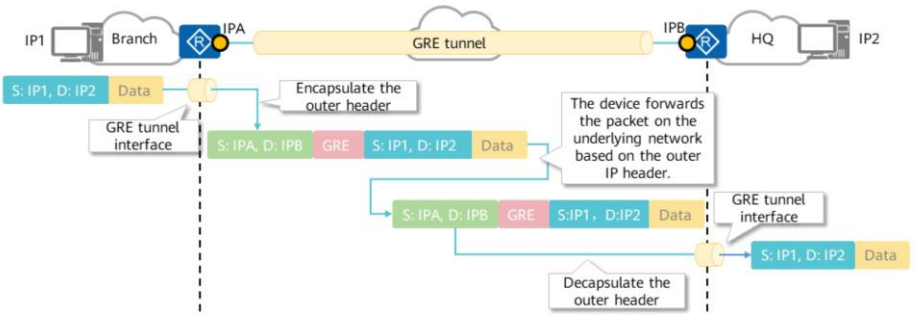


- Description of fields in a GRE header:

Field	Description
C	Checksum verification bit. The value 1 indicates that the Checksum field is inserted into the GRE header. The value 0 indicates that the GRE header does not contain the checksum field.
K	Key bit. The value 1 indicates that the Key field is inserted into the GRE header. The value 0 indicates that the GRE header does not contain the keyword field.
Recursion	Number of layers where GRE packets are encapsulated. The value of this field is increased by 1 after one GRE encapsulation is complete. If the number of encapsulation layers is greater than 3, the packet is discarded. This field is used to prevent packets from being encapsulated continuously.
Flags	Reserved field. The value must be 0.
Version	Version. The value must be 0.
Protocol Type	Type of the passenger protocol. A common passenger protocol is the IPv4 protocol, with the value of 0800. The protocol number of Ethernet over GRE is 0x6558.
Checksum	Checksum of the GRE header and the payload.
Key	Key used to authenticate the packet at the receive end.

GRE Fundamentals

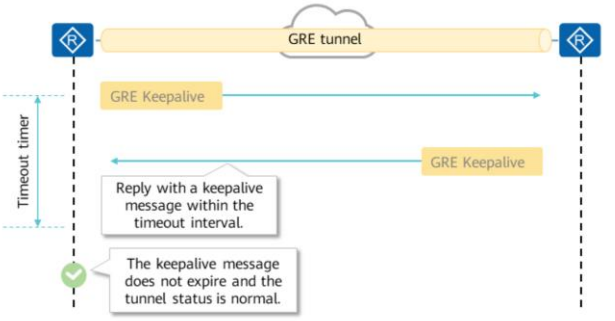
- The GRE tunnel is a Layer 3 tunnel and mainly carries IPv4/IPv6 packets. GRE encapsulates the outer IP header so that data can be transmitted on the public network. In this way, enterprise branches and the headquarters can communicate with each other.
- The following figure shows the process of forwarding packets over a GRE tunnel.



HUAWEI

Keepalive Detection

- The current GRE protocol does not have the link status detection function. If the remote interface is unreachable, the GRE tunnel cannot be terminated immediately. As a result, the source continuously forwards packets to the peer. The peer, however, cannot receive packets because the tunnel is unreachable. In this case, traffic is interrupted.
- The keepalive detection function monitors tunnel status to check whether the remote end is reachable.
- Keepalive timeout interval = Sending interval (5s by default) x Retry count (3 by default)



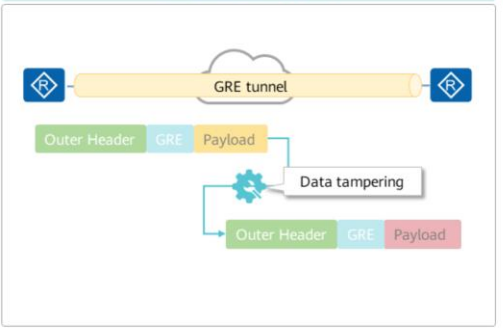
HUAWEI

- Keepalive detection functions as follows:
  - After being enabled on the source end of a GRE tunnel, the source end starts a timer to periodically send and count keepalive messages. The number of sent keepalive messages increases by one each time a keepalive message is sent.
  - The destination end sends a response message to the source end each time it receives a keepalive message from the source end.
  - If the source end receives a reply packet before the counter value reaches the preset value, it considers the remote end reachable. If the source end does not receive any response message before the counter reaches the preset value, specifically, the retry count, the source end considers the peer end unreachable and resets the counter. Then, the source end terminates the tunnel connection. In this case, the source interface still sends Keepalive messages to the remote interface. When the remote interface becomes Up, the source interface becomes Up and sets up a tunnel with the remote interface.

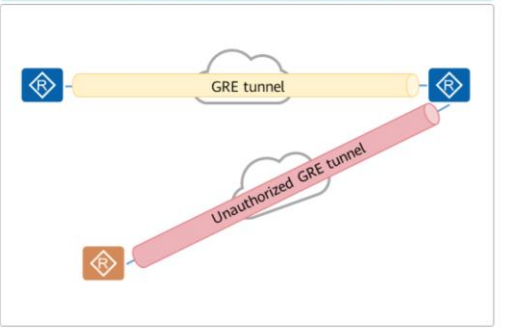
Security Threats to GRE Tunnels

- GRE tunnels are used to transmit data between branches and the HQ. Data is not encrypted and may be tampered with.
- There are potential risks in GRE tunnel establishment. Attackers can forge IP addresses to establish GRE tunnels between authorized and unauthorized devices.

GRE data tampering



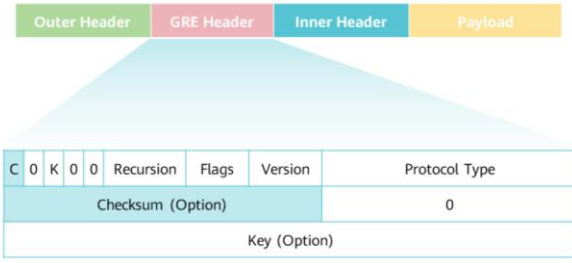
Unauthorized GRE tunnel setup



HUAWEI

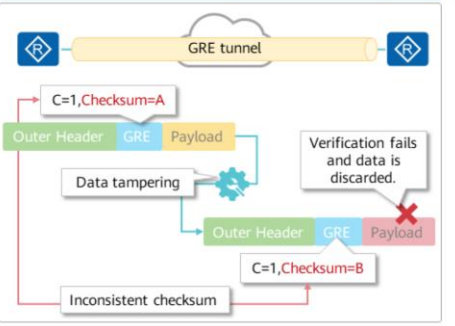
GRE Data Check and Verification

- Checksum verification is an end-to-end check on encapsulated packets.
- If the C bit in the GRE header is set to 1, the checksum is valid. The sender calculates the checksum based on the GRE header and payload. Then it sends out the packet that carries the checksum. After receiving the packet, the receiver also calculates the checksum and compares the result with the checksum carried in the packet. If they are the same, the receiver further processes the packet. Otherwise, it discards the packet.



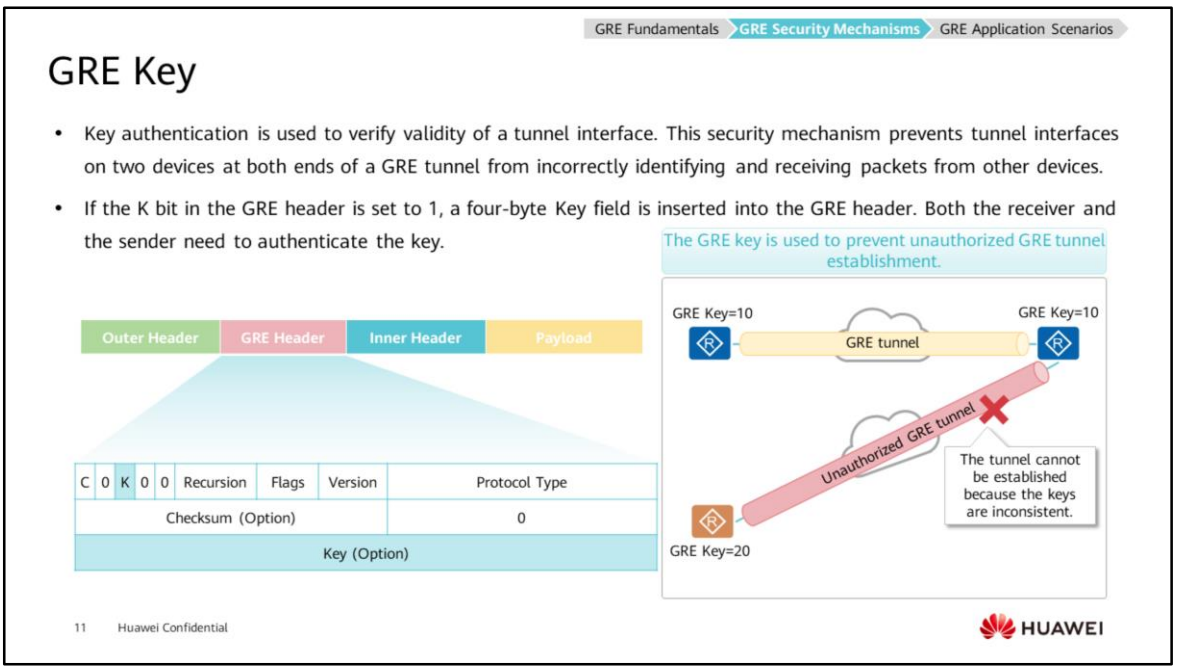
C	0	K	0	0	Recursion	Flags	Version	Protocol Type
Checksum (Option)								0
Key (Option)								

Data verification used to prevent data tampering



HUAWEI

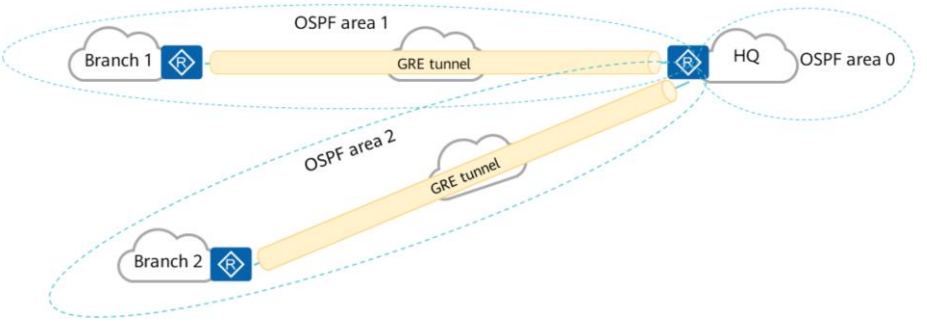
- You can enable or disable checksum verification on both ends of a tunnel in actual applications. If checksum verification is enabled on the local end and disabled on the remote end, the local end does not check checksum values of received packets, but checks checksum values of packets to be sent. If checksum verification is disabled on the local end and enabled on the remote end, the local end checks checksum values of received packets, but does not check checksum values of packets to be sent.



- This field identifies traffic in a tunnel. Packets of the same traffic use the same key. During packet decapsulation, GRE identifies data packets of the same traffic based on the key. Packets will pass verification only when the two ends of the tunnel use the same Key field. If packets fail the verification, they will be discarded. Successful authentication requires that both ends are either configured with the same Key field or not configured with the Key field.

Using GRE to Build an Intranet Between the HQ and Branches

- GRE tunnels can transmit IPv4/IPv6 unicast, multicast, and broadcast packets. Dynamic routing neighbor relationships can be configured between branches and the HQ through GRE tunnels, facilitating intranet interconnection between branches and the HQ.

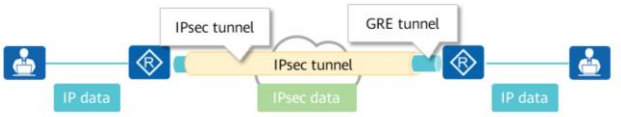


HUAWEI

GRE Over IPsec

- GRE is simple. However, data is transmitted over a GRE tunnel in cleartext and can be easily obtained. On the live network, GRE is usually used together with IPsec. The GRE technology is used to establish the internal network connection between the branch and headquarters, and the IPsec technology is used to encrypt GRE tunnel packets.

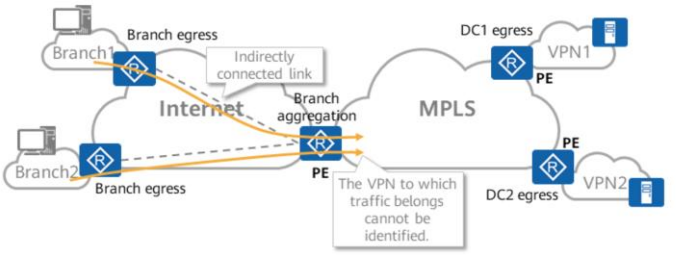
GRE over IPsec data encapsulation



HUAWEI

Limitations of MPLS VPN Networking

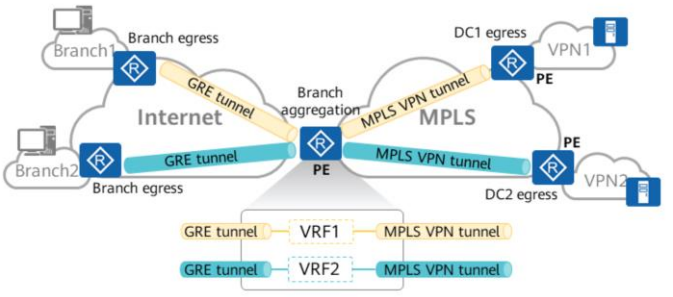
- To connect a CE to an MPLS VPN, you must use a physical link to directly connect the CE to a PE on the MPLS backbone network.
- In actual networking, however, not all CEs and PEs can be directly connected through physical links. For example, for multiple organizations that connect to the Internet or IP backbone network, their CEs may be far away from the PEs on the MPLS backbone network; therefore, they cannot be connected directly. Devices in these organizations cannot directly access the internal sites of the MPLS VPNs through the Internet or IP backbone network.



HUAWEI

Connecting to the MPLS VPN Through GRE Tunnels

- To enable a CE to access the MPLS VPN and ensure data transmission security, you can connect the CE and a PE through a public or private network, create a GRE tunnel between the CE and the PE, and associate the VPN with the GRE tunnel interface on the PE.



HUAWEI

Contents

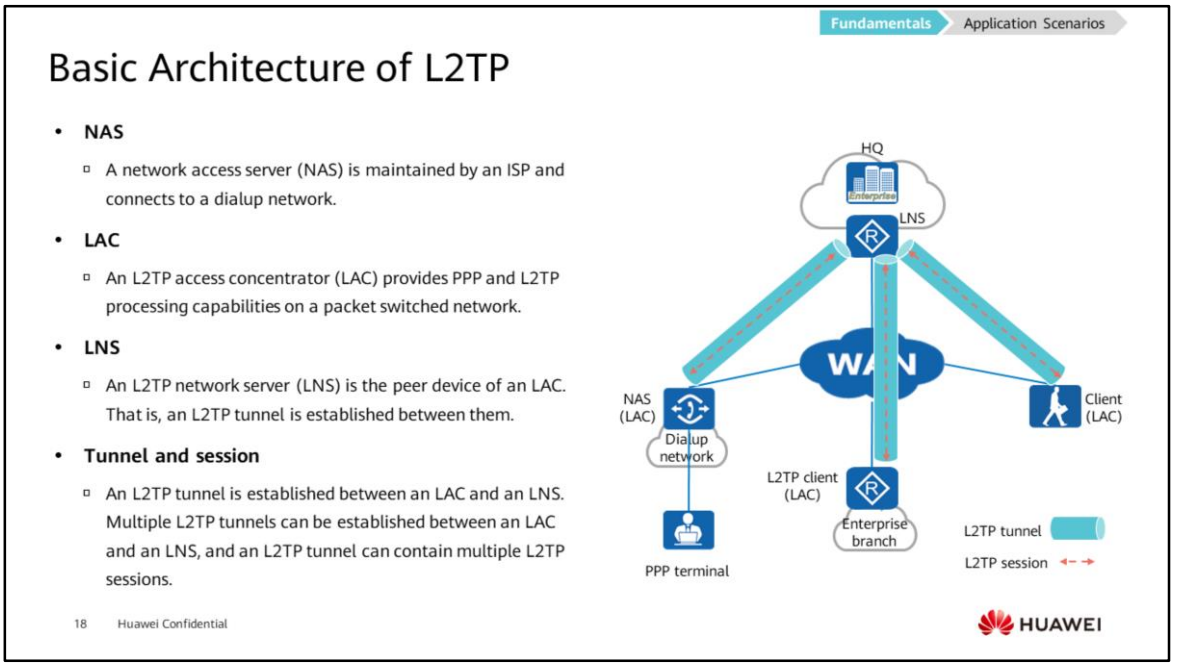
1. **Common Networking Technologies in WAN Interconnection**
   - GRE Fundamentals and Application Scenarios
   - L2TP Fundamentals and Application Scenarios
2. Security Technologies in WAN Interconnection
3. NAT Traversal Technologies in WAN Interconnection
4. Intelligent Traffic Steering Technologies in WAN Interconnection

HUAWEI

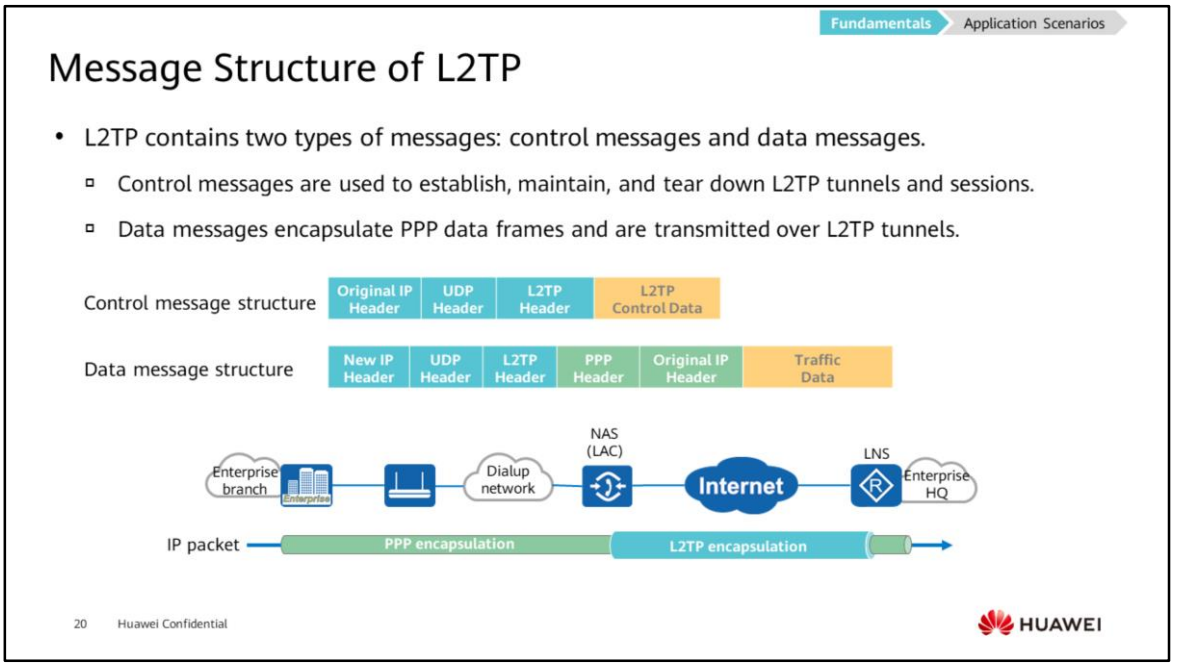
L2TP Overview

- By combining the advantages of the Layer 2 Forwarding (L2F) and Point-to-Point Tunneling Protocol (PPTP), Layer 2 Tunneling Protocol (L2TP) is an industry standard defined by the IETF. L2TP is a virtual private dialup network (VPDN) tunneling protocol that extends Point-to-Point Protocol (PPP) applications. It is an important VPN technology that provides access services for employees on the go to remotely access intranet resources.

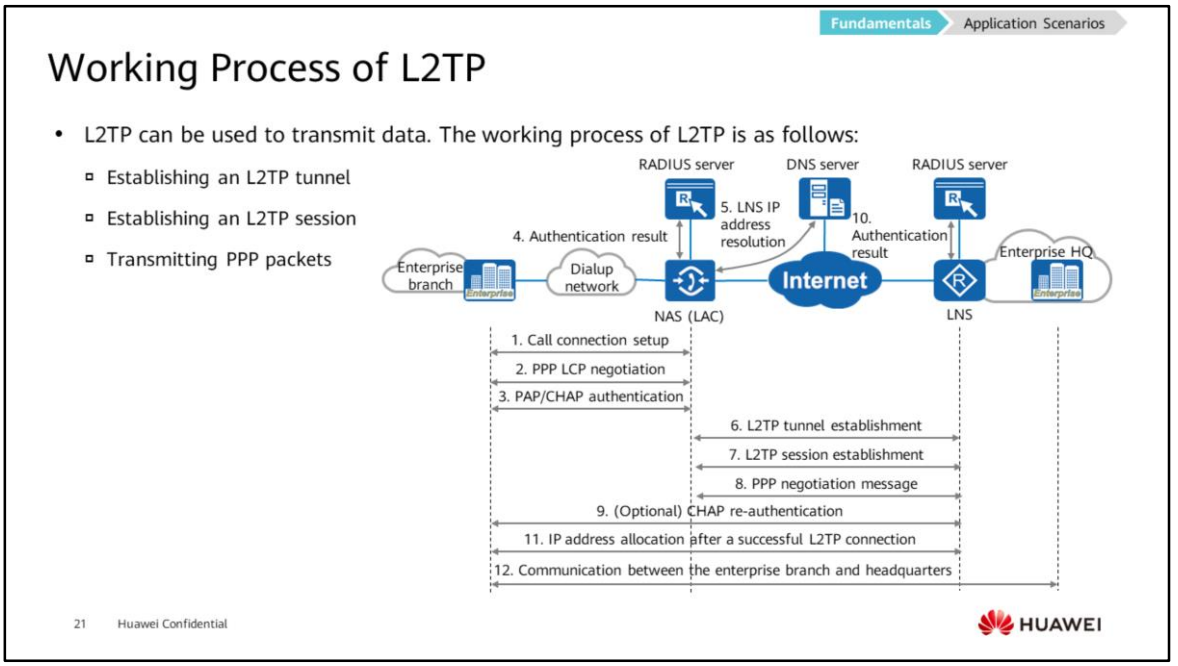
HUAWEI



- NAS
  - A network access server (NAS) is maintained by an ISP and connects to a dialup network. It is the nearest access point for PPP terminals. An NAS is used on a traditional dialup network. An ISP deploys an LAC on an NAS to provide L2TP services for remote dialup users and to establish tunnel connections with the enterprise headquarters.
- LAC
  - An L2TP access concentrator (LAC) provides PPP and L2TP processing capabilities on a packet switched network. An LAC establishes an L2TP tunnel connection with an L2TP network server (LNS) based on the user name or domain name carried in PPP packets to extend PPP negotiation to the LNS. Different networking environments can have different devices functioning as an LAC.
    - NAS-initiated scenario: On a traditional dialup network, an ISP deploys an LAC on an NAS. Alternatively, on the Ethernet of an enterprise branch, the ISP deploys a gateway for PPP terminals. The gateway functions as both a PPPoE server and an LAC.
    - L2TP client-initiated scenario: In an enterprise branch, an L2TP client functioning as an LAC is configured on the gateway to initiate an L2TP tunnel establishment request to an LNS. In this case, no dialup is required in the remote system to trigger L2TP tunnel establishment.
    - Client-initiated scenario: An employee on the go uses a PC or mobile terminal to access the Internet and uses the L2TP dialup software on the PC or mobile terminal. In this scenario, the PC or mobile terminal functions as an LAC.
  - An LAC can establish multiple L2TP tunnels to isolate data flows. That is, it can carry multiple L2TP connections.



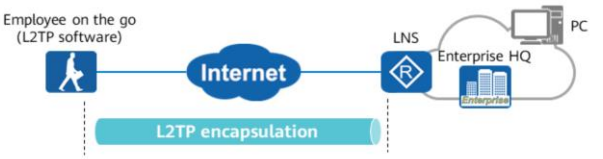
- Control message
  - Control messages are used to establish, maintain, and tear down L2TP tunnels and sessions. During the transmission of control messages, mechanisms such as retransmission of lost messages and periodic detection of tunnel connectivity are used to ensure the reliability of control message transmission. Traffic control and congestion control on control messages are supported.
  - Control messages are transmitted over an L2TP control channel. The control channel encapsulates control messages with L2TP headers and transmits them over an IP network.
- Data message
  - Data messages are used to encapsulate PPP frames, which are transmitted over tunnels, but such tunnels are unreliable. That is, a lost data message is not retransmitted, and traffic control and congestion control on data messages are not supported.
  - Data messages carrying PPP frames are transmitted over unreliable data channels. PPP frames are encapsulated using L2TP and then transmitted over the IP network.



- Establishing an L2TP tunnel
  - After receiving a PPP negotiation request from a remote user, the LAC sends an L2TP tunnel establishment request to the LNS. The LAC and the LNS exchange L2TP control messages to negotiate the tunnel ID and tunnel authentication information. After the negotiation succeeds, an L2TP tunnel is established between them and identified by the tunnel ID.
- Establishing an L2TP session
  - If an L2TP tunnel exists, the LAC and the LNS exchange control messages to negotiate the session ID. If no L2TP tunnel exists, the LAC and the LNS establish an L2TP tunnel first. The L2TP session carries LCP negotiation information and user authentication information of the LAC. After authenticating such information, the LNS notifies the LAC of the session establishment. The L2TP session is identified by a session ID.
- Transmitting PPP packets
  - After the L2TP session is established, the PPP terminal sends data packets to the LAC. The LAC encapsulates the L2TP packets based on information such as the L2TP tunnel and session ID and sends the packets to the LNS. The LNS decapsulates the L2TP packets and sends the packets to the destination host based on the routing and forwarding table.

Mobile Office

- Employees on the go connect to the intranet through L2TP. The LNS can authenticate access users and assign private IP addresses to them. If ACLs are configured, the LNS can also manage access rights of access users.



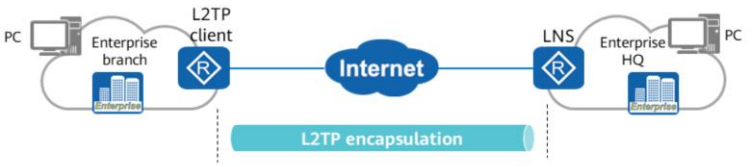
HUAWEI

- Employees on the go may need to communicate with the headquarters and access intranet resources of the headquarters at any time. Although they can access the headquarters gateway through the Internet, the headquarters gateway cannot identify and manage access users. To address this issue, configure the headquarters gateway as an LNS, so that virtual point-to-point connections can be established between the employees on the go and the headquarters gateway when the employees use the L2TP dialup software on the PC to initiate L2TP connections.

Fundamentals > Application Scenarios

Interconnection Between Branches and the Headquarters

- An enterprise sets up a VPN to connect its branch network to its headquarters network. After IP packets from branch users reach the L2TP client, the L2TP client forwards the packets to the virtual dialup interface. The virtual dialup interface forwards the packets to the LNS, which then forwards the packets to the destination host.



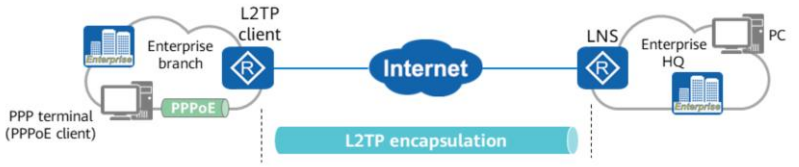
HUAWEI

- An enterprise has some branches located in other cities, and its branches use the Ethernet and have gateways deployed for branch users to access the Internet. The headquarters provides access services for branches. VPDN connections need to be established between branches and the headquarters gateway. Any branch user is allowed to access the headquarters network, and only the branch gateways need to be authenticated. In this case, the headquarters gateway functions as the LNS, and the branch gateways function as the L2TP clients. Virtual dialup is created on the branch gateways to trigger L2TP tunnel connections to the headquarters network. A virtual point-to-point connection is established between an L2TP client and the LNS. After IP packets of branch users reach an L2TP client, the L2TP client forwards the packets to the virtual dialup interface. The virtual dialup interface forwards the packets to the LNS, which then forwards the packets to the destination host.

Fundamentals > Application Scenarios

Interconnection Between Employees in Branches and the Headquarters

- Branch employees use PPPoE dialup software to connect to the headquarters network, and the branch gateway functions as the PPPoE server and LAC to forward call requests from branch users to the headquarters network.



HUAWEI

- An enterprise has some branches located in other cities, and its branches use the Ethernet and have gateways deployed for branch users to access the Internet. Headquarters users need to communicate with branch users, and the headquarters uniformly manages access of branch users. Therefore, L2TP is used to deploy the headquarters gateway as an LNS. Dialup packets of branch users cannot be transmitted directly over the Ethernet. Therefore, PPPoE dialup software needs to be deployed as a PPPoE client on the terminal that initiates the dialup packets, and the branch gateway functions as a PPPoE server and an LAC to forward call requests of branch users to the headquarters.

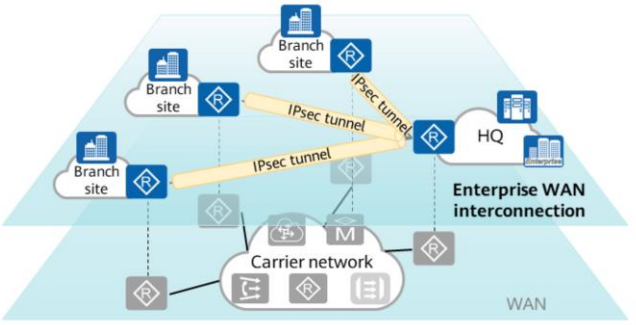
Contents

1. Common Networking Technologies in WAN Interconnection
2. **Security Technologies in WAN Interconnection**
   - Basic Concepts of IPsec
   - IPsec Fundamentals
   - IPsec Application Scenarios
3. NAT Traversal Technologies in WAN Interconnection
4. Intelligent Traffic Steering Technologies in WAN Interconnection

HUAWEI

IPsec Background

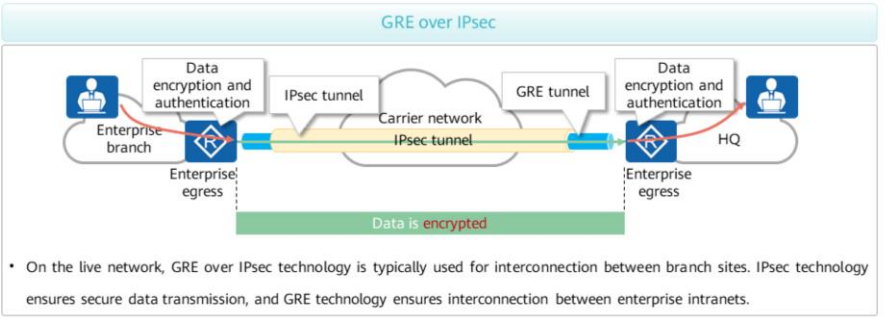
- Enterprise branches often need to communicate with each other. They can communicate using many methods, for example, using private lines or Internet links.
- Considering costs and requirements, some enterprises choose to use Internet links for interconnection. However, data may be intercepted when being transmitted on the Internet, posing security risks.
- IPsec technology encrypts data packets to secure enterprise interconnections.



HUAWEI

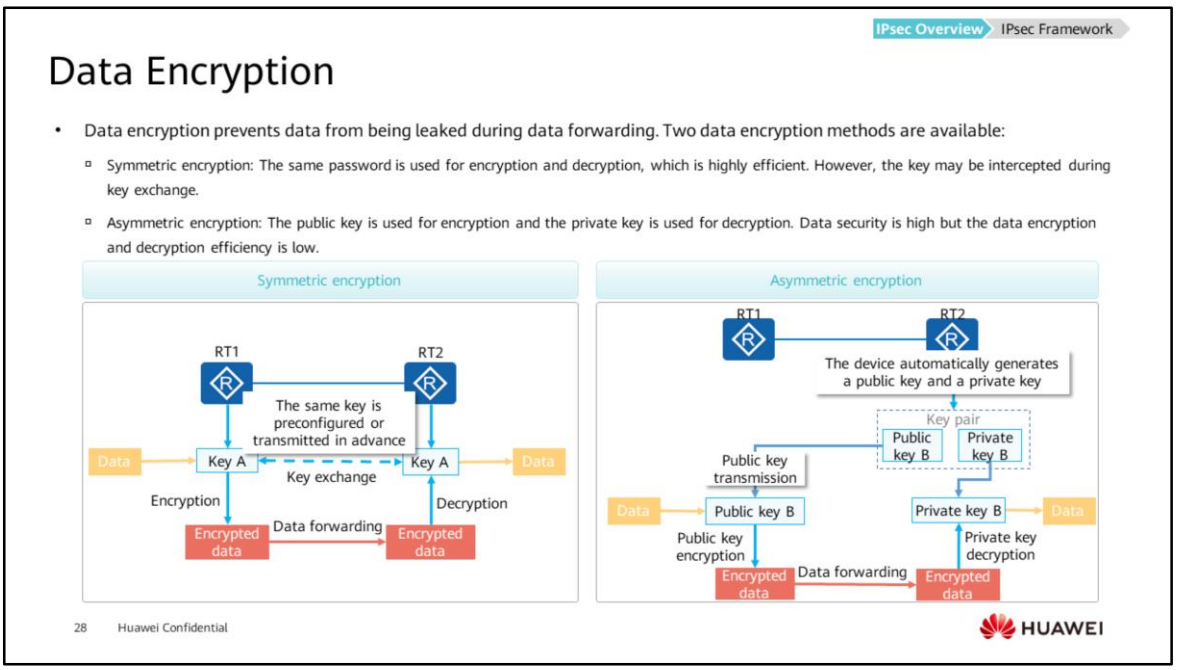
IPsec Overview

- The IPsec protocol suite is a series of security protocols developed by the Internet Engineering Task Force (IETF). It provides a cryptology-based, interoperable, and high-quality security protection mechanism for end-to-end IP packet exchange.
- IPsec encrypts and authenticates data to ensure secure data transmission on the Internet.
- IPsec VPN technology can be used with multiple VPN technologies to provide flexible and secure enterprise interconnections.



- On the live network, GRE over IPsec technology is typically used for interconnection between branch sites. IPsec technology ensures secure data transmission, and GRE technology ensures interconnection between enterprise intranets.

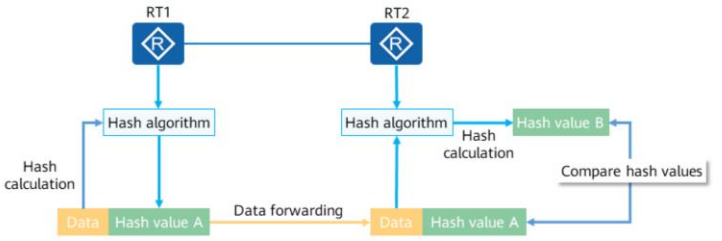
HUAWEI



- The symmetric encryption algorithm is also called traditional cryptographic algorithm, in which the encryption key can be calculated from the decryption key. The sender and receiver share the same key, which is used for both encryption and decryption. Symmetric key encryption is an effective method for encrypting a large amount of data. There are many algorithms for symmetric key encryption, and all of them aim to convert between cleartext (unencrypted data) and ciphertext. Because symmetric key encryption uses the same key for data encryption and decryption, data security depends on whether unauthorized users obtain the symmetric key. If two communicating parties want to use the symmetric key to encrypt data, they must exchange the key securely before exchanging the encrypted data.
- An asymmetric algorithm is also called public key algorithm, in which a public key is used for encryption and a private key for decryption. The two keys are mathematically related. In public key encryption, the public key can be publicly transmitted between two communicating parties or released in the public repository, but the private key is confidential. The data encrypted using the public key can be decrypted only using the private key. The data encrypted using the private key can be decrypted only using the public key.

Data Authentication

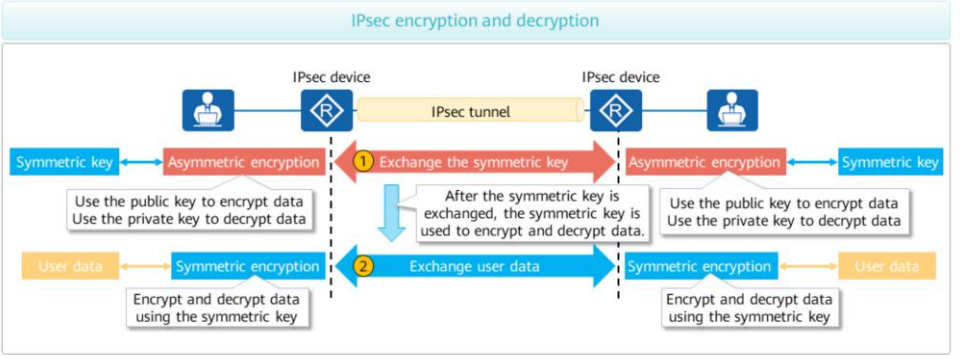
- The main purpose of data authentication is to check whether data is tampered with. Data authentication is mainly based on the hash algorithm.
  - A unique hash value is calculated based on the hash algorithm and then carried in the data before being forwarded to the peer device.
  - The peer device hashes the data again to obtain the hash value. It then compares the received hash value with the calculated one. If they are the same, the data is not tampered with.



HUAWEI

IPsec Encryption

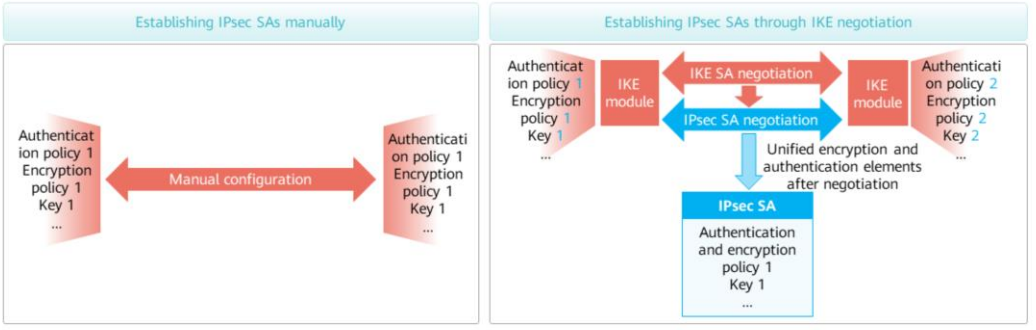
- IPsec uses both symmetric encryption and asymmetric encryption, ensuring data security and performance.
  - Uses an asymmetric algorithm to encrypt and transmit the key used for symmetric encryption.
  - Uses the exchanged symmetric key to encrypt data.



HUAWEI

SA

- A security association (SA) is an agreement between two IPsec peers on certain elements. For example, Data Encryption Standard (DES) is used as the encryption algorithm, Message Digest Algorithm 5 (MD5) is used as the authentication algorithm, and tunnel is used as the encapsulation mode.
- An IPsec SA can be established manually or through Internet Key Exchange (IKE) negotiation.

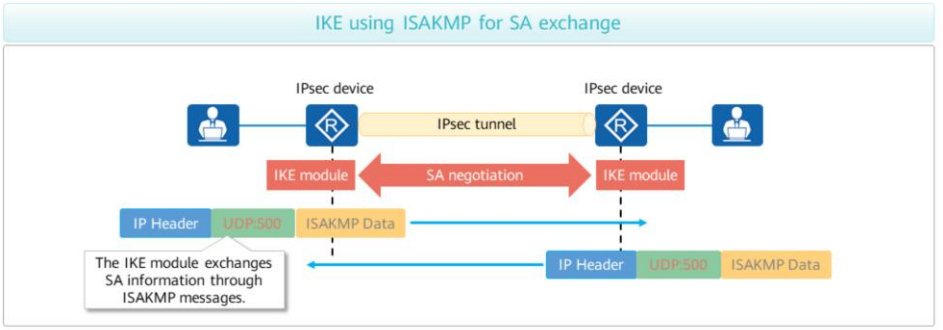


HUAWEI

- IPsec technology supports multiple data encryption, authentication, and encapsulation algorithms. When devices at both ends use IPsec for secure communication, they must use the same encryption and authentication algorithms. Therefore, a mechanism is required to help the devices negotiate these parameters.
- An IPsec SA can be established in either of the following ways:
  - Manual configuration: The management cost of manually established IPsec SAs is high. This is because the encryption and authentication modes need to be manually configured, SAs need to be manually updated, and SA information permanently exists, resulting in low security. This mode applies to small-scale networks.
  - IKE negotiation: The management cost of IPsec SAs established through IKE negotiation is low. The encryption and authentication modes are generated using the Diffie-Hellman (DH) algorithm, SA information is generated periodically, and SAs are dynamically updated. This mode applies to small-, medium-, and large-sized networks.
- An SA is uniquely identified by three parameters: security parameter index (SPI), destination IP address, and security protocol ID (AH or ESP).
- An IKE SA is used to establish a secure channel for exchanging IPsec SAs.

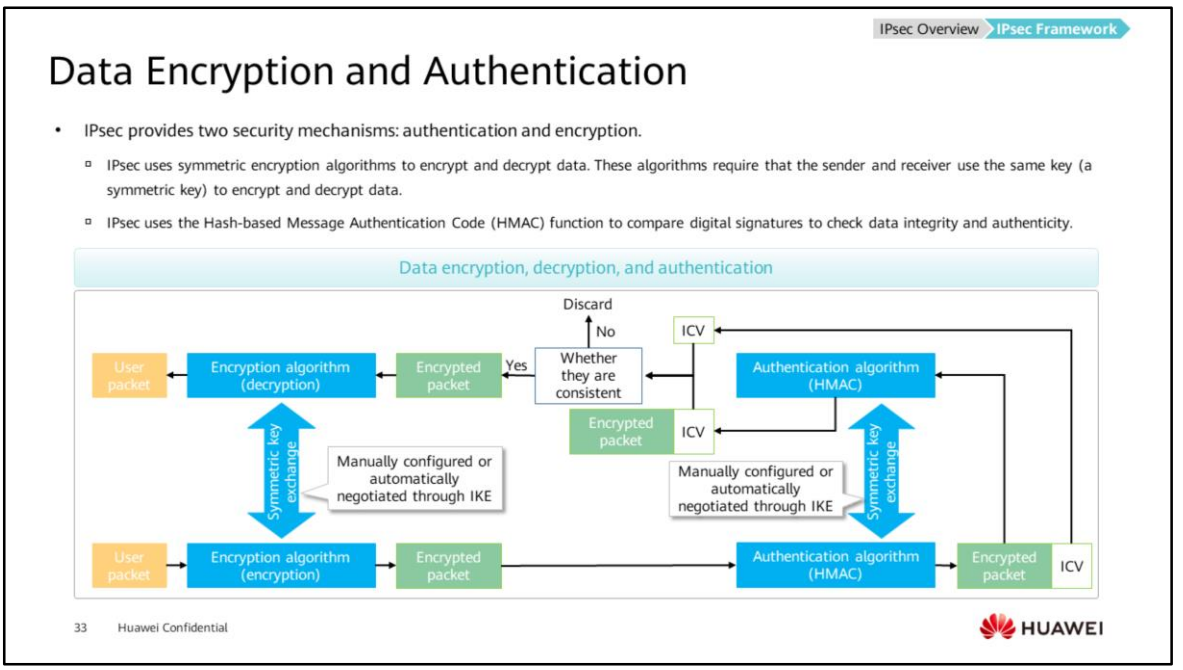
Key Exchange

- On the live network, the Internet Key Exchange (IKE) protocol is typically used to exchange symmetric keys.
- IKE is a UDP-based application-layer protocol. It is built upon the framework defined by the Internet Security Association and Key Management Protocol (ISAKMP). IPsec uses IKE for key auto-negotiation and IPsec SA establishment, simplifying IPsec configuration and maintenance.



HUAWEI

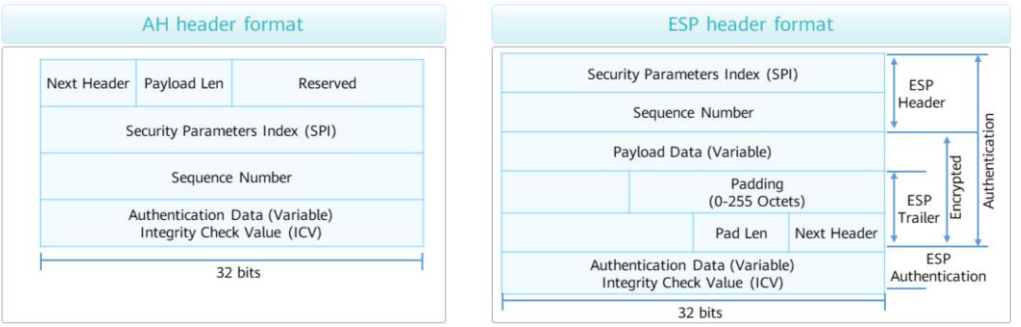
- IKE supports the following authentication algorithms including MD5, Secure Hash Algorithm 1 (SHA1), SHA2-256, SHA2-384, SHA2-512, and Senior Middle 3 (SM3).
- IKE supports the following encryption algorithms: DES, 3DES, AES-128, AES-192, AES-256, SM1, and SM4.
- ISAKMP is defined in RFC 2408, which defines the procedures for negotiating, establishing, modifying, and deleting SAs and defines the ISAKMP message format. ISAKMP provides a general framework for SA attributes and the methods of negotiating, modifying, and deleting SAs, without defining the specific SA format.
- ISAKMP messages can be transmitted using UDP or TCP through port 500. In most cases, ISAKMP messages are transmitted using UDP.



- Integrity check value (ICV) is used by the receiver for integrity check. Available authentication algorithms are MD5, SHA1, SHA2, and SM3.
- Common symmetric encryption algorithms used by IPsec include Data Encryption Standard (DES), Triple Data Encryption Standard (3DES), Advanced Encryption Standard (AES), and algorithms approved by State Cryptography Administration, such as SM1 and SM4. DES and 3DES are not recommended because they are insecure and pose security risks.
- Common authentication algorithms used by IPsec include MD5, SHA1, SHA2, and SM3. MD5 and SHA1 are not recommended because they are insecure and pose security risks.
- IPsec encryption cannot verify the authenticity or integrity of information after decryption. IPsec uses the HMAC function to compare digital signatures to check integrity and authenticity of data packets. In most cases, encryption and authentication are used together. The IPsec sender uses the authentication algorithm and symmetric key to generate a digital signature for the encrypted packet and sends the IP packet and digital signature to the receiver. The receiver uses the same authentication algorithm and symmetric key to process the encrypted packet and then generates a digital signature. Then the receiver compares the received and generated digital signatures to verify the data integrity and authenticity. If the packet passes the verification, the receiver decrypts it. Otherwise, the receiver discards it.

Security Protocols

- IPsec provides two transport layer protocols for authentication or encryption: Authentication Header (AH) and Encapsulating Security Payload (ESP).
  - AH provides only authentication but no encryption capabilities.
  - ESP provides both authentication and encryption.

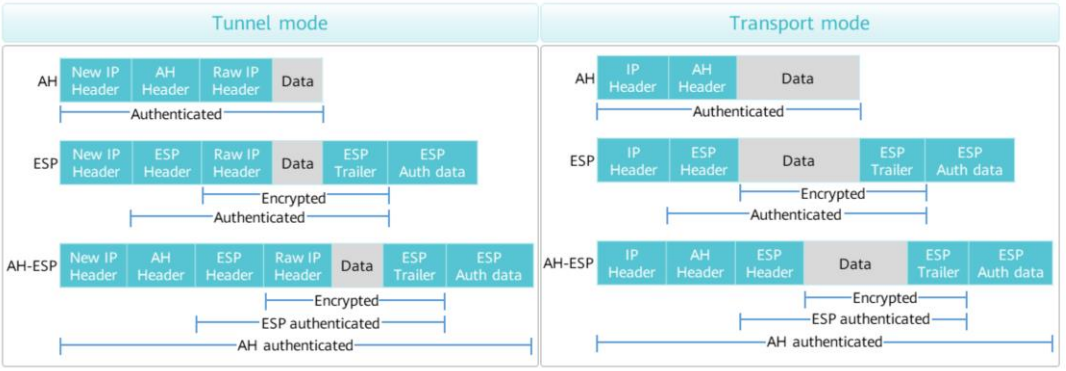


HUAWEI

- AH provides only authentication but no encryption capabilities. According to the AH protocol, an AH header is appended to the standard IP header in each packet. The sender performs hash calculation on packets and an authentication key. After packets carrying the calculation result arrive at the receiver, the receiver also performs hash calculation and compares the calculation result with the received calculation result. Any changes to the data during transmission will make the calculation result invalid. This implements data origin authentication and integrity verification. AH provides data integrity check on an entire IP packet.
- ESP provides both authentication and encryption. An ESP header is appended to the standard IP header in each data packet, and the ESP Trailer and ESP Auth data fields are appended to each data packet. In contrast to AH, ESP encrypts the payload before encapsulating it into a data packet to ensure data confidentiality, and protects the IP header only in tunnel mode.
- Key fields:
  - Sequence Number: This field is a counter that monotonically increases from 1. It uniquely identifies a packet to prevent replay attacks.
  - SPI: This field uniquely identifies an IPsec SA.
  - Authentication Data: This field contains the Integrity Check Value (ICV) and is used by a receiver for data integrity check. Available authentication algorithms are MD5, SHA1, SHA2, and SM3.

Encapsulation Modes

- IPsec encapsulation is a process of adding AH or ESP fields to original IP packets for packet authentication and encryption. This process is implemented in transport or tunnel mode.
- On the live network, the tunnel mode is often used for encapsulation.



HUAWEI

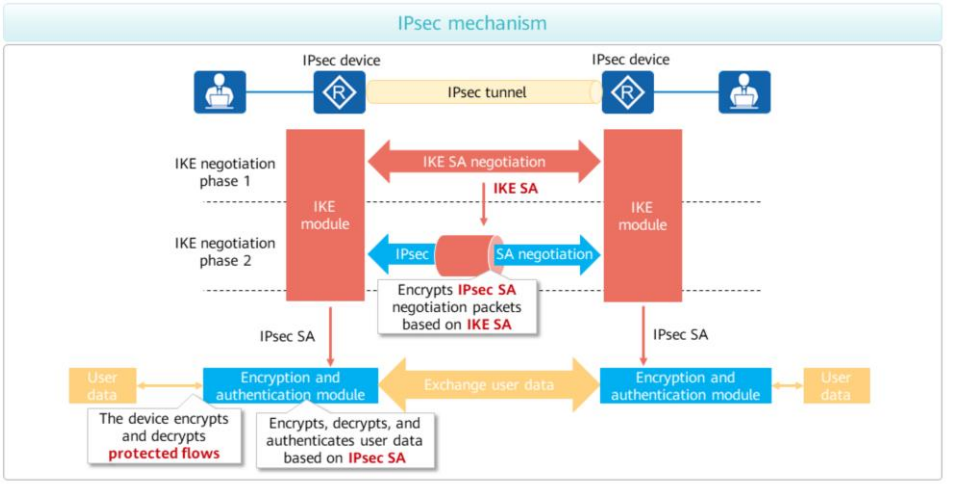
- In transport mode, an AH or ESP header is added between an IP header and a transport-layer protocol (TCP, UDP, or ICMP) header to protect the TCP, UDP, or ICMP payload. As no additional IP header is added, IP addresses in the original packets are visible in the IP header of the post-encrypted packet.
- In tunnel mode, an AH or ESP header is added before the raw IP header and then encapsulated into a new IP packet with a new IP header to protect the IP header and payload.

Contents

1. Common Networking Technologies in WAN Interconnection
2. **Security Technologies in WAN Interconnection**
   - Basic Concepts of IPsec
   - IPsec Fundamentals
   - IPsec Application Scenarios
3. NAT Traversal Technologies in WAN Interconnection
4. Intelligent Traffic Steering Technologies in WAN Interconnection

HUAWEI

IPsec Mechanism



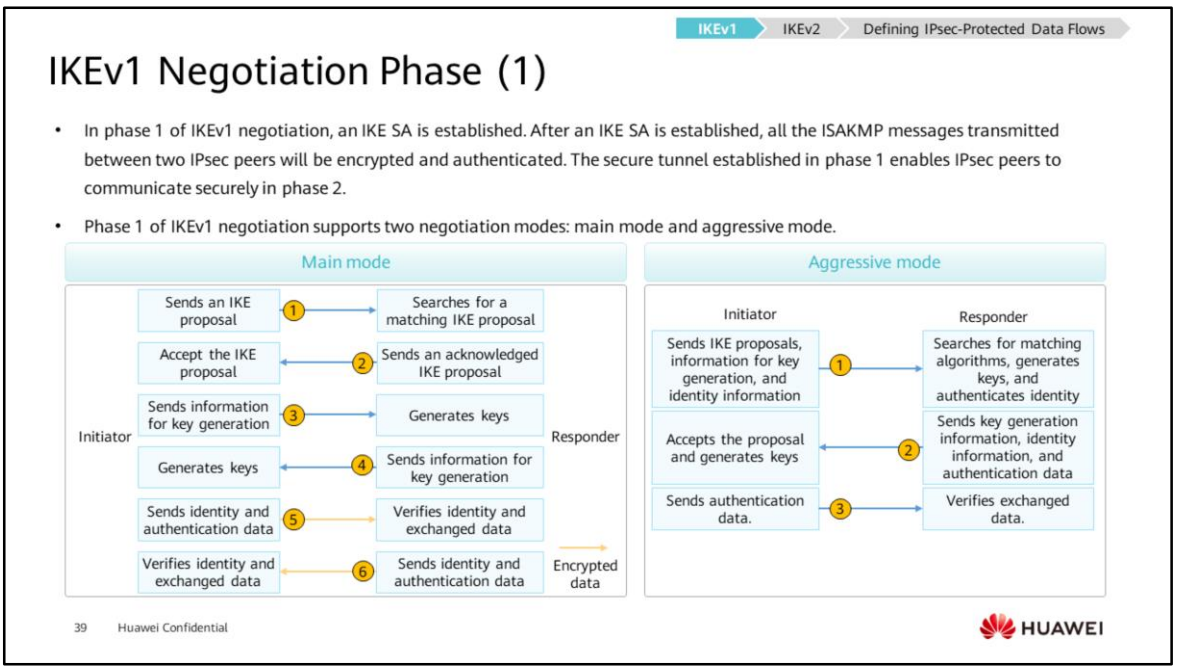
HUAWEI

- The IPsec mechanism is as follows:
  - An IKE SA is negotiated in the first phase of IKE negotiation.
  - The IKE SA is used to encrypt the packets in the second phase of IKE negotiation. That is, IPsec SAs are negotiated in the second phase of IKE negotiation.
  - IPsec SAs are used to encrypt data.

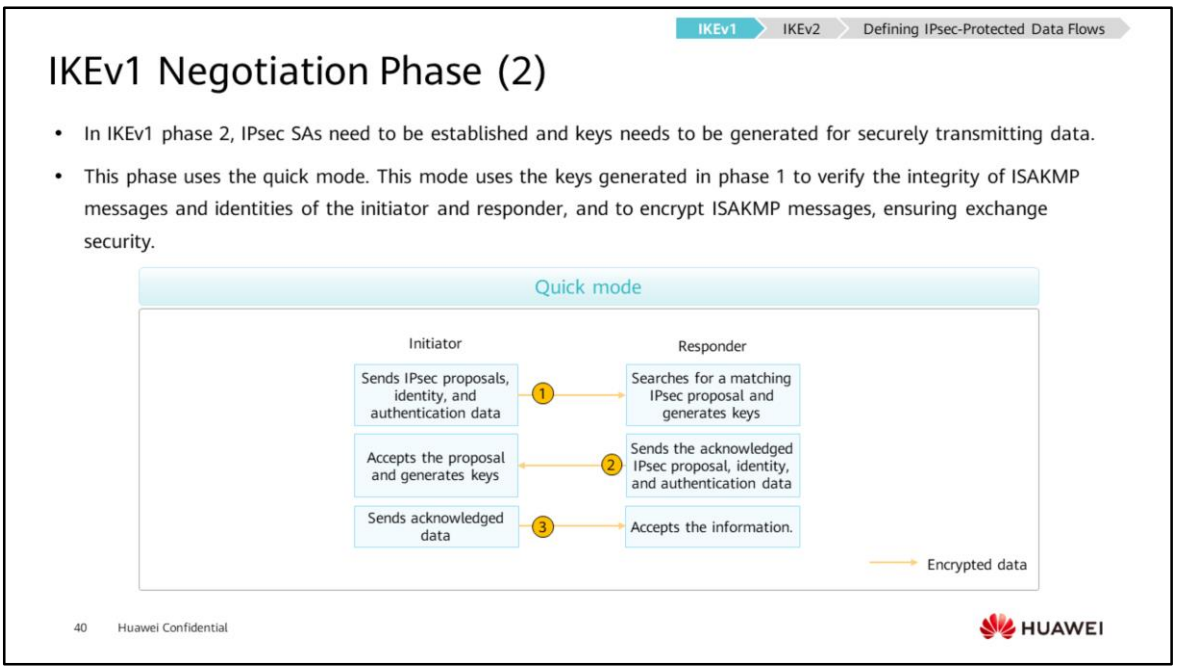
IKEv1

- IKEv1 negotiation goes through two phases: In phase 1, two IPsec peers negotiate and establish a secure tunnel (an IKE SA). In phase 2, the two IPsec peers establish a pair of IPsec SAs for secure data transmission through the secure tunnel established in phase 1.

HUAWEI



- The main mode requires three exchanges between the peers, totaling six ISAKMP messages. The three exchanges are described as follows:
  - Messages 1 and 2 are used for IKE proposal exchange.
    - The initiator sends one or more IKE proposals to the responder. The responder searches for the first matching IKE proposal and then sends it to the initiator. IKE proposals of the initiator and responder match if they have the same encryption algorithm, authentication algorithm, authentication method, and DH group identifier.
  - Messages 3 and 4 are used for key information exchange.
    - The initiator and responder exchange the DH public value and nonce value to generate the IKE SA authentication key and encryption key.
  - Messages 5 and 6 are used for identity and authentication information exchange. (Both parties use the generated keys to exchange information.)
    - The initiator and responder use the generated keys to authenticate each other and the information exchanged in main mode.
- The aggressive mode uses only three messages. Messages 1 and 2 are used to negotiate IKE proposals and exchange the DH public value, mandatory auxiliary information, and identity information. Message 2 also contains the identity information sent by the responder to the initiator for authentication. Message 3 is used by the responder to authenticate the initiator.
- Compared with the main mode, the aggressive mode reduces the number of exchanged messages and speeds up the negotiation. However, the aggressive mode does not encrypt identity information.



- In IKEv1 phase 2, two IPsec SAs are established through three ISAKMP messages:
  - Message 1 is used by the initiator to send local security parameters and identity authentication information to the responder.
    - Security parameters include protected data flows and parameters to be negotiated, such as an IPsec proposal. Identity authentication information includes the keys generated in phase 1 and keying materials generated in phase 2, and can be used to authenticate the peer again.
  - Message 2 is used by the responder to send acknowledged security parameters and identity authentication information, and to generate new keys.
    - The encryption key and authentication key used for secure data transmission over IPsec SAs are generated based on the keys generated in phase 1 and parameters such as the SPI and protocol. This ensures that each IPsec SA has unique encryption and authentication keys.
  - Message 3 is used by the initiator to send acknowledged information to communicate with the responder. IKEv1 negotiation then ends and IPsec SAs are established.

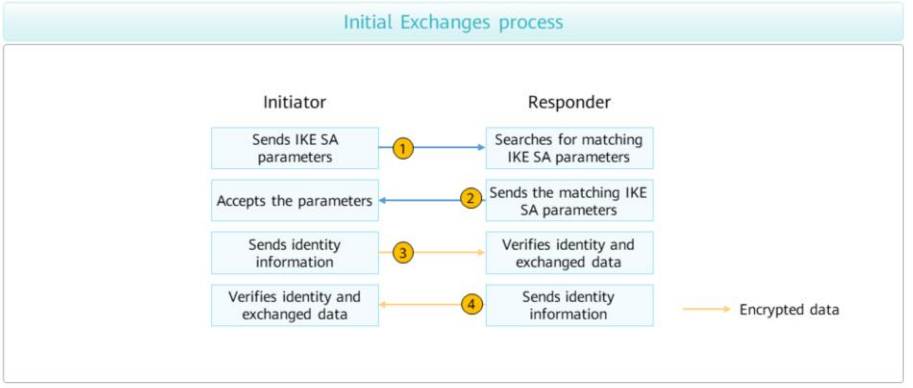
IKEv2

- The process of establishing SAs through IKEv2 negotiation is much simpler than that through IKEv1 negotiation. In normal cases, IKEv2 can establish a pair of IPsec SAs through only four messages in two exchanges. One additional Create_Child_SA Exchange can be used to establish another pair of IPsec SAs if required, during which only two messages are exchanged.
- IKEv2 defines three exchanges: Initial Exchanges, Create_Child_SA Exchange, and Informational Exchange.

HUAWEI

IKEv2 Initial Exchanges

- IKEv2 establishes the first pair of IPsec SAs through Initial Exchanges. Initial Exchanges involves four messages in two exchanges.

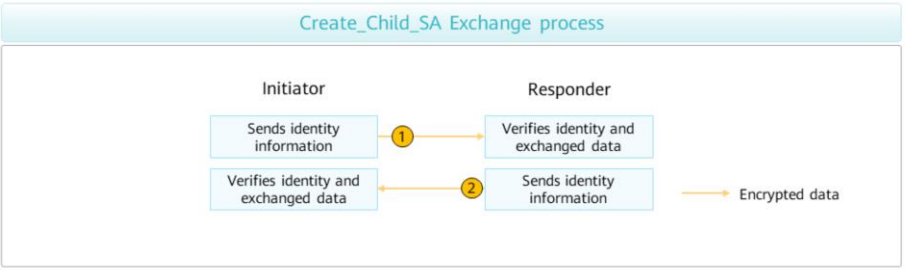


HUAWEI

- Messages 1 and 2 are used in exchange 1 (called IKE_SA_INIT). In exchange 1, IKE SA parameters are negotiated in plain text, including the encryption key, authentication key, random number, and DH key. After IKE_SA_INIT is complete, shared keying material is generated, from which all keys used by IPsec SAs are derived.
- Messages 3 and 4 are used in exchange 2 (called IKE_AUTH). In exchange 2, identities of the two parties and the first two messages are authenticated, and IPsec SA parameters are negotiated. IKEv2 supports Rivest-Shamir-Adleman (RSA) signature authentication, pre-shared key (PSK) authentication, and Extensible Authentication Protocol (EAP) authentication. The initiator omits the AUTH payload in message 3 to indicate that EAP authentication is required.

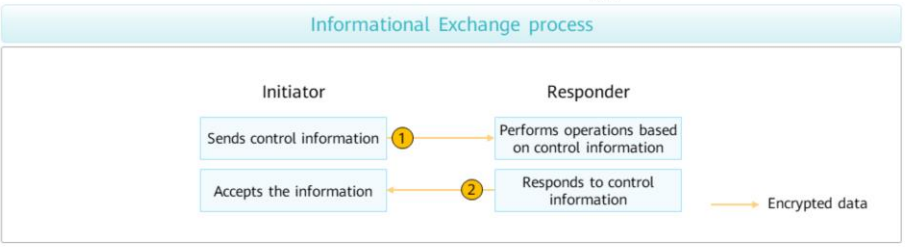
IKEv2 Create_Child_SA Exchange

- After one pair of IPsec SAs is established based on an IKE SA, Create_Child_SA Exchange can be performed to negotiate more pairs of IPsec SAs. In addition, Create_Child_SA Exchange can be performed for IKE SA re-negotiation.
- Create_Child_SA Exchange involves two messages in one exchange and corresponds to IKEv1 phase 2. The initiator in Create_Child_SA Exchange can be the initiator or responder in Initial Exchanges.



IKEv2 Informational Exchange

- IKEv2 peers perform Informational Exchange to exchange control information, including error information and notifications.
- Informational Exchange must be performed under the protection of an IKE SA. Specifically, Informational Exchange is performed after Initial Exchanges are complete. Control information may belong to an IKE SA or a child SA. Therefore, Informational Exchange must be protected by the IKE SA or the IKE SA based on which the child SA is established accordingly.



Defining IPsec-Protected Data Flows

- The data flows to be protected by IPsec can be defined using either of the following methods:
  - Use ACLs.
    - ACLs can be configured to define the data flows to be protected by an IPsec tunnel. The packets matching permit clauses in the ACLs will be protected.
  - Use routes.
    - Routes can be configured to define the data flows to be protected by an IPsec tunnel established through IPsec tunnel interfaces. All packets routed to these interfaces will then be protected.
- On the live network, GRE over IPsec typically defines protected flows based on routes.

HUAWEI

- The method of using routes has the following advantages:
  - Simplifies the IPsec configuration: IPsec-protected data flows are routed to tunnel interfaces, without the need to use ACLs to define the characteristics of traffic to be encrypted or decrypted.
  - Supports dynamic routing protocols.
  - Protects multicast traffic through GRE over IPsec.

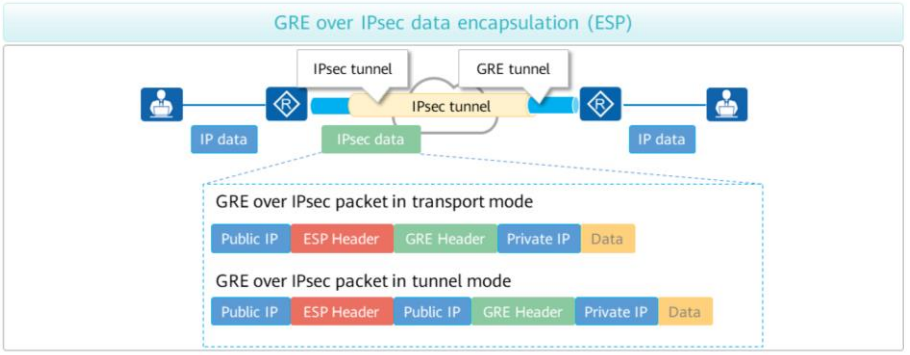
Contents

1. Common Networking Technologies in WAN Interconnection
2. **Security Technologies in WAN Interconnection**
   - Basic Concepts of IPsec
   - IPsec Fundamentals
   - IPsec Application Scenarios
3. NAT Traversal Technologies in WAN Interconnection
4. Intelligent Traffic Steering Technologies in WAN Interconnection

HUAWEI

GRE over IPsec

- Leveraging advantages of GRE and IPsec, GRE over IPsec encapsulates multicast, broadcast, and non-IP packets into ordinary IP packets and then securely transmits these IP packets through IPsec.
- GRE over IPsec encapsulates packets using GRE and then IPsec.

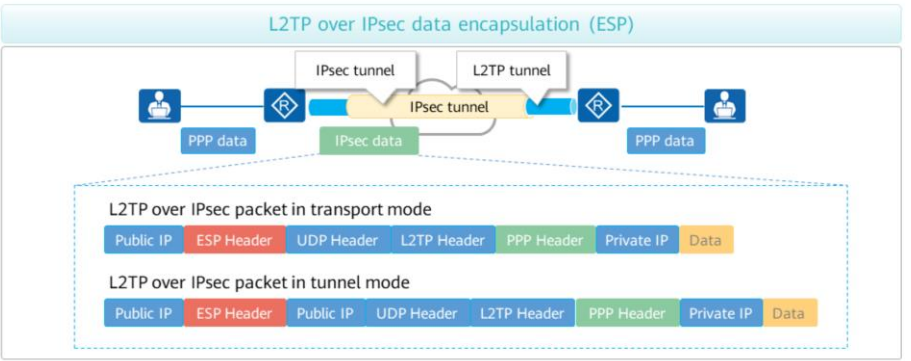


HUAWEI

- GRE over IPsec supports encapsulation in both tunnel and transport modes. An IPsec header needs to be added to packets if GRE over IPsec in tunnel mode is used, resulting in longer packets. In this case, packets are more likely to be fragmented. Therefore, GRE over IPsec in transport mode is recommended.
- In the IP header added during IPsec encapsulation, the source and destination addresses are the IP addresses of the local interface and remote interface to which an IPsec policy is applied.
- IPsec protects data flows from the GRE tunnel source to the GRE tunnel destination. In the IP header added during GRE encapsulation, the source and destination addresses are the source and destination addresses of a GRE tunnel.

L2TP over IPsec

- Layer 2 Tunneling Protocol (L2TP) over IPsec encapsulates packets using L2TP and then IPsec. It uses L2TP for user authentication and address allocation and uses IPsec for secure communication. L2TP over IPsec ensures that branches or traveling employees are securely connected to the headquarters.



HUAWEI

- L2TP encapsulation and then IPsec encapsulation are performed on packets transmitted over an L2TP over IPsec tunnel. In the IP header added during IPsec encapsulation, the source and destination addresses are the IP addresses of the local interface and remote interface to which an IPsec policy is applied.
- IPsec needs to protect the data flows from the L2TP tunnel source to the L2TP tunnel destination. In the IP header added to packets during L2TP encapsulation, the source and destination addresses are the source and destination addresses of an L2TP tunnel. When a branch connects to the headquarters, the source address of the L2TP tunnel is the IP address of the outbound interface on the L2TP access concentrator (LAC), and the destination address is the IP address of the inbound interface on the L2TP network server (LNS).
- A public IP header is added to packets during L2TP encapsulation, and another public IP header is added to packets if L2TP over IPsec in tunnel mode is used, resulting in longer packets, which are prone to being fragmented. Therefore, L2TP over IPsec in transport mode is recommended.
- The L2TP over IPsec negotiation process and packet encapsulation process are similar when traveling employees are remotely connected to the headquarters and when branch employees are connected to the headquarters. The difference is that, L2TP and IPsec encapsulation is performed on clients when traveling employees are remotely connected to the headquarters. The L2TP tunnel source address is the private address assigned to a client and can be any address in the IP address pool configured on the LNS. The L2TP tunnel destination address is the address of the inbound interface on the LNS.

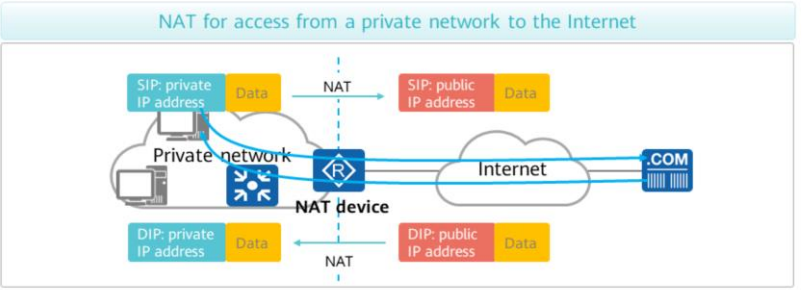
Contents

1. Common Networking Technologies in WAN Interconnection
2. Security Technologies in WAN Interconnection
3. **NAT Traversal Technologies in WAN Interconnection**
   - NAT Fundamentals
   - NAT Types
   - NAT Traversal Technologies
4. Intelligent Traffic Steering Technologies in WAN Interconnection

HUAWEI

Motivation Behind NAT

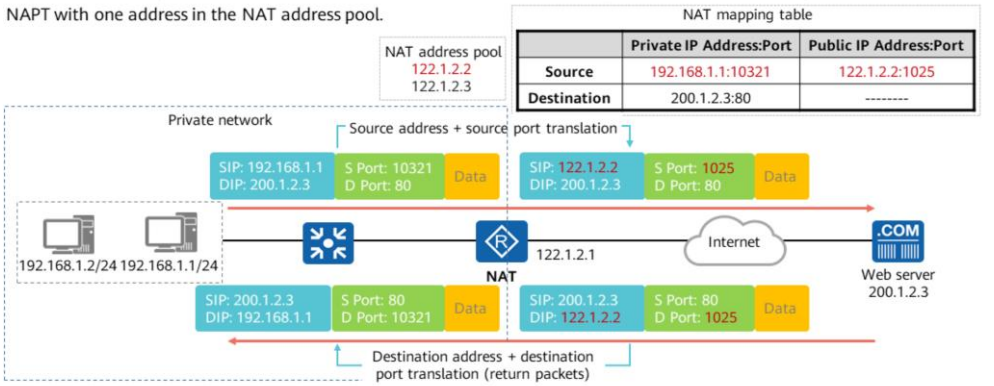
- Network address translation (NAT) was proposed in 1994. NAT can be used when some hosts on a private network have been assigned local IP addresses (that is, dedicated IP addresses used only on the private network) and want to communicate with hosts on the Internet (without encryption).
- NAT resolves the problem of insufficient public IP addresses and protects internal devices against external attacks.



HUAWEI

NAPT

- Network address and port translation (NAPT) translates both IP addresses and port numbers for addresses in an address pool. In this way, 1:n mapping between public and private addresses is implemented, which effectively improves public address utilization.
- Easy IP is a special type of NAPT. It maps all private network sessions to an egress address. That is, Easy IP can be understood as NAPT with one address in the NAT address pool.

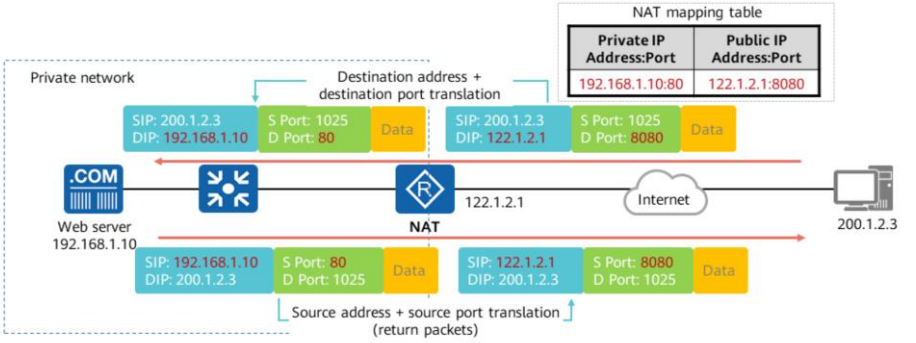


HUAWEI

- NAPT enables a public IP address to map multiple private IP addresses through ports. In this mode, both IP addresses and transport-layer ports are translated so that different private IP addresses with different source port numbers are mapped to the same public IP address with different source port numbers.
- Due to a limited number of TCP/UDP ports (0–65535), an address can be mapped to a limited number of private network sessions. Based on the NAT address pool, NAPT can randomly map a private network session to an address in the address pool, reducing the number of private network sessions.
- A NAT mapping entry is triggered by the first packet. When no traffic is transmitted for a period of time, the NAT mapping entry is automatically deleted to ensure security.
- NAPT and Easy IP are also known as source NAT because they change only the source address and port number of a packet.

NAT Server

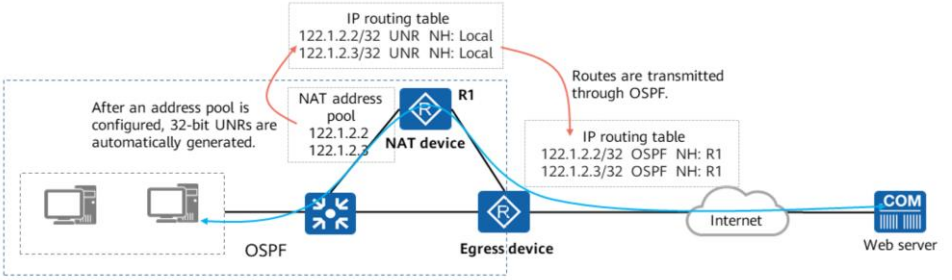
- NAT Server maps an internal server to a public network through a one-to-one mapping between a [**public IP address:port number**] and a [**private IP address:port number**]. This function is used when the internal server needs to provide services for the public network.
- An external host proactively accesses the [**public IP address:port number**] to communicate with the internal server.



HUAWEI

Route Advertisement in a NAT Address Pool

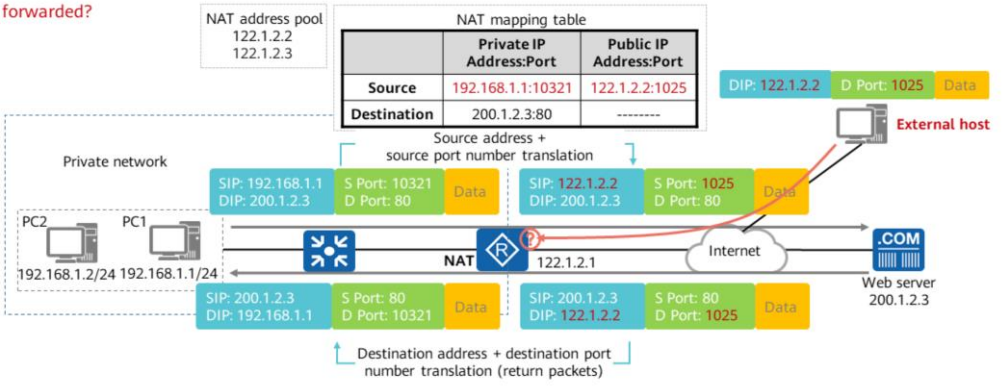
- In some cases, the NAT device may not be the egress device, so the return traffic needs to be diverted to the NAT device. To achieve this, you can import user network routes (UNRs) to a routing protocol.
- UNRs are generated by non-interface IP addresses and do not exist on any interface. In a NAT scenario, the device creates UNRs for addresses in the NAT address pool. For example, if the NAT address pool has addresses 122.1.2.2 and 122.1.2.3, the device automatically creates UNRs 122.1.2.2/32 and 122.1.2.3.3/32.



HUAWEI

Security Risks of the NAT Mapping Table

- After forwarding a data packet, the NAT device generates a NAT mapping table. When receiving a return packet, the NAT device considers the packet valid and forwards it only when a matching NAT mapping entry is found.
- If an external host uses the public IP address and port number in the NAT mapping table to send a packet, will the data be forwarded?



HUAWEI

- There are four types of NAT:
  - Full cone NAT
  - Restricted cone NAT
  - Port restricted cone NAT
  - Symmetric NAT

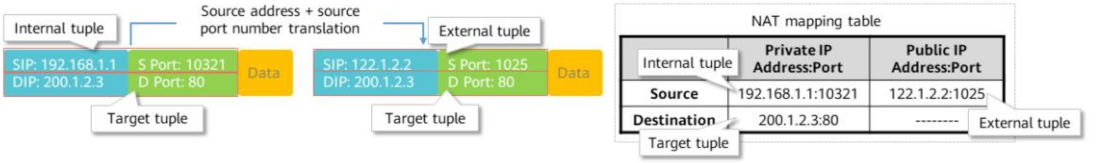
Contents

1. Common Networking Technologies in WAN Interconnection
2. Security Technologies in WAN Interconnection
3. **NAT Traversal Technologies in WAN Interconnection**
   - NAT Fundamentals
   - NAT Types
   - NAT Traversal Technologies
4. Intelligent Traffic Steering Technologies in WAN Interconnection

HUAWEI

Basic Concepts of NAT Types

- NAT can be classified into two types: cone NAT and symmetric NAT.
- Cone NAT is further classified into full cone NAT, restricted cone NAT, and port restricted cone NAT.
- Concepts related to NAT types
  - **Internal tuple**: a 2-tuple consisting of the private address and port number of an internal host, that is, the source address and source port number of a packet sent by the internal host.
  - **External tuple**: a 2-tuple consisting of the public address and port number obtained after an internal tuple is translated using NAT, that is, the post-NAT source address and source port number of a packet received by an external host.
  - **Target tuple**: a 2-tuple consisting of the address and port number of an external host, that is, the destination address and destination port number of a packet sent by an internal host.

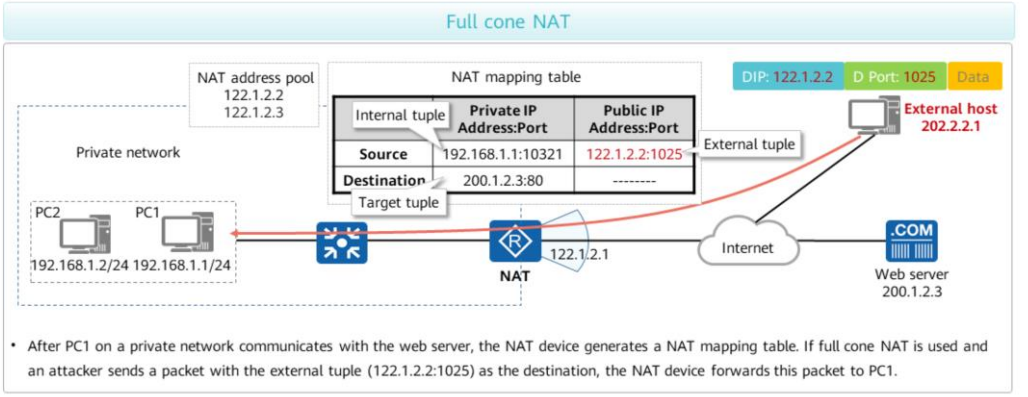


HUAWEI

- NAPT and Easy IP are also known as source NAT because they change only the source address and port number of a packet.

Full Cone NAT

- One internal tuple is translated into one external tuple by NAT. Any external host can send packets with the address and port number in the external tuple as the destination. When receiving such packets, the NAT device forwards them to the device corresponding to the mapping internal tuple. Full cone NAT is simple but has security risks.

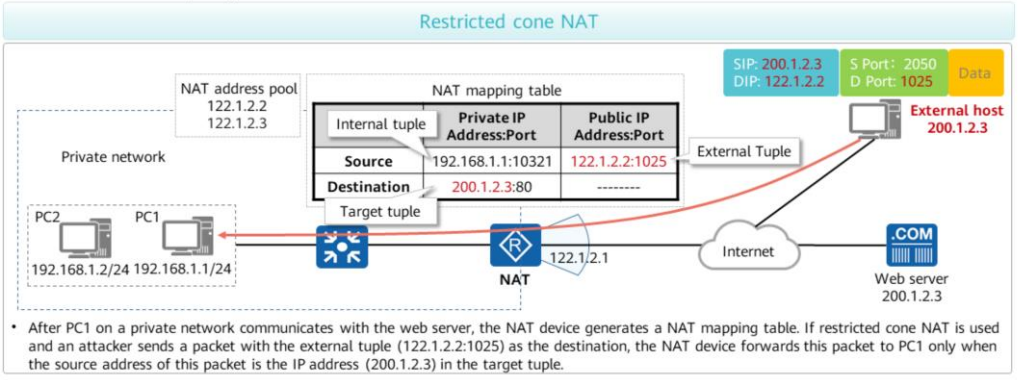


- After PC1 on a private network communicates with the web server, the NAT device generates a NAT mapping table. If full cone NAT is used and an attacker sends a packet with the external tuple (122.1.2.2:1025) as the destination, the NAT device forwards this packet to PC1.

HUAWEI

Restricted Cone NAT

- Restricted cone NAT is a restricted version of full cone NAT. When receiving a packet from an external host, the NAT device with restricted cone NAT enabled no only checks whether the destination address and port number of the packet are the same as those in the external tuple but also checks whether the source IP address of the packet is the same as the IP address in the target tuple. If both conditions are met, the packet is forwarded.

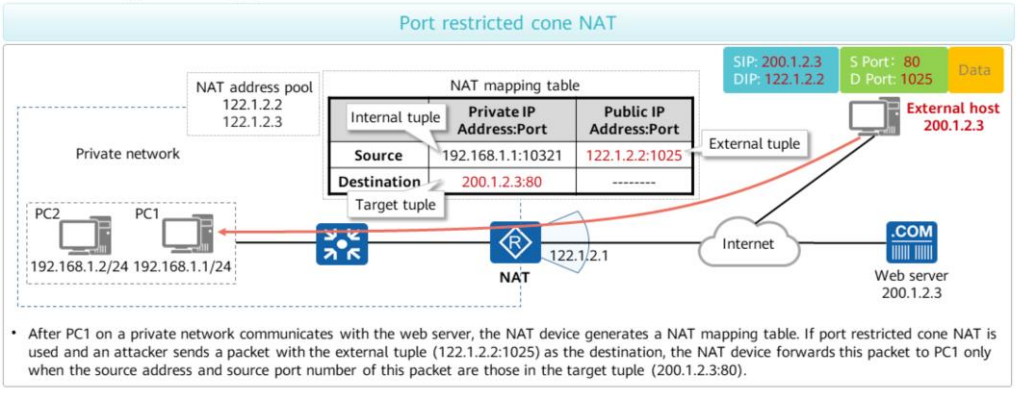


- After PC1 on a private network communicates with the web server, the NAT device generates a NAT mapping table. If restricted cone NAT is used and an attacker sends a packet with the external tuple (122.1.2.2:1025) as the destination, the NAT device forwards this packet to PC1 only when the source address of this packet is the IP address (200.1.2.3) in the target tuple.

HUAWEI

Port Restricted Cone NAT

- Port restricted cone NAT is a restricted version of restricted cone NAT. When receiving a packet from an external host, the NAT device with port restricted cone NAT enabled not only checks whether the source IP address and port number of the packet are the same as those in the external tuple but also checks whether the destination IP address and port number of the packet are the same as those in the target tuple. If both conditions are met, the packet is forwarded. This type of NAT is highly secure.

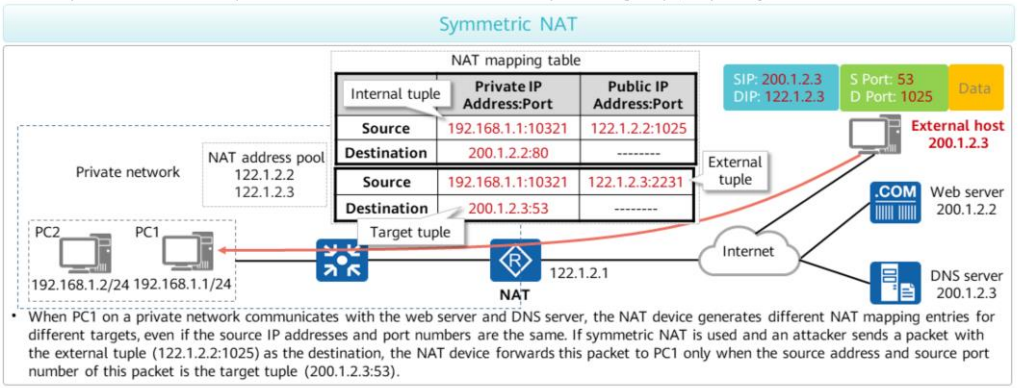


- After PC1 on a private network communicates with the web server, the NAT device generates a NAT mapping table. If port restricted cone NAT is used and an attacker sends a packet with the external tuple (122.1.2.2:1025) as the destination, the NAT device forwards this packet to PC1 only when the source address and source port number of this packet are those in the target tuple (200.1.2.3:80).

HUAWEI

Symmetric NAT

- When symmetric NAT is used, the same internal tuple and the same target tuple are translated into the same external tuple. However, if the internal or target tuple is different, they are translated into different external tuples. The NAT device with symmetric NAT enabled processes returned packets in a similar way to that with port restricted cone NAT enabled. Such a NAT device checks whether the source IP address + port number and the destination IP address + port number of a return packet are the same as those in the external tuple and target tuple, respectively.



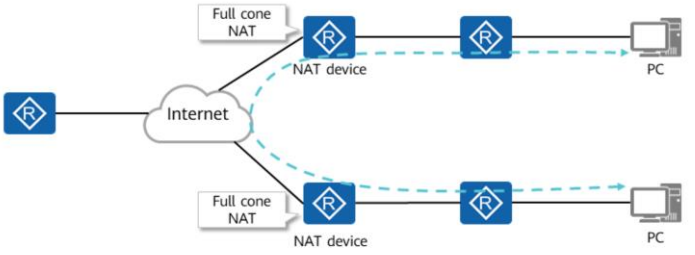
- When PC1 on a private network communicates with the web server and DNS server, the NAT device generates different NAT mapping entries for different targets, even if the source IP addresses and port numbers are the same. If symmetric NAT is used and an attacker sends a packet with the external tuple (122.1.2.2:1025) as the destination, the NAT device forwards this packet to PC1 only when the source address and source port number of this packet is the target tuple (200.1.2.3:53).

HUAWEI

NAT Application Scenarios

- On the live network, NAPT or Easy IP generally uses symmetric NAT.
- When NAT traversal is required on the live network, cone NAT can be used.

Application scenarios of NAT traversal



HUAWEI

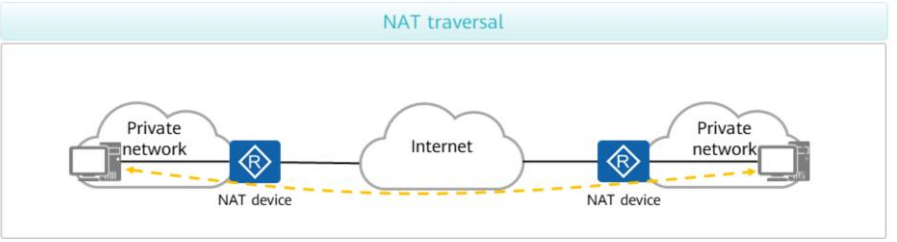
Contents

1. Common Networking Technologies in WAN Interconnection
2. Security Technologies in WAN Interconnection
3. **NAT Traversal Technologies in WAN Interconnection**
   - NAT Fundamentals
   - NAT Types
   - NAT Traversal Technologies
4. Intelligent Traffic Steering Technologies in WAN Interconnection

HUAWEI

Motivation Behind NAT Traversal

- Although NAT enables users on a private network to access a public network, it has the following defects:
  - NAT generates NAT mapping entries for the traffic from the private network to the public network. These entries have aging time. If a session between the two ends is silent for a long time, the connection is interrupted.
  - The private IP addresses of users are translated into the same public IP address by NAT. However, because servers may restrict the access frequency of the same IP address to prevent DoS attacks, some users may fail to access the servers. Besides, some applications cannot effectively trace the original IP devices, making it difficult for network management and fault locating.
- Multiple NAT traversal technologies are developed to solve the problems encountered by end-to-end IP applications in the NAT environment.



HUAWEI

Existing NAT Traversal Technologies

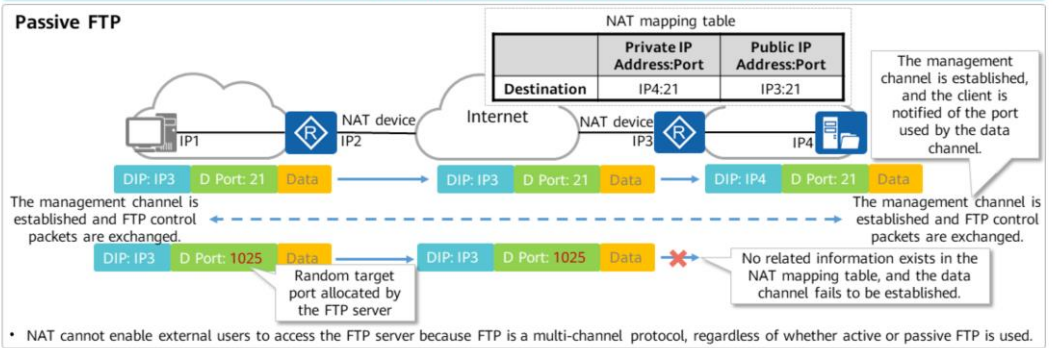
- Currently, the mainstream NAT traversal technologies are as follows:
  - Universal Plug and Play (UPnP)
  - Application Level Gateway (ALG)
  - Middlebox Communications (MIDCOM)
  - Full Proxy
  - Session Traversal Utilities for NAT (STUN)
  - Traversal Using Relay NAT (TURN)
  - Interactive Connectivity Establishment (ICE)
  - STUN and TCP too (STUNT)
- This course describes ALG and STUN.

HUAWEI

Motivation Behind NAT ALG

- Servers that provide various services (such as HTTP, DNS, and FTP) on the live network are usually deployed on a private network. Therefore, the NAT Server function needs to be deployed on the NAT device so that external users can access these services. However, some services use multi-channel protocols, such as FTP, DNS, and SIP. Deploying NAT Server cannot allow external users to access these services. For these multi-channel protocols, you can use NAT ALG to solve the problem.

Challenges of multi-channel protocols to NAT (FTP)

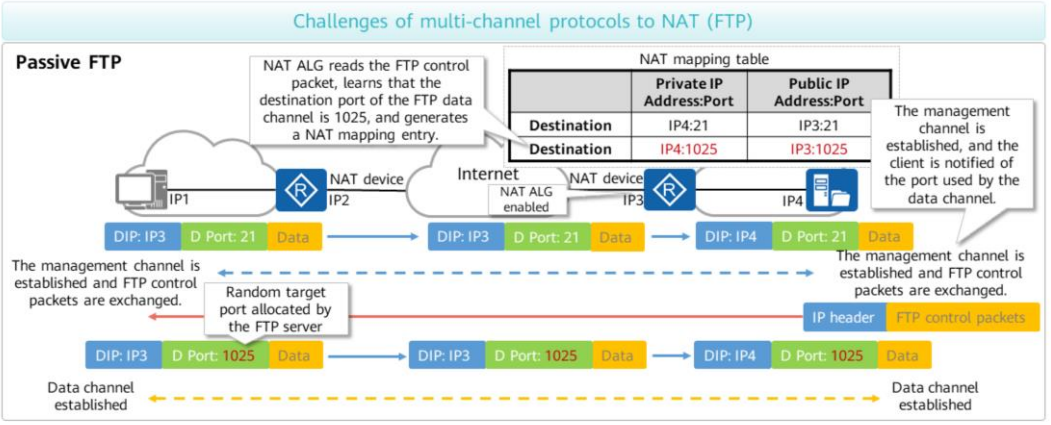


- NAT cannot enable external users to access the FTP server because FTP is a multi-channel protocol, regardless of whether active or passive FTP is used.

HUAWEI

Fundamentals of NAT ALG

- NAT ALG can parse the payload of packets, identify and translate important information (such as the destination port of the FTP data channel) in the payload, and generate NAT mapping entries based on the important information to enable external hosts to access the servers on a private network.



HUAWEI

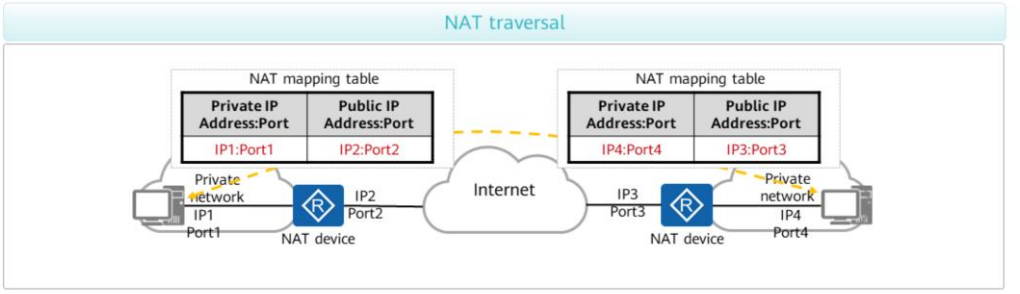
Disadvantages of NAT ALG

- NAT ALG needs to read protocol packets. For new protocols, the NAT ALG feature needs to be upgraded to support them. NAT ALG does not support proprietary multi-channel protocols.
- A large number of devices on the live network do not support NAT ALG. As a result, the deployment cost is high.
- NAT ALG solves the problems facing multi-channel protocols in the NAT Server scenario, but does not support the communication between hosts on private networks.

HUAWEI

Overview of STUN

- In addition to NAT ALG, cone NAT can also be used for NAT traversal. It creates NAT mapping entries on NAT devices in advance and establish connections between private networks based on the NAT mapping entries.
- STUN is mainly used to obtain the mapping between the private IP address + port and the post-NAT public IP address + port on the NAT device. The data channel for NAT traversal needs to be established in other modes.

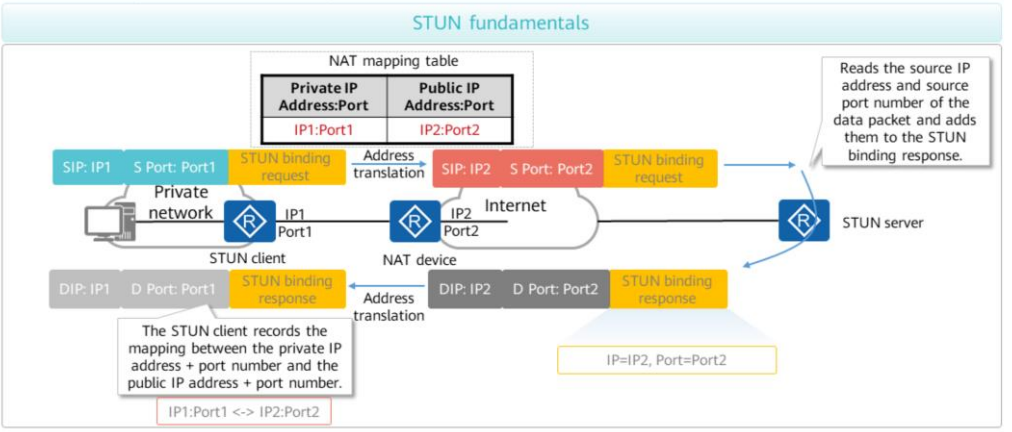


HUAWEI

- In RFC 3489, STUN is a complete NAT traversal solution and its full name is Simple Traversal of UDP Through NATs.
- In the new RFC 5389 revision, STUN is positioned to provide a tool for NAT traversal rather than a complete solution. The full name of STUN is changed to Session Traversal Utilities for NAT. Besides the full name difference, STUN in RFC 5389 differs from STUN in RFC 3489 in that STUN in RFC 5389 supports NAT traversal for TCP.

Fundamentals of STUN

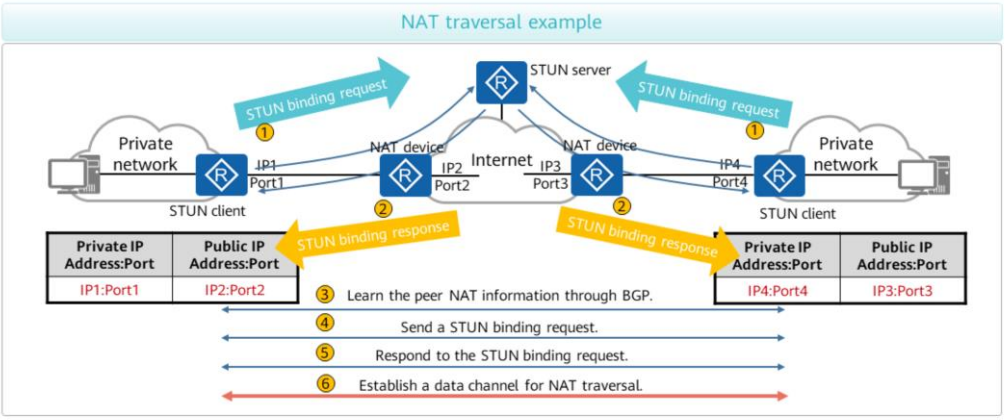
- STUN uses the client/server model. The STUN client and STUN server exchange packets to discover the NAT device and determine the IP address and port number allocated by the NAT device.



HUAWEI

Application Example of STUN

- In an SD-WAN scenario, STUN can be used to interconnect post-NAT devices and establish data channels.



HUAWEI

- A STUN client sends a STUN binding request to the STUN server.
- After receiving the STUN binding request, the STUN server obtains the source IP address and port number, constructs a STUN binding response, and sends the response to the client.
- The STUN client obtains an IP address and port number from the binding response, and compares the obtained IP address and port number with the source IP address and port number carried in the binding request. If they are different, a NAT device is used in front of the STUN client.
- STUN clients use BGP to learn each other's NAT information (IP addresses and port numbers before and after NAT).
- The local STUN client uses the local pre-NAT IP address and port number and the pre-NAT IP address and port number of the peer STUN client to construct a STUN binding request and sends it to the peer STUN client. In addition, the local STUN client uses the local pre-NAT IP address and port number and the post-NAT IP address and port number of the peer STUN client to construct a STUN binding request and sends it to the peer STUN client. The peer STUN client performs the same operations.
- After receiving the STUN binding request, the peer STUN client sends a STUN binding response to the local STUN client. The local STUN client performs the same operations.
- After the preceding STUN messages are exchanged, a data channel is established between the STUN clients so that packets can traverse the NAT devices.

Contents

1. Common Networking Technologies in WAN Interconnection
2. Security Technologies in WAN Interconnection
3. NAT Traversal Technologies in WAN Interconnection
4. **Intelligent Traffic Steering Technologies in WAN Interconnection**
   - SAC
   - SRP

HUAWEI

Service Reliability

- In the cloud computing era, network reliability cannot meet user requirements. Users want to understand the live network status based on applications and adjust the network based on the application status.
- Such requirement poses the following challenges to traditional networks:
  - Traditional networks cannot accurately identify applications.
  - Traditional networks cannot be adjusted based on applications.
- To cope with the challenges, two technologies are developed:
  - Smart Application Control (SAC): This technology can flexibly identify applications.
  - Smart Policy Routing (SPR): This technology can switch forwarding paths based on the network or application status.

HUAWEI

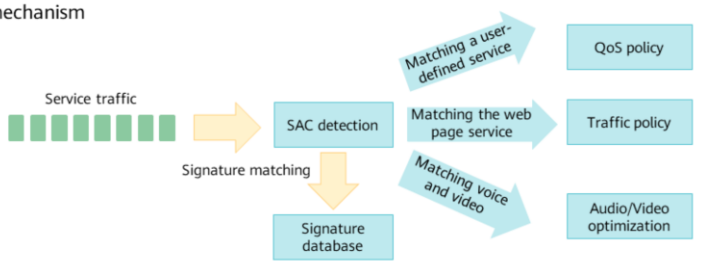
Overview of SAC

- Traditional networks are managed based on traffic. However, in the cloud computing era, services are becoming increasingly important. Networks need to be managed and monitored based on applications instead of Five-tuple information.
- Traditional routing and switching devices cannot identify application-layer information. Therefore, it is difficult to manage networks based on applications. Smart Application Control (SAC) technology helps routing and switching devices identify classified applications.
- SAC uses service awareness (SA) and first packet identification (FPI) technologies to detect and identify Layer 4 to Layer 7 information (such as HTTP and RTP) in packets.

HUAWEI

SAC Signature Database

- Signature identification is a basic function of SA technology. Different applications typically use different protocols, and different application protocols have their own signatures. A signature that can identify a protocol is known as a signature code. The system analyzes service flows passing through a device, and compares the analysis result with the signature database on the device. It identifies an application by detecting the signature code in data packets.
- SAC signature databases include FPI and SA signature databases. FPI signatures refer to signatures for identifying FPI applications, and SA signatures refer to signatures for identifying SA applications.
- SAC working mechanism

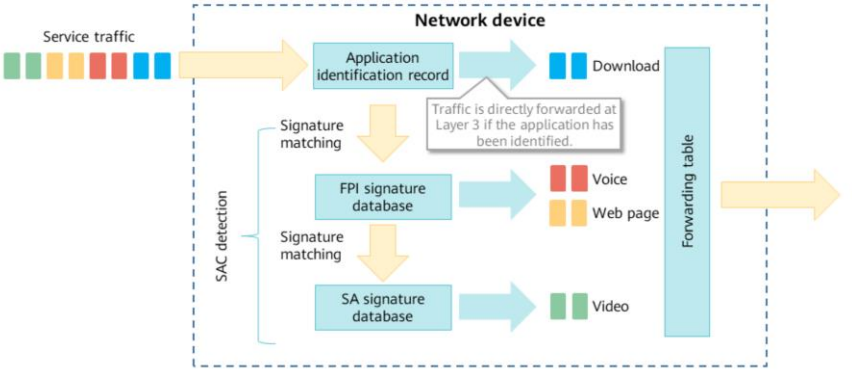


HUAWEI

- The SAC signature database file can only be updated through upgrades and cannot be manually modified.
- The SAC signature database can be updated in either of the following modes:
  - Online update: The SAC signature database can be updated through the security center platform or intranet update server.
  - Local update: The upgrade package is downloaded from the security center platform and uploaded to the device through FTP for the update of the SAC signature database.

SAC Application Identification Process

- During SAC application identification, the system checks whether an application is identified. If the application is not identified, the system checks the FPI signature database and SA signature database in sequence.

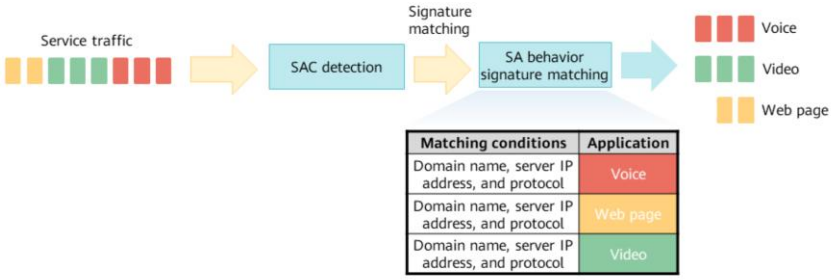


HUAWEI

- After a packet enters the device, the device determines whether the corresponding application has been identified based on the 5-tuple information carried in the packet. If the application has been identified, the device forwards the packet at Layer 3 without identifying the application again. If the application has not been identified, the device performs the SAC application identification process. The device then processes the packet based on the SAC identification result and forwards the packet at Layer 3. The SAC application identification process is as follows: The device identifies an application based on the ACL rules defined in FPI. If the application cannot be identified, the device identifies the application based on the DNS entries defined in FPI. If the application still cannot be identified, the device identifies the application based on the protocol and port mapping table defined in FPI. If the application still cannot be identified, the device starts the SA identification process.

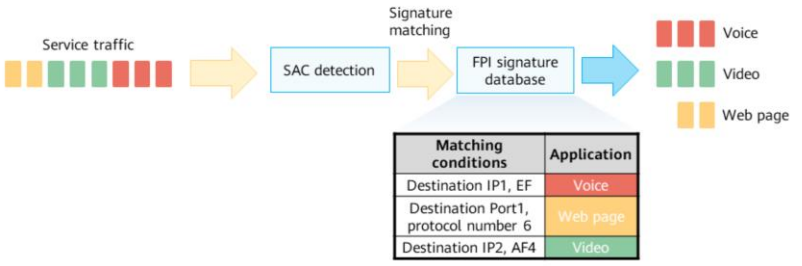
SA

- After receiving data, the device can use service awareness (SA) technology to match applications.
- SA uses the SA signature database to detect services. The existing SA signature database is embedded with more than 6000 applications, ensuring a high identification rate for public applications. In most cases, the SA signature database can only be updated online or locally and cannot be manually modified.



FPI

- There is a problem in matching applications based on the SA signature code. That is, the application corresponding to the first several packets may fail to be identified based on the SA signature code. As a result, the processing on the first and subsequent packets may be inconsistent. First packet identification (FPI) enables a device to identify an application by matching the first packet of a flow.
- FPI identifies applications based on 5-tuple information, DSCP values, protocols, and DNS domain names. The system provides a predefined FPI signature database to help SAC identify applications. You can also define FPI applications to identify new applications.



HUAWEI

- FPI applications are classified into the following types:
  - Predefined and user-defined FPI applications based on the protocol and port number: These two types of applications are identified using entries that are generated based on the protocol and port number carried in packets. The difference is as follows: Packets of a predefined FPI application contain common protocols and port numbers, while packets of a user-defined FPI application contain the protocols and ports that you define.
  - Predefined and user-defined FPI applications based on the DNS domain name: These two types of applications are identified using DNS entries generated through association between FPI and DNS. The difference is as follows: Packets of a predefined FPI application contain common DNS domain names, while packets of a user-defined FPI application contain the DNS domain names that you define.
  - User-defined FPI application based on 5-tuple and DSCP information. This application is identified based on the user-defined 5-tuple and DSCP information using advanced ACL rules.
- Identification process of FPI applications based on the DNS domain name
  - FPI applications based on the DNS domain name are identified using DNS entries generated through association between FPI and DNS. The FPI signature database contains the mappings between domain names and applications. DNS response packets contain the mappings between domain names and IP addresses. Based on the mappings, a device generates DNS entries, which contain the mappings between IP addresses and applications. The device searches for DNS entries based on the IP address carried in the application protocol packets to identify the corresponding application.

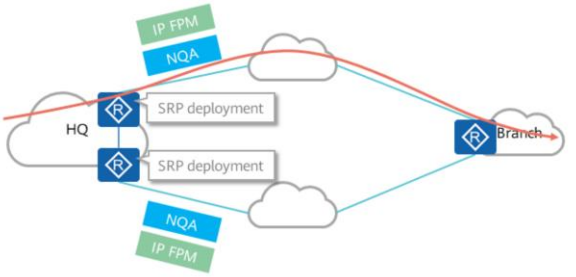
Contents

1. Common Networking Technologies in WAN Interconnection
2. Security Technologies in WAN Interconnection
3. NAT Traversal Technologies in WAN Interconnection
4. **Intelligent Traffic Steering Technologies in WAN Interconnection**
   - SAC
   - SRP

HUAWEI

Overview of SPR

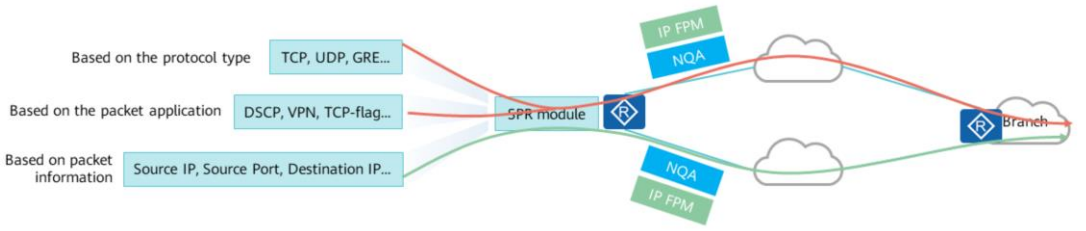
- In the cloud computing era, more users shift their attention from network connectivity to service availability, such as service response speed and service quality. However, traditional networks cannot detect link quality and service requirements, resulting in poor user experience.
- Smart Policy Routing (SPR) addresses this problem. It actively detects the link quality and matches service requirements to select an optimal link to forward service data. SPR prevents network blackholes and flappings.



HUAWEI

SPR Service Differentiation

- SPR differentiates traffic based on the protocol type, packet application, and packet information.
- Different link quality parameter thresholds can be set for different services. You can set the delay (D), jitter (J), packet loss rate (L), and composite measure indicator (CMI).
- CMI is calculated based on the delay, jitter, and packet loss rate.
- SPR selects routes based on the CMI.

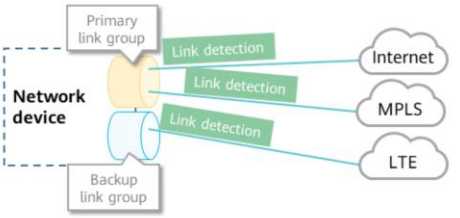


HUAWEI

- SPR classifies services based on the following attributes:
  - Protocol types: IP, TCP, UDP, GRE, IGMP, IPINIP, OSPF, and ICMP
  - Packet applications: DSCP, ToS, IP precedence, fragment, VPN, and TCP-flag
  - Packet fields: Source IP Address, Destination IP Address, Protocol, Source Port, Destination Port, Source IP Prefix, Destination IP Prefix
- When SPR selects routes for services based on the NQA detection result, the CMI is calculated using the following formula:
  - CMI = 9000 – CMI-method. The default value of CMI-method is D + J + L.
  - If NQA is used, a larger CMI value indicates better link quality.
- When SPR selects routes for services based on the IP FPM detection result, the CMI is calculated using the following formula:
  - CMI = D + J + L
  - If IP FPM is used, a smaller CMI value indicates better link quality.

SPR Detection Link and Link Group

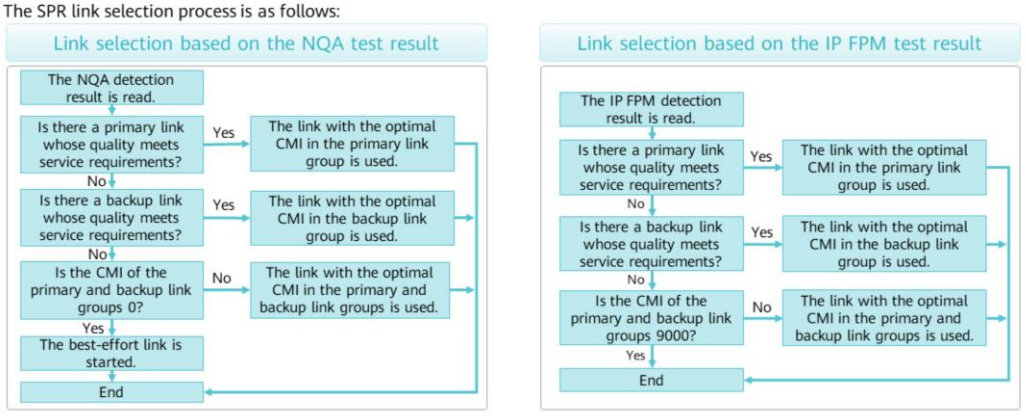
- SPR obtains quality indicators of detection links through probes (NQA or IP FPM) and then selects an optimal link.
- A link group can contain one or more detection links.
- SPR defines three roles for links: primary link group, backup link group, and best-effort link. When no suitable link is available in the primary and backup link groups, SPR activates the best-effort link to forward service data.



HUAWEI

SPR Link Selection

- SPR periodically obtains the NQA or IP FPM detection result to determine whether a link meets service requirements. If the link does not meet service requirements, a link switchover is triggered.
- The SPR link selection process is as follows:

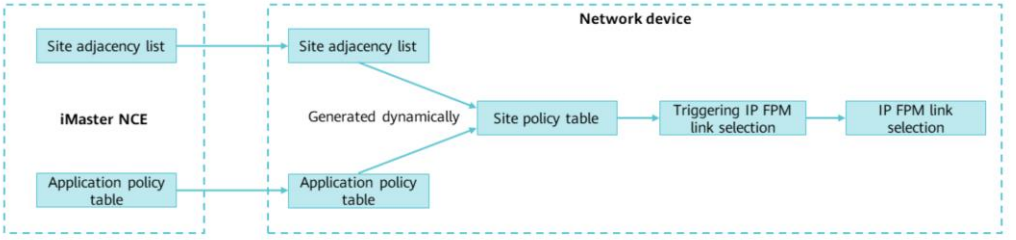


HUAWEI

- When a network is unstable, SPR triggers link switchovers frequently, which degrades service experience. SPR provides the flapping suppression function to address this problem.
- The flapping suppression function is disabled by default, and the flapping suppression period is configurable. After traffic is switched to a new link, SPR starts the flapping suppression timer. Within a flapping suppression period, SPR does not perform a link switchover even if it does not obtain the NQA test result indicating that the link meets service requirements within a switchover period. After the flapping suppression timer expires, if SPR still does not obtain the NQA test result indicating that the link meets service requirements within a switchover period, SPR performs a link switchover. If SPR obtains the NQA test result indicating that the link meets service requirements within a switchover period, SPR retains traffic on the link without performing a link switchover.

Using iMaster NCE to Implement SPR

- During SPR deployment through iMaster NCE, to improve the site specifications on the entire network, separation between sites and application policies and traffic-triggered link selection are used together.
- iMaster NCE maintains site adjacency information and application policies, and SPR is configured on routers. Traffic-triggered link selection allows for on-demand generation of SPR configurations. This prevents a large number of configurations from being created on the device and reduces the impact of link selection (based on IP FPM) on the CPU, significantly reducing the burden on the device.
- iMaster NCE can use SAC to classify service traffic based on applications.



Quiz

1. (Multiple-answer question) Which of the following are private line technologies? ( )
   A. SDH
   B. L2TP
   C. MPLS VPN
   D. IPsec VPN
2. (True or false) SRP technology can flexibly select egress links based on the link quality. ( )
   A. True
   B. False

HUAWEI

1. AC
2. A

Summary

- GRE over IPsec is used for WAN interconnection, and L2TP over IPsec is used for remote intranet access.
- Based on IKE SA information, IPsec SA information can be securely transmitted on the public network. Based on IPsec SA information, service data can be securely transmitted on the public network.
- To establish a WAN connection between two private network devices, you need to use a NAT traversal technology. The common NAT traversal technology is STUN.
- To ensure WAN reliability, the link quality needs to be detected first, and then a proper egress or egress link is selected based on the link quality through specific technologies.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

SD-WAN Solution Planning and Design

HUAWEI

Foreword

- After years of development and evolution, WAN interconnection has undergone significant changes. In the past, WAN interconnection was centered on the network, and there were few applications. As the main component of WAN interconnection, the network took the most important position. However, with the rise of cloud computing, the potential of applications is fully exploited, and WAN interconnection gradually become application-centric.
- To address the challenges posed by cloud computing to WAN interconnection, Huawei launches the SD-WAN Solution.
- This course will provide insights into the functions available in the SD-WAN Solution and the design scheme for this solution.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the challenges faced by enterprise WAN interconnection.
  - Describe the architecture and functions of Huawei SD-WAN Solution.
  - Describe the SD-WAN networking design scheme.
  - Describe the SD-WAN service design scheme.
  - Describe the SD-WAN reliability design scheme.
  - Describe the SD-WAN security design scheme.

HUAWEI

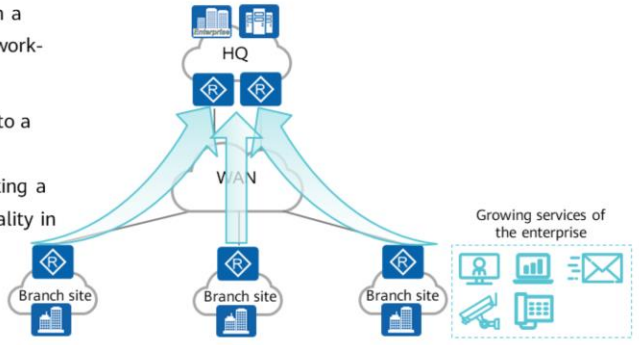
Contents

1. **Development Trends and Challenges Facing Enterprise WAN Interconnection**
2. Overview of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. Service Design for Huawei SD-WAN Solution
5. Reliability and Security Design for Huawei SD-WAN Solution

HUAWEI

Challenges Brought by Cloud Computing to Enterprise WAN Interconnection

- Before cloud computing is widely used, there are only a small number of network applications. The network service quality depends only on the bandwidth, and service traffic does not need to be managed in a refined manner. The Internet is built on a network-centric design.
- The advent of the cloud computing era leads to a significant increase in the number of network applications. Enterprises have difficulty in striking a balance between the line price and service quality in the face of soaring traffic.

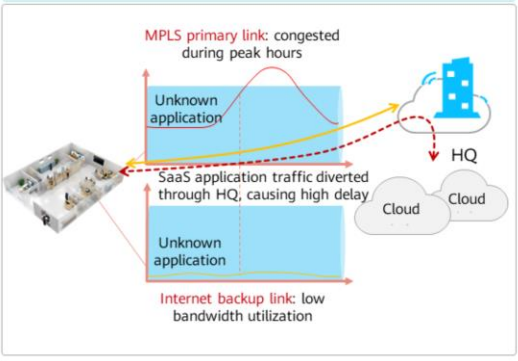


HUAWEI

Challenges Brought by Multiple Services to Enterprise WAN Interconnection

- Enterprises are unable to detect service quality in real time and therefore cannot effectively guarantee key services. In addition, enterprises cannot monitor service traffic in real time, and therefore are unable to quickly adjust service traffic.

No application visibility, causing difficulty in traffic scheduling



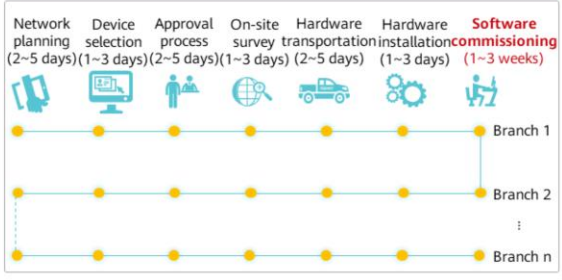
Difficulty in managing key services such as voice, video, and SaaS applications

HUAWEI

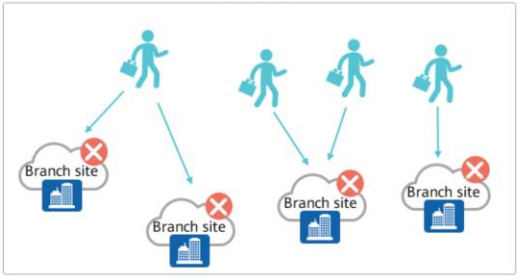
Challenges Brought by Large Numbers of Branches to Enterprise WAN Interconnection

- With the development of an enterprise, it will have more and more cross-city, cross-province, and cross-country branches, causing the following problems in managing branch site networks:
  - Too many branches result in high O&M costs.
  - It takes a long time to provision new services in branches.
  - It is difficult to rectify faults on branch networks.

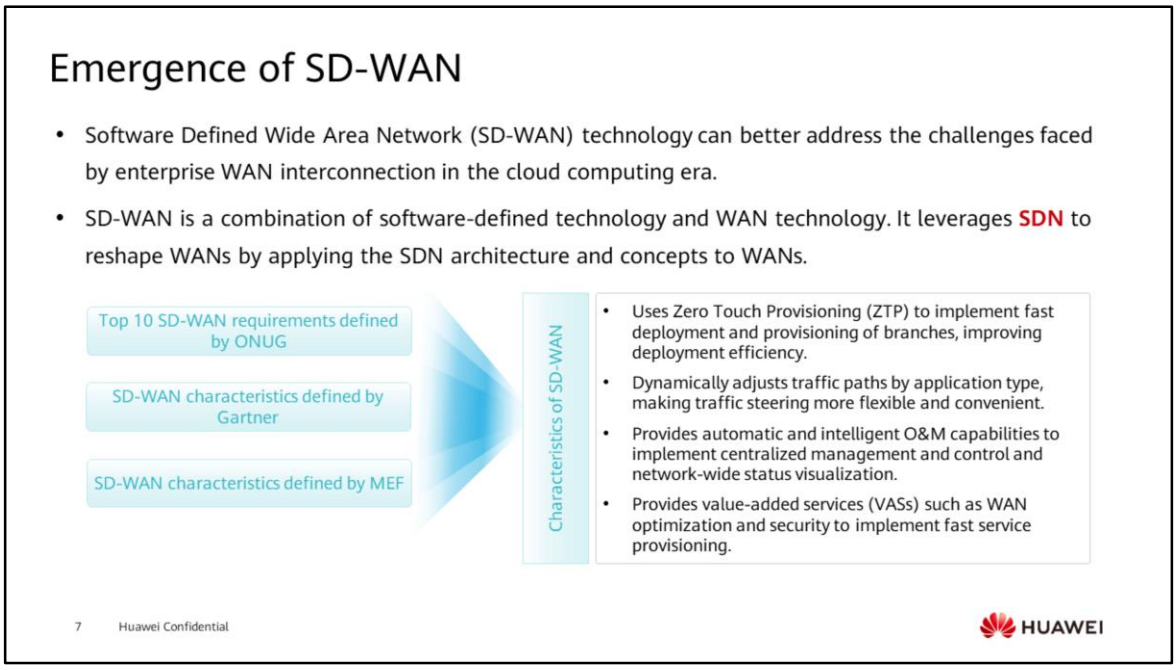
Time-consuming service provisioning in branches



Difficult troubleshooting on branch networks, resulting in high O&M costs



HUAWEI



- ONUG: refers to Open Networking User Group. ONUG is an influential user organization led by large enterprises and comprised mainly of IT users. It was founded by IT technical executives of well-known large enterprises in North America and is dedicated to driving IT implementation and network technology transformation for large enterprises. ONUG members include large enterprises in industries such as finance, insurance, medical care, and retail. ONUG serves as a platform for high-end customers in North America to discuss and communicate IT requirements.
- Gartner: It is the world's most authoritative IT research and advisory company. Its research scope covers all IT industries. It provides objective and fair demonstration reports and market research reports for customers in terms of IT research, development, evaluation, application, and market, thereby assisting customers in market analysis, technology selection, project demonstration, and investment decision-making.

Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. **Overview of Huawei SD-WAN Solution**
   - Huawei SD-WAN Solution
   - Major Functions of Huawei SD-WAN Solution
   - Application Scenarios of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. Service Design for Huawei SD-WAN Solution
5. Reliability and Security Design for Huawei SD-WAN Solution

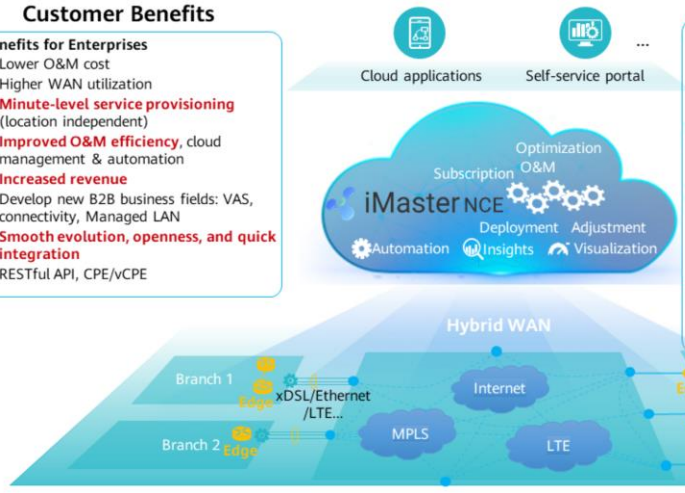
HUAWEI

Overview of Huawei SD-WAN Solution

Customer Benefits

Benefits for Enterprises

- Lower O&M cost
- Higher WAN utilization
- **Minute-level service provisioning** (location independent)
- **Improved O&M efficiency**, cloud management & automation
- **Increased revenue**
  Develop new B2B business fields: VAS, connectivity, Managed LAN
- **Smooth evolution, openness, and quick integration**
  RESTful API, CPE/vCPE



Key Technologies

5G uplink: 5G supported by all CPE series products

- Large bandwidth: 230 Mbit/s for uplink, 900 Mbit/s for downlink
- Support for all network generations: 5G/4G/3G/2G networks

High performance: no congestion during forwarding

- CPU+NP heterogeneous forwarding architecture

Optimal experience: application-based intelligent traffic steering, ensuring experience of key applications

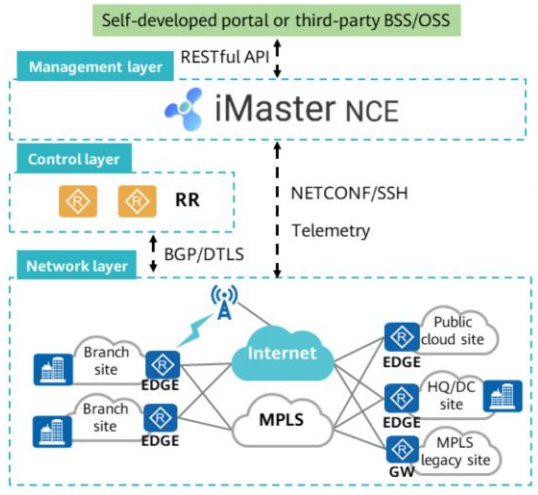
- Application-based intelligent traffic steering, enabling on-demand 5G+fiber scheduling.

Simplified O&M: full-process automation and plug-and-play

- Visibility of application, branch, device, and link status, and centralized management

HUAWEI

Architecture of Huawei SD-WAN Solution

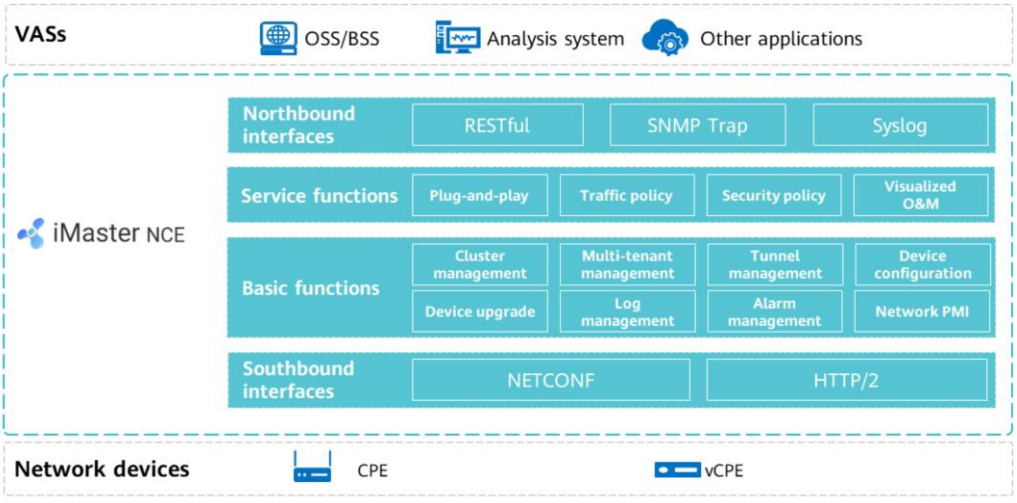


Role	Funcation	Product Form
Management layer	1) Network service orchestration 2) Network performance monitoring and visualization 3) Network O&M 4) Network device management	iMaster NCE
Control layer	1) Routing and tunnel information distribution 2. IPsec key exchange 3) VPN topology definition 4) NAT Stun service	RR (AR Router)
Network layer	1) EDGE: egress CPEs at enterprise branches, headquarters, DCs, and cloud sites 2) GW: multi-tenant gateway device	AR Router

HUAWEI

- The overall architecture of the SD-WAN solution consists of the network layer, control layer, and orchestration layer. The layers are associated with each other through standard interfaces and communication protocols.
    - Layer: The Agile Controller-Campus manages enterprise interconnection services in a refined manner. The Agile Controller-Campus manages network devices through NETCONF in the southbound direction. The Agile Controller-Campus provides standard RESTful interfaces to interwork with third-party applications in the northbound direction.
    - Control layer: The controller works with distributed control components to transfer routes between sites in an area and implement inter-area network interconnection.
    - Network layer: Cost-effective network devices and overlay technologies are used between branches, headquarters, and cloud platforms to build on-demand full-network connections based on any link, such as the Internet and traditional private lines.
- iMaster NCE-WAN/iMaster NCE-Campus can be used as Huawei SD-WAN controllers.
- Enterprise branches, headquarters, data centers, and IT infrastructure deployed on the cloud are collectively referred to as enterprise sites.

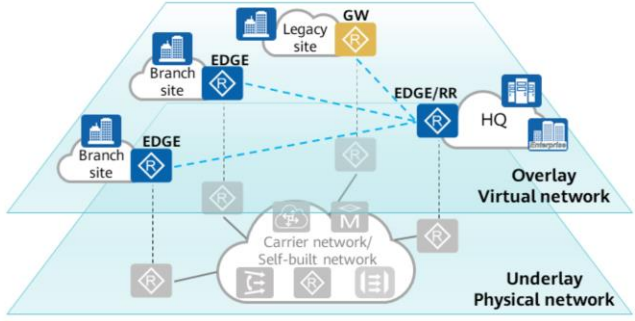
Architecture of Huawei iMaster NCE-WAN



HUAWEI

Network Layer Overview

- An enterprise SD-WAN network can be divided into two layers: physical network (underlay) and virtual network (overlay), which are completely decoupled from each other.
  - Physical network: refers to the underlay WAN provided by a carrier or built by the enterprise, including the private line network and the Internet.
  - Virtual network: also called the overlay network. Huawei SD-WAN Solution uses the IP overlay virtualization technology to build one or more virtual overlay networks on top of the physical network. Service policies are deployed on virtual networks and are decoupled from the physical network, thereby separating services from the WAN.

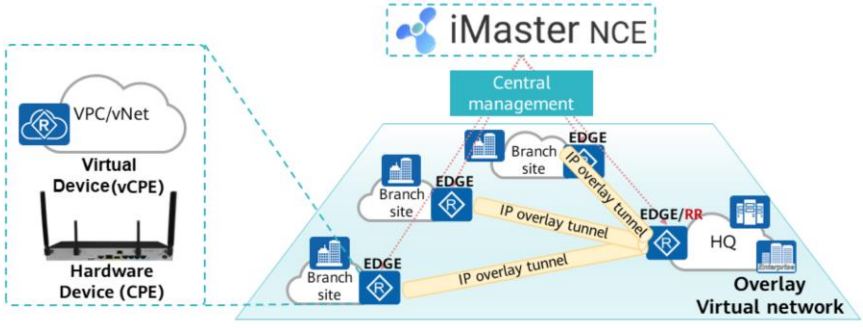


HUAWEI

- Multiple virtual networks can be deployed to provide different services for the same tenant (for example, services for multiple departments) or provide different services for different tenants.
- In terms of network device functions, the network layer of the SD-WAN Solution consists of two types of NEs: CPE and gateway (GW).

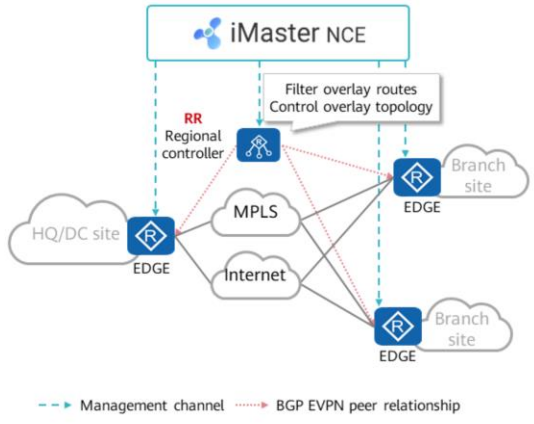
EDGE Overview

- A EDGE is essentially an edge node of the SD-WAN network. EDGEs are interconnected using the IP overlay tunneling technology.
- Traditional hardware EDGEs are typically used on the HQ and branch sites, and virtual devices can be deployed at sites on the public cloud.
- All SD-WAN EDGEs of an enterprise are centrally managed and maintained in iMaster NCE-WAN by the tenant administrator.



HUAWEI

RR Overview



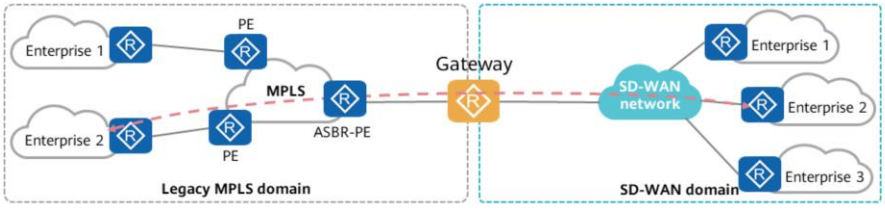
- A route reflector (RR) is used to transmit BGP routes.
- In Huawei SD-WAN Solution, RRs also control routes and network topologies. Therefore, RRs are also called regional controllers in this solution.
- Both RRs and EDGEs at edge sites are managed by iMaster NCE-WAN.
- Control channels are established between RRs as well as between RRs and edge sites.
- RRs are managed by iMaster NCE-WAN and control route sending and receiving at edge sites based on the overlay network topology model. In this way, sites can communicate with each other based on the user-configured overlay topology model.

HUAWEI

- RR site: The CPE at the site functions as an RR and distributes EVPN routes between CPE gateways at edge sites based on VPN topology policy.
- If the tenant administrator assigns the role of "gateway + RR" to an egress CPE when adding the CPE, the site where the CPE resides is an RR site. If no device at a site is assigned the "gateway + RR" role, the site is an edge site.
- An edge site can establish IBGP peer relationships with two RRs that back up each other.
- Multiple RRs can be deployed for a tenant. All RRs are connected in full-mesh mode on the control plane.

Gateway Overview

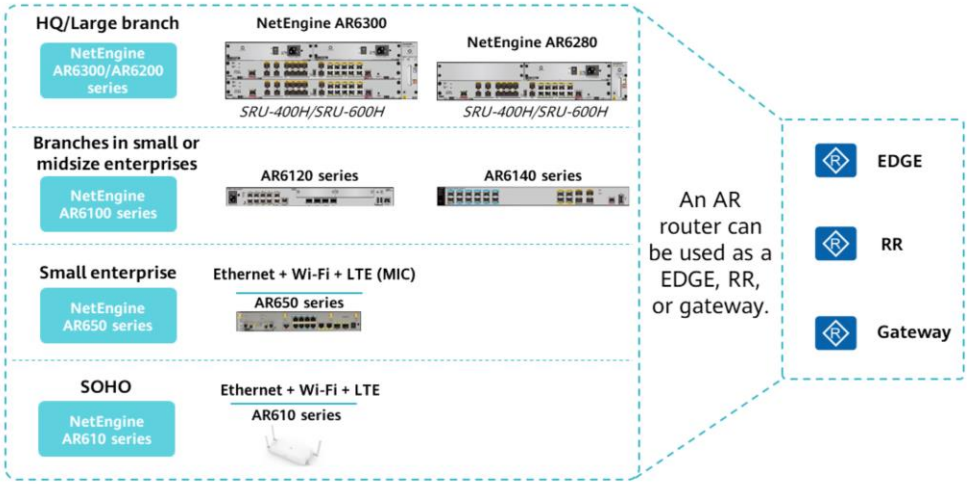
- New SD-WAN sites of an enterprise need to communicate with its legacy sites or third-party services. Some legacy sites are interconnected through MPLS VPN, and SD-WAN sites are interconnected through IP overlay tunnels. Therefore, the legacy network and SD-WAN network cannot directly communicate with each other.
- An SD-WAN gateway can connect to both the SD-WAN and legacy networks. It can function as an intermediate gateway to implement interconnection between SD-WAN and legacy networks.



HUAWEI

- A gateway has different roles in different service scenarios. For example, a gateway connected to a legacy site may be referred to as an interworking gateway (IWG), and a gateway connected to the cloud may be called a cloud gateway. These gateways can extend functions by interconnecting with each other to establish a Point of Presence (PoP) network, where these gateways are referred to as PoP gateways.

Huawei CPE Devices: NetEngine AR Routers



HUAWEI

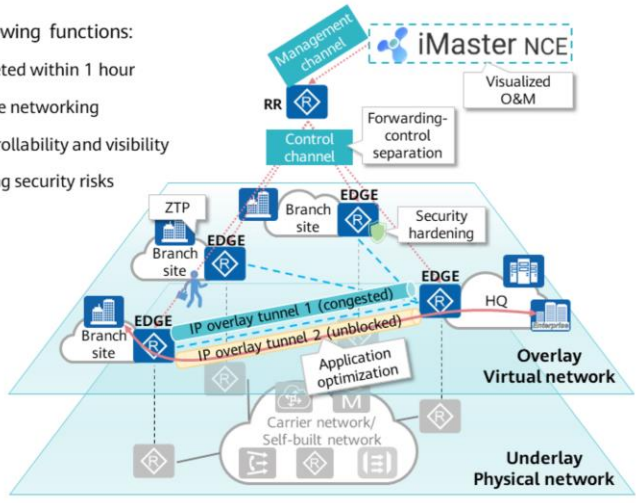
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. **Overview of Huawei SD-WAN Solution**
   - Huawei SD-WAN Solution
   - Major Functions of Huawei SD-WAN Solution
   - Application Scenarios of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. Service Design for Huawei SD-WAN Solution
5. Reliability and Security Design for Huawei SD-WAN Solution

HUAWEI

Major Functions of Huawei SD-WAN Solution

- Huawei SD-WAN Solution provides the following functions:
  - ZTP, enabling service provisioning to be completed within 1 hour
  - Forwarding-control separation, enabling flexible networking
  - Application optimization, enabling service controllability and visibility
  - Complete security protection system, eliminating security risks
  - Visualized O&M for quick fault locating



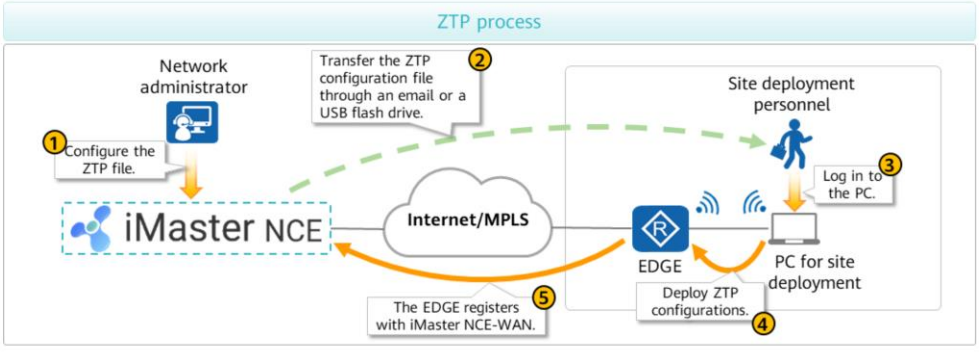
HUAWEI

- ZTP: Multiple ZTP modes are available to enable EDGEs to quickly register with iMaster NCE-WAN.
- Forwarding-control separation enables flexible networking: Each EDGE establishes a management channel with iMaster NCE-WAN through NETCONF, and iMaster NCE-WAN delivers configurations to EDGEs to establish IP overlay tunnels between the EDGEs.
- Application optimization for service controllability and visibility: The service awareness (SA) technology is used to identify applications. TCP Flow Performance Measurement (FPM) and IP FPM technologies are used to implement application-based quality measurement. The IP FPM technology can also be used for link quality measurement. Smart Policy Routing (SPR) technology provides intelligent link switchover based on the application quality.
- Complete security protection system ensures service security: Multiple VPN technologies, such as IPsec and MPLS, are leveraged to provide end-to-end protection. The firewall function is supported to provide comprehensive security protection at the hardware, pipe, and application levels.
- Visualized O&M for quick fault locating: iMaster NCE-WAN collects network-wide data and displays key indicators, helping O&M personnel quickly locate faults.

ZTP | Flexible networking | Application optimization | Security

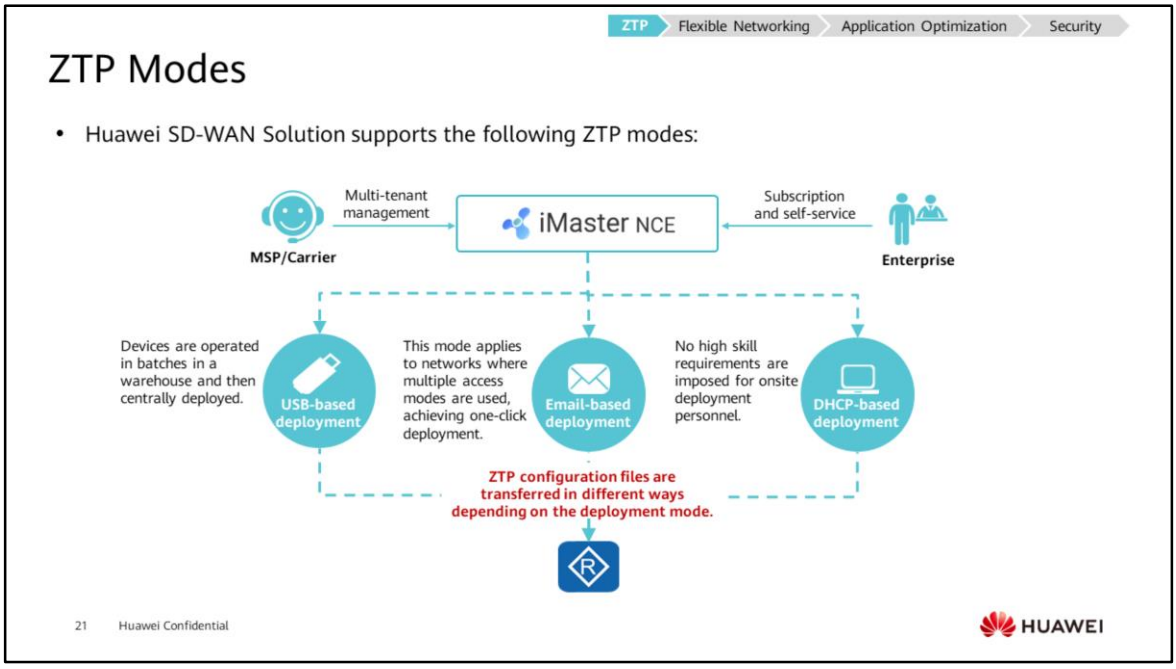
ZTP Overview

- With the development of network technologies such as SDN and cloud computing, a growing number of enterprise networks are using cloud-based management mode, but most sites still need to be deployed by technical engineers onsite, leading to high deployment costs and long deployment periods. To address these problems, Huawei develops the Zero Touch Provisioning (ZTP) function.

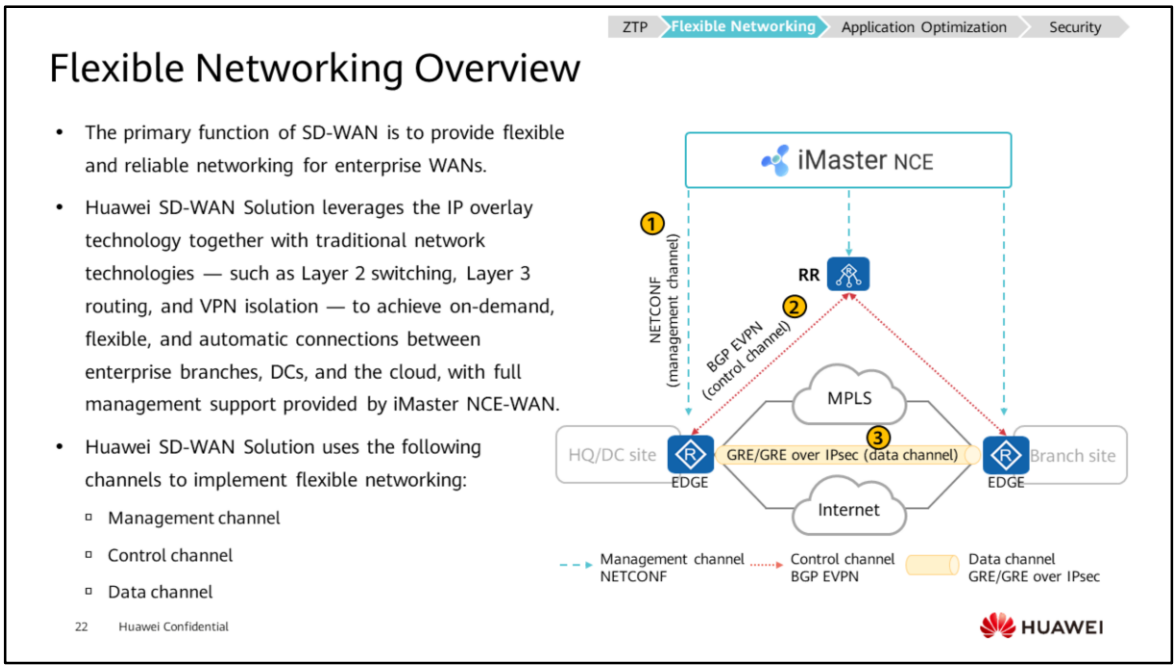


HUAWEI

- Huawei SD-WAN Solution supports the following ZTP modes:
  - Email-based deployment
  - DHCP-based deployment
  - USB-based deployment



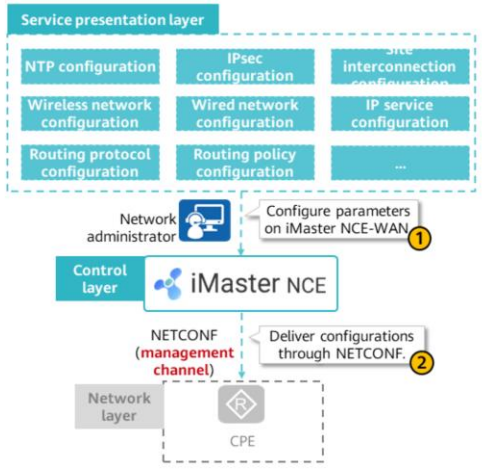
- Files for USB-based, email-based, and DHCP-based deployment can be generated through iMaster NCE-WAN.
- For details about each deployment mode, learn the course *Management and O&M*.



- Management channel:
  - iMaster NCE-WAN sets up management channels with all devices through NETCONF, so as to manage NEs and orchestrate services on the entire network.
- Control channel:
  - EDGEs set up control channels with the RR.
  - The RR centrally controls and distributes service routes between branch sites.
  - The enhanced BGP EVPN protocol is used to implement separate transmission of tenants' VPN route and next hop information, and IPsec SA negotiation.
- Data channel:
  - EDGEs set up data channels with each other.
  - EDGEs forward data based on GRE or GRE over IPsec tunnels. The extended GRE header carries VN IDs to differentiate tenants or departments, thereby transmitting data of multiple VNs over the same tunnel.

Management Channel

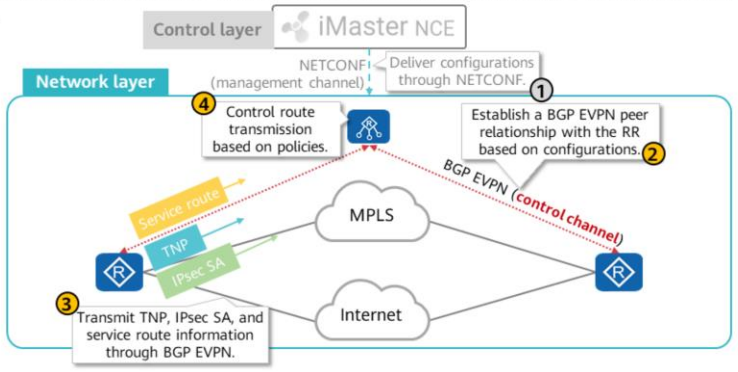
- Huawei iMaster NCE-WAN establishes management channels with CPEs through NETCONF.
- iMaster NCE-WAN delivers configurations through control channels to achieve the following functions:
  - Unified management of CPEs, automatic service delivery, and unified control of overlay networks
  - Application visualization and automatic application optimization
  - Network security services



HUAWEI

Control Channel

- After iMaster NCE-WAN delivers configurations to a CPE through the management channel, the CPE establishes a control channel with an RR through BGP EVPN.
- The control channel is used to transmit transport network port (TNP) information, IPsec SA information, and service routes.
- After the control channel is established, iMaster NCE-WAN controls route transmission and overlay topology establishment by deploying policies on the RR.

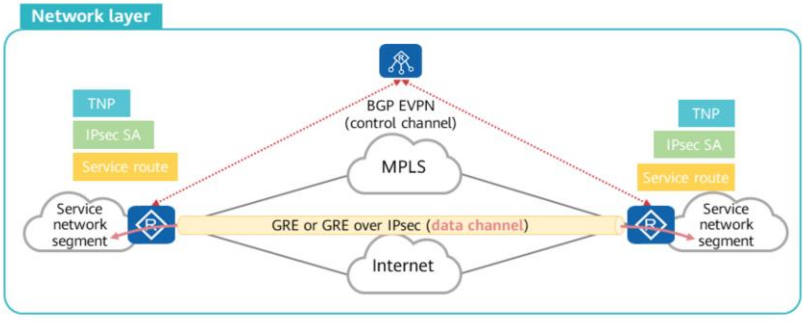


HUAWEI

- A TNP is a WAN port on a CPE used for connecting to a transport network. The key TNP information includes the site ID, CPE router ID, transport network ID, public IP address, private IP address, and tunnel encapsulation mode.

Data Channel

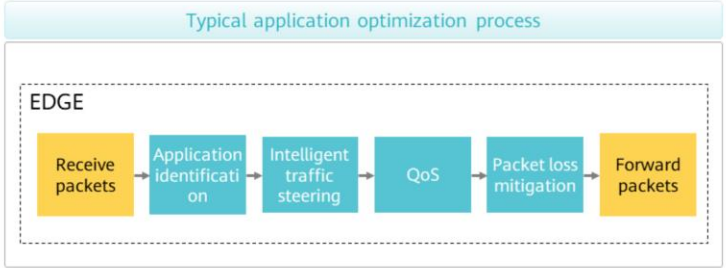
- Huawei SD-WAN Solution uses GRE or GRE over IPsec to establish data channels.
- CPEs establish GRE or GRE over IPsec tunnels based on the TNP and IPsec SA information transferred through BGP EVPN.
- CPEs forward data based on the service routes transferred through BGP EVPN.



HUAWEI

Application Optimization Overview

- The traditional WAN has the following disadvantages that make it unable to meet network requirements:
  - Applications of different values are carried on the same link.
  - Dynamic path selection cannot be implemented when link quality deteriorates.
  - No effective measure is available when link quality deteriorates.
- To address these problems, Huawei SD-WAN Solution provides the application experience optimization solution that offers the following functions:
  - Application identification
  - Intelligent traffic steering
  - QoS
  - Packet loss mitigation

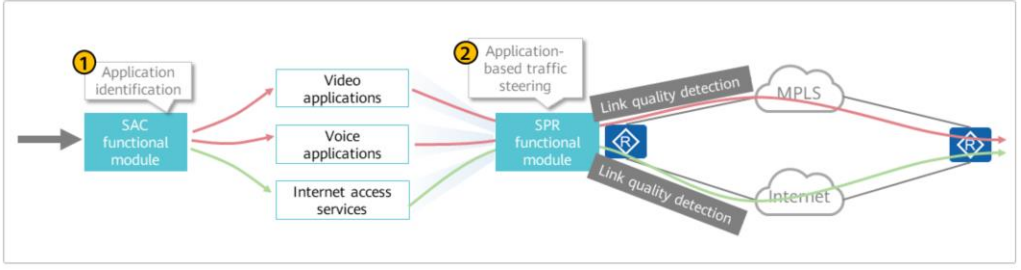


HUAWEI

Application Identification and Traffic Steering

- Huawei SD-WAN Solution uses Smart Application Control (SAC) to identify applications and uses SPR to implement application-based traffic steering.
  - SAC enables a device to identify applications and groups application traffic through SA and first-packet identification (FPI).
  - SPR enables a device to measure the link quality based on link quality detection packets and determine forwarding paths for traffic.

Application identification and traffic steering processes

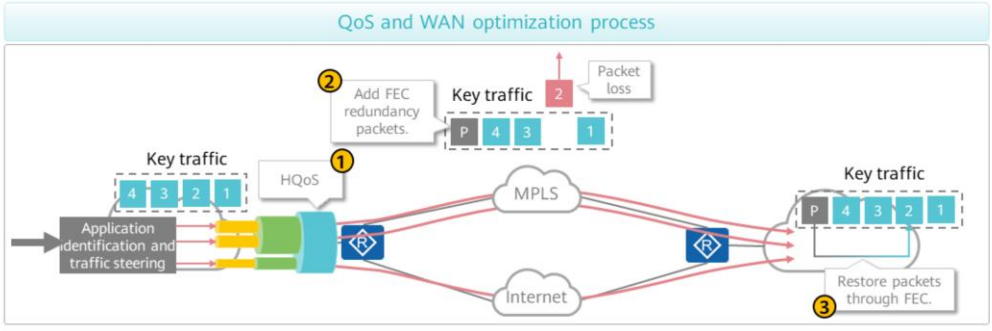


HUAWEI

- For details about SAC and SPR, learn the course *HA Technologies*.

QoS and WAN Optimization

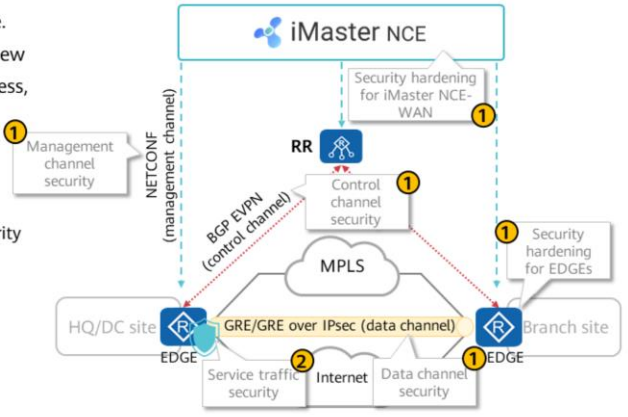
- Huawei SD-WAN Solution uses HQoS for bandwidth control and scheduling, and uses Forward Error Correction (FEC) or Adaptive FEC (A-FEC) for WAN traffic optimization.
  - HQoS implements hierarchical scheduling based on multi-level queues and differentiates services and users, implementing refined QoS.
  - FEC or A-FEC optimization enables the local device to adjust related parameters based on packet loss on the network to generate redundant packets. The peer device then verifies and reassembles the packets.



HUAWEI

Security Overview

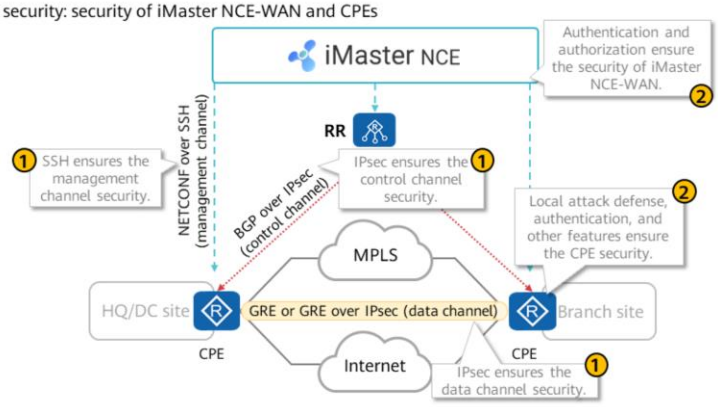
- The advent of SD-WAN stimulates the transition of the enterprise WAN architecture from the traditional closed architecture to an open one. This increases the attack surface and brings new security challenges, such as unauthorized access, data leakage, and network attacks.
- The security of Huawei SD-WAN Solution involves two aspects:
  1. System security: includes inter-component security and component security.
  2. Service security: includes firewall, IPS, and URL filtering.



HUAWEI

SD-WAN System Security Hardening

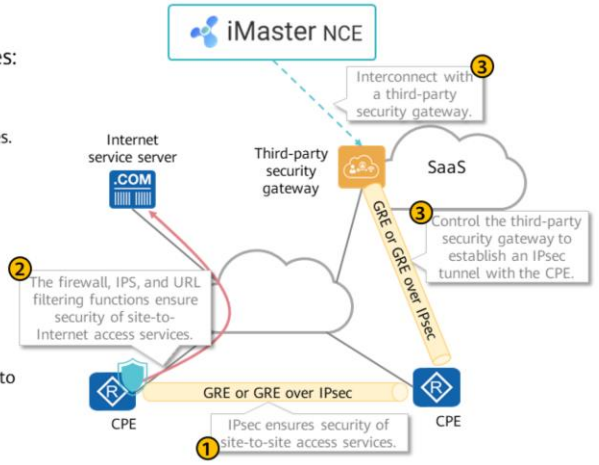
- SD-WAN system security includes:
  - Inter-component security: security of management, control, and forwarding channels
  - Component security: security of iMaster NCE-WAN and CPEs



HUAWEI

SD-WAN Service Security

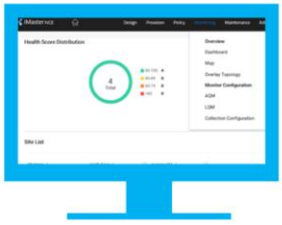
- From the perspective of traffic, SD-WAN services are classified into the following types:
  - Site-to-site access service
    - IPsec ensures security of site-to-site access services.
  - Site-to-Internet access service
    - The built-in firewall, IPS, and URL filtering functions of CPEs ensure the security of site-to-Internet access services.
  - Site-to-cloud access service
    - iMaster NCE-WAN is interconnected with a third-party security gateway and controls this gateway to provide security services.



HUAWEI

Visualized O&M Overview

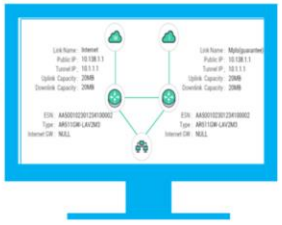
Quickly obtain abnormal traffic



Real-time alarm monitoring

- Customized dashboard (by role or preference)
- Network-wide real-time alarms (minute-level)

Quickly locate faulty devices or sites



Visualized topology status

- Displays topology based on sites and links
- Provides real-time status and performance of sites and links

Optimize WAN investment and configuration policies

45+ customized views

Site, link, application, device, and health status views

- Bandwidth utilization for a site
- Top *N* sites by throughput
- Top *N* applications by traffic
- Link throughput trend...

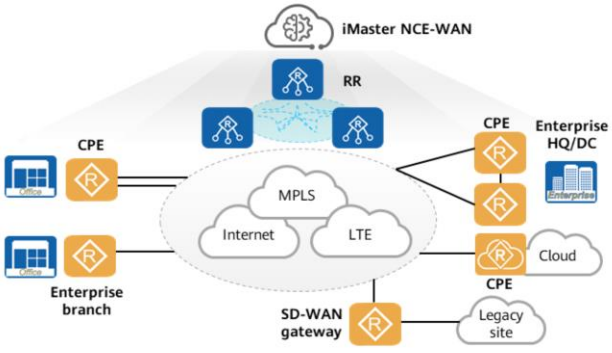
HUAWEI

Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. **Overview of Huawei SD-WAN Solution**
   - Huawei SD-WAN Solution
   - Major Functions of Huawei SD-WAN Solution
   - Application Scenarios of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. Service Design for Huawei SD-WAN Solution
5. Reliability and Security Design for Huawei SD-WAN Solution

HUAWEI

SD-WAN Business Model: Enterprise-Built

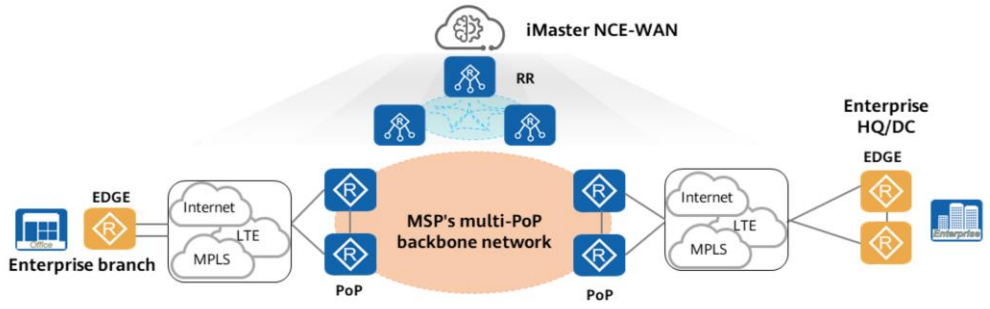


Products Involved	Functions
iMaster NCE-WAN	As the core of the SD-WAN Solution, it centrally manages CPEs, automatically delivers services, and uniformly controls the overlay network.
RR	It is the distributed control component that distributes VPN routes between EDGEs based on VPN topology policies.
EDGE	Egress EDGEs of enterprise branches, headquarters, and DCs.
SD-WAN gateway	An intermediate gateway device that connects an SD-WAN network with a non-SD-WAN network.

- SD-WAN: iMaster NCE-WAN is deployed on the WAN to centrally manage EDGEs and implement ZTP, thereby shortening the service provisioning time. This helps enterprises cope with challenges brought by cloud services and change services on demand.
- Large enterprises with a vast number of branches, such as financial institutions, retail chains, and gas stations, can deploy iMaster NCE-WAN at the headquarters to set up SD-WAN networks and manage SD-WAN services on their own.

HUAWEI

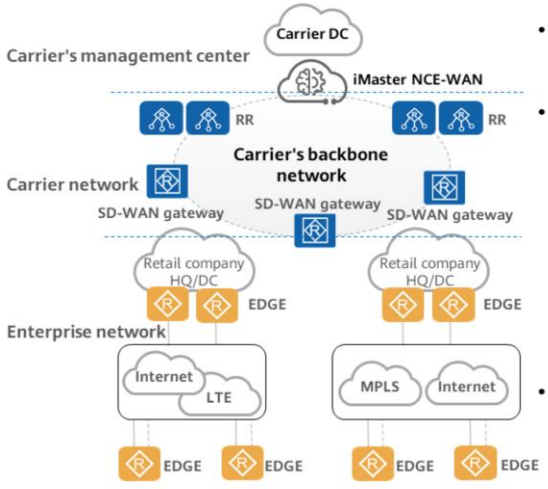
SD-WAN Business Model: MSP Resale



- A managed service provider (MSP) provides a unified SD-WAN controller (iMaster NCE-WAN) to offer SD-WAN services for multiple enterprises. The MSP builds a PoP backbone network, to which an enterprise can connect through the nearest PoP provided by the MSP, thereby achieving high-quality enterprise interconnection. A PoP supports multiple tenants and provides PoP access for multiple enterprises at the same time.
- Enterprises, as tenants, lease SD-WAN services provided by MSPs. An enterprise tenant can manage the SD-WAN services of all sites belonging to it, but it cannot view the SD-WAN services of other tenants.

HUAWEI

SD-WAN Business Model: Carrier Resale



- A carrier provides SD-WAN services for multiple enterprises through iMaster NCE-WAN.
- Enterprises, as tenants, lease SD-WAN services provided by carriers. An enterprise tenant can manage the SD-WAN services of all sites belonging to it, but it cannot view the SD-WAN services of other tenants. Enterprises either manage and control their SD-WAN services based on the tenant permissions assigned by carriers, or they can entrust their SD-WAN services to carriers for management and control.
- SD-WAN gateways are used to implement flexible interconnection and fast compatibility between SD-WAN networks and carriers' legacy backbone networks.

HUAWEI

Section Summary

- SD-WAN solves the problems of difficult management, and expensive construction and maintenance costs in multi-branch interconnection scenarios. It also guarantees the quality of key services.
- Huawei SD-WAN Solution provides the following functions:
  - ZTP, enabling service provisioning to be completed within 1 hour
  - Forwarding-control separation, enabling flexible networking
  - Application optimization, enabling service controllability and visibility
  - Complete security protection system, eliminating security risks
  - Visualized O&M for quick fault locating
- Huawei SD-WAN Solution is offered in three business models: enterprise-built, MSP resale, and carrier resale.

HUAWEI

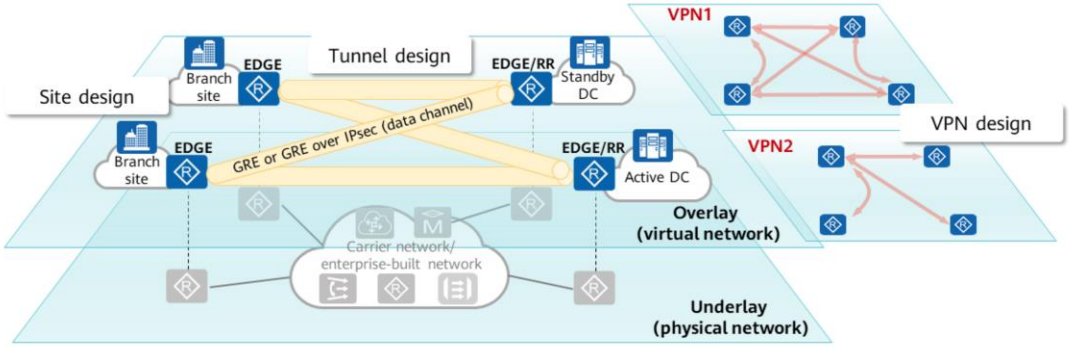
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. **Networking Design for Huawei SD-WAN Solution**
   - SD-WAN Networking Design Overview
   - SD-WAN Site Design
   - SD-WAN Tunnel Design
   - SD-WAN VPN Design
4. Service Design for Huawei SD-WAN Solution
5. Reliability and Security Design for Huawei SD-WAN Solution

HUAWEI

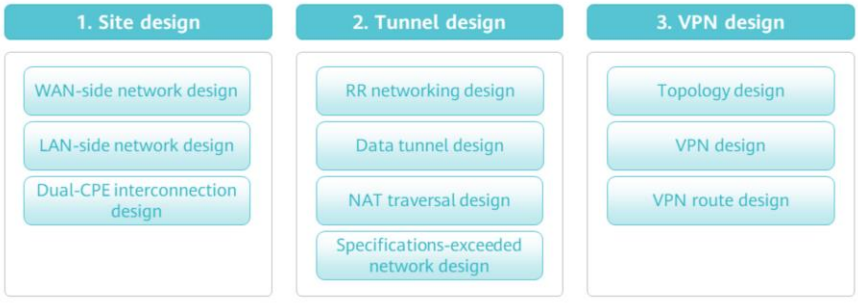
Overview of Networking Design for Huawei SD-WAN Solution

- SD-WAN networking design includes site design, tunnel design, and VPN design.



HUAWEI

Networking Design Process for Huawei SD-WAN Solution



HUAWEI

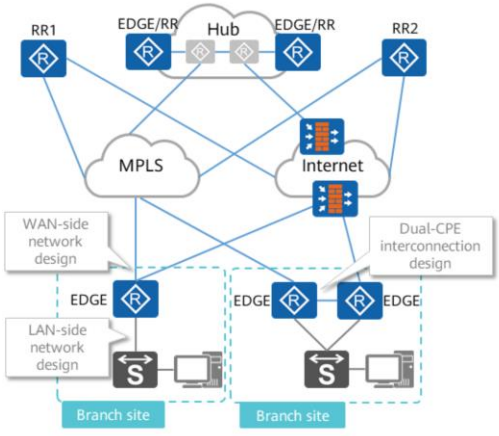
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. **Networking Design for Huawei SD-WAN Solution**
   - SD-WAN Networking Design Overview
   - SD-WAN Site Design
   - SD-WAN Tunnel Design
   - SD-WAN VPN Design
4. Service Design for Huawei SD-WAN Solution
5. Reliability and Security Design for Huawei SD-WAN Solution

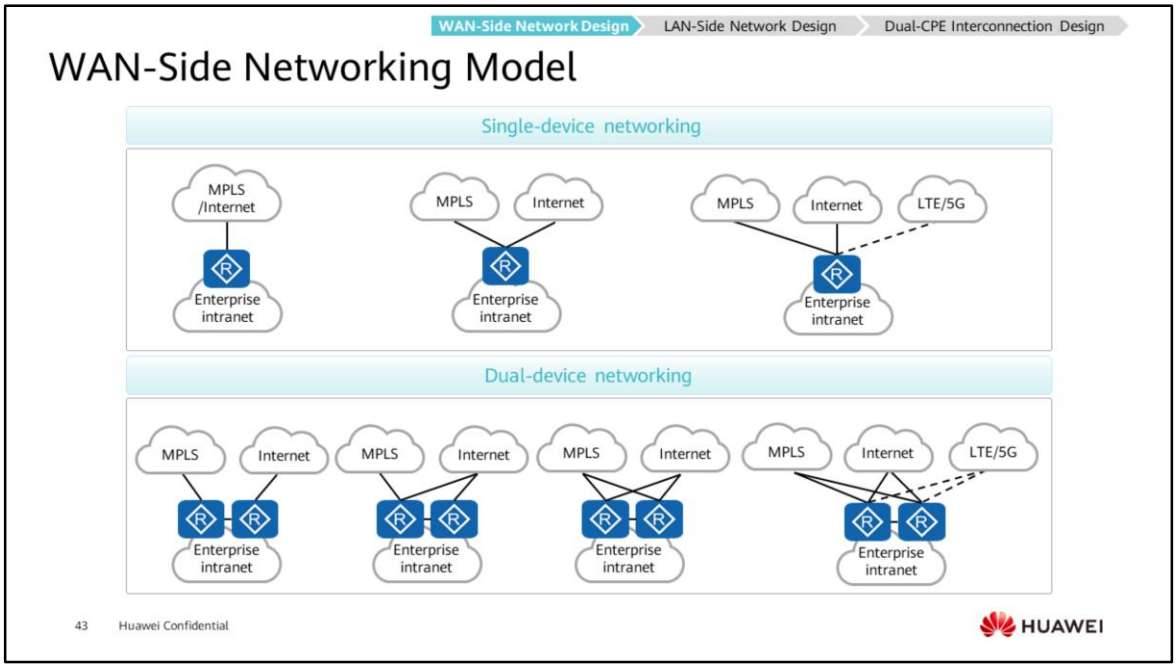
HUAWEI

Site Design Panorama

- SD-WAN sites refer to sites for which SD-WAN is deployed for interconnection, and are managed and monitored by iMaster NCE-WAN.
- Site design generally covers:
  - WAN-side network design
  - LAN-side network design
  - Dual-CPE interconnection design



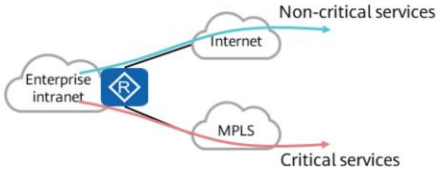
HUAWEI



- An SD-WAN site can be deployed with a single CPE or dual CPEs. For small sites, a single CPE can be deployed. For sites with high reliability requirements, dual CPEs are recommended to provide device-level redundancy.
- It is recommended that multiple links be deployed on the WAN side of a CPE. These links can back up each other, offering link-level reliability. In addition, services can easily select the active and standby links among these WAN-side links with differentiated quality.
- A maximum of 10 WAN links can be deployed for each CPE at an SD-WAN site. During actual deployments, to enhance reliability and facilitate O&M, it is recommended that a maximum of three WAN links be deployed for a single CPE at a site, and a maximum of six WAN links be deployed for a site with two CPEs.

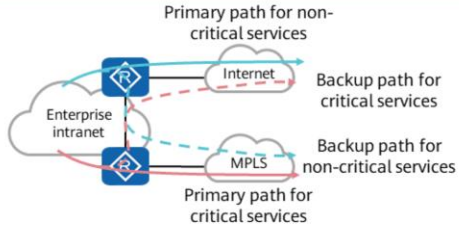
Typical WAN-Side Networking

Single-CPE multi-link networking



- For small and midsize sites or branch sites with low reliability requirements, only one CPE needs to be deployed. Multiple WAN links can be deployed and enterprise application traffic of different importance is transmitted over different links.
- Typically, one MPLS link and one Internet link are deployed on the WAN side. The MPLS link has guaranteed SLA and is used to carry enterprises' critical applications. Non-critical applications are carried on the Internet link.

Dual-CPE multi-link networking

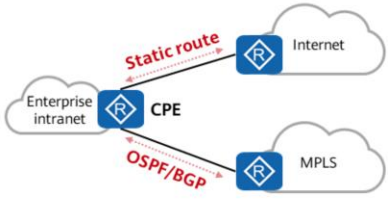


- In large sites or branch sites with high reliability requirements, dual CPEs are generally deployed. To ensure service transmission reliability, at least two WAN links are deployed for each CPE.
- Two CPEs are deployed at a branch site. The MPLS link is deployed for one CPE, and the Internet link is deployed for the other CPE. To ensure reliability, an interlink is deployed between the two CPEs to prevent data forwarding failures when a CPE fails.

HUAWEI

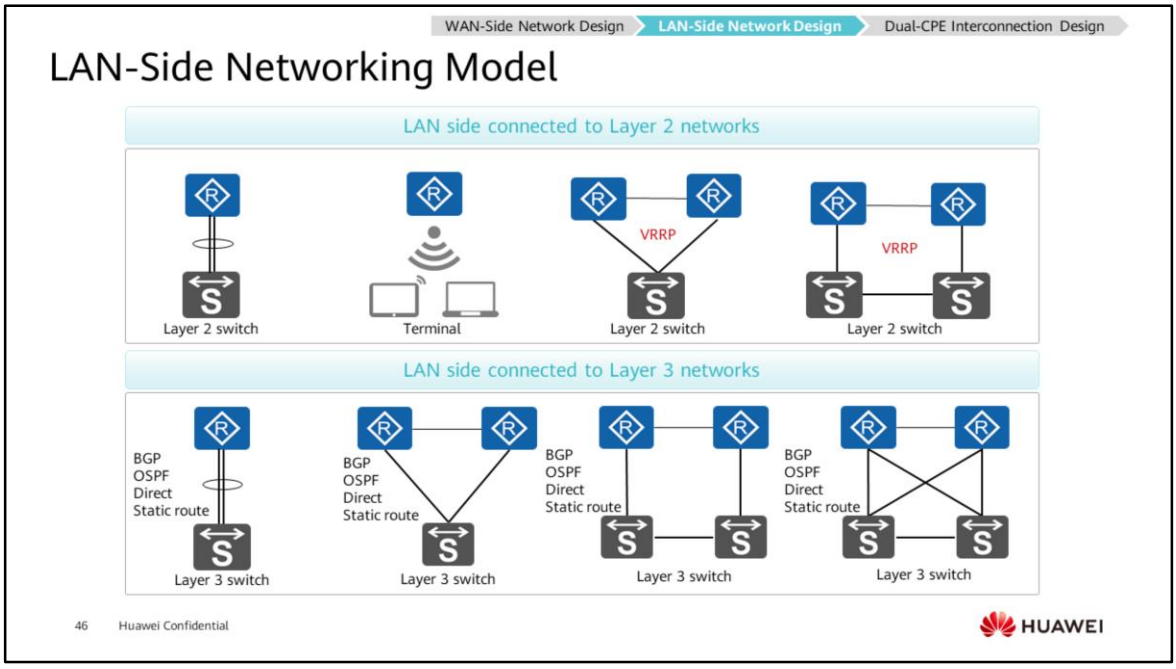
WAN-Side Underlay Route Design

BGP used to access WANs

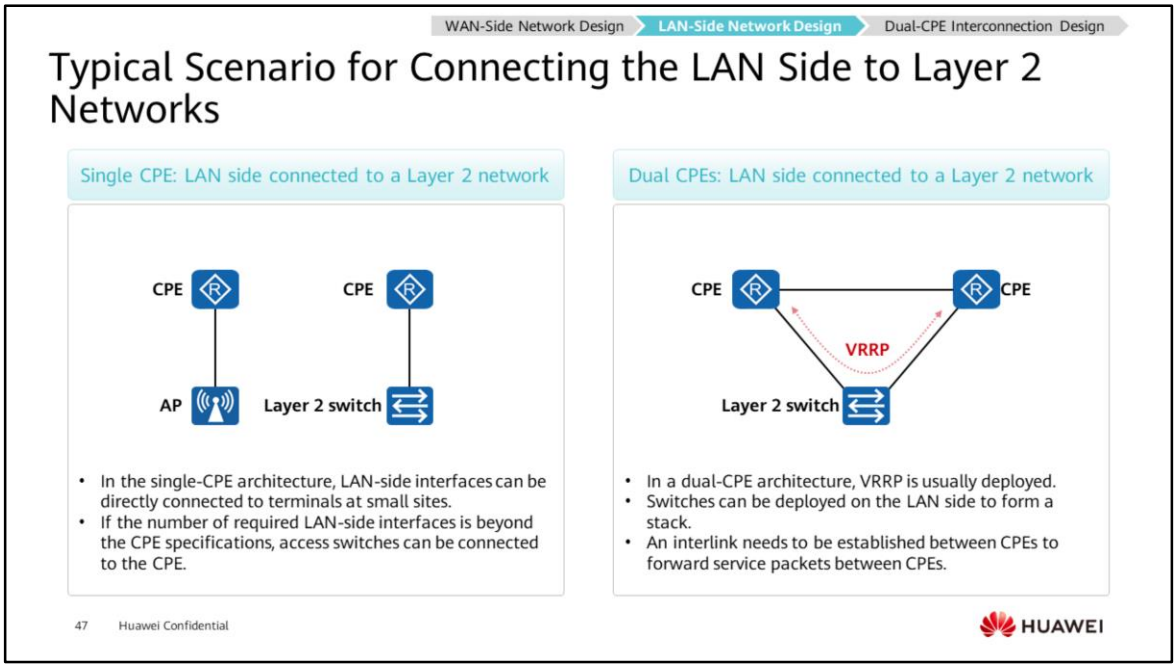


- WAN interfaces of CPEs support OSPF, EBGP, and static routing protocols. The routing protocol to be used must be the same as that of the LAN-side network device.
- On the live network, if static routes are used, services may be interrupted because WAN network faults on indirect links cannot be detected.

HUAWEI



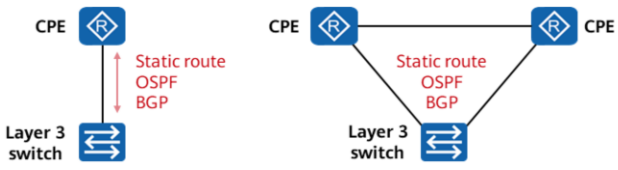
- LAN side connected to Layer 2 networks:
  - At small sites with a simple intranet structure, and CPEs typically connect to the intranet of the site at Layer 2.
- LAN side connected to Layer 3 networks:
  - In Layer 3 interconnection scenarios, SD-WAN routers support the direct connection, dual-homing, and partial-ring networking.
  - In such scenarios, routing protocols simply need to be configured for CPEs based on the requirements of the LAN device.



- In the single-CPE structure, the LAN connection is simple.
  - For small sites, for example, SOHO sites, LAN-side interfaces can be directly connected to terminals at the sites.
  - If the number of required LAN-side interfaces is beyond the CPE specifications, access switches can be connected to the CPE. In this case, access switches can be connected to the CPE in one-armed mode.
- In the dual-CPE architecture, VRRP is usually configured for the CPEs to prevent the dual-CPE architecture from affecting the LAN side.
  - Multiple switches can be deployed on the LAN side to form a stack. If two CPEs are deployed at a site, they can be interconnected directly or through the LAN.
  - CPEs can be directly connected through an Eth-Trunk link. In addition, an interlink needs to be established between the active and standby CPEs to forward service packets between the CPEs.

Typical Scenario for Connecting the LAN Side to Layer 3 Networks

Single or dual CPEs: LAN side connected to a Layer 3 network



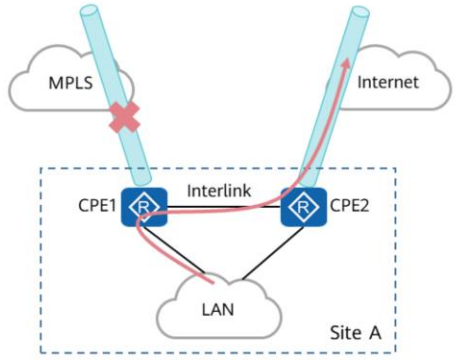
- In the Layer 3 interconnection scenario, if only one CPE is deployed, the network structure is simple. In such a scenario, only the routing protocol needs to be configured on the LAN side based on requirements of LAN-side devices.

HUAWEI

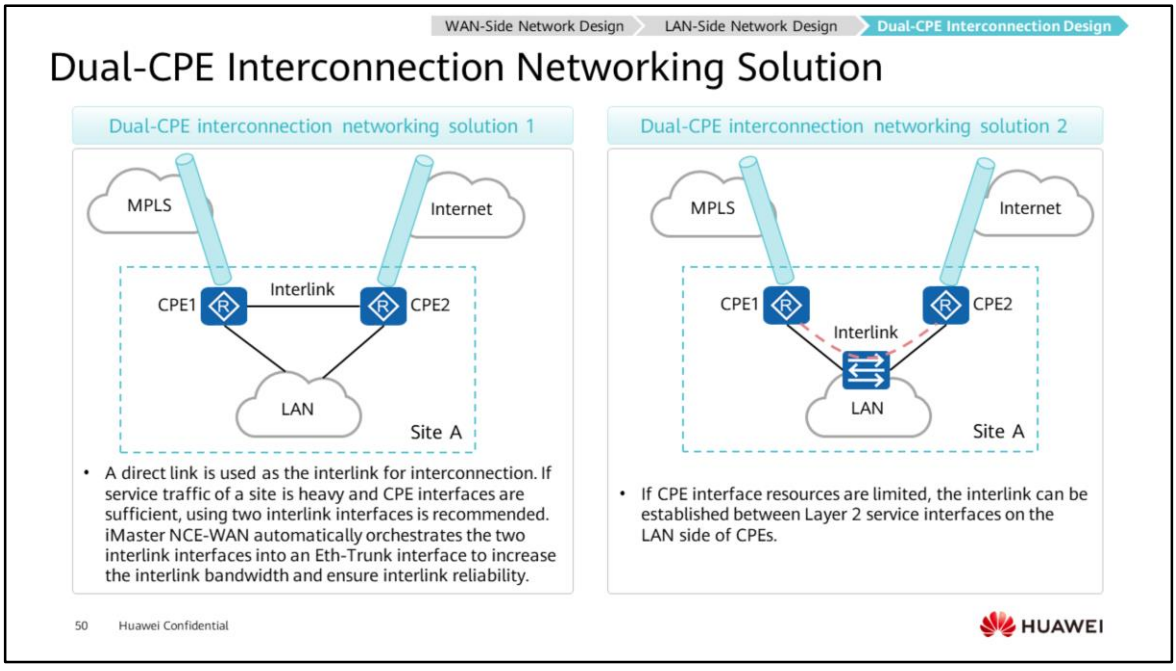
- For large enterprise sites, the network structure is complex as Layer 3 core devices are deployed on the network. Therefore, egress routers must support interconnection with Layer 3 devices. They can be connected to a Layer 3 network directly or in dual-homing mode. BGP, OSPF, and static routes are supported.
- In the Layer 3 interconnection scenario, if only one CPE is deployed, the network structure is simple. In such a scenario, only the routing protocol needs to be configured on the LAN side based on requirements of LAN-side devices.

Dual-CPE Interconnection Networking Model

- In dual-CPE dual-link networking, if an uplink fails, the interlink between CPEs is used to transmit data.
- In addition to exchanging routes between two CPEs at a site, the interlink is also used to synchronize key information between the two CPEs, including link information, tunnel connection information, SLA information about tunnel connections, IPsec SA, and application identification result.
- With such information, resources on the two CPEs at each SD-WAN site can be regarded as a whole. Then services such as cross-device traffic steering can be implemented.



HUAWEI



- Solution 1 is recommended. In this solution, the interlink and service links are independent of each other. When WAN-side links are adjusted, the interlink will not be affected, and the service flow direction is clear.

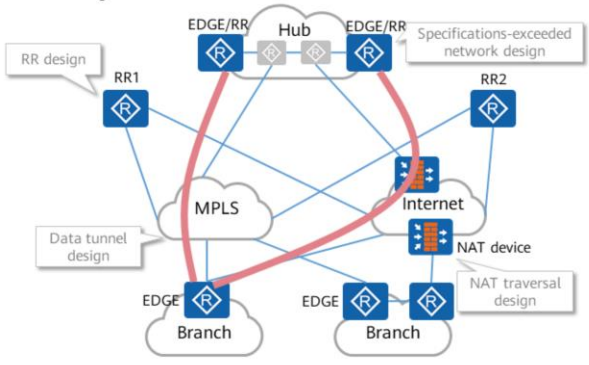
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. **Networking Design for Huawei SD-WAN Solution**
   - SD-WAN Networking Design Overview
   - SD-WAN Site Design
   - SD-WAN Tunnel Design
   - SD-WAN VPN Design
4. Service Design for Huawei SD-WAN Solution
5. Reliability and Security Design for Huawei SD-WAN Solution

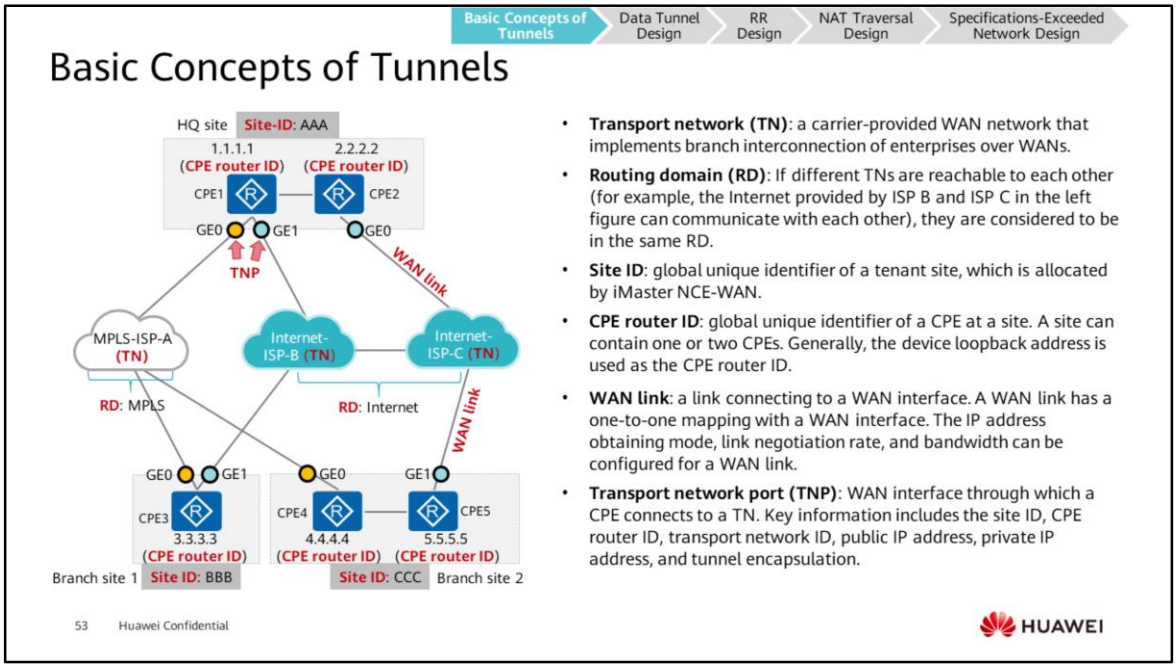
HUAWEI

Tunnel Design Overview

- Tunnels in the SD-WAN Solution can be classified into: management tunnel, control tunnel, and data tunnel. Control tunnels are established between RRs and EDGEs, and data tunnels are established between EDGEs. Data tunnels carry services.
- Tunnel design generally covers:
  - Data tunnel design
  - RR design
  - NAT traversal design
  - Specifications-exceeded network design



HUAWEI

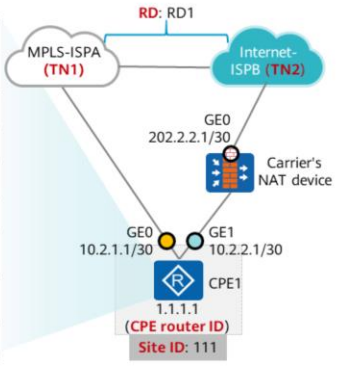


- The TN and RD are used to set up overlay tunnels in enumerated mode.
- The site ID is used as the next hop of a routing entry.
- The CPE router ID is used to establish BGP peer relationships between different sites.
- TNPs are used to establish tunnels.

TNP

- A TNP mainly describes WAN link information of a site and is mainly used to establish control and data tunnels.
- The main information about a TNP is as follows:

TNP Information	Example		Description
Site ID	111		Site of the TNP.
CPE router ID	1.1.1.1		CPE to which the TNP belongs.
TNP ID	1	2	TNP ID.
Transport network	TN1	TN2	TN to which the TNP belongs.
Routing Domain	RD1	RD1	RD to which the TNP belongs.
Public IP	10.2.1.1	202.2.2.1	WAN-side public IP address (post-NAT).
Private IP	10.2.1.1	10.2.2.1	WAN-side private IP address (pre-NAT).
Encapsulation	GRE	GRE	Encapsulation mode.

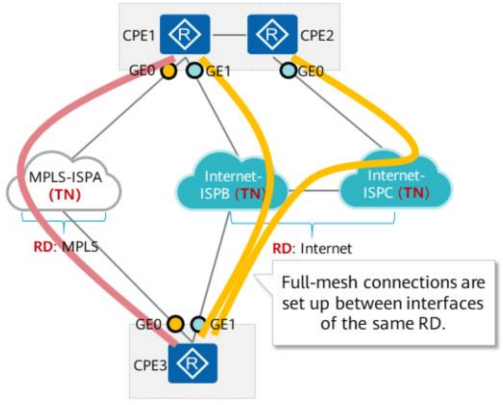


HUAWEI

- TNs and RDs are mainly used to enumerate tunnels.
  - Tunnel enumeration: All tunnels that can be established are enumerated.
- Router IDs are mainly used to establish control channels.
- A site ID is used as the next hop for data forwarding.
- A TNP ID can be considered as the interface number.
- The public and private IP addresses are used as the source or destination IP addresses of control and data channels.
  - Some CPEs are deployed behind the NAT device. To establish data channels between CPEs, you need to know the post-NAT public IP address.
    - CPEs typically use the Session Traversal Utilities for NAT (STUN) technology to detect public IP addresses.

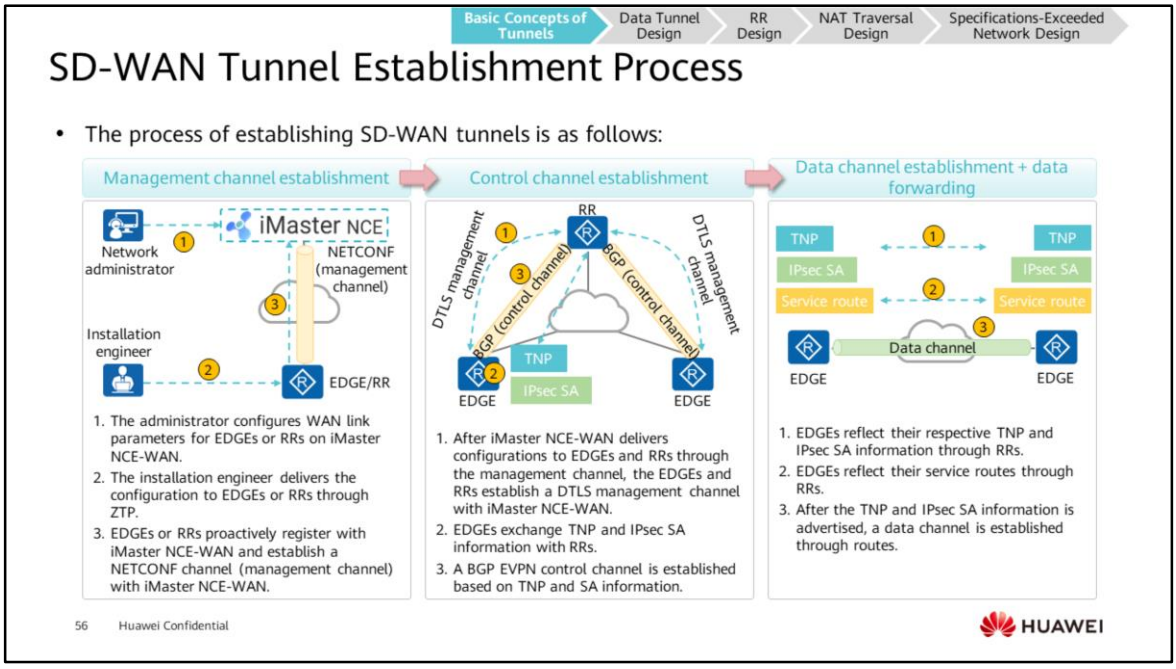
Tunnel Enumeration

- CPEs enumerate tunnels based on the **RD** and **TN** in TNP information.
- Tunnel enumeration ensures SD-WAN network reliability and guarantees service quality.
- Tunnel enumeration rules:
  - Connections can be enumerated only between TNPs in the same RD. Full-mesh connections are set up between interfaces of the same RD.
  - Connections cannot be enumerated between TNPs in different RDs.



HUAWEI

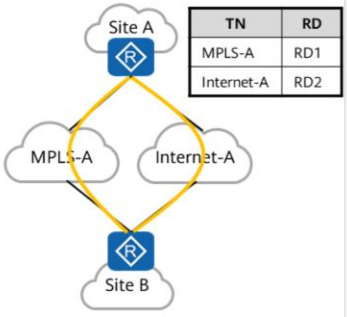
- Data tunnels are enumerated before being established to ensure that all available data tunnels are established.
- Tunnels can be enumerated only when the following conditions are met:
  - The CPE has learned service routes of the peer site.
  - The CPE has learned the TNP information of the peer site.



- The management channel is used to establish control channels and deliver basic configurations.
- Control channels are used to establish data channels.
  - Datagram Transport Layer Security (DTLS) is a security protocol for the packet transport layer, which is used to ensure TCP data security.
- Data channels are used to transfer user data.
- TNP information is exchanged twice:
  - During the control channel establishment phase, TNP information is exchanged to exchange information about channel establishment between RRs and EDGEs.
  - During the data channel establishment phase, TNP information is exchanged to exchange information about channel establishment between EDGEs.

Tunnel Networking Model

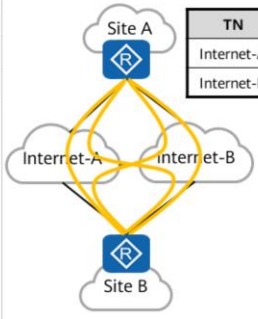
Different RDs



TN	RD
MPLS-A	RD1
Internet-A	RD2

- Sites A and B are connected through two links and are isolated on the underlay network. Different RDs are planned. Two tunnels are established between the sites.

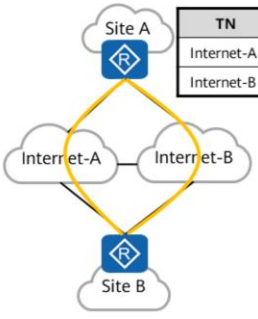
Different TNs, same RD



TN	RD
Internet-A	RD1
Internet-B	RD1

- Sites A and B are connected through two links and can communicate with each other on the underlay network. They are planned in the same RD and set up a full-mesh network.

Network in the same RD, different RDs created logically



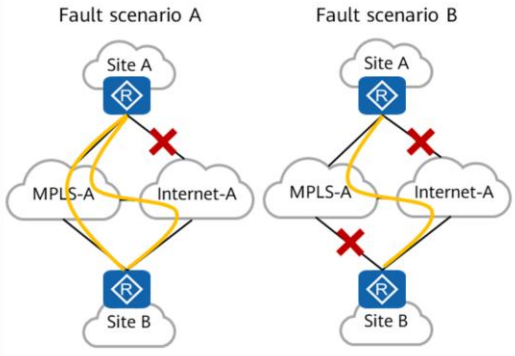
TN	RD
Internet-A	RD1
Internet-B	RD2

- Sites A and B are connected through two links and can communicate with each other on the underlay network. They are planned in different RDs. Two tunnels are established between the sites.

HUAWEI

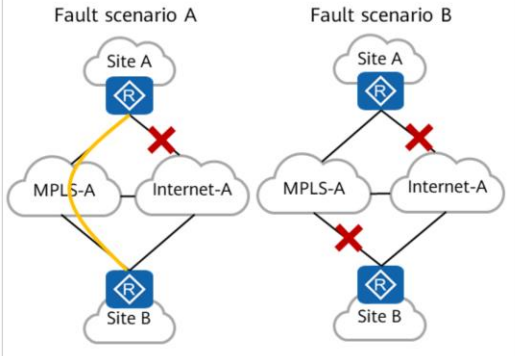
Tunnel Networking Reliability

Different TNs, same RD



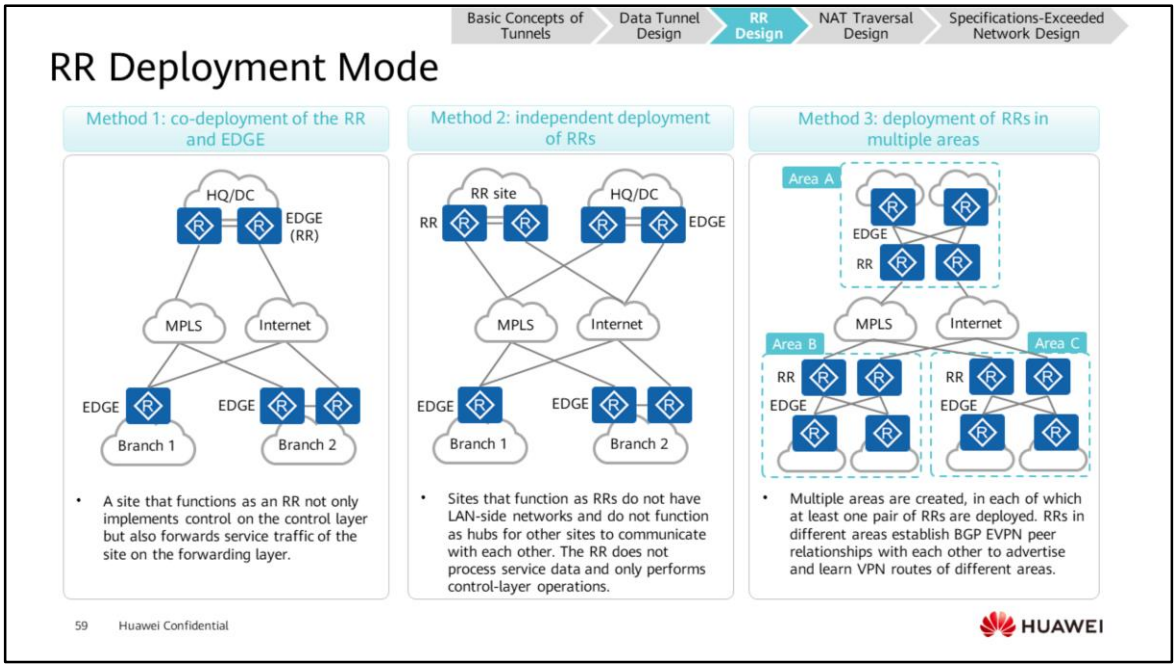
- **Same RD for different TNs**: The networking reliability is enhanced, but more virtual connection resources of the device are consumed.

Different RDs and TNs



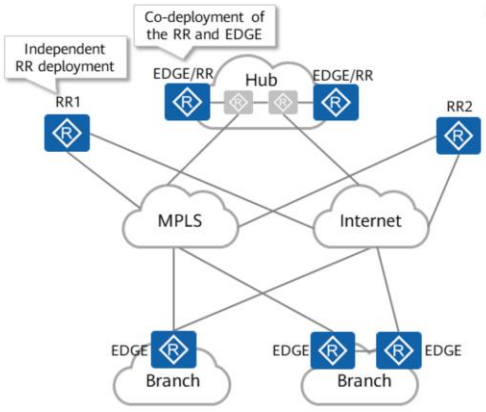
- **Different RDs for different TNs**: The number of connections is reduced, and the networking scale is expanded. However, the networking is less reliable.

HUAWEI



- RR deployment suggestions are as follows:
  - To prevent site network adjustment from affecting the stability of RRs, you are advised to use method 2, that is, independent deployment of RRs.
  - Use high-performance devices as RRs. For details about the devices that can function as RRs, see the specifications list.
  - Configure a public IP address for an RR, or deploy a NAT device before the RR. Only 1:1 static NAT is supported.
  - An EDGE site can connect to a maximum of two RR sites. Two EDGEs can be deployed at each RR site deployed for a tenant. If there are a large number of EDGEs, multiple RRs can be deployed, and each RR serves some EDGE sites.
  - When one EDGE is connected to two RR sites, the EDGE establishes a BGP connection with each RR at the RR sites.
  - If a branch site has a standby link, for example, the branch or the RR has a standby link, and the active link is normal, no control channel is established for the standby link. When all active links from the branch site to the RR are down, the standby link is involved in the establishment of control channels.

RR Deployment Principles



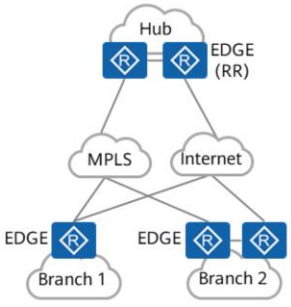
- Considering the importance of RRs, follow the principles below when deploying RRs:
  - Redundancy must be implemented for RRs. That is, at least two RRs must be deployed on the live network. This prevents network-wide services from being interrupted because the only RR deployed on the network fails.
  - An edge site must be connected to two RRs to implement egress backup.
  - It is recommended that RRs be independently deployed to ensure stability.
  - If RRs cannot be deployed independently, select routers at core positions as RRs and ensure that the device performance can meet the requirements.
  - Use the RR models recommended in the specifications list.

HUAWEI

- A EDGE can be connected to a maximum of two RR sites (four RRs).
- For small networks (for example, a network with fewer than 50 sites), RRs and hubs can be deployed in co-located mode.
- RRs require strong BGP connection capabilities (number of BGP peers), large number of EVPN connections, and high route reflection capability and efficiency. In actual deployments, select the RR models recommended in the specifications list, for example, AR6300/AR6280.

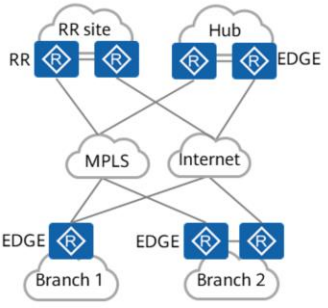
Typical RR Networking: Enterprise-Built

Co-deployment of the RR and EDGE



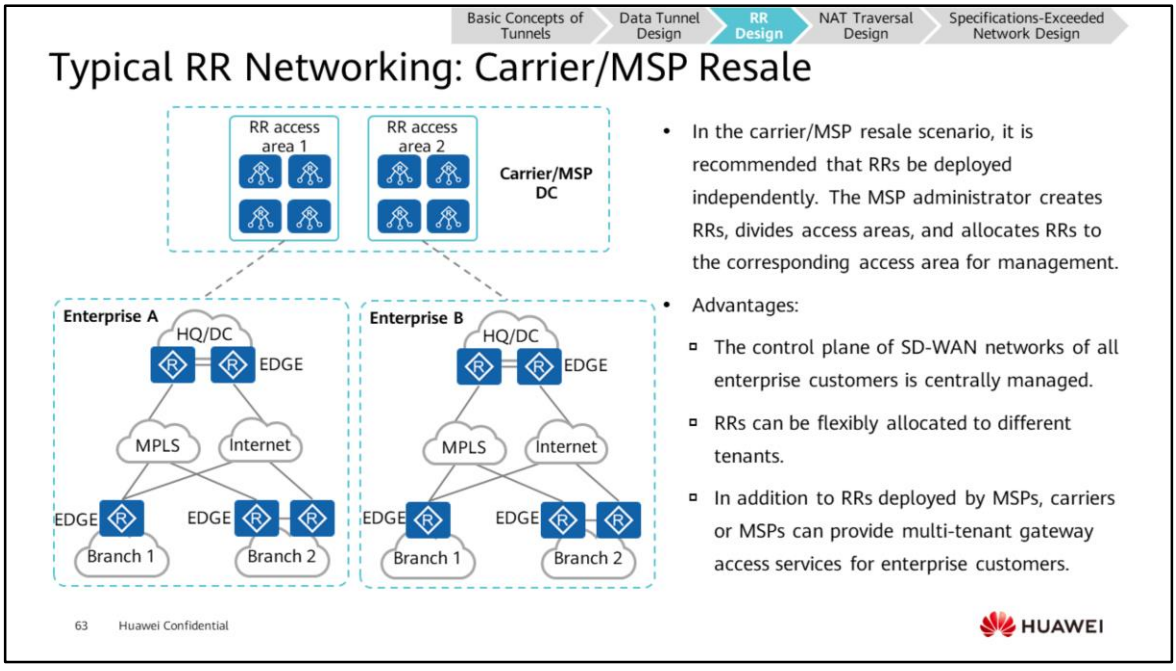
- Small and midsize enterprises have a small number of branch sites. If the traffic between branches and the HQ/DC is not heavy, it is recommended that the RR be co-deployed with the EDGE. The hub site functions as the RR to carry the control plane.

Independent RR deployment



- Large enterprises have a large number of sites. If these enterprises have high network reliability requirements, independent RR deployment is recommended.

HUAWEI

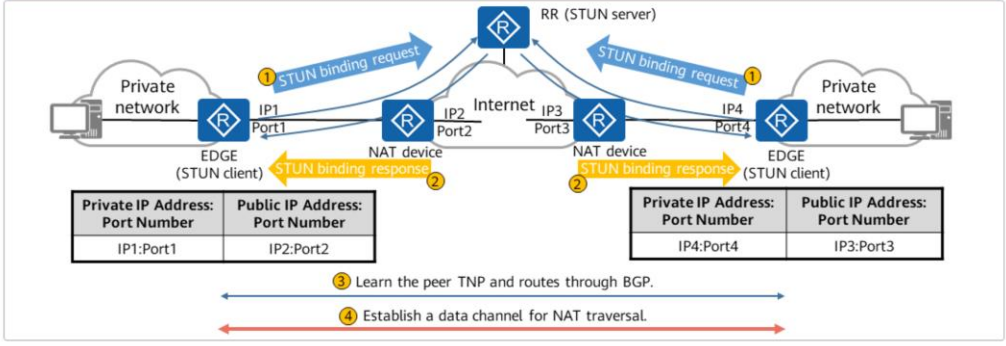


- An MSP administrator deploys independent RRs as a service provided by carriers or MSPs for enterprise users to access. Two RR service modes are available: sharing and exclusive. In sharing mode, one RR is shared by multiple tenants. In exclusive mode, one RR is exclusively used by a tenant.
    - MSP RR sharing by multiple tenants: For small and midsize enterprises, it is recommended that one RR be shared by multiple tenants, if there are a small number of sites. Sites of different enterprise tenants are connected to the same multi-tenant RR, which carries the control plane.
    - MSP RR exclusive to a tenant: Carriers can specify an independent RR in an access area for a large enterprise with a large number of sites to use exclusively. Based on actual scenarios, carriers or MSPs can deploy RRs in co-deployment mode, independently, and based on areas for tenants.

NAT Traversal

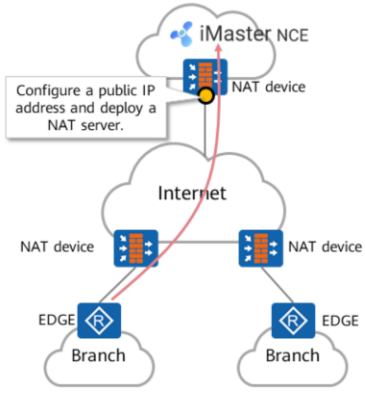
- On the live network, some enterprise branches use home broadband to access the network.
  - In this case, traffic needs to pass through the NAT device. Therefore, tunnels cannot be directly established for such enterprise branches.
- In SD-WAN scenarios, NAT traversal technology is required for interconnection between such enterprise branches.

Establishing a data channel for NAT traversal



HUAWEI

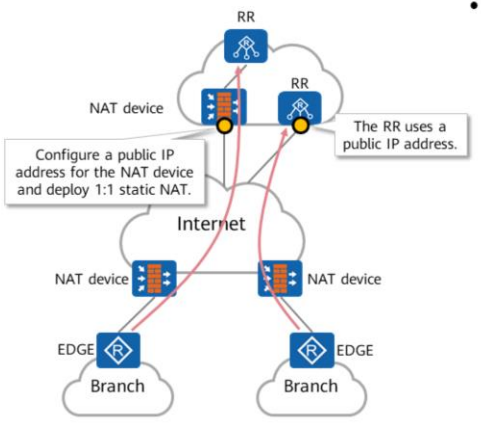
NAT Traversal: Connection Between EDGEs and iMaster NCE-WAN



- A EDGE proactively sends a registration request to iMaster NCE-WAN. To implement this, the southbound IP address of iMaster NCE-WAN must be reachable.
- On the live network, iMaster NCE-WAN may be deployed behind a NAT device. In this case, the NAT Server needs to be deployed on the NAT device and the following ports need to be mapped:
  - 10020: southbound port, which is used by devices to register with iMaster NCE-WAN.
  - 10031: southbound port, which is used by devices to connect to iMaster NCE-WAN through HTTP.
  - 18008: northbound port, which is used for web login.
  - 80: northbound port, which is used for web login.
  - 18018: northbound port, which is used by PCs to upload files to iMaster NCE-WAN.
  - 18021: file server port, which is used by iMaster NCE-WAN to update files.
- The NAT Server can be deployed. Therefore, NAT traversal is not involved in this scenario.

HUAWEI

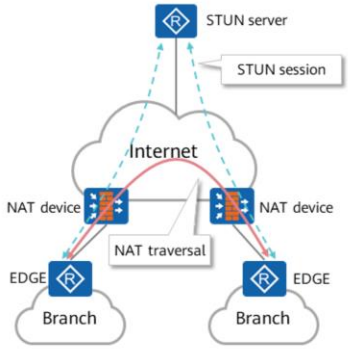
NAT Traversal: Connection Between EDGEs and RRs



- An RR must have a public IP address or be configured with 1:1 static NAT. Therefore, NAT traversal is not involved in communication between CPEs and RRs.

HUAWEI

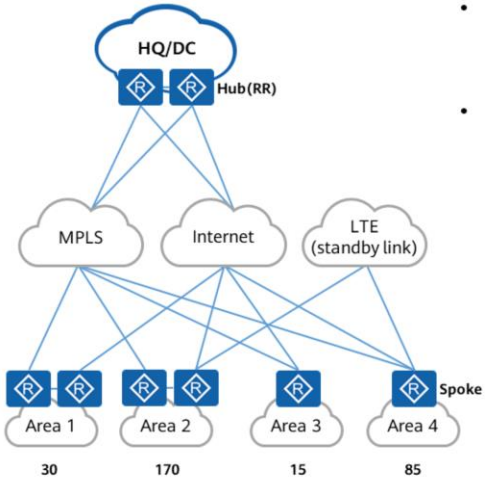
NAT Traversal: Connection Between EDGEs



- A data channel is established between EDGEs, and traffic between sites is transmitted through the data channel. If a EDGE is deployed on the private network behind the NAT device, NAT traversal is required for communication with another EDGE, especially in scenarios where two EDGEs at two sites are deployed on the private network behind the NAT device.

HUAWEI

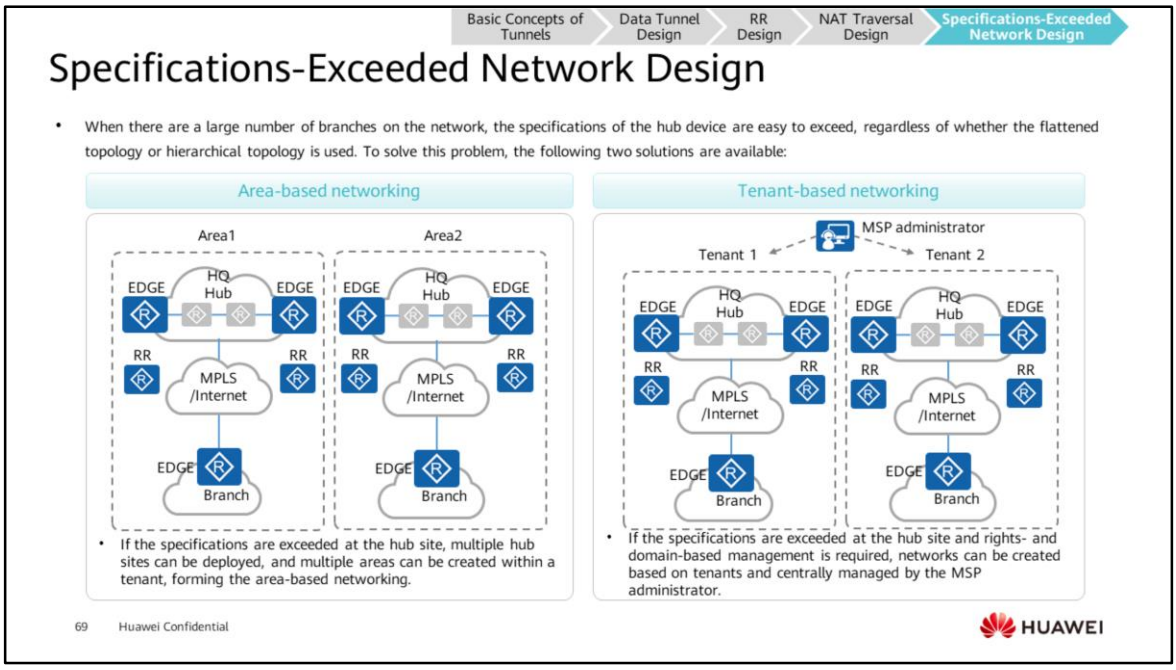
Networking Specification Calculation



- When planning a network, consider the device performance specifications. In the SD-WAN Solution, the number of BGP peers, number of tunnels, and bandwidth also need to be considered.
- Networking specification calculation
  - Number of BGP peers of hubs/RRs = Number of sites with dual gateways x 2 + Number of sites with a single gateway
    - Number of BGP peers = 200 x 2 + 100 = 500
  - Number of hub tunnels = Total number of tunnels of each gateway at a hub site
    - The two hubs at a dual-hub site share the tunnel specification. For example, if a single EDGE supports 1000 tunnels, the two hubs at a dual-hub site share the tunnel specification.
    - Number of hub tunnels = 300 x 4 = 1200
  - Network bandwidth of a hub site = Number of branch sites x Bandwidth required by sites
    - Total network bandwidth required by the hub = 300 x 10 Mbps = 3 Gbps
  - iMaster NCE-WAN can manage a large number of sites. In most cases, you do not need to pay attention to the networking specification calculation.

HUAWEI

- Different device models have different BGP peer specifications and tunnel specifications.
- For details about the product specifications, see the product documentation.



- Area-based networking:
  - Multiple areas are created under a tenant, and multiple hub sites are deployed in the HQ/DC. Each area is associated with one or two hub sites.
  - Branch sites are added to corresponding hub sites by area.
  - RRs can be independently deployed. Each pair of RRs are associated with sites in an area.
  - Traffic between inter-area sites is transmitted through the LAN side of the hub.
- Tenant-based networking:
  - The MSP administrator creates multiple tenants, and multiple hub sites are deployed in the HQ/DC. Each tenant is associated with one or two hub sites.
  - Branch sites are grouped by their geographical areas and are added to different tenants.
  - RRs can be independently deployed. Each pair of RRs are associated with sites in an area.
  - Traffic between inter-area sites is transmitted through the LAN side of the hub.

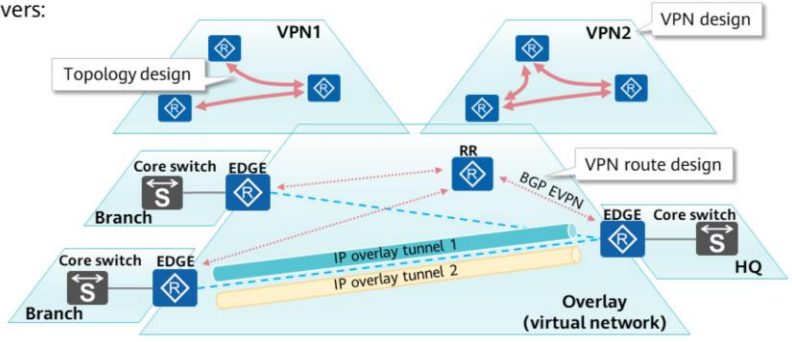
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. **Networking Design for Huawei SD-WAN Solution**
   - SD-WAN Networking Design Overview
   - SD-WAN Site Design
   - SD-WAN Tunnel Design
   - SD-WAN VPN Design
4. Service Design for Huawei SD-WAN Solution
5. Reliability and Security Design for Huawei SD-WAN Solution

HUAWEI

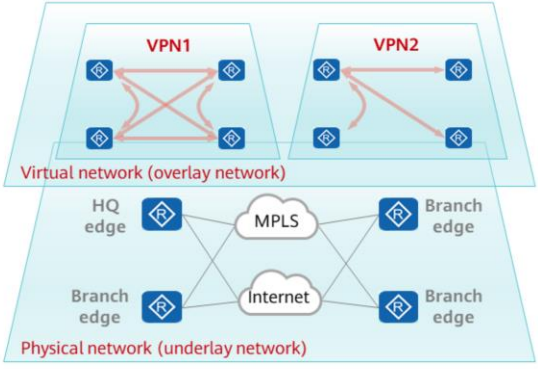
VPN Design Panorama

- The SD-WAN Solution uses VPNs to isolate services for multiple departments of a single tenant. In this case, the network of each department under a tenant is an independent service VPN.
- VPN design generally covers:
  - VPN design
  - Topology design
  - VPN route design



HUAWEI

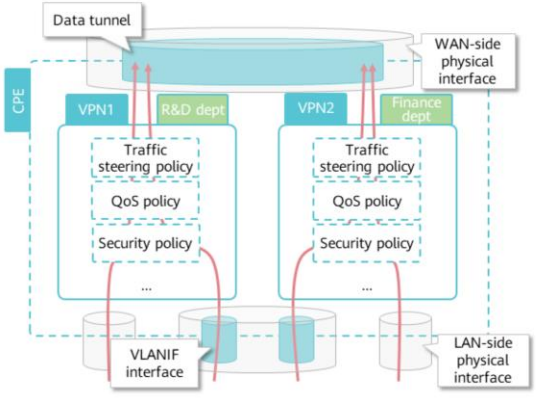
SD-WAN VPN



- VPN is short for virtual private network on SD-WAN. Each VPN is an independent IP Layer 3 private network. Multiple VPNs are logically isolated from each other, from the tunnel established at a site to the EDGE at the site. Therefore, these VPNs cannot directly communicate with each other.
  - Each VPN can use an independent topology model, including hub-spoke, full-mesh, partial-mesh, and hierarchical networking.
  - Service policies, such as traffic steering policies and QoS policies, can be independently configured for each VPN.

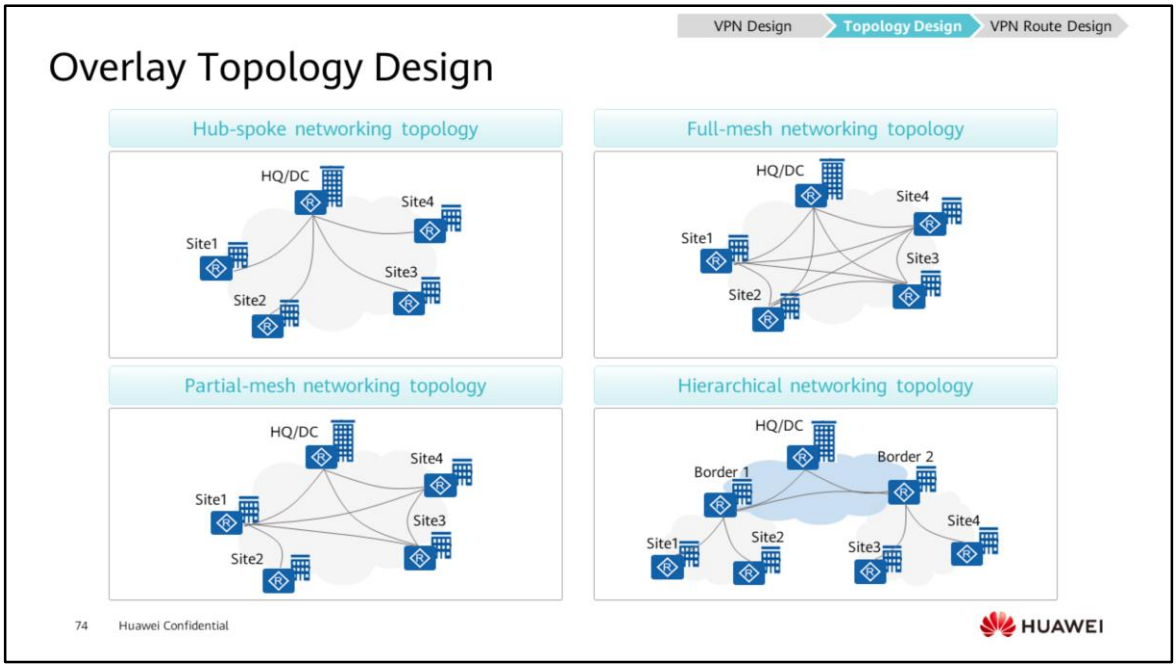
HUAWEI

VPN Planning and Design



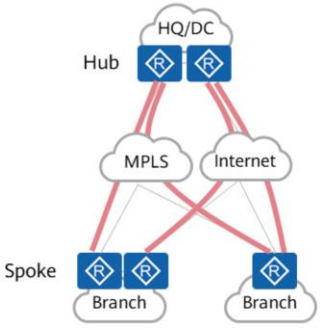
- VPN planning suggestions:
  - Determine the number of departments, for example, R&D department, finance department, and marketing department, whose services need to be isolated based on service isolation requirements.
  - Assign LAN-side physical interfaces and VLANs to different departments. One department can have an independent LAN-side physical interface, or all departments share one physical interface and are isolated through VLANs.
  - Plan initial policies for different departments, including traffic steering policies, Internet access policies, QoS policies, legacy site access policies, and URL, IPS, and firewall policies.

HUAWEI



- Hub-spoke:
  - Generally, the enterprise HQ/DC functions as a hub site, and enterprise branches function as spoke sites. Server applications deployed in the HQ/DC are accessed through the WAN in a centralized manner.
- Full-mesh:
  - In the topology, different branches can directly communicate with each other, without the need to divert traffic through intermediate nodes.
- Partial-mesh:
  - Partial-mesh can be considered as a special type of full-mesh networking. If an underlay network is available between two sites, traffic can be directly transmitted between the two sites. Otherwise, the two sites communicate with each other through a redirect site.
- Hierarchical networking:
  - The hierarchical networking model can be considered as the combination of the single-layer networking model. A WAN is divided into multiple areas, which are interconnected through a centralized backbone area to implement inter-area communication between sites.

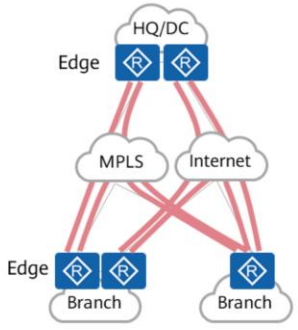
Hub-Spoke Networking



- **Solution Overview**
  - Generally, the enterprise HQ/DC functions as a hub site, and enterprise branches function as spoke sites. Server applications deployed in the HQ/DC are accessed through the WAN in a centralized manner.
  - If branches of an enterprise need to communicate with each other, traffic between them is transmitted through the hub site. All external access traffic of branch sites is first sent to the hub site.
- **Application Scenario**
  - This mode is generally applicable to enterprises where traffic is mainly sent from branch sites to the hub site.
  - Enterprise applications are centrally stored on servers at the HQ/DC. The main service traffic is from branch sites to the hub site.
  - Only small traffic is transmitted between branch sites.
  - A typical example is chain stores. The major traffic of a chain store is destined to the HQ/DC, and there is almost no traffic between chain stores.

HUAWEI

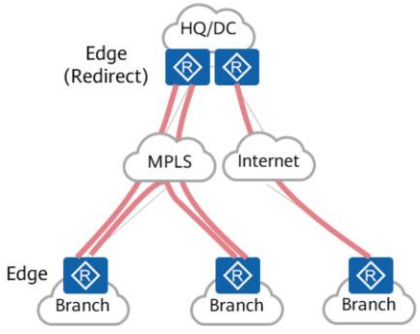
Full-Mesh Networking



- **Solution Overview**
  - In the full-mesh topology, branches can directly communicate with each other, without the need to divert traffic through intermediate nodes.
- **Application Scenario**
  - The full-mesh topology is applicable to small enterprises with a small number of sites or large enterprises whose branches need to collaborate with each other.
  - Collaborative services, for example, high-value applications including VoIP and video conferencing, have stringent requirements on network performance such as the packet loss rate, delay, and jitter. To meet requirements of such services, branches are recommended to directly communicate with each other.
  - The full-mesh topology is simple, and service communication is efficient. However, the full-mesh topology has average network scalability and applies to a network with 10 to 100 branch sites.

HUAWEI

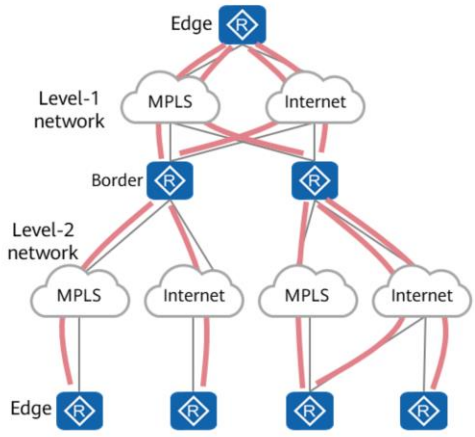
Partial-Mesh Networking



- **Solution Overview**
  - A partial-mesh network can be considered a type of special full-mesh network. If direct underlay network connections are available between sites, traffic is directly transmitted between the sites. Otherwise, traffic between sites is forwarded through a redirect site, to which both sites are connected.
- **Application Scenario**
  - Branch sites cannot directly communicate with each other over the underlay network and use a redirect site for communication instead.
  - In full-mesh networking, branch sites directly communicate with each other over the underlay network. To enhance reliability, a redirect site can be deployed. When the underlay network is faulty and sites cannot directly communicate with each other, traffic between branch sites is transmitted through the redirect site.

HUAWEI

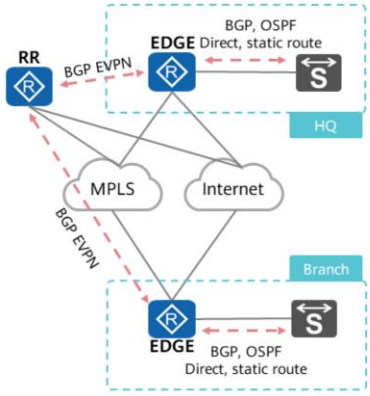
Hierarchical Networking



- **Solution Overview**
  - The hierarchical networking model can be considered as the combination of the single-layer networking model. A WAN is divided into multiple areas, which are interconnected through a centralized backbone area to implement inter-area communication between a large number of sites.
  - For example, a multinational enterprise can be divided into multiple areas (China, Europe, America, etc.) based on its management structure. Each area uses a single-layer networking model, which can be hub-spoke or full-mesh. In addition, each area uses one or more sites as their border sites. The border sites of each area form the backbone area for interconnection between areas, that is, the level-1 network. In this way, border sites connect to both the level-2 area network and level-1 network.
- **Application Scenario**
  - The hierarchical networking is applicable to enterprises with a large number of sites or multinational enterprises with widely distributed sites. The hierarchical networking has a clear network structure and good network scalability.

HUAWEI

VPN Route Design



- The VPN routes of the local EDGE are collected using the LAN-side routing protocol and then sent to the peer EDGE through the BGP EVPN protocol of the WAN.
  - LAN-side routes: LAN-side interfaces of EDGEs support OSPF, BGP, and static routing protocols. The routing protocol to be used must be the same as that of the LAN-side network device.
  - WAN-side overlay routes: Generally, BGP EVPN is deployed to establish BGP peer relationships with RRs and advertise overlay network domain routes. The configuration is automatically orchestrated by the controller.
  - LAN- and WAN-side routes can be filtered using routing policies.

HUAWEI

Section Summary

- SD-WAN networking design includes site design, tunnel design, and VPN design.
  - Site design: includes WAN/LAN networking design and dual-CPE design.
  - Tunnel design: includes data tunnel design, RR design, NAT design, and specifications-exceeded network design.
  - VPN design: includes VPN design, topology design, and VPN route design.

HUAWEI

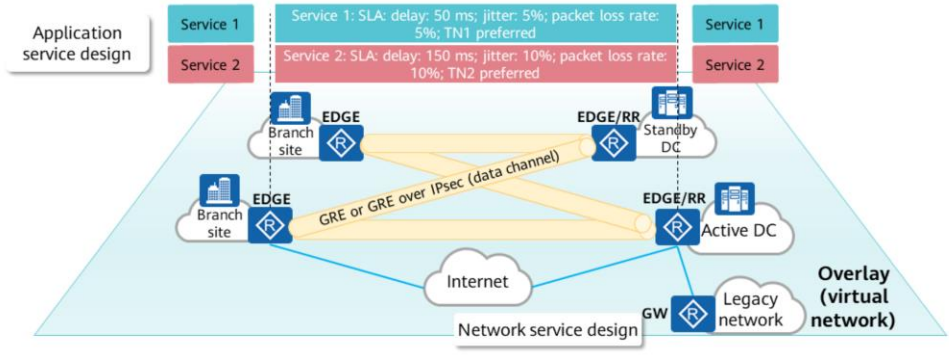
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. **Service Design for Huawei SD-WAN Solution**
   - SD-WAN Service Design Overview
   - SD-WAN Application Service Design
   - SD-WAN Network Service Design
5. Reliability and Security Design for Huawei SD-WAN Solution

HUAWEI

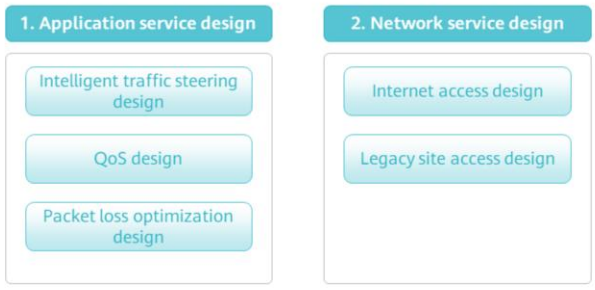
Overview of Service Design for Huawei SD-WAN Solution

- SD-WAN service design includes application service design and network service design.



HUAWEI

Service Design Process for Huawei SD-WAN Solution



HUAWEI

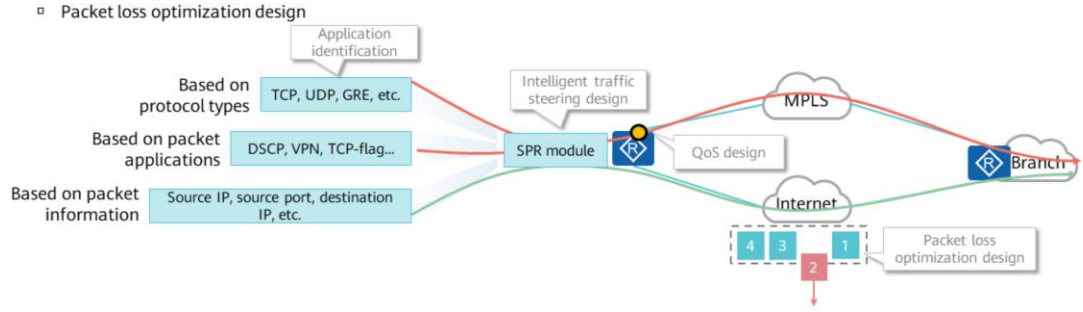
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. **Service Design for Huawei SD-WAN Solution**
   - SD-WAN Service Design Overview
   - SD-WAN Application Service Design
   - SD-WAN Network Service Design
5. Reliability and Security Design for Huawei SD-WAN Solution

HUAWEI

Application Service Design Panorama

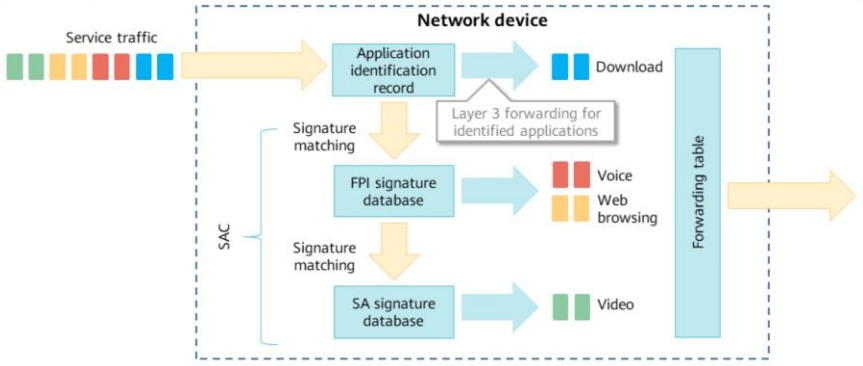
- Enterprises have diverse applications, such as production, coordination, and entertainment applications. Different applications have different requirements on bandwidth and link quality.
- Application service design generally covers:
  - Intelligent traffic steering design
  - QoS design
  - Packet loss optimization design



HUAWEI

Application Identification

- Traditional routing and switching devices cannot identify application-layer information. Therefore, it is difficult to manage networks based on applications. Smart Application Control (SAC) technology helps routing and switching devices identify and classify applications.
- SAC first checks whether an application has been identified. If an application has not been identified, SAC checks the first-packet inspection (FPI) signature database and service awareness (SA) signature database in sequence.

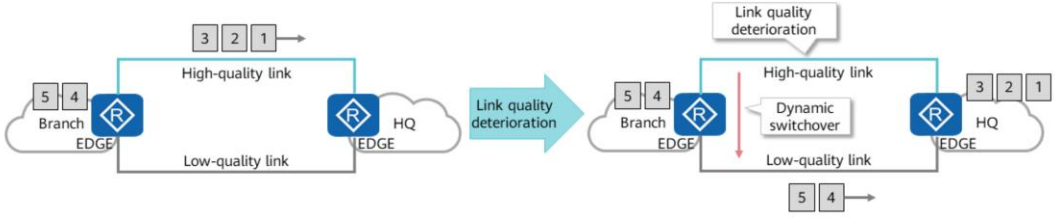


HUAWEI

- After a packet enters the device, the device determines whether the corresponding application has been identified based on the 5-tuple information carried in the packet. If the application has been identified, the device forwards the packet at Layer 3 without identifying the application again. If the application has not been identified, the device performs the SAC application identification process. The device then processes the packet based on the SAC identification result and forwards the packet at Layer 3.
- SAC identifies an application as follows: SAC first identifies an application according to ACL rules in FPI. If the application fails to be identified, the application is identified according to the DNS association table in FPI. If the application fails to be identified, the device identifies the application based on the protocol port table in the FPI. If the application still fails to be identified, the device attempts to identify the application using SA.

Intelligent Traffic Steering Design: Link Quality-based Traffic Steering

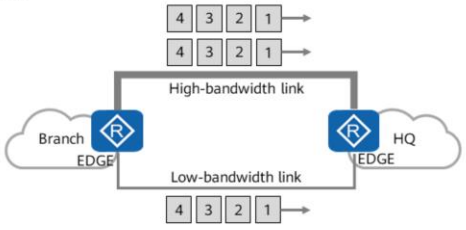
- Different applications have different requirements on link quality. For example, voice and video services have stringent requirements on the delay and packet loss rate. Generally, the delay must be within 150 ms and the packet loss rate must be less than 1% for these services.
- To guarantee voice and video services, the MPLS link that offers good quality can be configured as the active link for the services, with the Internet link functioning as the standby link. In addition, you can configure SLA requirements for the services to implement traffic steering based on the link SLA.



HUAWEI

Intelligent Traffic Steering Design: Load Balancing-based Traffic Steering

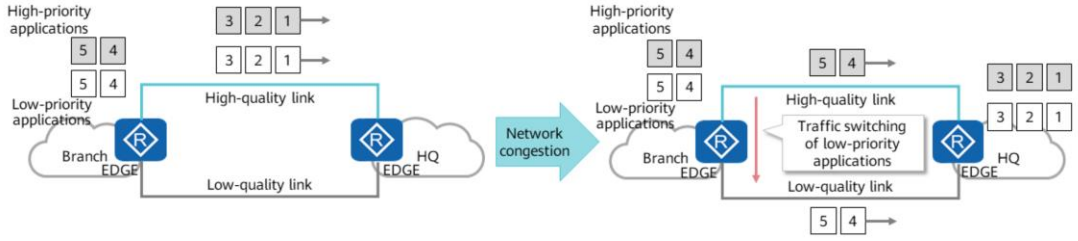
- If an enterprise with multiple links wants to fully utilize the link bandwidth resources and steer traffic in load balancing mode based on the link bandwidth, load balancing-based traffic steering can be configured.
- For example, the two MPLS links can be configured as primary links for the voice service. If the quality of both links meets the SLA requirements of the voice service, voice service flows can be carried over the two MPLS links in load balancing mode. Through real-time bandwidth utilization monitoring, the bandwidth utilization can reach 85%, fully utilizing the link bandwidth.



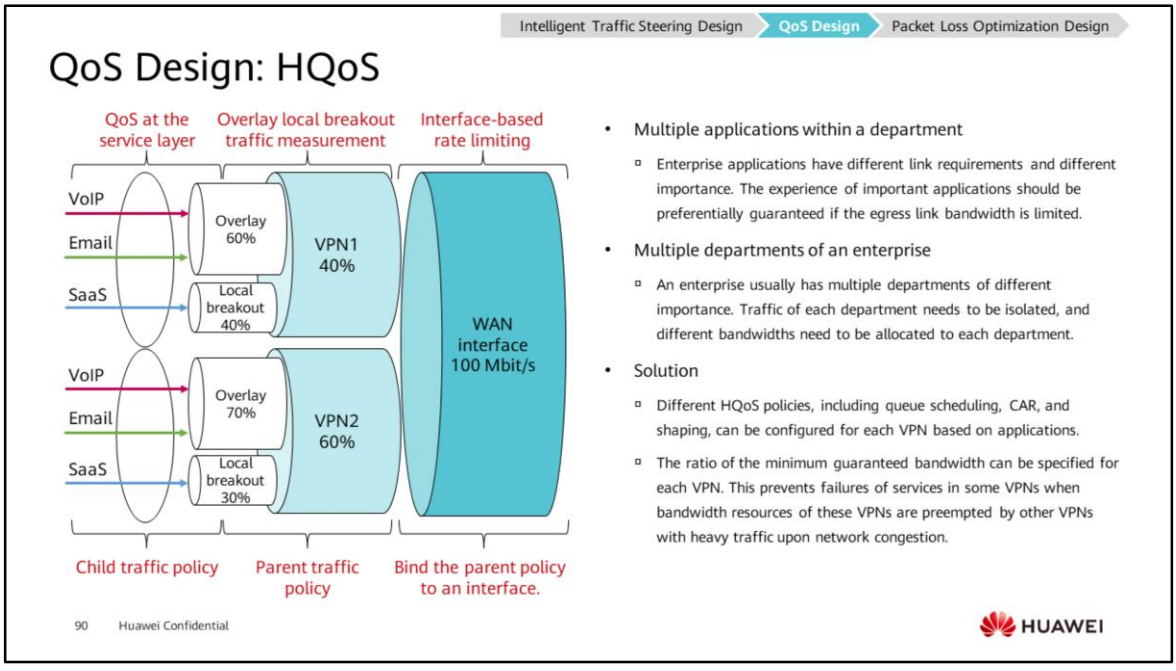
HUAWEI

Intelligent Traffic Steering Design: Application Priority-based Traffic Steering

- If multiple types of service packets are transmitted on the same link, application priority-based traffic steering can be used, so that traffic of high-priority applications is preferentially processed when congestion occurs, ensuring experience of high-priority applications.
- For example, voice, video, and file transfer services are carried on an MPLS link. If the link bandwidth is insufficient, the experience of the voice and video services is preferentially guaranteed.



HUAWEI



- An enterprise usually has multiple departments of different importance. Traffic of each department needs to be isolated, and different bandwidths need to be allocated to each department.
    - A specified bandwidth quota is assigned to each department to meet its service requirements.
    - If some departments do not fully use their bandwidth quotas, the idle bandwidth resources can be used by other departments with insufficient bandwidth.
    - Internet access traffic and traffic for communication with legacy sites needs to be controlled separately.

Packet Loss Optimization Design

- Forward error correction (FEC) technology can mitigate packet loss. During data transmission, FEC adds redundant packets carrying check information at the transmit end and performs check at the receive end. If a packet is lost or damaged on the network, the redundant packet can be used to recover it.
- The following FEC modes are supported:
  - Adaptive forward error correction (A-FEC): A CPE automatically detects the packet loss rate and adjusts the number of redundant packets in real time based on the packet loss rate. This mode reduces the number of redundant packets and saves bandwidth when the packet loss rate is low. When the packet loss rate increases, redundant packets are increased to improve the packet loss mitigation performance. However, if the packet loss rate suddenly increases greatly, it takes a period of time for the CPE to detect the event and adjust the number of redundant packets. As a result, packet loss occurs for a short period of time.
  - Determined forward error correction (D-FEC): The number of redundant packets is calculated based on the configured packet loss rate. This mode can recover the lost packets within the configured maximum packet loss rate. Although the redundancy of this mode is greater than that of A-FEC, packet loss does not occur when the packet loss rate suddenly changes.
- After FEC is deployed and the network runs for a period of time, you can view the packet loss mitigation effect on iMaster NCE-WAN to determine whether the FEC design is proper.

HUAWEI

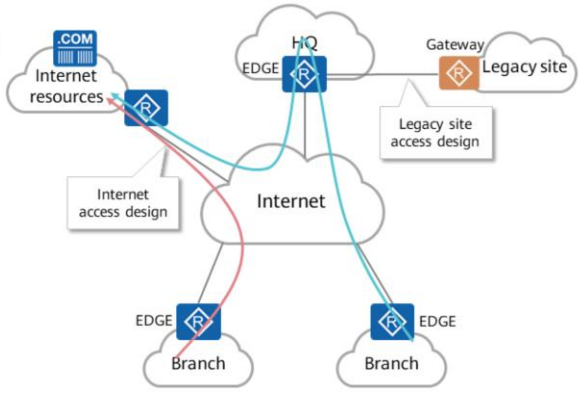
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. **Service Design for Huawei SD-WAN Solution**
   - SD-WAN Service Design Overview
   - SD-WAN Application Service Design
   - SD-WAN Network Service Design
5. Reliability and Security Design for Huawei SD-WAN Solution

HUAWEI

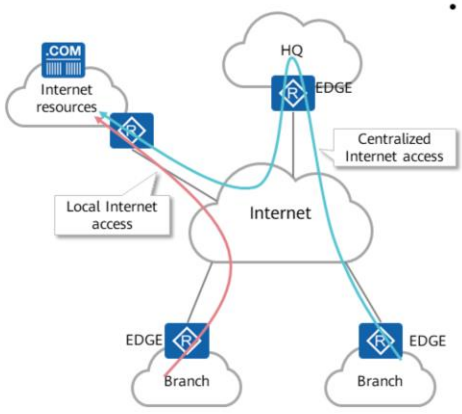
Network Service Design Panorama

- SD-WAN sites may also need to access the Internet or legacy sites. Therefore, specific network services need to be deployed.
- Network service design generally covers:
  - Internet access design
  - Legacy site access design



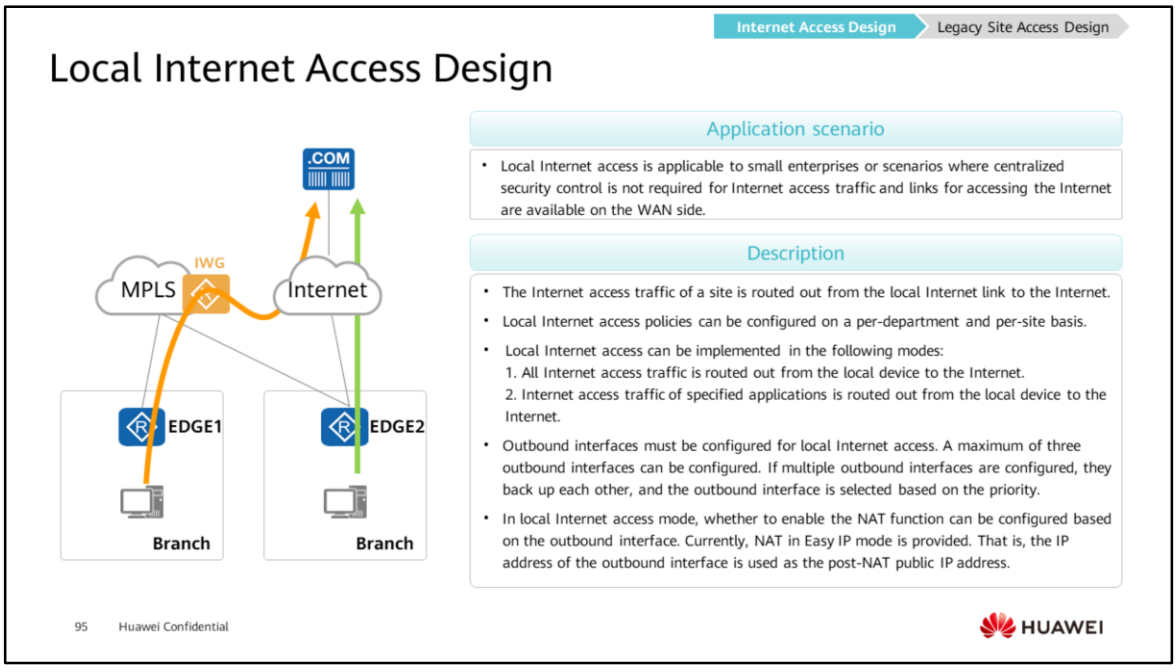
HUAWEI

Internet Access Solution



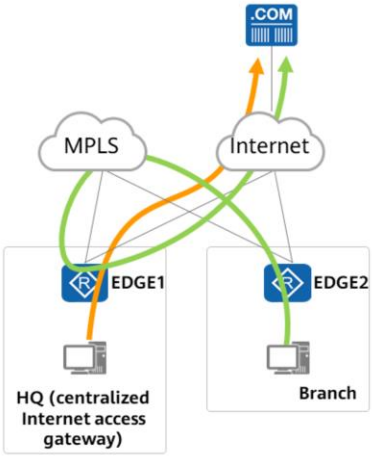
- The SD-WAN Solution provides the following Internet access modes:
  - Local Internet access: The Internet access traffic of a site is routed from the local EDGE to the Internet.
  - Centralized Internet access: The Internet access traffic of all sites is diverted to the centralized Internet access site and then to the Internet.
  - Hybrid Internet access: combination of local Internet access and centralized Internet access.
    - Local Internet access (default) + centralized Internet access: By default, all Internet access traffic is routed out from the local device. When the local Internet access interface is faulty, Internet access traffic is forwarded through the centralized gateway.
    - Centralized Internet access (default) + Local Internet access for specified traffic: By default, Internet traffic is routed out through the centralized Internet access site. Traffic of some specified services is directly routed to the Internet through the local WAN-side link.

HUAWEI



- Traffic cannot be transmitted to the Internet through multiple links in load balancing mode. The links can work only in active/standby mode based on their priorities.
- If local Internet access is configured for specified application traffic and centralized Internet access is also configured, local Internet access for specified application traffic is implemented by orchestrating policy-based routing (PBR).
- If local Internet access is enabled, the default route on the underlay WAN needs to be configured. The default route can be a static route (mainly for Internet access through the Internet network interface) or BGP/OSPF route (mainly for Internet access through the MPLS network interface).

Centralized Internet Access Design



Application scenario

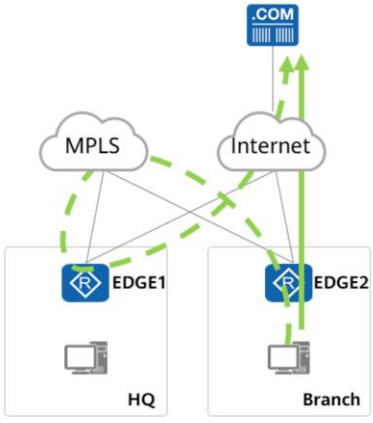
- This mode is applicable to scenarios where no link is available for a site to access the Internet or centralized security control is required for Internet access traffic. A centralized Internet access gateway is configured so that traffic from other sites can be forwarded to the centralized Internet access gateway through the overlay network.

Description

- All sites of a tenant access the Internet through a centralized Internet access site.
- Either of the following solutions can be used for Internet access through the centralized gateway site:
  - If the centralized Internet access site has the Internet egress on the LAN side, all Internet access traffic is routed out through the LAN side of the centralized Internet access site. In this mode, you need to configure a default route or dynamic routing protocol on the LAN side so that the default route can be learned from the LAN side.
  - If the centralized Internet access site accesses the Internet through the WAN-side interface, all Internet access traffic is transmitted to the Internet through the WAN side of the centralized Internet access site. (Note: For the site that functions as a centralized gateway, the local Internet access function must also be enabled.)

HUAWEI

Hybrid Internet Access Design: Local Internet Access (Default) + Centralized Internet Access



Application scenario

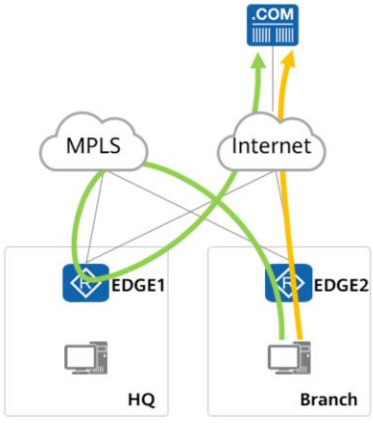
- This mode applies to small enterprises or scenarios where Internet access traffic does not need to be centrally managed.
- The WAN side of the site has links to access the Internet, and the site has high requirements on Internet access reliability.

Description

- Both local Internet access and centralized Internet access are configured on the SD-WAN network.
- Local Internet access is configured for all traffic at a site.
- If a site has two Internet access links, that is, one for local Internet access and one for centralized Internet access, the link for local Internet access has a higher default route preference.
- By default, all Internet access traffic is routed out from the local Internet access interface. When the local Internet access interface is faulty, Internet access traffic is diverted to the centralized gateway and then routed out.

HUAWEI

Hybrid Internet Access Design: Centralized Internet Access (Default) + Local Internet Access for Specified Traffic



Application scenario

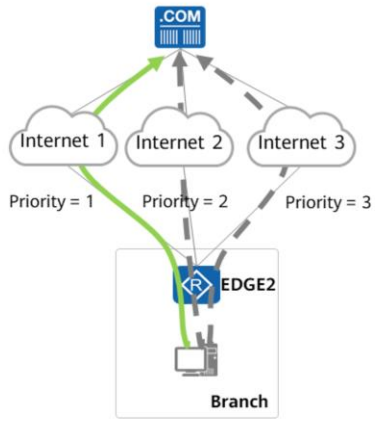
- This mode applies to the scenario where a site has Internet access links, most Internet access traffic needs to be centrally managed, and traffic of some applications can be directly routed out to meet SLA requirements.

Description

- Configure both local Internet access and centralized Internet access on the SD-WAN network.
- Local Internet access is configured at the centralized Internet access site, where all traffic is routed out in local Internet access mode.
- Local Internet access is enabled at other sites and is configured for specified applications.
- By default, Internet access traffic is routed out through the centralized Internet access site, and specific experience-sensitive service traffic is directly transmitted to the Internet through the local WAN-side link.

HUAWEI

Internet Access Link Reliability



- **Standby link solution design:**
  - Links for Internet access are configured on a per-site basis. A maximum of three WAN links can be configured for Site-to-Internet access.
  - A priority is configured for each WAN link. Currently, each link has a unique priority. That is, only one WAN link is used as the active link for accessing the Internet at a time.
  - When a WAN link with the highest priority fails, traffic to the Internet is automatically switched to the link with the second highest priority.
  - The standby link and hybrid Internet access mode can be used together.

HUAWEI

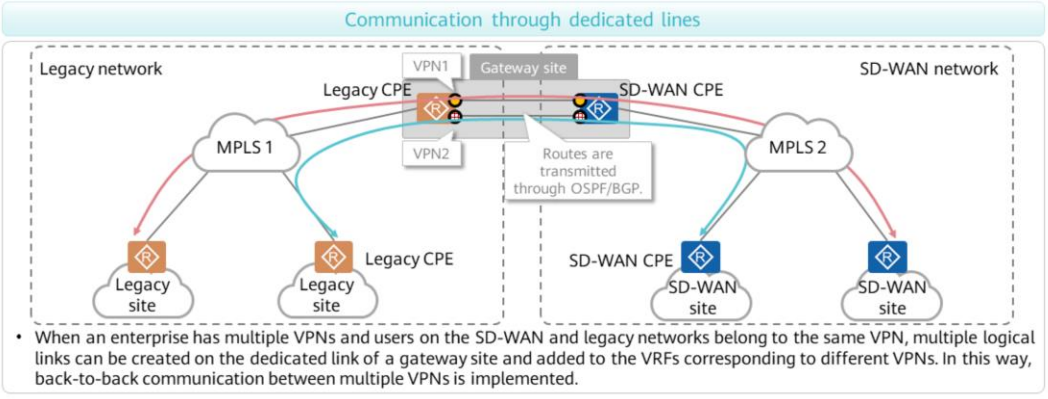
Legacy Site Communication Solution

- Before deploying the SD-WAN Solution, enterprises may have legacy sites that are interconnected through traditional WAN private lines (such as MPLS). After the SD-WAN Solution is deployed, SD-WAN sites and legacy sites may need to communicate with each other. To implement this, a solution for communication with legacy sites must be deployed.
- Communication between SD-WAN sites and legacy sites involves the following scenarios:
  - Communication through enterprise sites: Enterprise sites are directly connected to legacy networks. Two communication solutions are available in this scenario:
    - Communication through dedicated lines
    - Local access
  - Communication through the IWG: Carriers use this mode to provide communication with legacy sites for multiple enterprises. Generally, three communication solutions are available in this scenario:
    - Inter-AS Option B
    - Inter-AS Option A (Layer 3 VXLAN)
    - Inter-AS Option A (Layer 3 VLAN)

HUAWEI

Communication Through Enterprise Sites: Communication Through Dedicated Lines

- When the underlay network connected to an SD-WAN site cannot communicate with a legacy MPLS network, select an SD-WAN CPE and a legacy CPE, which are logically considered as a gateway site.



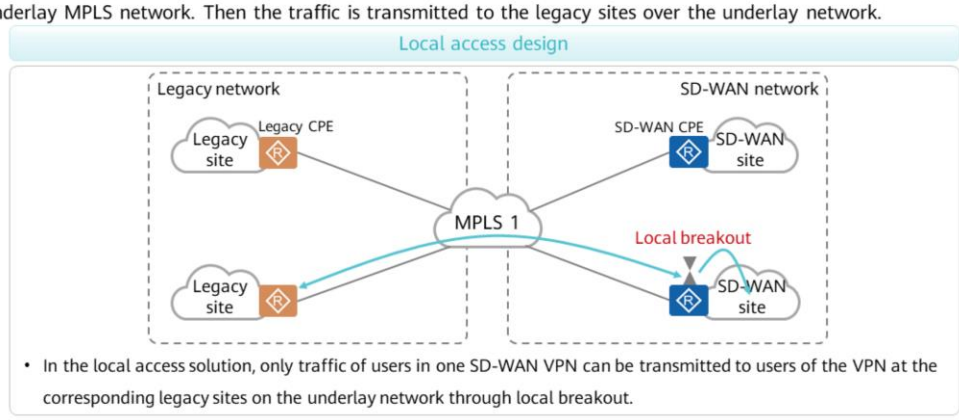
- When an enterprise has multiple VPNs and users on the SD-WAN and legacy networks belong to the same VPN, multiple logical links can be created on the dedicated link of a gateway site and added to the VRFs corresponding to different VPNs. In this way, back-to-back communication between multiple VPNs is implemented.

HUAWEI

- A dedicated link is established between user-side interfaces on both the legacy CPE and SD-WAN CPE. The dedicated link runs a protocol such as BGP or OSPF to exchange routes between the legacy MPLS network and SD-WAN network. In this way, users on the two networks can communicate with each other through the dedicated link.

Communication Through Enterprise Sites: Local Access

- When the underlay network connected to an SD-WAN site is the legacy MPLS network or can communicate with the legacy MPLS network, local breakout technology can be used to transmit user traffic of the SD-WAN site to the underlay MPLS network. Then the traffic is transmitted to the legacy sites over the underlay network.



- In the local access solution, only traffic of users in one SD-WAN VPN can be transmitted to users of the VPN at the corresponding legacy sites on the underlay network through local breakout.

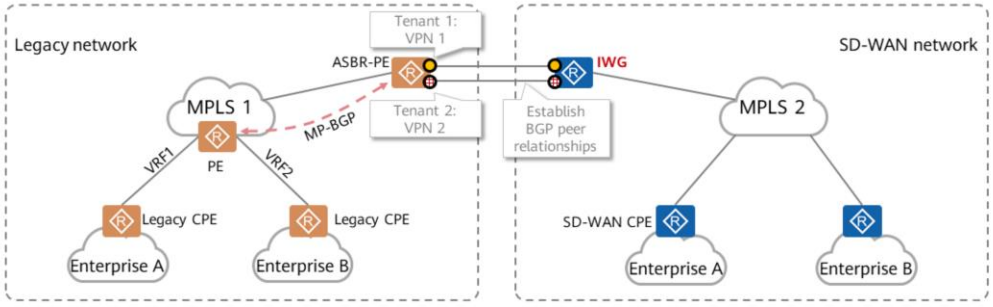
HUAWEI

- Multiple traffic models are supported in this scenario, and you can choose one based on your service requirements.
  - Distributed local access: This model applies if all SD-WAN sites can access legacy sites over the underlay MPLS network through local breakout. In this model, traffic of each site is directly forwarded through the local site, without the need of being forwarded through overlay tunnels.
  - Centralized local access: If some SD-WAN sites cannot access legacy sites through local breakout, you can configure a site that can communicate with the legacy sites as the centralized access site. Traffic from other SD-WAN sites is sent to the centralized access site through overlay tunnels, and then forwarded to the legacy sites through local breakout.
  - Hybrid local access: The SD-WAN Solution enables multi-link sites using the distributed local access model to use local access preferentially, with centralized local access as a backup. This enhances reliability. Traffic from a site that uses the distributed local access model is preferentially transmitted to a legacy site through local breakout. If the MPLS link for local access fails, traffic is automatically switched to the overlay tunnel of another link and transmitted to the centralized access site. The centralized access site then forwards the traffic to legacy sites.

IWG

- The IWG can connect both SD-WAN sites and legacy MPLS VPN sites for multiple enterprise tenants. Each time a tenant is added, only one MP-EBGP peer needs to be configured to interconnect with the peer ASBR-PE.s

Solution design for multiple tenants to communicate with legacy sites

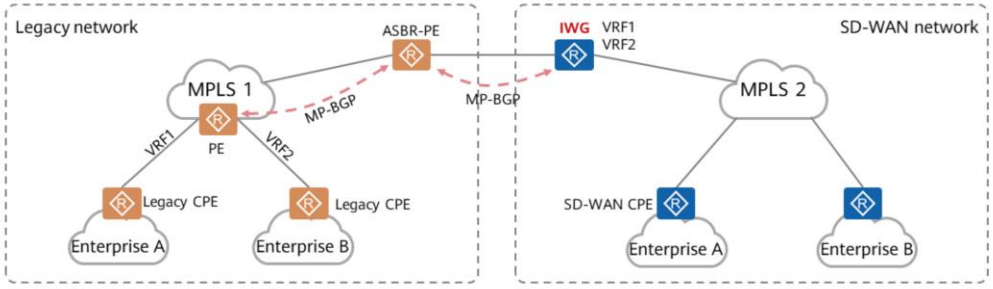


HUAWEI

Communication Through the IWG: Inter-AS Option B

- Similar to the inter-AS Option B solution in MPLS VPN, in the inter-AS Option B solution in SD-WAN, MP-EBGP is used between the IWG and ASBR-PE on legacy networks to exchange labeled VPN-IPv4 routes received from PEs in their respective ASs.

Inter-AS Option B



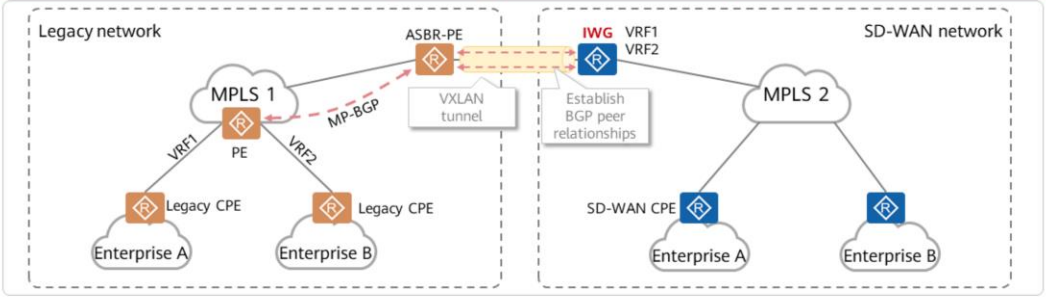
- A pair of public interfaces are configured on the IWG and ASBR-PE, and MPLS is enabled. MP-EBGP is used to exchange labeled VPN-IPv4 routes. BGP is used to transmit inter-AS labels. Therefore, LDP is not required between the IWG and ASBR-PE.

HUAWEI

Communication Through the IWG: Inter-AS Option A (Layer 3 VXLAN)

- Similar to the inter-AS Option A solution in MPLS VPN, in the inter-AS Option A solution for SD-WAN, IWGs are interconnected with ASBR-PEs through Layer 3 VXLAN tunnels, and the EBGP or IBGP peer relationship is established at both ends of VXLAN tunnels to exchange routes between SD-WAN sites and legacy sites.

Inter-AS Option A (Layer 3 VXLAN) solution design

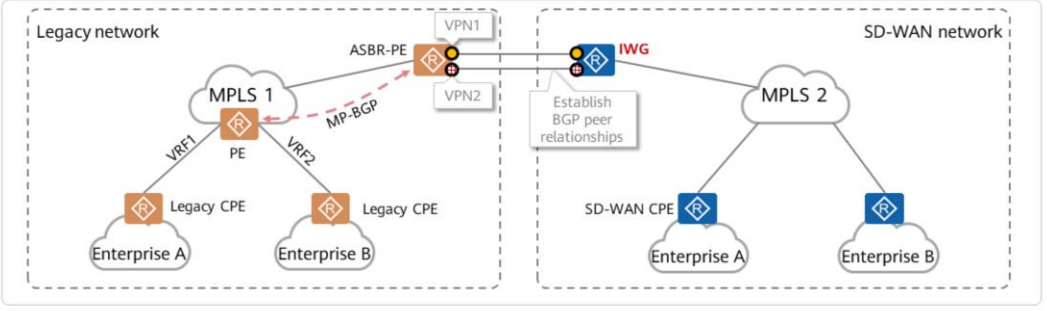


HUAWEI

Communication Through the IWG: Inter-AS Option A (Layer 3 VLAN)

- Between the IWG of the SD-WAN network and the PE of the legacy MPLS network, a pair of specific interfaces (VLANIF interfaces or Ethernet sub-interfaces) are configured for each user VPN and are added to the corresponding VRF. In addition, EBGP is configured based on the VPN instance to exchange routes between the SD-WAN and MPLS domains.

Inter-AS Option A (Layer 3 VLAN) solution design



HUAWEI

Section Summary

- SD-WAN service design includes application service design and network service design.
  - Application service design: includes intelligent traffic steering design, QoS design, and packet loss optimization design.
  - Network service design: includes Internet access design and legacy site access design.

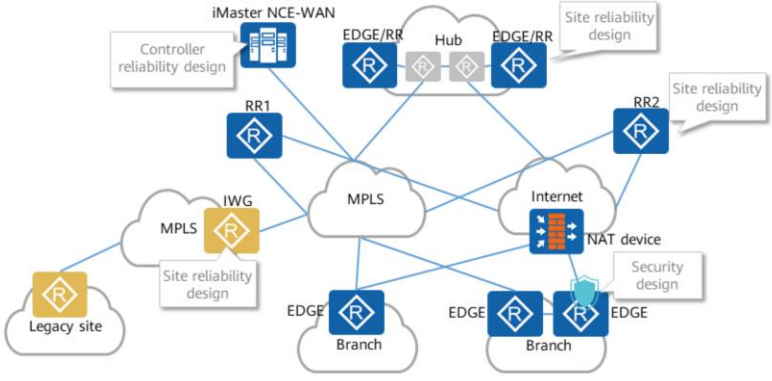
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. Service Design for Huawei SD-WAN Solution
5. **Reliability and Security Design for Huawei SD-WAN Solution**
   - Reliability and Security Design Overview
   - Site Reliability Design
   - Controller Reliability Design
   - Security Design

HUAWEI

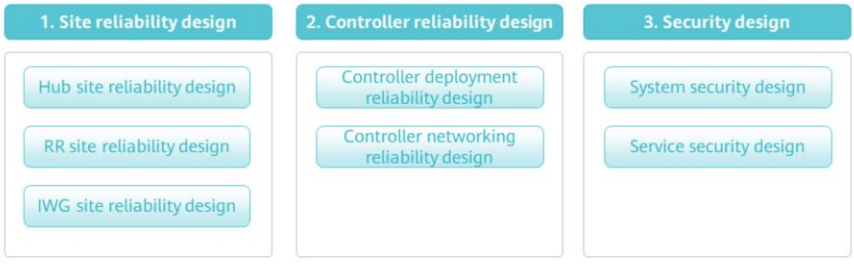
Overview of Reliability and Security Design for Huawei SD-WAN Solution

- Reliability and security design for Huawei SD-WAN Solution mainly includes three parts: site reliability design, controller reliability design, and security design.



HUAWEI

Reliability and Security Design Process for Huawei SD-WAN Solution



HUAWEI

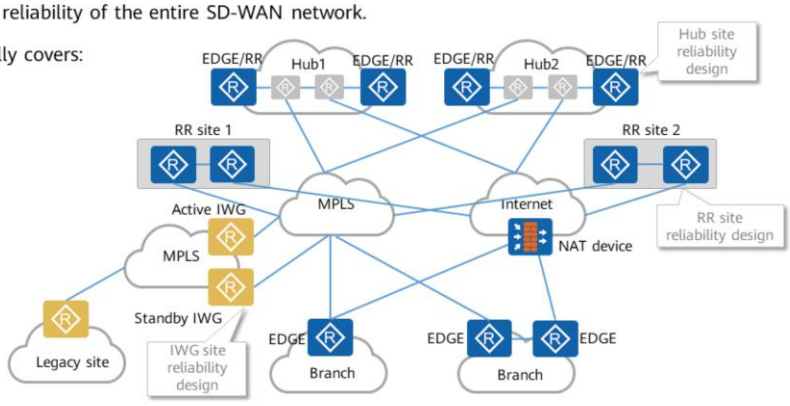
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. Service Design for Huawei SD-WAN Solution
5. **Reliability and Security Design for Huawei SD-WAN Solution**
   - Reliability and Security Design Overview
   - Site Reliability Design
   - Controller Reliability Design
   - Security Design

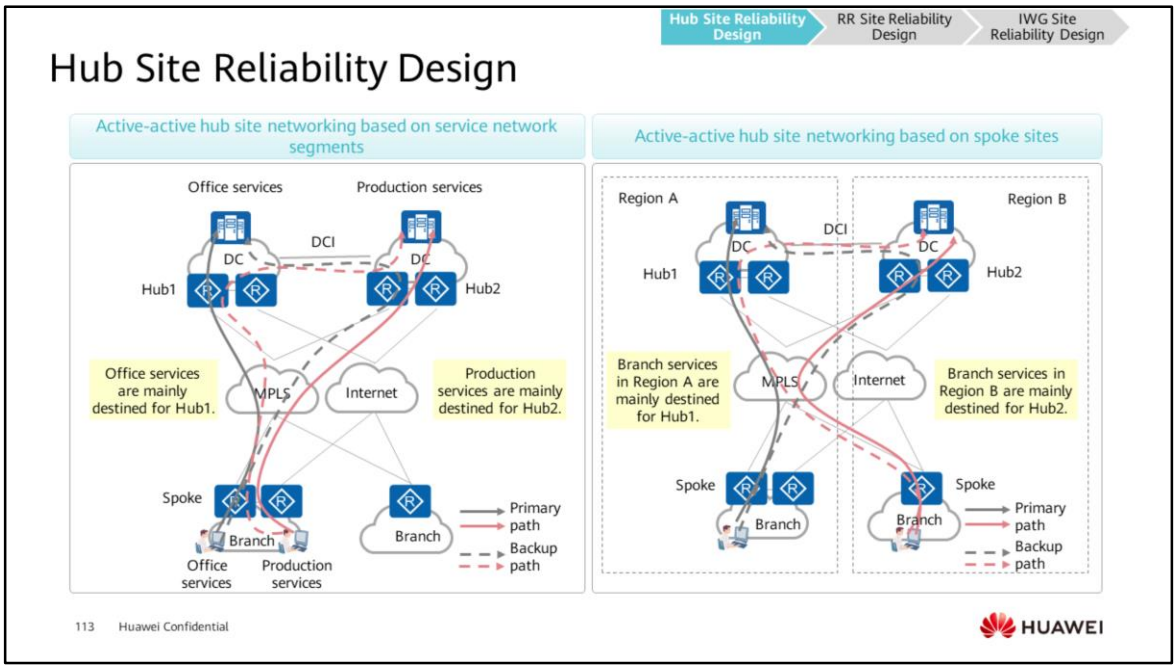
HUAWEI

Site Reliability Design Overview

- SD-WAN sites are typically classified into edge sites, hub sites, and RR sites, as well as gateway sites connected to traditional networks. In Huawei SD-WAN Solution, hub sites, RR sites, and gateway sites are key sites, and their reliability greatly affects the reliability of the entire SD-WAN network.
- Site reliability design generally covers:
  - Hub site reliability design
  - RR site reliability design
  - IWG site reliability design



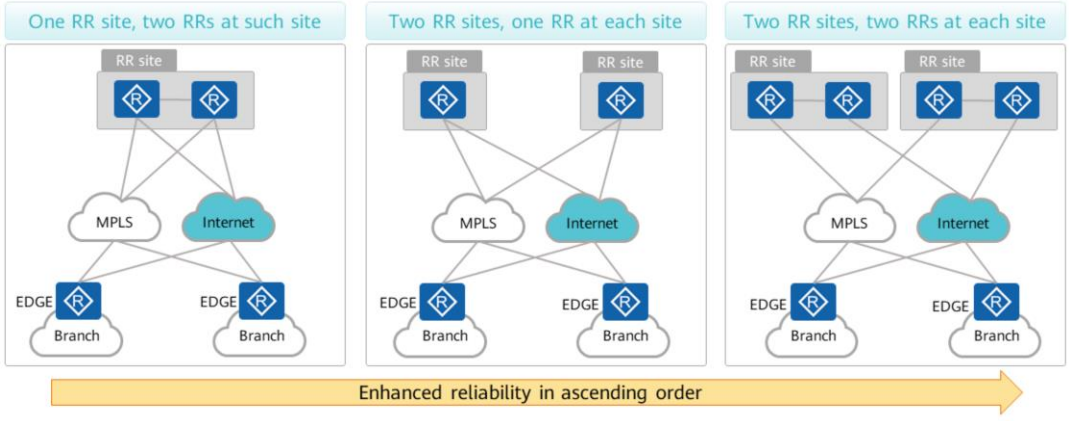
HUAWEI



- Active-active hub site networking based on service network segments:
  - If branch sites' access to DCs can be distinguished based on service network segments, active-active hub sites can be implemented based on the service network segments. Each of the hub sites functions as the active hub site for one service or standby hub site for another different service.
- Active-active hub site networking based on spoke sites:
  - If an enterprise's WAN spans a large physical distance and the enterprise's two DCs are far away from each other, active-active hub sites based on spoke sites are recommended. Each spoke site connects to both hub sites deployed in the DCs and preferentially accesses the nearer hub site for a better DC service access experience. The nearer hub site is the primary one, while the farther hub site is the backup one.

RR Site Reliability Design: RR Redundancy

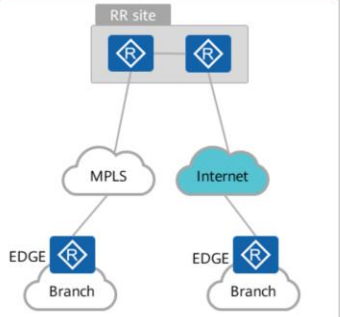
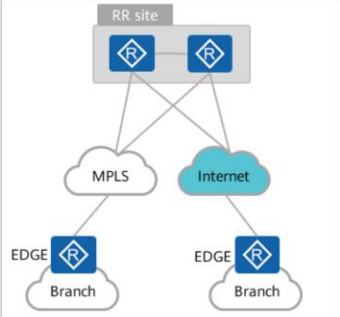
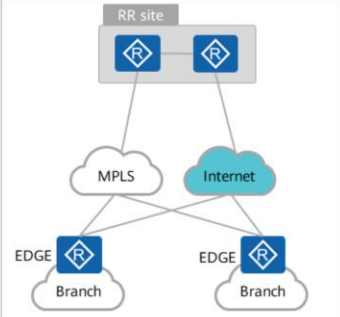
- One EDGE can be connected to two RR sites (four RRs) at most. The following figures from left to right show the RR site deployment modes in three scenarios, where reliability is enhanced in ascending order.



HUAWEI

RR Site Reliability Design: RR Link Redundancy

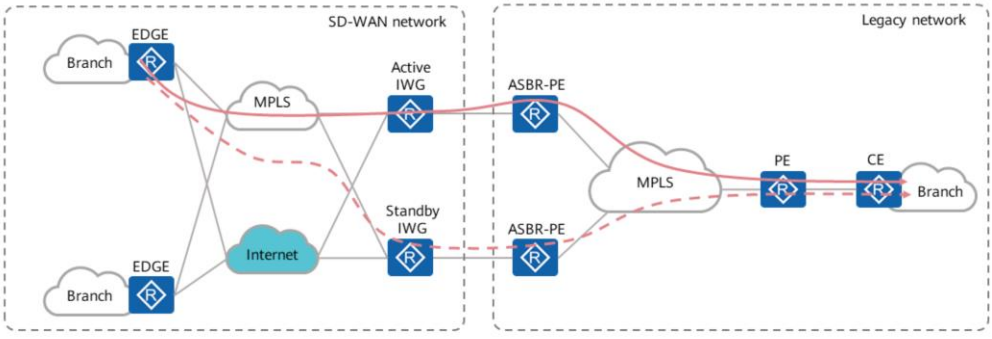
- The edge site needs to be dual-uplinked to the RR site.

Single link	Multi-link (topology 1)	Multi-link (topology 2)
		
• Low reliability: If one RR is faulty, some branch sites will lose connection with the RR site, causing service interruptions.	• Multiple links are deployed on the RR side. This ensures reliable connections between EDGEs and RRs.	• Multiple links are deployed at the branch site. This ensures reliable connections between EDGEs and RRs.

HUAWEI

IWG Reliability Design

- Branch sites are specified with active and standby IWGs to establish IWG tunnels. On the active and standby IWGs, BGP routing policies are automatically orchestrated to control the preference of BGP routes originated from the legacy network.
- The preference of BGP routes on the active IWG is higher than that of BGP routes on the standby IWG.



HUAWEI

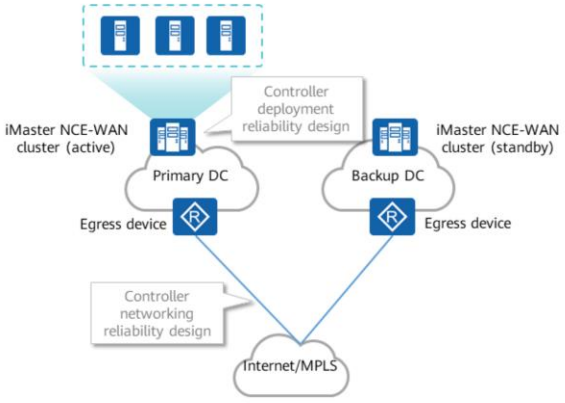
Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. Service Design for Huawei SD-WAN Solution
5. **Reliability and Security Design for Huawei SD-WAN Solution**
   - Reliability and Security Design Overview
   - Site Reliability Design
   - Controller Reliability Design
   - Security Design

HUAWEI

Controller Reliability Design Overview

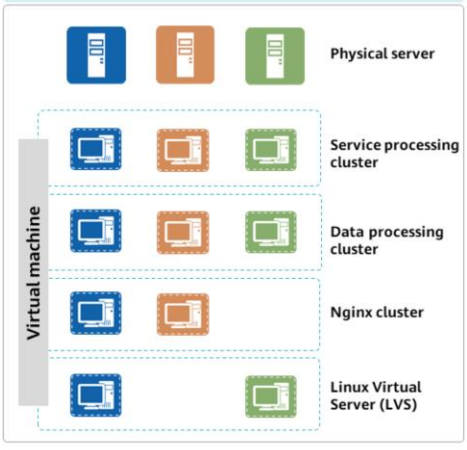
- The controller iMaster NCE-WAN is the core component of Huawei SD-WAN Solution. All services in Huawei SD-WAN Solution are orchestrated through the controller. Therefore, controller reliability directly affects network services.
- Controller reliability design generally covers:
  - Controller deployment reliability design
  - Controller networking reliability design



HUAWEI

Controller Deployment Design in a Site

Controller deployment in distributed cluster architecture



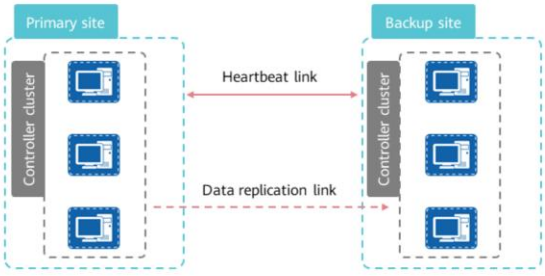
- The controller is deployed in **distributed cluster architecture** to provide high reliability and load balancing capabilities. When a node in the cluster is faulty, other nodes take over services, without affecting services.
- **Northbound load balancing:** External requests are distributed to all nodes in the cluster instead of being centrally processed on just one node. Doing so fully utilizes cluster capabilities and improves reliability.
- **Southbound load balancing:** Controller nodes are dynamically assigned to network devices based on the loads of each controller node.

Cluster	Function Description
Service processing cluster	Provides service processing capabilities, such as CPE management, overlay network configuration delivery, and traffic policy configuration.
Data processing cluster	Provides functions such as CPE performance data storage and data aggregation.
Nginx cluster	Serves as a high-performance HTTP proxy server that forwards concurrent connection requests and implements L4-L7 load balancing for northbound traffic.
LVS	Works as a load balancing component and implements L1-L4 load balancing for southbound and northbound traffic.

HUAWEI

Controller DR Deployment Design

- The geo-redundant disaster recovery (DR) system consists of two sets of iMaster NCE-WAN, which are deployed at two geo-redundant sites.



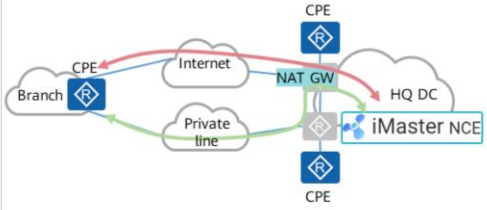
- Geo-redundant DR ensures backup between two controller clusters. The number of nodes in the active cluster must be the same as that in the standby cluster.
- The active and standby controller clusters are both running. However, only the active cluster can provide services, while the standby cluster does not provide services. Data in the active cluster is synchronized to the standby cluster in real time to ensure data consistency.
- The same domain name solution is used for northbound and southbound interfaces of the active and standby controller clusters. Tenants and devices access the active controller cluster through this unified domain name. After active/standby switchover, this domain name is mapped to the new active controller cluster.
- Active/standby switchover is performed manually and can be completed within 10 minutes. In the case of component-level faults in the controller cluster, services are still guaranteed through reliability technologies, including clustering and fault tolerance.
- Geo-redundant DR network requirements: latency ≤ 20 ms; bandwidth ≥ 125 Mbit/s.

HUAWEI

Reliability Design for Single-Controller Networking

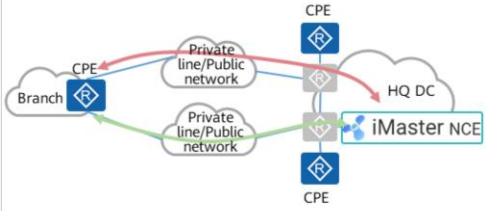
- To ensure the reliability of SD-WAN management channels, dual gateways and dual links generally need to be deployed for highly reliable communication between EDGEs/RRs and the controller.

Hybrid networking (public and private network)



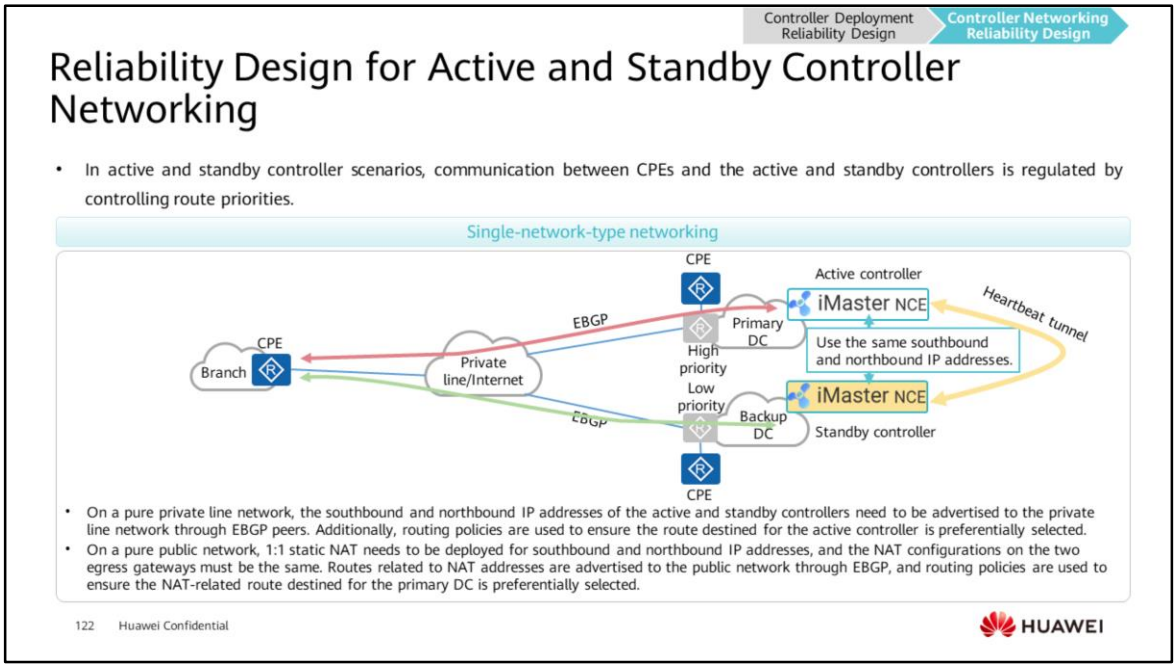
- The controller is deployed inside the HQ DC, and uses public IP addresses to provide services for CPEs.
- 1:1 static NAT is deployed on the public network egress device at HQ.
- NAT-related public IP addresses need to be advertised into the private line network.
- Communication between CPEs and the controller through the public network or private line must traverse the NAT device on the public network.

Single-network-type networking



- The controller is deployed inside the HQ DC.
- The private line network needs to send the network segment where the controller resides to VPN networks.
- On the public network, 1:1 static NAT must be deployed on the public network egress device. Static NAT uses the same public IP address.

HUAWEI



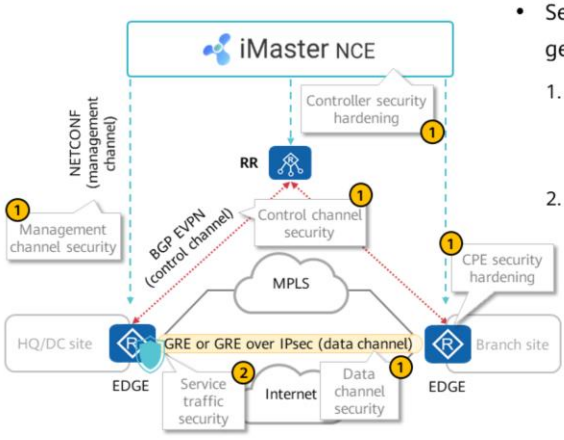
- The active controller is deployed in the primary DC, and the standby controller in the backup DC. A heartbeat tunnel is established between the active and standby controllers to synchronize data and verify the controller status.
- The active and standby controllers use the same southbound and northbound IP addresses. In public network scenarios, the same NAT address must also be configured.

Contents

1. Development Trends and Challenges Facing Enterprise WAN Interconnection
2. Overview of Huawei SD-WAN Solution
3. Networking Design for Huawei SD-WAN Solution
4. Service Design for Huawei SD-WAN Solution
5. **Reliability and Security Design for Huawei SD-WAN Solution**
   - Reliability and Security Design Overview
   - Site Reliability Design
   - Controller Reliability Design
   - Security Design

HUAWEI

Security Design Overview

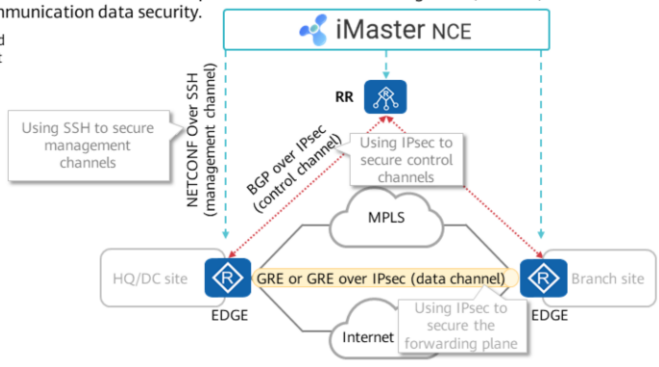


- Security design for Huawei SD-WAN Solution generally covers:
  1. System security: The security requirements of the SD-WAN Solution must be met. In this way, the entire system can run securely and stably.
  2. Service security: The security requirements of services carried by the SD-WAN Solution must be met. Appropriate security protection measures are flexibly selected as required, thereby ensuring secure and reliable running of user services.

HUAWEI

System Security Design: Inter-Component Communication Security

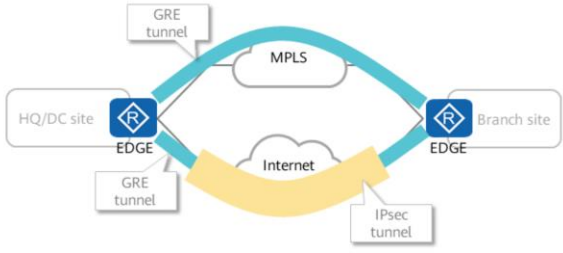
- The communication channels between components of Huawei SD-WAN Solution are classified into three types: management channels, control channels, and data channels. EDGEs/RRs communicate with the controller through management channels, CPEs communicate with RRs through control channels, and EDGEs communicate with each other through data channels.
- The components of Huawei SD-WAN Solution use secure communication protocols to establish management, control, and data channels, thereby ensuring inter-component communication data security.
  - Security mechanisms for management channels and control channels are enabled by default. You do not need to configure them.
  - Security mechanisms for data channels need to be designed flexibly as required.



HUAWEI

System Security Design: Data Channel Security

- Data channels are overlay tunnels established between EDGEs. IPsec is used on such tunnels to ensure data confidentiality and integrity during transmission.
  - IPsec must be deployed when data channels are established over the Internet.
  - IPsec can be deployed as required when data tunnels are established over MPLS or private networks.



HUAWEI

System Security Design: Administrator Authentication and Authorization

- The system supports local and remote authentication for administrators. Strict identity authentication ensures that only authorized administrators can log in to the system.
- When a tenant administrator logs in to iMaster NCE-WAN, two-factor authentication (account/password + SMS verification code) is supported. The tenant administrator can be authenticated in either of the following modes:
  - Local authentication: When a tenant administrator logs in to iMaster NCE-WAN, iMaster NCE-WAN authenticates the tenant administrator.
  - LDAP server authentication: iMaster NCE-WAN connects to an LDAP server (such as a general-purpose LDAP server or Windows AD server). When a tenant administrator logs in to iMaster NCE-WAN, the LDAP server authenticates the tenant administrator.
- When a device administrator logs in to a CPE through the CLI, the device administrator can be authenticated in either of the following modes:
  - Local authentication: The tenant administrator can set passwords for device administrators on iMaster NCE-WAN, which then delivers the passwords to CPEs. When a device administrator logs in to a CPE, the CPE authenticates the device administrator.
  - TACACS authentication: A CPE connects to a TACACS server through the underlay network. When a device administrator logs in to the CPE, the TACACS server authenticates the device administrator.

HUAWEI

System Security Design: CPE Security

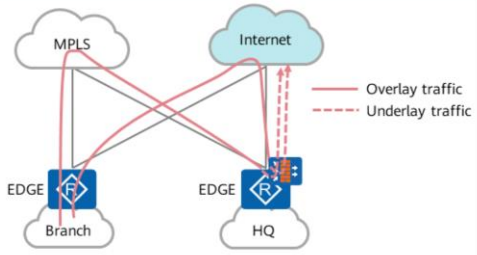
- A CPE must have a secure system architecture and support multiple security protection measures for defending against various security threats.
- A CPE must have multiple security protection capabilities on the control plane, management plane, and forwarding plane, including but not limited to the following:
  - Physical security: A CPE must be able to disable the service ports, serial ports, and services that are not in use, thereby preventing attacks.
  - Data security: A CPE must be able to encrypt sensitive information, such as service data, user names, and passwords, to prevent sensitive information leakage. A CPE must also control data access permissions to prevent unauthorized access to data.
  - Authentication: A CPE must provide system permission control and account permission management functions to implement strict identity authentication and permission control on login behaviors. It also must support security mechanisms such as account/password protection, password complexity check, and anti-brute force cracking of passwords.
  - Attack defense: A CPE must be able to defend against various network attacks, such as IP flood attacks, ICMP flood attacks, malformed packet attacks, and fragment attacks.
  - Security audit: A CPE must have a comprehensive log system to log all configuration operations and abnormal status during system running for future audit.

HUAWEI

Service Security Design: Internet Access Security

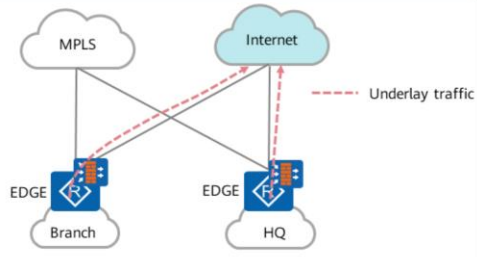
- Huawei SD-WAN Solution uses the built-in security capabilities of CPEs — including ACL filtering, firewall, IPS, and URL filtering — to ensure service security, thereby flexibly meeting service security requirements in different scenarios.

Centralized Internet access



- Internet access traffic from all branch sites is diverted to the HQ site and then goes to the Internet.
- Enable the firewall function on the CPE at the HQ site to isolate internal and external networks.

Local Internet access



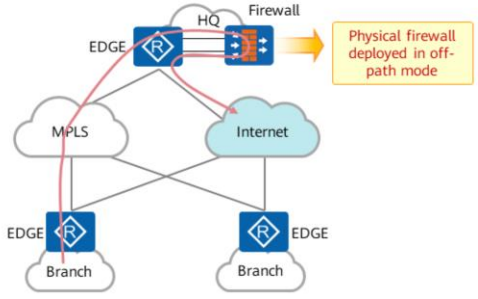
- Internet access traffic from the branch site and HQ site directly goes to the Internet through their respective local CPE.
- Enable the firewall function on the CPEs at the branch site and HQ site to isolate internal and external networks.

HUAWEI

Service Security Design: VAS Advanced Security

- In the centralized Internet access scenario, when a site accesses the Internet through a WAN-side outbound interface, a physical firewall can be deployed in off-path mode to provide VAS-based advanced security protection functions.

VAS advanced security



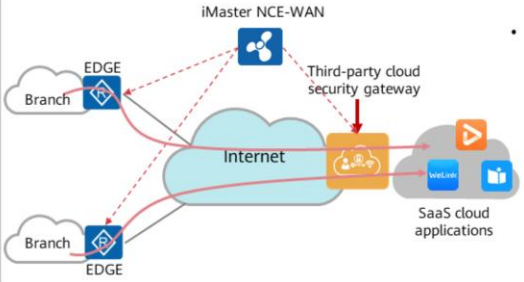
- A physical firewall is deployed in off-path mode. It provides advanced security protection functions for centralized Internet access traffic from the site.
- After VASs are deployed, centralized Internet access traffic is diverted to the physical firewall on the LAN side of the HQ CPE. After being processed by the firewall, centralized Internet access traffic then goes to the Internet through the underlay network.

HUAWEI

Service Security Design: Third-Party Cloud Security

- Huawei SD-WAN Solution can connect to third-party cloud security gateways (Zscaler and Forcepoint). In this way, enterprises' public cloud access traffic and SaaS traffic are sent to such gateways. Third party cloud security gateways perform security check for the traffic and provide access control, threat prevention, and data protection functions, implementing security protection.

Application scenario



- To improve network reliability, the third-party cloud security gateway generally provides two access points to establish tunnels with the EDGE, and the EDGE uses GRE tunnels to connect to such access points.
  - Single-gateway scenario
    One CPE establishes two GRE tunnels (active and standby) with the two access points of the third-party cloud security gateway.
  - Dual-gateway scenario
    Two CPEs establish four GRE tunnels (active and standby) with the two access points of the third-party cloud security gateway.

HUAWEI

Section Summary

- Reliability and security design for Huawei SD-WAN Solution covers site reliability design, controller reliability design, and security design.
  - Site reliability design: hub site reliability design, RR site reliability design, and IWG site reliability design.
  - Controller reliability design: controller deployment reliability design and controller networking reliability design.
  - Security design: system security design and service security design.

HUAWEI

Quiz

1. (Multiple-answer question) Which of the following overlay network topology types are supported by Huawei SD-WAN Solution? ( )
   A. Hub-spoke
   B. Full-mesh
   C. Partial-mesh
   D. Hierarchical networking
2. (True or false) The intelligent traffic steering function can dynamically adjust data forwarding links based on link quality. ( )
   A. True
   B. False

HUAWEI

1. ABCD
2. A

Summary

- Huawei SD-WAN Solution efficiently addresses the challenges facing today's WANs.
- Huawei SD-WAN Solution provides abundant functions, including ZTP, flexible networking, intelligent traffic steering, and service security.
- Design for Huawei SD-WAN Solution mainly covers three parts:
  - Networking design: site design, tunnel design, and VPN design
  - Service design: application service design and network service design
  - Reliability and security design: site reliability design, controller reliability design, and security design

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Enterprise Bearer WAN Solution

HUAWEI

Foreword

- A wide area network (WAN) is a computer network that connects local area networks (LANs) or metropolitan area networks (MANs) in different regions. A WAN allows information and network resources to be shared in a large scope.
- An enterprise IP bearer WAN is a backbone WAN used to implement cross-region communication inside an enterprise. In enterprise network scenarios, various sectors, such as government, finance, education, and power, widely use IP bearer WANs to connect sites and clouds in different geographical locations, facilitating digitalization.
- This course first describes basic WAN concepts and the evolution of WAN bearer technologies, and then introduces Huawei's CloudWAN solution and key technologies.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the basic concepts of the WAN.
  - Describe the trend, challenges, and evolution of the IP bearer WAN.
  - Describe Huawei's CloudWAN solution.
  - Describe the key technologies of Huawei's CloudWAN solution.
  - Describe the typical industry application scenarios of Huawei's CloudWAN solution.

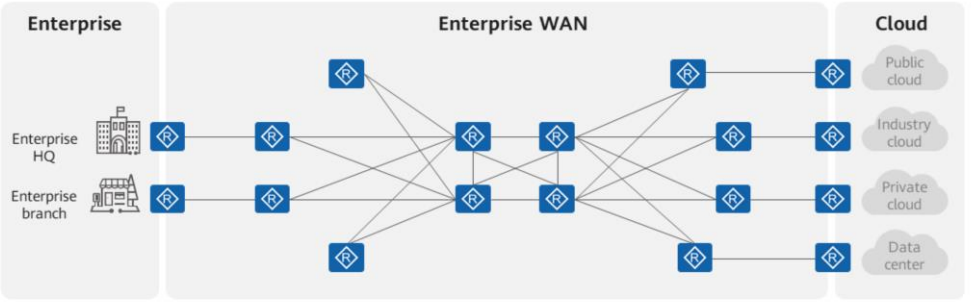
HUAWEI

Contents

1. **Enterprise IP Bearer WAN Overview**
2. CloudWAN Solution Overview
3. Typical Application Scenarios of the CloudWAN Solution

HUAWEI

Enterprise WAN Overview



Enterprise WAN definition

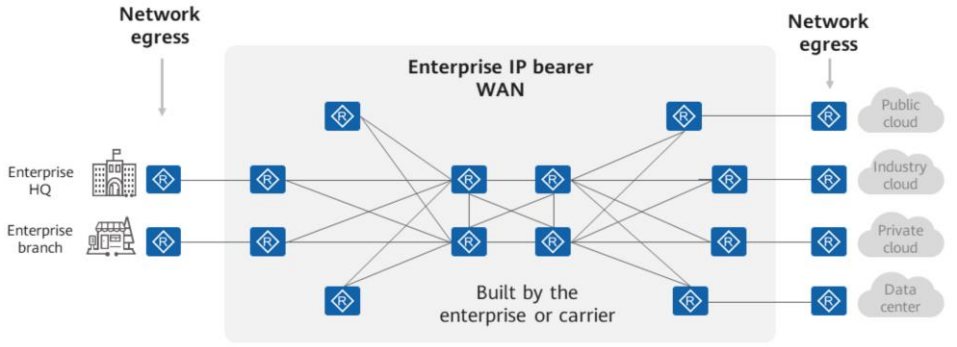
- The enterprise WAN is used to implement cross-region communication inside an enterprise.
- The enterprise WAN provides interconnection between the enterprise HQ and branches, between the enterprise and clouds, and between clouds.

Classification by purpose

- Self-built, for internal use
- Self-built, for external use

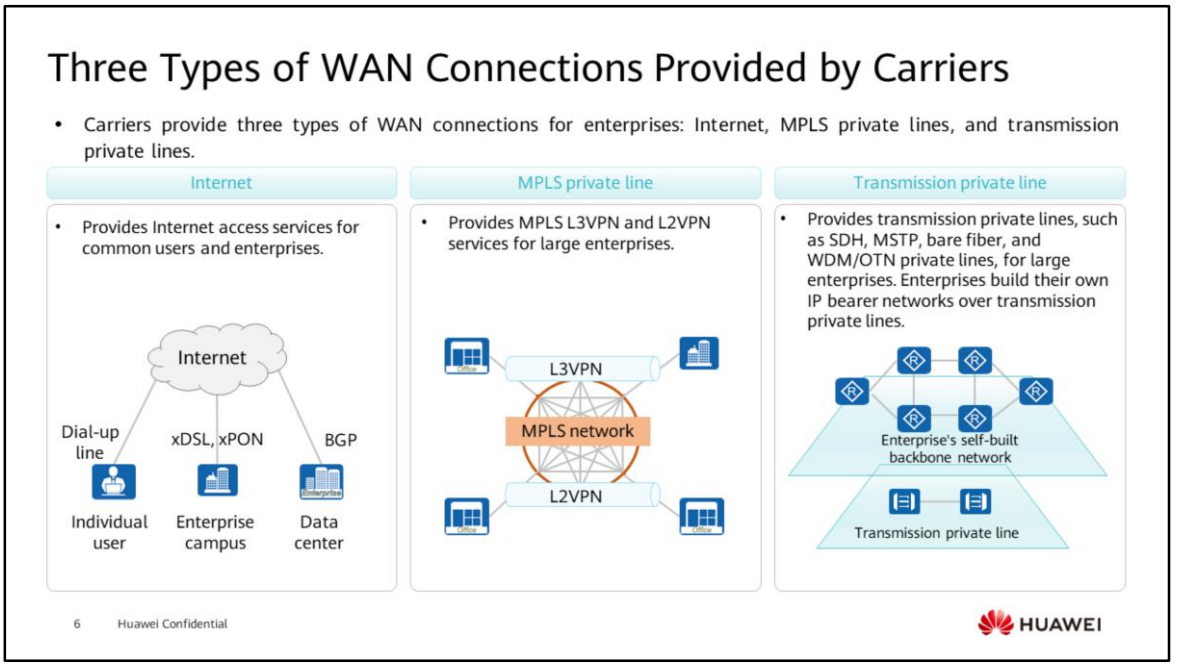
HUAWEI

Enterprise IP bearer WAN



Enterprise WAN = Enterprise network egress + enterprise IP bearer WAN (built by the enterprise or carrier)

HUAWEI

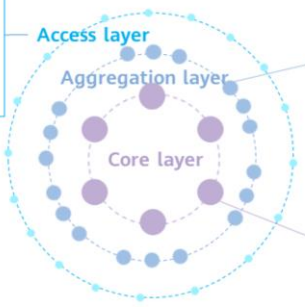


- Internet:
  - Site-to-Internet private line: Ethernet private line/xPON private line/xDSL private line. The access is restricted by geographic locations. It applies to inter-enterprise communication over Internet-based encrypted tunnels.
  - Dial-up connection: low bandwidth and low tariff. The access is not restricted by geographical locations. It applies to individual users.
  - BGP: applies to data center Internet egresses.
- As technologies are constantly developing, historical WAN technologies such as T1/E1, PSTN, ATM, and frame relay are not described here.

Typical Logical Architecture of the Enterprise IP Bearer WAN

- The typical architecture of the enterprise bearer WAN is divided into three layers: access layer, aggregation layer, and core layer.

Access layer (CE-PE):
- Provides access for data centers in different cities.
- Provides access for branches in different cities.
- Provides access for external services in different cities.



Aggregation layer (PE-P):
- Aggregates and transmits different types of services to the core layer based on physical locations.

Core layer (P):
- Generally adopts the Full-mesh + dual-plane architecture.
- It forwards traffic between different regions over stable, reliable, and service quality-guaranteed connections.

HUAWEI

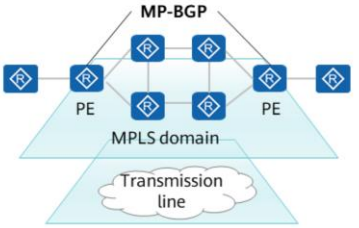
Typical Bearer Technologies of the Enterprise IP Bearer WAN

IP bearer

- All services are carried over the same IP network, and IP addresses are used to differentiate services.
- This bearer mode applies to small- and medium-sized enterprises or networks that do not have specific service isolation, differentiated SLA, or traffic engineering requirements.

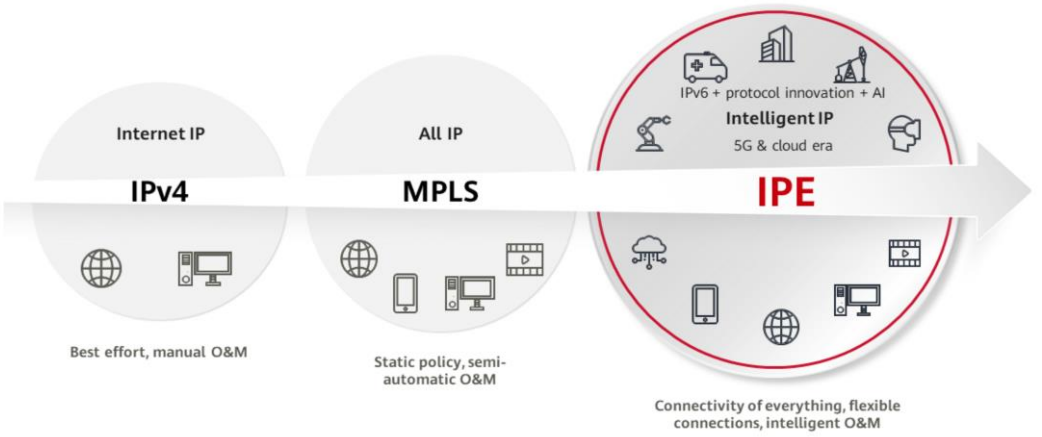
MPLS bearer

- Data is encapsulated into MPLS packets and forwarded based on labels.
- VPNs are used to differentiate services. Service isolation and SLA assurance are supported.
- The WAN VPN architecture consists of the data plane and control plane. The data plane uses MPLS, and the control plane uses MP-BGP. This bearer mode applies to large enterprises.



HUAWEI

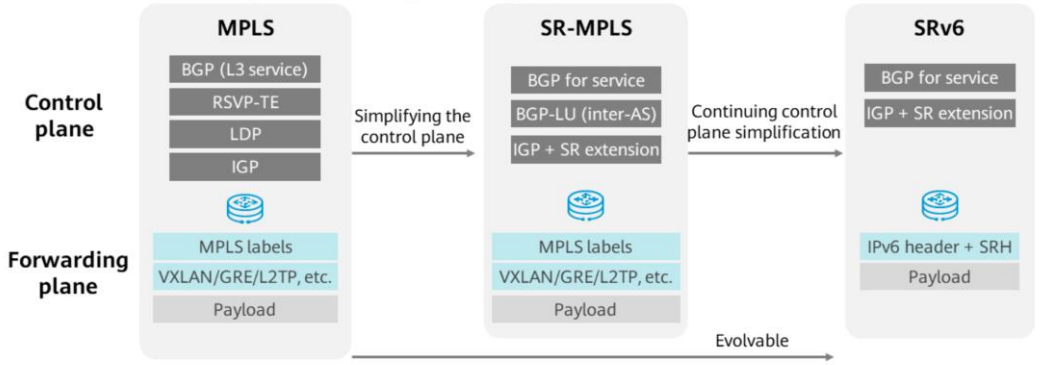
Enterprise IP Bearer WAN Moving Towards the IPv6+ Era



HUAWEI

Enterprise IP Bearer WAN Technologies Evolving Towards SRv6

- With the development of technologies and service requirements, VPN becomes the mainstream bearer technology adopted by the WAN. The control and forwarding plane technologies of the WAN keep evolving.
- The bearer WAN continuously evolves towards segment routing (SR) and IPv6.



HUAWEI

- BGP Labeled Unicast (BGP-LU) (RFC 3017) is both an inter-AS and an intra-AS routing protocol.

Contents

1. Enterprise IP Bearer WAN Overview
2. **CloudWAN Solution Overview**
3. Typical Application Scenarios of the CloudWAN Solution

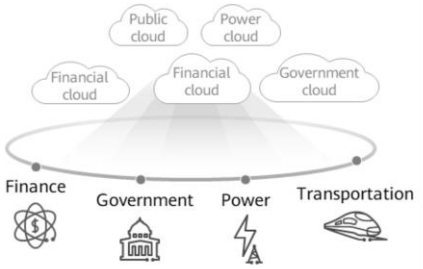
HUAWEI

Two Major Changes Brought by Digital Transformation

1. Cloudification of millions of enterprises

Cloud adoption by enterprises
Local processing > Service cloudification

Multi-cloud adoption by enterprises
Private cloud > Hybrid cloud



2. IP-based production network

Multiple TDM private networks > One IP bearer network
IEC and UIC propose IP-based transformation for power and transportation sectors, respectively.

Power relay protection
Traditional relay protection > Wide-area relay protection

- **Delay < 5 ms**
- **Multicast technology**

Train control and dispatching
Manual monitoring > Over-the-horizon monitoring

- **Bandwidth > 100 Mbit/s**
- **Clock precision < 3 µs**

Fuel pipe safety detection
Manual inspection > UAV inspection

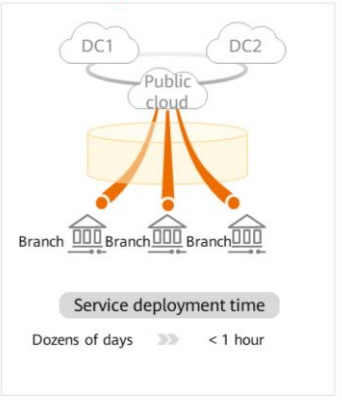
- **Any access, flexible connection**

HUAWEI

3 Challenges Facing the WAN in the Cloud Era

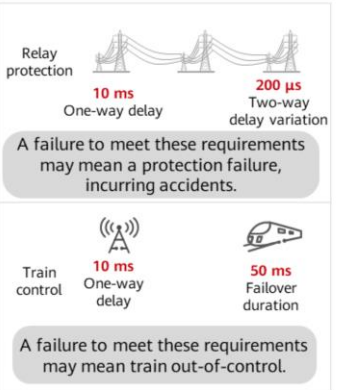
Cloudification of millions of enterprises

How can networks be as agile as clouds?



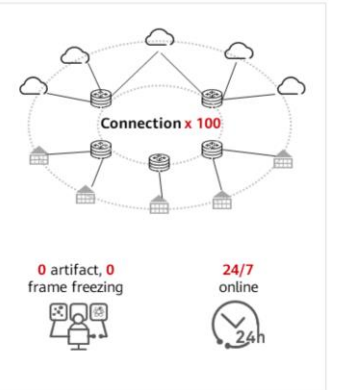
Production service bearer

Can IP provide deterministic experience?



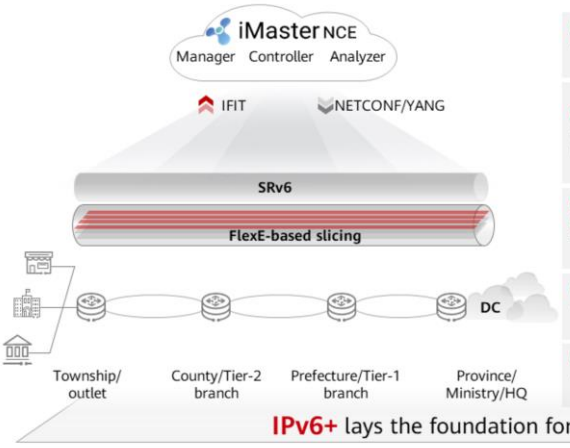
Connection scale x 100 ↑

How can O&M be simpler and networks more reliable?



HUAWEI

CloudWAN 3.0: Leading WANs into the Intelligent Cloud-Network Era



One-hop cloud access: flexible cloud-network connection
- SRv6 enables service provisioning within minutes and agile service cloudification.

One-fiber multipurpose transport: deterministic experience
- Hierarchical slicing
- Patented fingerprint-based slicing technology, simplifying deployment

One-click fast scheduling: cloud-network coordinated scheduling
- SDN + intelligent cloud-map algorithm, improving cloud-network resource utilization

One-network wide connection: network digitalization
- Hop-by-hop detection technology, real-time visualization of network-wide status, troubleshooting within minutes

Integrated security, all-round security protection
- Qiankun security cloud service, proactively identifying cyber security threats

IPv6+ lays the foundation for the digital infrastructure.

HUAWEI

CloudWAN 3.0: Management, Control, and Analysis Platform + Intelligent Universal Service Routers for the Cloud Era

Network management

iMaster NCE = U2000 (management) + Controller (control) + uTraffic (analysis)

Metro router

NetEngine 9000 Backbone router

NetEngine 9000-20

NetEngine 40E universal service router

NetEngine40E-X16A

NetEngine40E-X8A

NetEngine 8000 Smart router

NetEngine 8000 M8

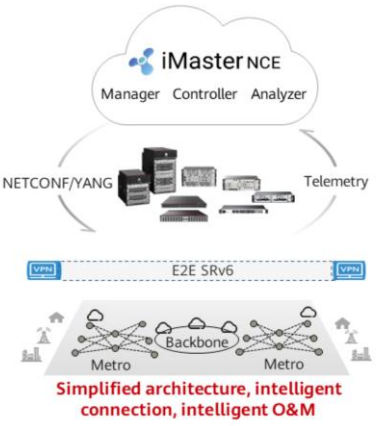
NetEngine 8000 M6

NetEngine 8000 F1A

NetEngine 8000 M1A/M1C

HUAWEI

CloudWAN 3.0: Management, Control, and Analysis Platform iMaster NCE-IP



Simplified architecture, intelligent connection, intelligent O&M

Management

- **NE management:** topology, alarm, configuration, and inventory management
- **Service management**: tunnel and VPN service management

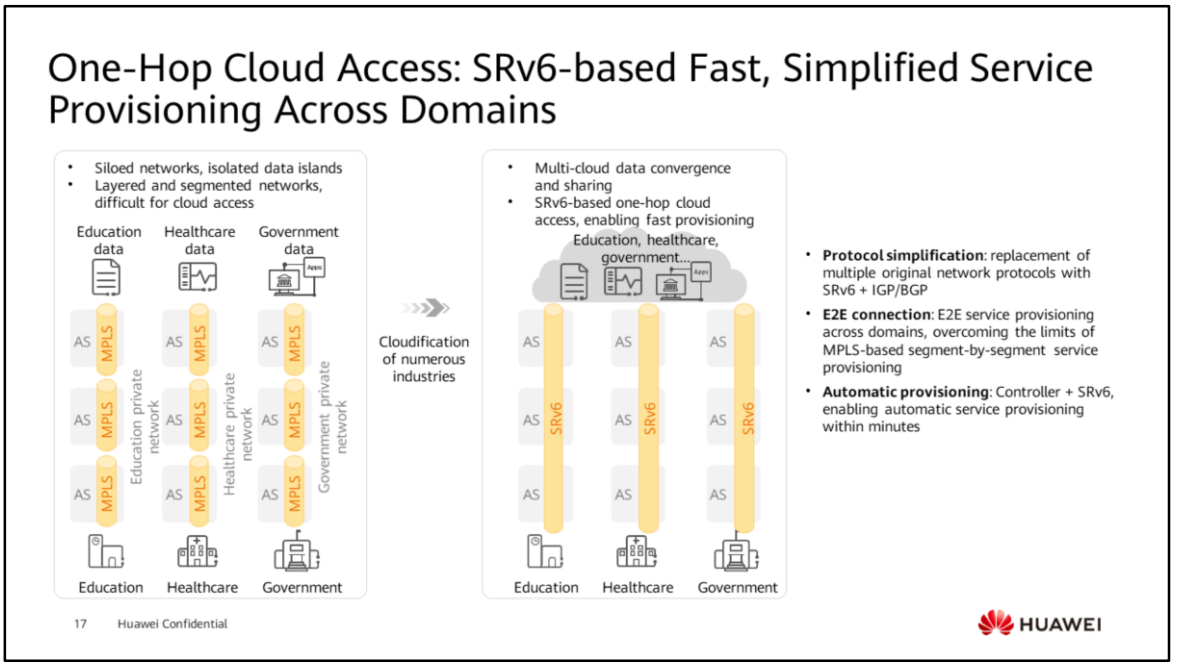
Controller

- **Centralized path computation**: path computation based on multiple constraints
- **Logical topologies**: cost, delay, and bandwidth topologies
- **Network optimization**: service path adjustment and optimization

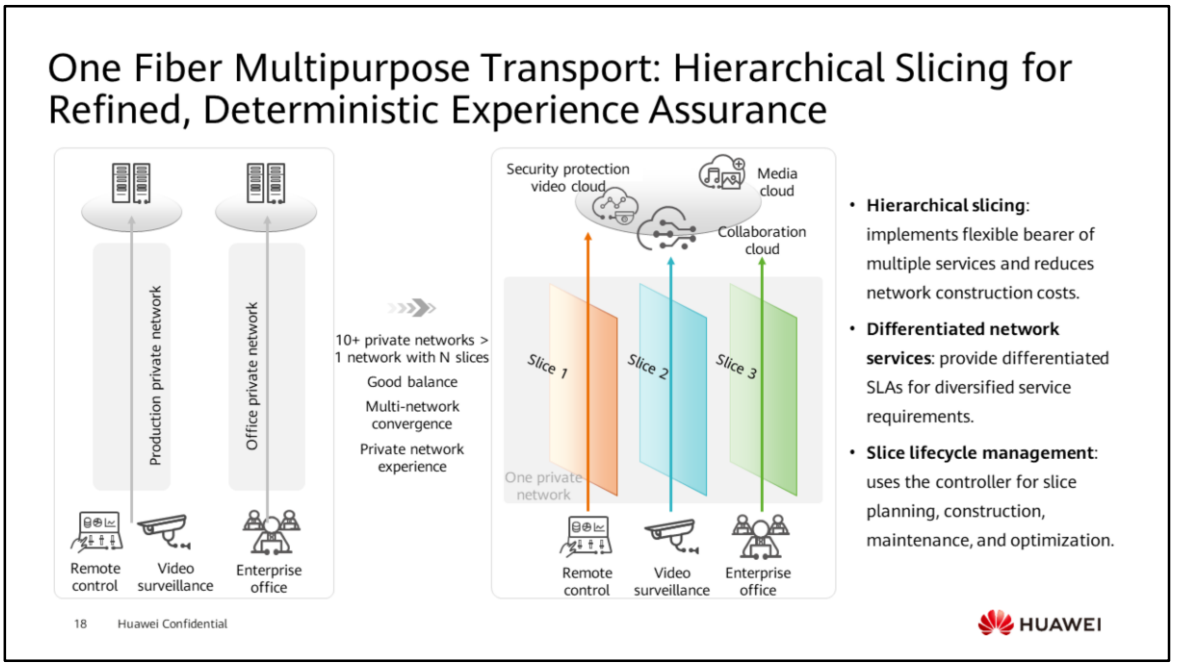
Analysis

- **Basic network analysis**: display and analysis of performance, traffic, and quality
- **Analysis-based prediction**: traffic, fault, and exception prediction

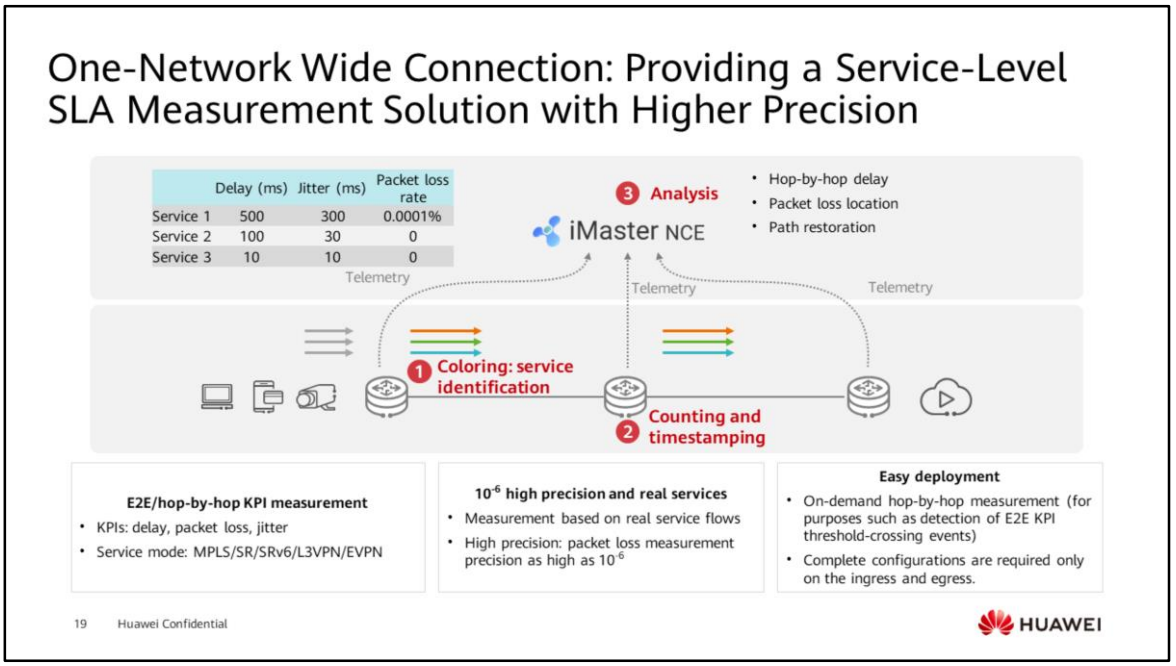
HUAWEI



- In the past, most of our networks were siloed private networks, such as education, healthcare, and government private networks. These networks were independent physical private networks and could not communicate with each other. The handling of some services may involve multiple private networks. Moreover, a service may be deployed segment by segment even on one private network. For example, multiple ASs may exist on a network due to the division of administrative domains, and one network service may be deployed across ASs (on a common network, service data is generally carried over MPLS). In this situation, a large number of device configurations and personnel communication are required. The network administrator needs to perform a large number of configurations on AS boundary devices, and it takes a long time to migrate the service to the cloud. The acceleration of enterprise digital transformation drives alignment between networks and clouds.
- Now, increasingly more industries are deploying data to the cloud, making it easier to converge or share data. The introduction of SRv6 can remove process barriers and accelerate service provisioning. Simply put, SRv6 can be deployed on both ends of an SRv6 tunnel to implement one-hop cloud access.
- In the MPLS era, a large number of control-plane protocols, such as IGP, BGP, LDP, and RSVP-TE, are required to carry VPN services on a network or implement traffic engineering. On the forwarding plane, there are protocols such as MPLS, GRE, and L2TP or native IP. The network configuration and configuration modification are complex. Huawei's CloudWAN solution simplifies network deployment by replacing multiple network protocols with SRv6+IGP/BGP. SRv6 uses IPv6 as the forwarding plane protocol. On a WAN where IPv6 is deployed, it is easy to deploy an end-to-end tunnel, even in inter-AS scenarios. The SDN controller can be used to implement automated SRv6 service provisioning within minutes.

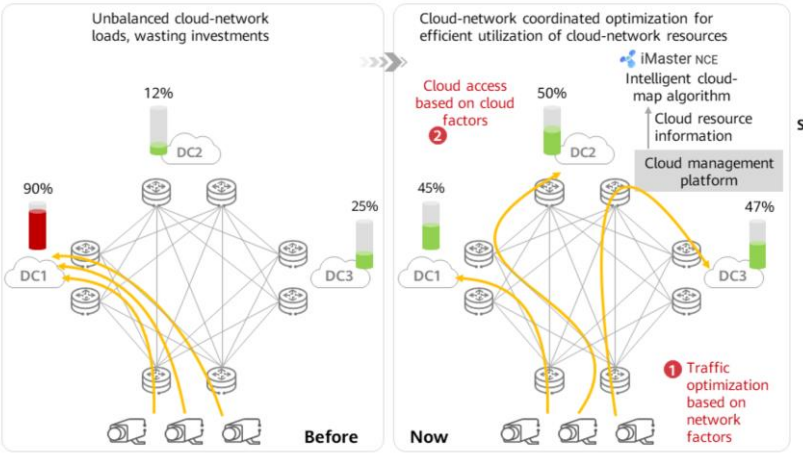


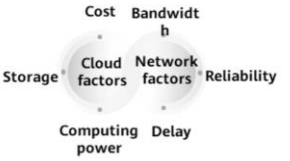
- Huawei uses the hierarchical slicing technology to power IP-based production networks, ensuring deterministic SLAs for production services.
- In the past, production and office services were carried over multiple independent private networks. Repeated network construction resulted in high investment costs and complex O&M of multiple networks. By deploying multiple slices on one IP bearer network, production services such as remote industrial control and video surveillance are directly isolated from office services, delivering 100% bandwidth guarantee for mission-critical services.



- iFIT integrates the RFC 8321 coloring technology and in-band detection technology to directly measure service packets. It works with second-level telemetry data collection and iMaster NCE for unified management, computation, and visualization. In this way, it implements real-time visualization and proactive monitoring of network quality SLAs and fast fault demarcation and locating.

One-Click Fast Scheduling: Cloud-Network Coordinated Scheduling, Improving Cloud-Network Resource Utilization





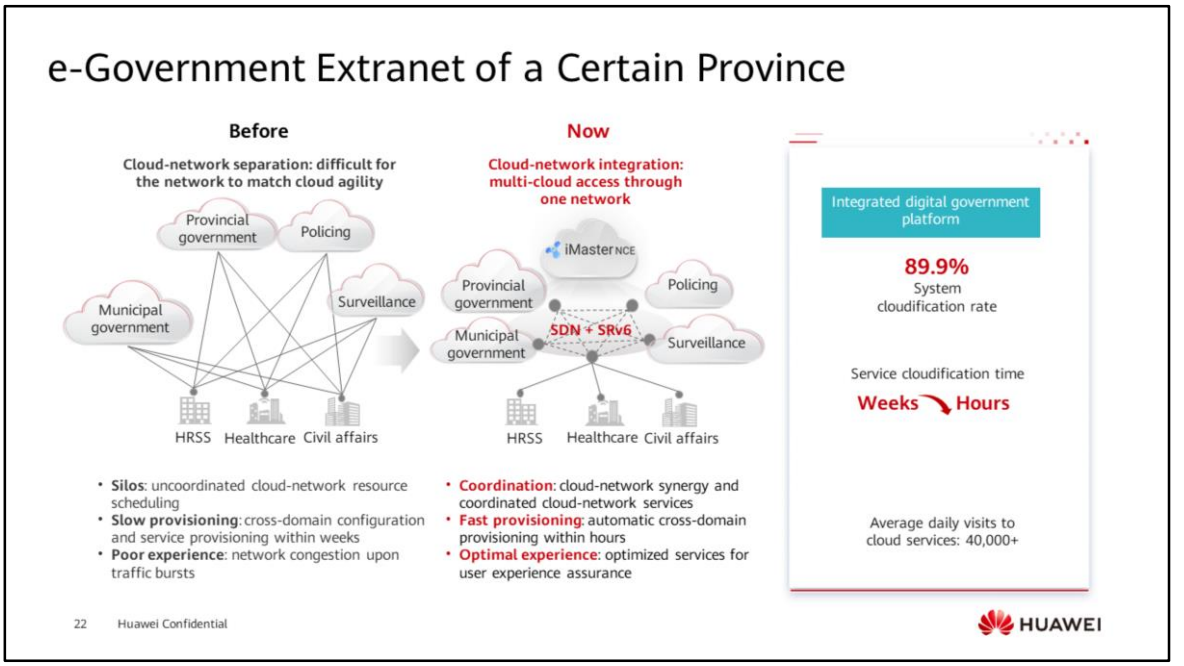
- The intelligent cloud-map algorithm combines the cloud and network factors.
- Load balancing is performed based on service conditions and cloud loads. Services are intelligently scheduled to the most appropriate cloud based on SRv6 and SDN technologies.

HUAWEI

Contents

1. Enterprise IP Bearer WAN Overview
2. CloudWAN Solution Overview
3. **Typical Application Scenarios of the CloudWAN Solution**

HUAWEI

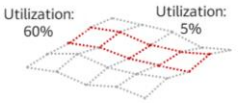


- Quick Network Adjustment upon Cloud Changes, Integrated Service Provisioning, Cloud-based Data Sharing.

Intelligent Traffic Optimization on an Enterprise WAN

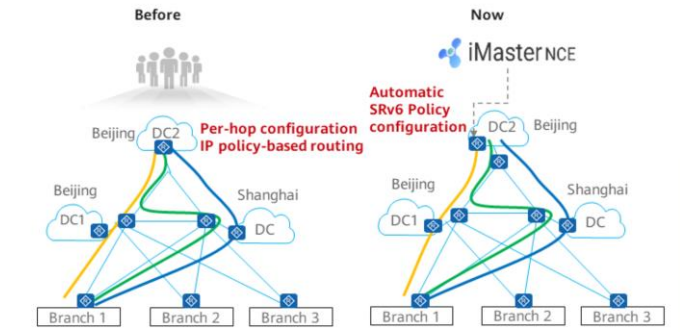
Service scenarios

1. The average traffic between the HQ and branches increases by about **20%** annually, but the annual private line leasing budget is only allowed to increase by **5%**.
   The average private line utilization is only about **25%**.
2. Enterprise data synchronization services **run at night** and require **ultra-large bandwidth**. On traditional networks, traffic load balancing is difficult to implement due to the **shortest path first principle**.



"We hope that traffic optimization can be automatically or manually triggered based on factors such as time range, traffic threshold, and traffic burst."
— Senior network architect

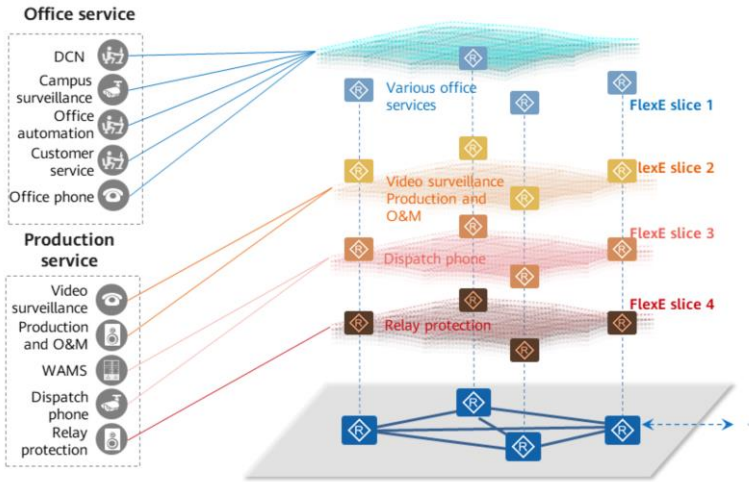
Solution



1. Policy-based routing is manually configured hop by hop. A single optimization operation takes more than 2 hours and is prone to errors.
2. Traffic optimization often needs to be performed at night and requires the attendance of dedicated personnel.

HUAWEI

Multi-Service Bearer Through Network



Physical network slicing: TDM-based FlexE divides a physical network into multiple slices on demand to isolate different services and ensure service SLAs.

Flexible, convenient service deployment

Unified bearer: NCE is used to deploy network resources in a unified manner for power service transport, achieving intelligent optimization.

Support for network evolution over the next **10** years:
Smooth upgrade to 100/200 Gbit/s
Smooth evolution to SRv6/SR/EVPN

HUAWEI

Quiz

1. (Single-answer question) In Huawei's CloudWAN solution, which of the following technologies can absolutely ensure the bandwidth of key services?( )
   A. SRv6 Policy
   B. FlexE-based network slicing
   C. iFIT
   D. Telemetry

HUAWEI

1. B

Summary

- An enterprise IP bearer WAN is a backbone WAN used to implement cross-region communication inside an enterprise. In enterprise network scenarios, various sectors, such as government, finance, education, and power, widely use IP bearer WANs to connect sites and clouds in different geographical locations, facilitating digitalization.
- Bearer WAN technologies evolve from MPLS to SRv6. In the cloud era, networks are expected to meet requirements regarding visualization, awareness, optimization, deterministic delay, openness, and programmability.
- Huawei's CloudWAN solution meets all the preceding requirements. We will explore more about this solution in subsequent learning.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Enterprise Bearer WAN Architecture and Key Technologies

HUAWEI

Foreword

- The IP bearer WAN, which usually covers a country, is a data communication network that provides interconnection between multiple LANs or branch networks across regions.
- This course introduces the concepts, principles, and applications of the enterprise bearer WAN's typical architecture, bearer technologies, VPN services, traffic optimization, SLA, reliability, and network management and analysis. To introduce these key aspects, this course uses a large enterprise with three data centers in two cities and multiple branches in different regions as an example.

HUAWEI

Objectives

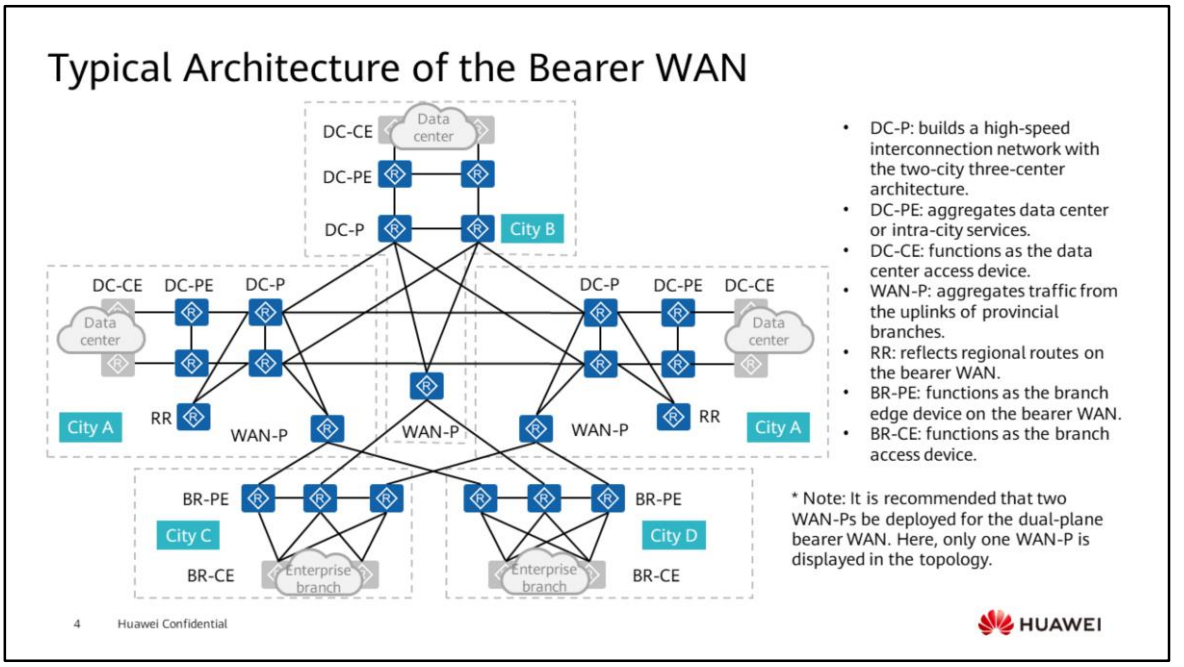
- On completion of this course, you will be able to:
  - Describe the typical architecture of the bearer WAN.
  - Describe the basic concepts of Multiprotocol Label Switching (MPLS), Segment Routing-Multiprotocol Label Switching (SR-MPLS), and Segment Routing IPv6 (SRv6).
  - Describe the principles of WAN traffic optimization.
  - Describe the SLA model and packet processing flow.
  - Describe the network reliability design.
  - Describe the key protocols for network management and analysis.

HUAWEI

Contents

1. **Bearer WAN Architecture**
2. Bearer WAN Basics
3. VPN Service
4. Network Traffic Optimization
5. SLA
6. Network Reliability
7. Network Management and O&M

HUAWEI

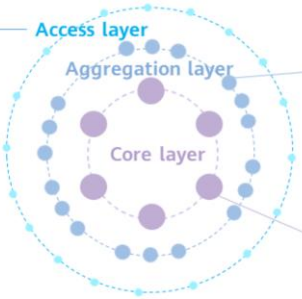


- The Ps for intra-city data center are directly connected using WDM or bare optical fibers. The link bandwidth can reach 10 Gbit/s. To reduce costs, consider connecting local data centers with remote data centers through carriers' MSTP links.

Typical Architecture of the Bearer WAN

- The typical architecture of a bearer WAN is divided into three layers: access layer, aggregation layer, and core layer.

Access layer (CE-PE):
- Provides access for data centers in different cities.
- Provides access for branches in different cities.
- Provides access for external services in different cities.

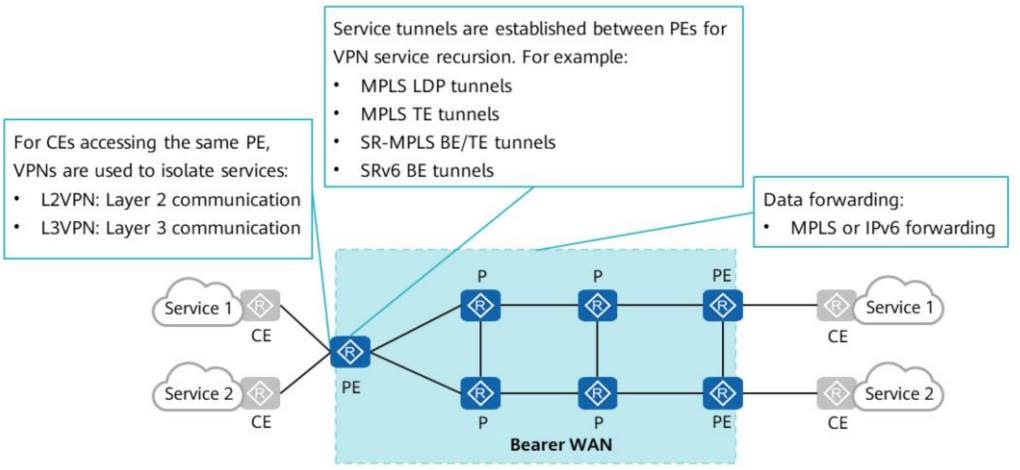


Aggregation layer (PE-P):
- Aggregates and transmits different types of services to the bearer WAN based on physical locations.

Core layer (P):
- Generally adopts the full-mesh + dual-plane architecture.
- Provides stable, reliable high-speed traffic forwarding between regions.

HUAWEI

Technologies Used by the Typical Bearer WAN Architecture



HUAWEI

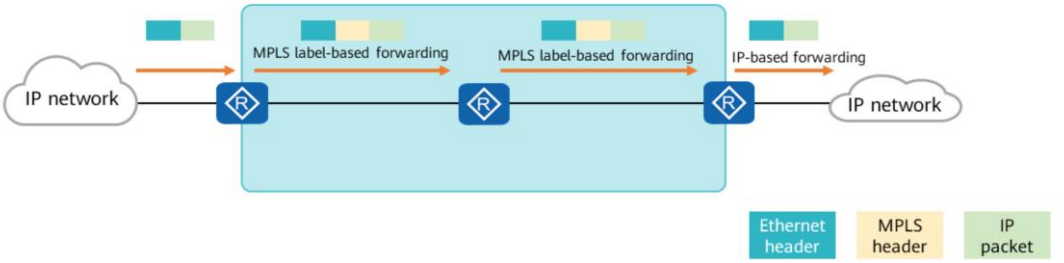
Contents

1. Bearer WAN Architecture
2. **Bearer WAN Basics**
3. VPN Service
4. Network Traffic Optimization
5. SLA
6. Network Reliability
7. Network Management and O&M

HUAWEI

MPLS Overview

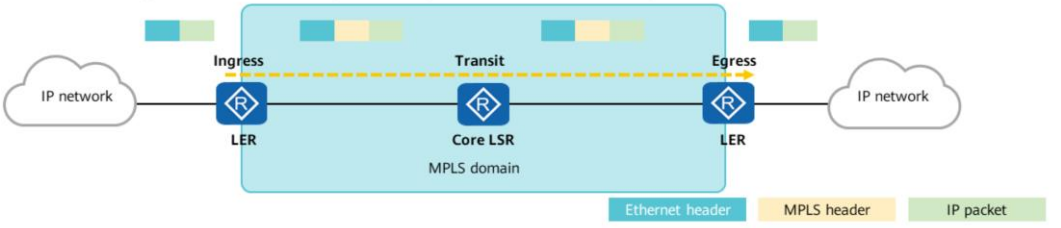
- MPLS is located between the data link layer and the network layer in the TCP/IP protocol stack and can provide services for all network layers.
- An MPLS header is added between a data-link-layer header and a network-layer header, and data can be forwarded quickly based on the MPLS header.



HUAWEI

MPLS Terms (1)

- MPLS domain: consists of a series of consecutive network devices that run MPLS.
- Label switching router (LSR): a routing device, such as a router or switch, that runs MPLS. An LSR that resides at the edge of an MPLS domain and connects to a non-MPLS network is called a label edge router (LER). An LSR that resides inside an MPLS domain is called a core LSR.
- The path that MPLS packets take in an MPLS network is called a label switched path (LSP). The LSP is a unidirectional path that transmits traffic from the ingress to the egress.
- The start node of an LSP is called the ingress, an intermediate node of the LSP is called the transit node, and the end node of the LSP is called the egress. An LSP has one ingress, one egress, and zero, one, or multiple transit nodes.

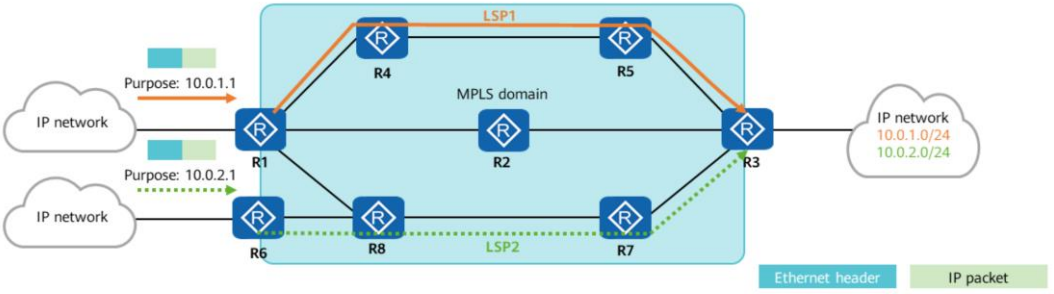


HUAWEI

MPLS SR-MPLS SRv6

MPLS Terms (2)

- Forwarding equivalence class (FEC): a set of packets with similar or identical characteristics and forwarded in the same way by LSRs. In traditional IP forwarding that uses the longest match algorithm, all packets that match the same route belong to the same FEC.
- An LSP is composed of an ingress LSR, an egress LSR, and a variable number of transit LSRs. Therefore, an LSP can be considered as an ordered set of these LSRs.
- An LSP must be established before a packet is forwarded; otherwise, the packet fails to traverse an MPLS domain.
- An LSP is a unidirectional path from the start point to the end point. If bidirectional data communication is required, an LSP for return traffic needs to be established between the two ends.



HUAWEI

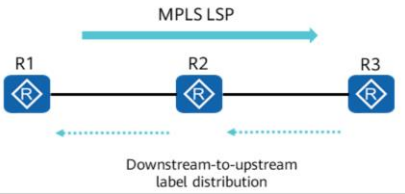
- For more information about MPLS, see HCIP-Datacom-Advanced Routing & Switching Technology.

MPLS Terms (3)

- LSPs can be statically configured or dynamically established.
- There are two common protocols for dynamically establishing LSPs: LDP and Resource Reservation Protocol-Traffic Engineering (RSVP-TE).

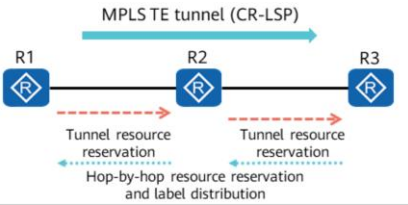
LDP: used to establish common LSPs

- LDP distributes labels from downstream routers to upstream routers based on the routing table to set up common LSPs.
- A common LSP is usually the shortest path calculated by an IGP, which does not factor in aspects such as bandwidth, tunnel protection, and traffic optimization.



RSVP-TE: used to establish CR-LSPs

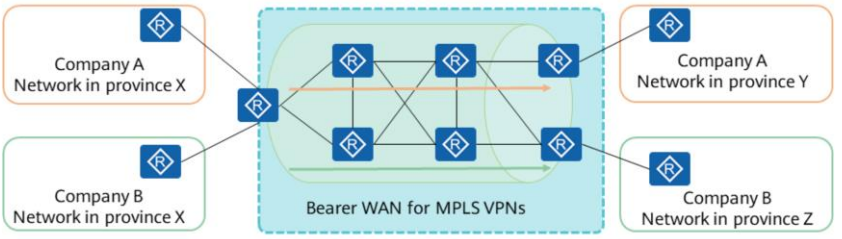
- RSVP-TE establishes tunnels by applying for and reserving tunnel resources end to end.
- An LSP that is set up based on bandwidth or path constraints is called a constraint-based routed label switched path (CR-LSP).



HUAWEI

MPLS LDP Overview

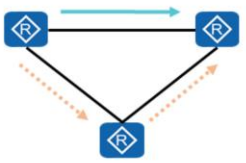
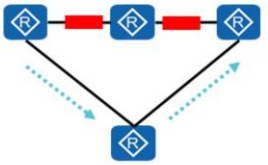
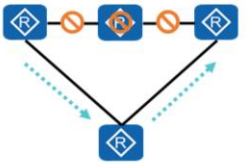
- LDP is a control protocol of MPLS and provides functions such as FEC classification, label distribution, and LSP establishment and maintenance. LDP defines the messages used in label distribution as well as the message processing procedures.
- LDP is easy to configure and maintain and is widely used to create LSPs in BGP/MPLS IP VPN scenarios. As shown in the figure, the carrier builds a bearer WAN for MPLS VPNs to provide inter-provincial L3VPN services for customers.
- MPLS LDP LSPs are established based on the shortest IP paths and do not support the planning of tunnel forwarding paths.



HUAWEI

MPLS TE Overview

- MPLS TE, as its name suggests, is a combination of MPLS and TE. It provides functions such as path planning, traffic optimization, and fault protection for MPLS VPN services.
- Compared with MPLS LDP, MPLS TE enhances VPN traffic control and protection.

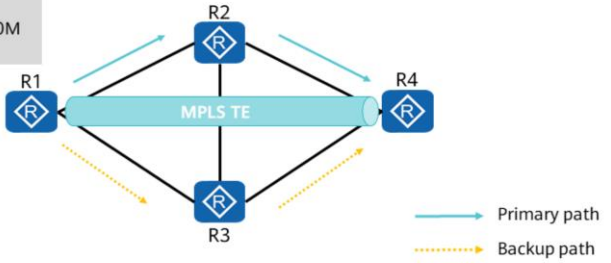
Path planning	Traffic optimization	Fault protection
		
• Different paths are planned for different services.	• Traffic is evenly distributed to idle links when unbalanced traffic distribution occurs due to major events, such as live broadcast.	• Fast protection switching is performed in the case of a device or link fault.

HUAWEI

MPLS TE Tunnel

- MPLS TE often associates multiple LSPs with a virtual tunnel interface, and such a group of LSPs is called an MPLS TE tunnel.
- An MPLS TE tunnel provides SLA assurance, but requires complex configuration and manual planning.

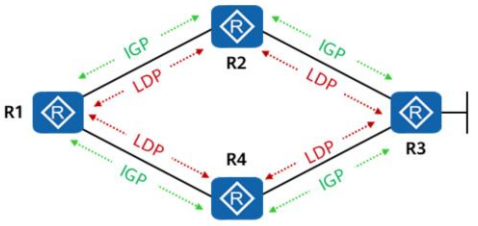
```
[R1] interface tunnel1
[R1-Tunnel1] ip address ...
[R1-Tunnel1] tunnel-protocol mpls te
[R1-Tunnel1] destination R4
[R1-Tunnel1] mpls te bandwidth 50M
...
```



HUAWEI

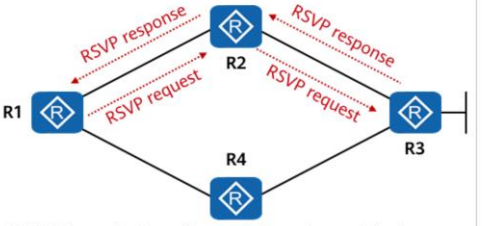
Issues with MPLS LDP and RSVP-TE

MPLS LDP



- LDP itself does not have the path computation capability and requires an IGP for path computation.
- Both the IGP and LDP need to be deployed for the control plane, and devices need to exchange a large number of packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- If LDP-IGP synchronization is not achieved, data forwarding may fail.

RSVP-TE

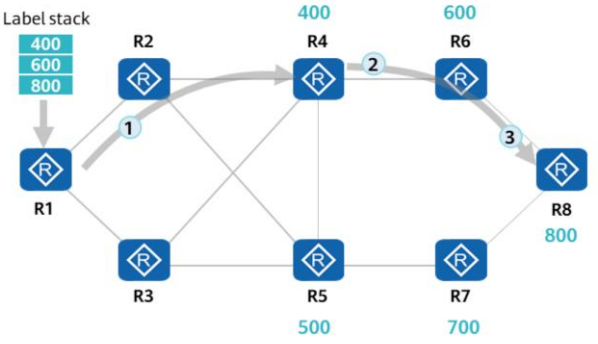


- RSVP-TE is complex to configure and does not support load balancing.
- To implement TE, devices need to exchange a large number of RSVP packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- RSVP-TE uses a distributed architecture, so that each device only knows its own state and needs to exchange signaling packets with other devices.

HUAWEI

SR-MPLS Overview

- SR is designed to forward data packets on a network using the source routing model.
- SR-MPLS, as its name suggests, is SR based on MPLS label forwarding.



- Source routing:
  - The source node selects a path and pushes an ordered label stack into the packet.
  - Other nodes on the network forward the packet according to the label stack encapsulated into the packet.
- SR has the following characteristics:
  - Extends existing protocols (e.g. IGP) to facilitate network evolution.
  - Supports both centralized controller-based control and distributed forwarder-based control, providing a balance between the two control modes.
  - Enables networks to quickly interact with upper-layer applications through the source routing technology.

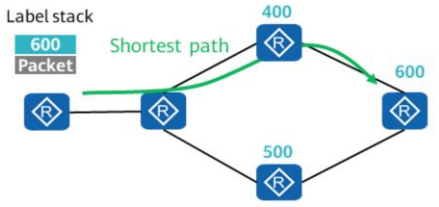
HUAWEI

SR-MPLS BE and SR-MPLS TE

- The SR-MPLS tunneling technology can be implemented in either SR-MPLS BE or SR-MPLS TE mode.

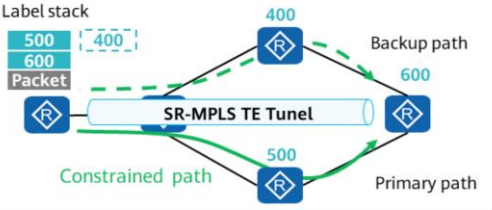
SR-MPLS BE

- Forwarding path: Similar to an MPLS LDP LSP, an SR-MPLS BE LSP is calculated using the IGP shortest path first (SPF) algorithm. An SR-MPLS BE LSP has only one label layer (destination node).
- In the production environment, SR-MPLS BE is generally used as the DR solution for SR-MPLS TE. For example, if a controller fault causes a tunnel delivery failure, the IGP can be used to generate forwarding tunnels.



SR-MPLS TE

- Forwarding path: An SR-MPLS TE path is created using SR based on TE constraints. An SR-MPLS TE tunnel generally uses multiple layers of labels to implement path control and supports primary and backup paths.
- SR-MPLS TE is usually used with a controller. After the controller globally computes a path, it delivers a label stack to the corresponding ingress.



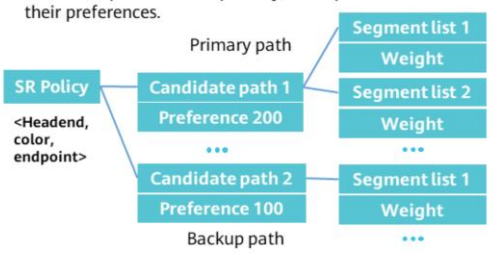
HUAWEI

SR-MPLS Policy

- The SR-MPLS Policy, also called the SR-MPLS TE Policy, is one of the mainstream SR-MPLS implementation modes.
- As defined in the corresponding RFC, an SR Policy is identified by <headend, color, endpoint> and contains multiple candidate paths.

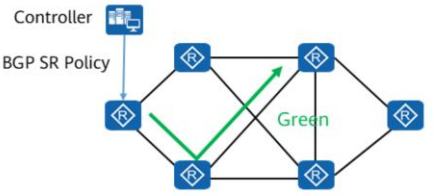
SR Policy model

- A candidate path can contain multiple segment lists, and load balancing can be implemented among these segment lists based on weights.
- Candidate paths work in primary/backup mode based on their preferences.



SR Policy application scenario

- The controller collects information such as the global topology, network bandwidth, and link delay. After computing an SR Policy path based on service requirements, the controller uses BGP to deliver the path to the ingress.

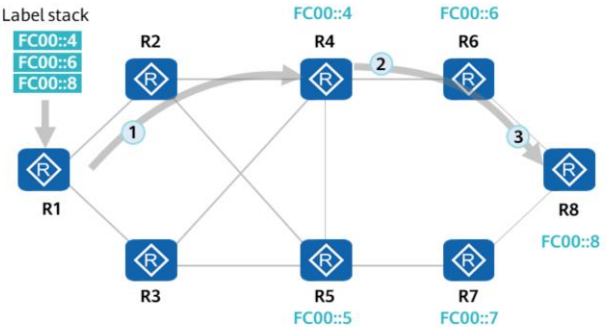


HUAWEI

- Based on MPLS and IPv6 forwarding technologies, SR Policies can be classified into SR-MPLS and SRv6 Policies.

SRv6 Overview

- SRv6 is designed to forward data packets on an IPv6 network using the source routing model.
- Both SRv6 and SR-MPLS comply with the SR architecture. Their main difference lies in data plane instructions. The former is based on the IPv6 network and uses IPv6 addresses as instructions. In contrast, the latter is based on the MPLS network and uses MPLS labels as instructions.



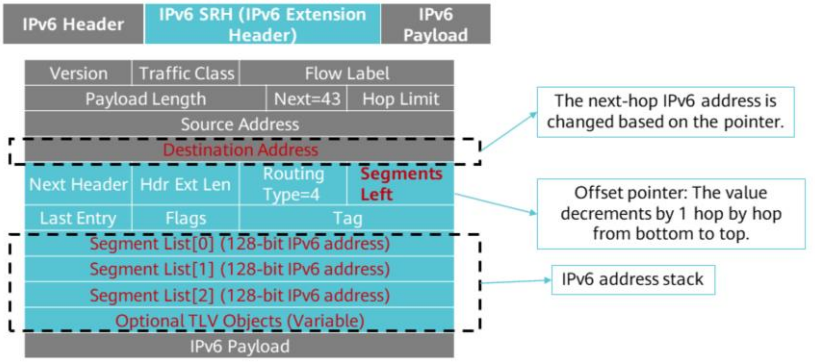
Source routing:

- The source node selects a path and pushes an ordered label stack into the packet.
- Other nodes on the network forward the packet according to the label stack encapsulated into the packet.

HUAWEI

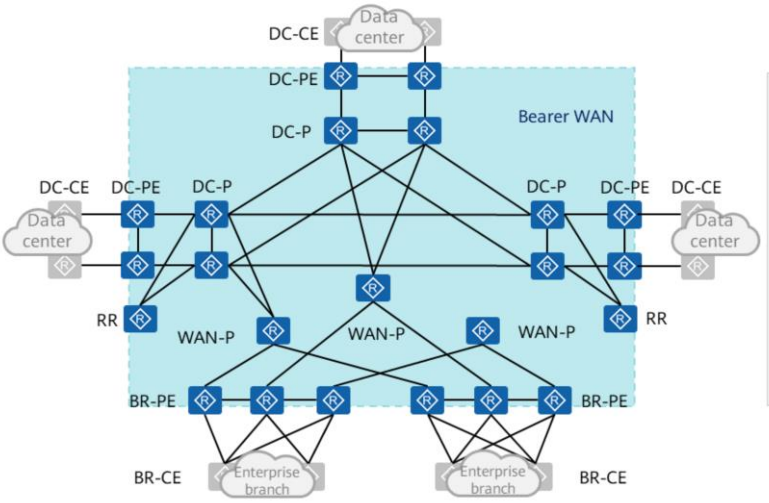
SRv6 Extension Header

- SRv6 adds a segment routing header (SRH) to IPv6 packets. The SRH contains an explicit IPv6 address stack. During the forwarding process, SRv6 nodes continuously update the destination address and offset the address stack to complete hop-by-hop forwarding.



HUAWEI

Summary: WAN Bearer Technologies



Enterprise bearer WANs can be roughly classified into two types: MPLS network and IPv6 network.

- MPLS network:
  - MPLS LDP
  - MPLS TE
  - SR-MPLS
- IPv6 network:
  - SRv6

HUAWEI

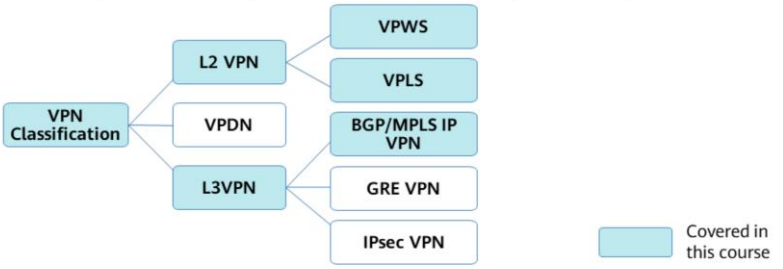
Contents

1. Bearer WAN Architecture
2. Bearer WAN Basics
3. **VPN Service**
   - WAN VPN Overview
     - Tunnel Management Overview
4. Network Traffic Optimization
5. SLA
6. Network Reliability
7. Network Management and O&M

HUAWEI

VPN Classification

- The VPN technology is widely used as a virtual private tunneling technology. VPN can be classified into various types from different perspectives. For example, VPN can be classified into Layer 3 VPN (L3VPN), Layer 2 VPN (L2VPN), and Virtual Private Dial-up Network (VPDN) by implementation layer.
- VPWS, VPLS, and BGP/MPLS IP VPN are more widely used on bearer WANs.
- GRE VPN and IPsec VPN, which are mainly used on the Internet, are beyond the scope of this course.



HUAWEI

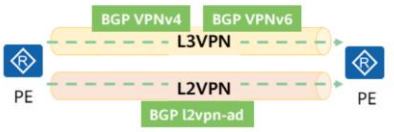
- Traditional switching networks, such as asynchronous transfer mode (ATM) and frame relay (FR) networks, are integrated with IP or MPLS networks. As a result, Layer 2 virtual private network (L2VPN) emerges. L2VPN includes Virtual Pseudo Wire Service (VPWS) and Virtual Private LAN Service (VPLS):
  - VPWS is a P2P L2VPN technology that emulates the basic behaviors and characteristics of services such as ATM and frame relay.
  - VPLS provides P2MP L2VPN services so that sites are connected as if they were on the same LAN.
- Virtual Private Dial-up Network (VPDN) is a virtual private network constructed on the public network. It uses a dedicated network encryption communication protocol to provide access services for international organizations and mobile workforce of enterprises. There are multiple VPDN tunneling protocols, among which Layer Two Tunneling Protocol (L2TP) is the most widely used. Strictly speaking, L2TP is also a type of L2VPN, but its network structure and protocol design are quite different from those of other types of L2VPN. In addition, L2TP uses the dial-up mode. Therefore, L2TP is classified as VPDN.
- L3VPN is also called Virtual Private Routing Network (VPRN), including RFC 2547-based BGP/MPLS IP VPN as well as IPsec VPN and GRE VPN carried over IPsec or GRE tunnels.

WAN VPN Service Overview

- An enterprise establishes a bearer WAN to provide wide-area interconnection for its internal and external services, such as production, office, external connection, and test services. These services are logically isolated but share the same physical network resources. Therefore, these services are called WAN VPN services.
- WAN VPN can be classified into L2VPN and L3VPN.

Traditional L2VPN and L3VPN technologies

- Traditional L2VPN includes VPLS and VPWS, which can use LDP or BGP to establish virtual links.
- Traditional L3VPN uses VPN instances to isolate services and uses BGP VPNv4/v6 to transmit Layer 3 information.



EVPN

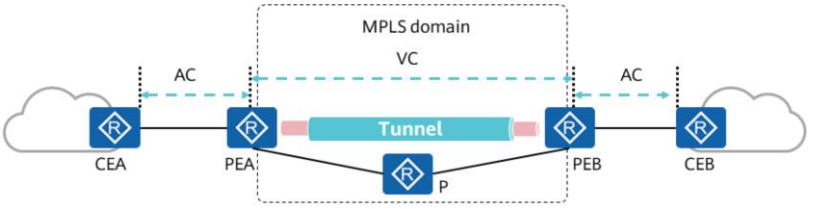
- EVPN provides both Layer 2 and Layer 3 capabilities and functions as a control plane protocol to transmit Layer 2 and Layer 3 information (MAC/IP addresses). EVPN can be used together with the traditional VPN technology.
- EVPN supports EVPN VPLS, EVPN VPWS, EVPN L3VPN, etc.

HUAWEI

- A traditional L2VPN does not have any control plane and does not transmit service route information (MAC addresses). It uses BGP as the signaling protocol to establish VCs.
- For details about VPN classification, see the book SRv6 Network Programming: Ushering in a New Era of IP Networks.

Traditional WAN L2VPN Overview

- The traditional WAN L2VPN is based on the MPLS network. VPWS provides a point-to-point Layer 2 network, and VPLS provides a point-to-multipoint Layer 2 network.
- The basic MPLS L2VPN architecture is composed of the attachment circuit (AC), VC, and tunnel.
  - AC: independent physical or virtual circuit connecting a CE and a PE. An AC interface can be either a physical or logical interface.
  - VC: logical connection between two PEs. A VC is established using a signaling protocol, such as BGP AD.
  - Tunnel: used to transparently transmit service data. Typical tunnels include MPLS LDP tunnels and MPLS TE tunnels.
- VPLS and VPWS are widely used on carrier networks to provide MPLS Layer 2 private line services for enterprises.

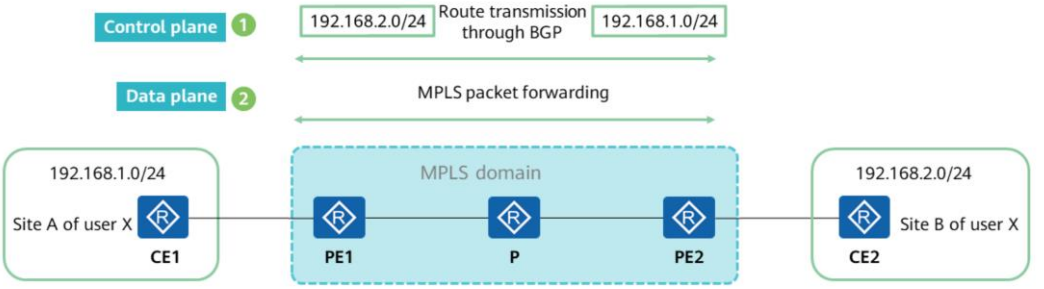


HUAWEI

- VCs are also called pseudo wires (PWs) in some documents.

Traditional WAN L3VPN Overview

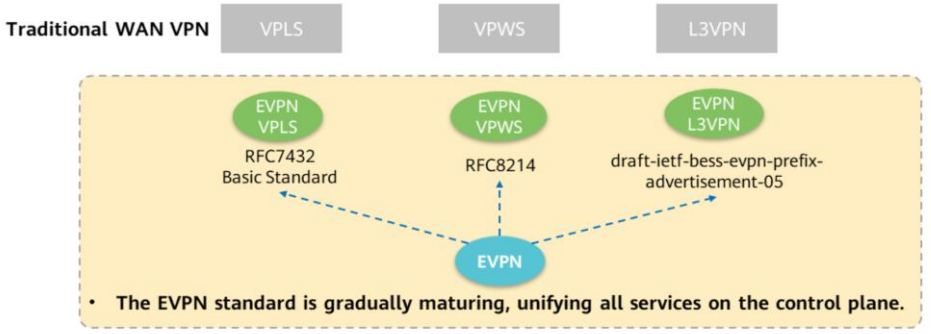
- Traditional WAN L3VPN generally refers to BGP/MPLS L3VPN.
- It distinguishes the control plane from the data plane. The control plane uses MP-BGP to advertise VPN routes, and the data plane uses MPLS LSPs to forward VPN packets.
- BGP/MPLS L3VPN is widely used on carrier networks to provide MPLS Layer 3 private line services for enterprises.



HUAWEI

WAN EVPN Overview

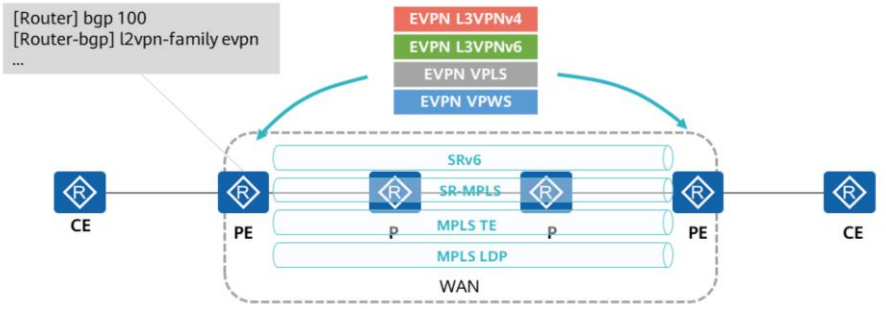
- EVPN was initially designed as an L2VPN technology based on BGP extensions. With the development of protocol extensions, EVPN can also support L3VPN now.
- EVPN can well serve as the control plane protocol for WAN VPN. It can be used with traditional VPN technologies to provide EVPN VPLS, EVPN VPWS, and EVPN L3VPN.



HUAWEI

WAN EVPN Application

- On a WAN, EVPN can be used with multiple tunneling technologies to support multiple application scenarios.
- EVPN, as a control plane protocol, can work with MPLS LDP, MPLS TE, SR-MPLS, and SRv6 tunnels, as shown in the figure.
- On Huawei devices, EVPN L2VPN and L3VPN services share the same address family.



HUAWEI

Summary: WAN VPN

- The VPN technology is widely used in various enterprise scenarios. VPNs can be built over either the Internet or a private network.
- L2VPN (VPLS and VPWS), L3VPN (BGP/MPLS IP VPN), and L2/L3 EVPN can be deployed on a bearer WAN built by an enterprise.
- EVPN, as a control plane protocol, can work with different bearer technologies (such as MPLS LDP, MPLS TE, SR-MPLS, and SRv6) to provide integrated and unified VPN services for enterprises.
- When multiple tunneling technologies are deployed on an enterprise bearer WAN, VPNs must recurse to tunnels based on tunnel policies.

HUAWEI

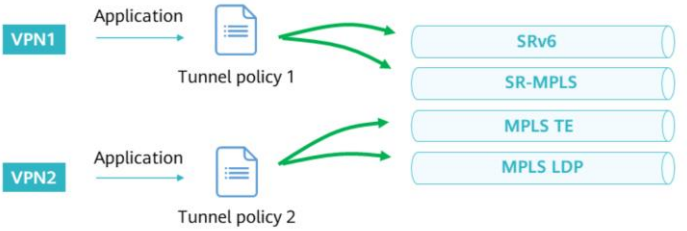
Contents

1. Bearer WAN Architecture
2. Bearer WAN Basics
3. **VPN Service**
   - WAN VPN Overview
   - Tunnel Management Overview
4. Network Traffic Optimization
5. SLA
6. Network Reliability
7. Network Management and O&M

HUAWEI

Tunnel Management

- Huawei devices use a tunnel management (TNLM) module to manage tunnels. It selects a certain tunnel for an application according to specific configurations and notifies the application of the tunnel's status.
- Common VPN tunnels include LSPs (MPLS LDP), MPLS TE tunnels, GRE tunnels, SR-MPLS Policies, and SRv6 Policies.
- Tunnel management configuration includes two parts: configuring a tunnel policy and applying a tunnel policy to a VPN.



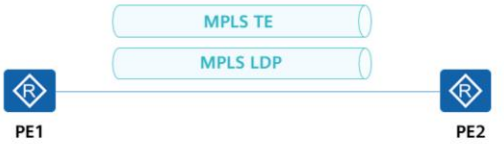
HUAWEI

- GRE: GRE can be applied to both L2VPN and L3VPN. Generally, the bearer WAN for MPLS VPN uses LSPs as public network tunnels. If the bearer WAN (P devices) has only IP functions but not MPLS functions, and the PEs at the network edge have MPLS functions, the LSPs cannot be used as public network tunnels. In this case, GRE tunnels can be used to replace LSPs to provide L3VPN or L2VPN solutions on the bearer WAN.
- SR-MPLS Policy: a type of SR-MPLS tunnel.
- SRv6 Policy: a type of SRv6 tunnel.
- You can configure tunnel policies or tunnel policy selectors for tunnel management. This course uses tunnel policy configuration as an example. Tunnel policy selectors apply to inter-AS VPN scenarios. For details, see the product documentation for NetEngine products.

Configuring a Tunnel Policy

- A tunnel policy determines the types and sequence of tunnels to be selected.
- By default, a VPN service recurses only to one LSP. If multiple LSPs are available, a tunnel policy can be used for load balancing among these LSPs.
- In this example, to implement load balancing between MPLS LDP LSPs and TE tunnels, you need to configure a tunnel policy for the VPN and apply the tunnel policy to the VPN. The tunnel policy policy1 requires tunnels to be selected in the sequence of first CR-LSPs and then LSPs, and the number of tunnels for load balancing is 2. The system preferentially selects two CR-LSPs. If only one or no CR-LSP is available on the network, the system selects one or two LSPs, respectively, for service transmission. In the scenario where only one CR-LSP is available, it works together with the selected LSP.

```
[PE1] tunnel-policy policy1
[PE1-tunnel-policy-policy1] tunnel select-seq cr-lsp lsp load-balance-number 2
```



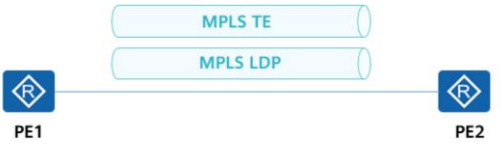
HUAWEI

- The configuration of tunnel policy parameters involves many details. For example, CR-LSP-based tunnels include RSVP-TE tunnels and SR-MPLS TE tunnels. The system determines the priorities of these tunnels based on their up time. For details, see "VPN Tunnel Management Configuration" in the product documentation for Huawei NetEngine routers.

Applying a Tunnel Policy to a VPN

- After being configured, a tunnel policy needs to be applied to a VPN. The mode in which a tunnel policy is applied to a VPN varies according to the VPN type.
- This example shows how to apply a tunnel policy to an L3VPNv4 instance.

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpnb] ipv4-family
[PE1-vpn-instance-vpnb-af-ipv4] tnl-policy policy1
```

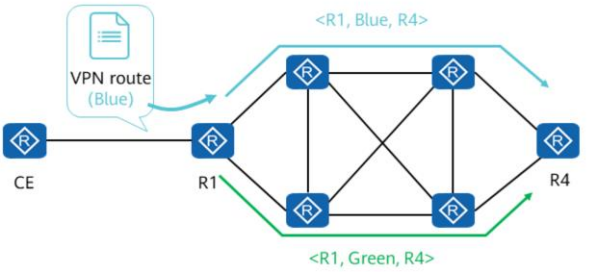


HUAWEI

- For the application of other types of VPN, such as VPNv6, L2VPN, and EVPN, see the product documentation for NetEngine routers.

Extension: SR Policy-based Traffic Diversion

- Tunnel policies are designed by Huawei for VPN service recursion in the MPLS era. They effectively decouple tunnel establishment from tunnel selection. In this way, the traffic of a VPN can be directed to multiple tunnels for load balancing. After the SR technology is introduced, the implementation mode changes. SR Policies integrate tunnel establishment (SR forwarding path) and tunnel policies (color-based traffic diversion by default). SR Policies cannot be selected together with other types of tunnels.
- An SR Policy is identified by <headend, color, endpoint> and can contain multiple forwarding paths. A VPN service selects an SR Policy based on the color attribute.



- There are two SR Policies (blue and green) between R1 and R4.
- R1 queries its VPN routing table after receiving traffic from the CE.
- If the color of the corresponding VPN route is Blue, R1 selects an SR Policy with the same color for route recursion.

HUAWEI

- For details, see "SR Policy" in 2. Terminology in RFC 8402.
- SR Policy traffic diversion can be based on the binding SID, color, and DSCP value. Details are not provided here.

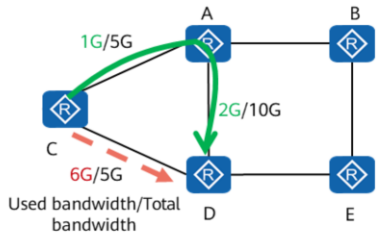
Contents

1. Bearer WAN Architecture
2. Bearer WAN Basics
3. VPN Service
4. **Network Traffic Optimization**
5. SLA
6. Network Reliability
7. Network Management and O&M

HUAWEI

Network Congestion Background

Drawback of fixed bandwidth-based route selection and related solution

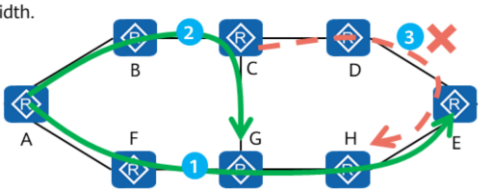


Used bandwidth/Total bandwidth

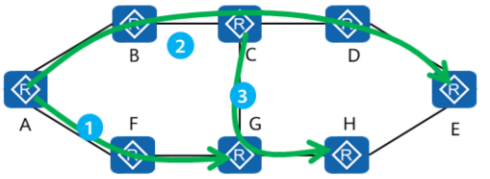
The network computes forwarding paths based on bandwidth. The link from Router C to Router D is the shortest forwarding path. The rate of service traffic from Router C to Router D exceeds the link bandwidth, and packet loss occurs. Although other links are idle, the algorithm still selects the shortest path for traffic forwarding. From a global perspective, the optimal traffic forwarding path is C -> A -> D in this situation.

Drawback of sequenced tunnel establishment and related solution

Tunnels are established in the following sequence: A-E > A-G > C-H. Tunnel C-H, however, fails to be established due to insufficient bandwidth.



Global path computation for optimal tunnel path adjustment:



HUAWEI

Network Traffic Optimization Overview

- Network traffic optimization is to perform global analysis on network congestion, obtain the path computation result based on a proper optimization policy (ensuring the SLA of critical services), and apply the computation result to the network for congestion elimination.
- Network traffic optimization can be divided into three phases: network information collection, path computation for network traffic optimization, and optimization result delivery.

1. Network information collection

The controller collects global network information, including:

- Network topology
- Network bandwidth
- Link delay
- Network traffic and other information

2. Path computation for network traffic optimization

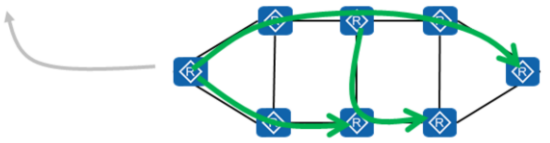
The controller computes paths based on the optimization target. Optimization targets include:

- Least path cost
- Shortest path delay
- Maximum link bandwidth utilization
- ...

3. Optimization result delivery

The controller delivers the computation result to network devices in any of the following modes:

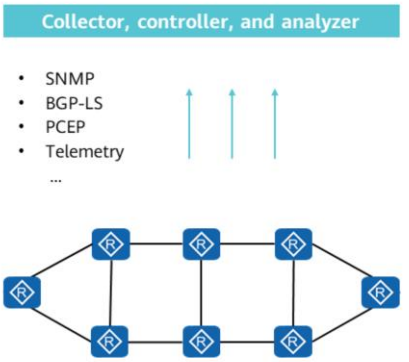
- NETCONF
- PCEP
- BGP SR Policy
- ...



HUAWEI

Network Information Collection

- The collection of network information, including the network topology, interface bandwidth, link delay, and traffic statistics, is the prerequisite for network traffic optimization.



Network topology and interface bandwidth collection

- In the industry, SNMP is generally used to collect basic network topology and device information.
- BGP-LS is used to collect IGP and TE topology information (including interface bandwidth).
- PCEP and BGP SR Policy are used to collect TE tunnel information.

Link delay collection

- The link delay is collected using TWAMP, flooded in the IGP domain, and then reported to the controller through BGP-LS.

Traffic statistics collection

- To determine bandwidth sufficiency and perform traffic optimization, the controller needs to collect interface and tunnel traffic statistics in real time.
- Mainstream traffic statistics collection technologies include SNMP, telemetry, and NetStream.

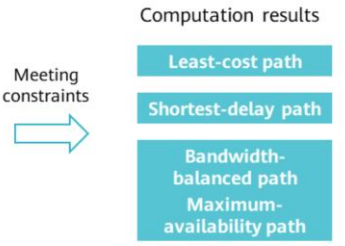
HUAWEI

- Different collection protocols may be used in different solutions. For example, PCEP is used to collect TE tunnel information on Huawei MPLS networks, and BGP SR Policy is used to collect TE tunnel information on SRv6 networks.

Network Optimization Computation

- Network optimization computation refers to the computation of global or local optimal paths based on service requirements through the corresponding algorithms.
- When computing paths, the controller needs to ensure that computed paths meet the related constraints. If the constraints cannot be met, the controller retains the original paths. The following table lists some constraints.

Constraint	Description
Priority	This constraint specifies the priorities of different types of tunnels and enables a tunnel with a higher priority to preempt the bandwidth resources of a tunnel with a lower priority.
Bandwidth	This constraint requires paths to be computed based on tunnel bandwidth requirements.
Hop count	This constraint requires paths to be computed based on hop count requirements. For example, the path length of an SR-TE tunnel is limited by the maximum stack depth (MSD) of the ingress node.
Explicit path	This constraint requires paths to be computed in either strict or loose mode. You can specify the links or nodes that are to be included or excluded.
Delay threshold	This constraint requires paths to be computed within the threshold range specified for path computation.
Affinity	This constraint supports the include-all, include-any, and exclude modes.

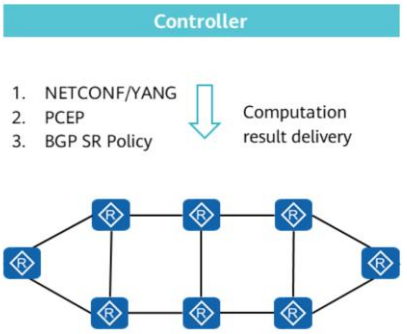


HUAWEI

- Bandwidth-balanced path: path with more remaining bandwidth among all paths that meet the constraints and have the same cost.
- Maximum-availability path: path with the maximum availability among all paths that meet the constraints.

Optimization Result Delivery

- After the controller computes a network path, you can choose whether to apply the computation result to the network. There are multiple implementation modes:



1 NETCONF/YANG

- The controller delivers the computation result to network devices as configurations.
- The YANG model is standardized and provides good compatibility with different vendors.

2 PCEP

- The controller delivers PCEP messages to create or update LSPs.
- The PCEP standard does not define a tunnel model, and vendor-specific protocols cannot interoperate with each other.

3 BGP SR Policy

- The controller uses BGP extensions to deliver tunnels.
- The RFC defines the tunnel model and data packet structure in a unified manner, facilitating product interoperability between different vendors.

HUAWEI

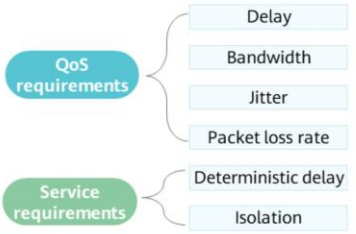
Contents

1. Bearer WAN Architecture
2. Bearer WAN Basics
3. VPN Service
4. Network Traffic Optimization
5. **SLA**
6. Network Reliability
7. Network Management and O&M

HUAWEI

New Requirements for Bearer Network SLA Assurance

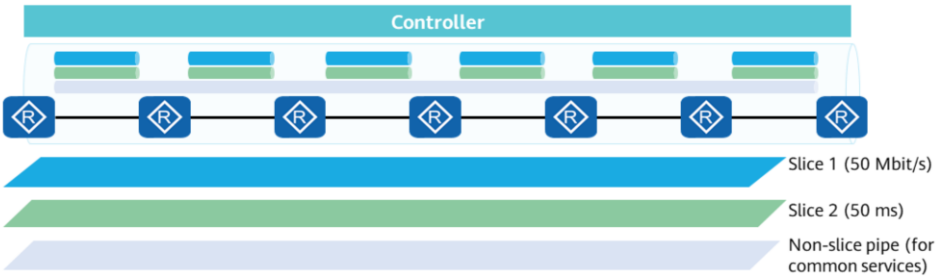
- An SLA is a formal commitment between a service provider and a customer. In the WAN service field, in addition to basic connectivity requirements, the SLA also focuses on deterministic delay, bandwidth, reliability, and isolation (security).
- Generally, a bearer path carries the traffic of multiple types of services. When different types of service traffic are transmitted on the same path, differentiated bearer needs to be provided based on SLA requirements. Traditional QoS uses statistical multiplexing to set different priorities for specific services to ensure smooth experience of high-priority services. However, QoS falls short of meeting the isolation and deterministic delay requirements.
- Network slicing can divide a bearer network into virtual networks of different service levels and provide dedicated logical channels for services with high quality and security requirements.



- For example, fund settlement services between financial enterprises require high security and stable delay; the file transfer service within an enterprise requires high bandwidth; and enterprise voice services require low jitter.

Slice-based Bearer Network: One Network for Multiple Purposes (Carrying Multiple Services), Lowering Costs

- The IP network uses statistical multiplexing to greatly improve network utilization and reduce per-bit transmission costs. However, statistical multiplexing brings uncertainty to the quality assurance levels of different services. Moreover, it is inappropriate to prepare resources based on the highest SLA requirements to meet the requirements of all types of private lines and customers. The converged bearer network needs to balance multi-service isolation and statistical multiplexing to meet the SLA requirements of each service.
- Slice resource reservation technologies, such as FlexE, channelized sub-interfaces, and QoS queues, can be used to direct services to respective service slices. These slices are isolated from each other and do not affect each other, providing different SLA levels.



HUAWEI

QoS Overview

- QoS provides differentiated service quality for different applications.
- Generally, QoS provides three service models: best-effort service, integrated service (IntServ), and differentiated service (DiffServ).
- DiffServ is the most widely used QoS model on IP networks.

Best-effort service

- Applications can send any number of packets at any time.
- The network tries its best to send packets and does not provide delay or reliability guarantee.

IntServ

- An application needs to apply for specific services before sending packets.
- The network reserves network resources for the application using a protocol such as RSVP.

DiffServ

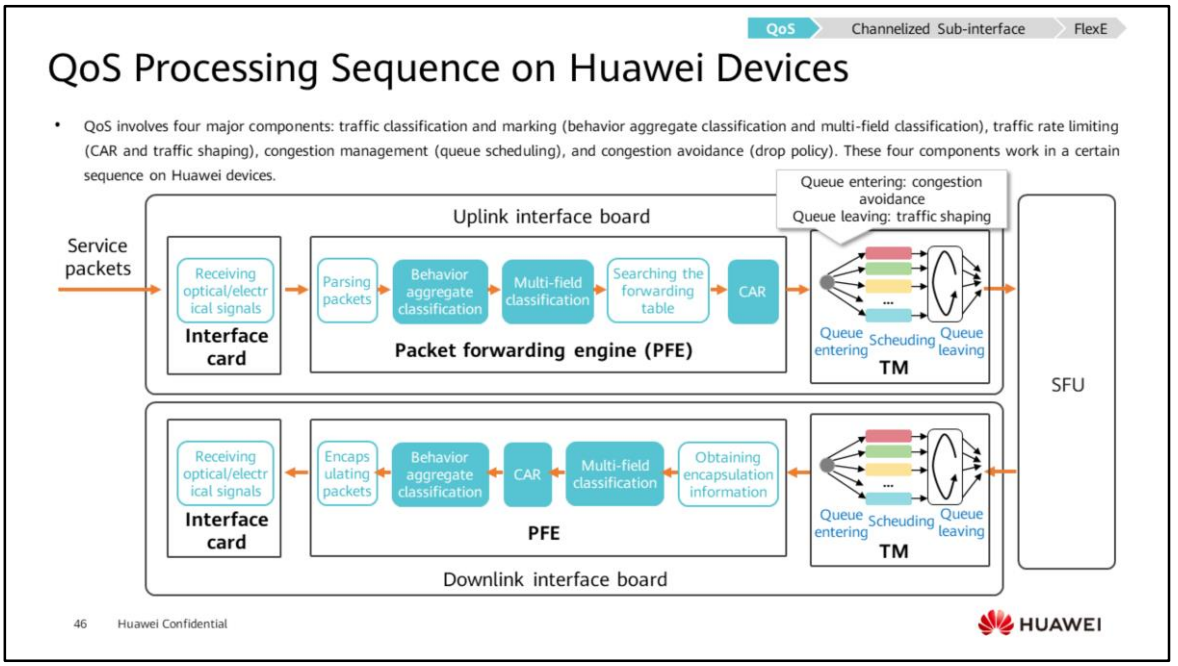
- Traffic on the network is divided into multiple classes. A processing behavior is defined for each class, so that each class has a different forwarding priority, packet loss rate, and delay.
- It is most widely used on IP networks.

HUAWEI

DiffServ Model-based QoS Components

- DiffServ model-based QoS consists of four components: traffic classification and marking, traffic policing and shaping, congestion management, and congestion avoidance.
  - Traffic classification and marking: Dividing data packets into different classes or configuring different priorities for packets is the prerequisite for implementing differentiated services. This component classifies data packets into different types, with each type of traffic being a traffic class. Traffic classification does not change the original data packets. In comparison, marking, which sets different priorities for data packets, changes the original data packets.
  - Rate limiting (traffic policing and shaping): This component limits the rate of service traffic. It does this by discarding excess traffic when the service traffic exceeds the rate limit. Traffic policing controls the traffic receiving rate, and traffic shaping controls the traffic sending rate.
  - Congestion management: This component buffers packets in queues upon network congestion and determines the forwarding order using a specific scheduling algorithm.
  - Congestion avoidance: This component monitors network resource use. When congestion becomes severe, some packets are discarded to prevent network overload.
- Traffic classification and marking is the prerequisite and foundation for implementing differentiated services.
- Traffic policing, traffic shaping, congestion management, and congestion avoidance are used to control network traffic from different aspects.

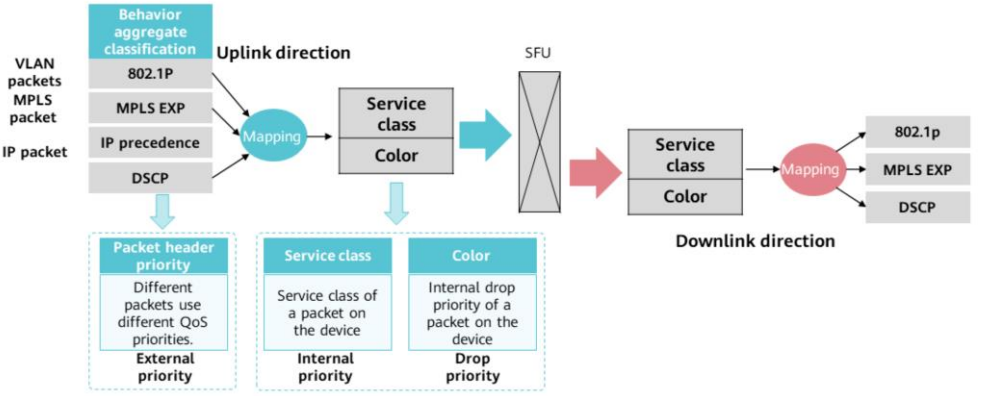
HUAWEI



- For the basic packet forwarding process, see HCIP-Datacom-Core Technology-01 Introduction to Network Devices.
- Behavior aggregate classification: Packets are roughly classified based on the IP precedence or DSCP value of IP packets, TC value of IPv6 packets, EXP value of MPLS packets, and 802.1p value of VLAN packets to identify traffic with different priorities or service levels and implement internal-external priority mapping for the traffic.
- Multi-field classification elaborately classifies packets based on complex rules, such as the 5-tuple (source address, source port number, protocol number, destination address, and destination port number).
- Packet Forwarding Engine (PFE): After a router is powered on, it runs a routing protocol to learn the network topology and generate a routing table. If the interface board registers successfully, the main control board can generate forwarding entries according to the routing table and deliver entries to the interface board. In this manner, the router can forward packets according to the forwarding table. The component that forwards data packets is a chip located on an interface board and is called a packet forwarding engine (PFE).

Behavior Aggregate Classification

- Behavior aggregate classification allows a device to roughly classify packets based on simple rules.



HUAWEI

DSCP/IP Precedence/ 802.1p/EXP Value Mapping

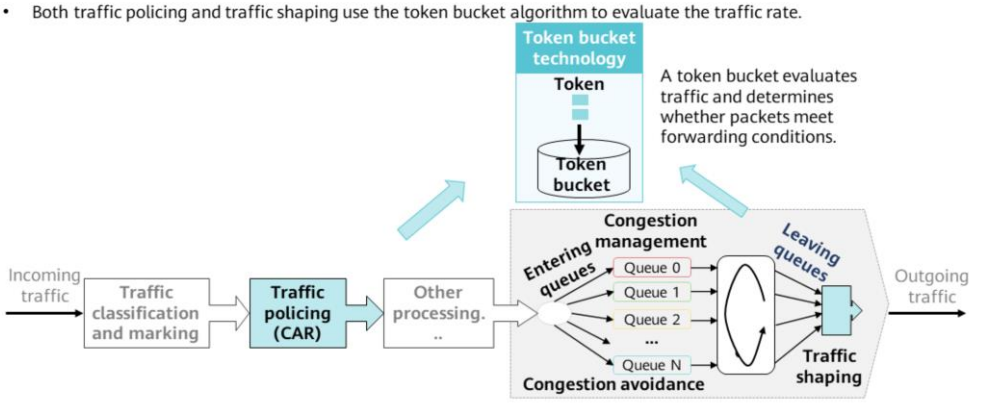
Priorities in descending order

802.1P	MPLS Exp	IP-Precedence	DSCP	DSCP Name				
7	7	7	56-63	CS	CS7 (56)			
6	6	6	48-55		CS6 (48)			
5	5	5	40-47	EF	EF (46)			
4	4	4	32-39	AF	AF4	AF41 (34)	AF42 (36)	AF43 (38)
3	3	3	24-31		AF3	AF31 (26)	AF32 (28)	AF33 (30)
2	2	2	16-23		AF2	AF21 (18)	AF22 (20)	AF23 (22)
1	1	1	8-15		AF1	AF11 (10)	AF12 (12)	AF13 (14)
0	0	0	0-7	BE	BE (0)			

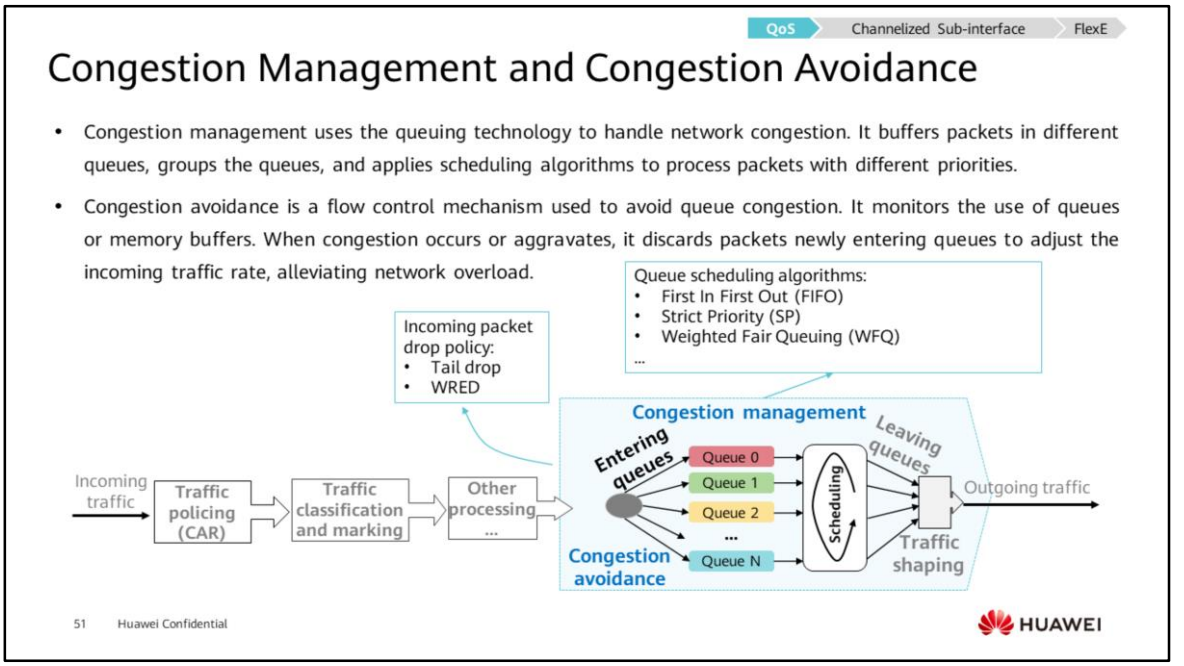
HUAWEI

Traffic Policing and Traffic Shaping

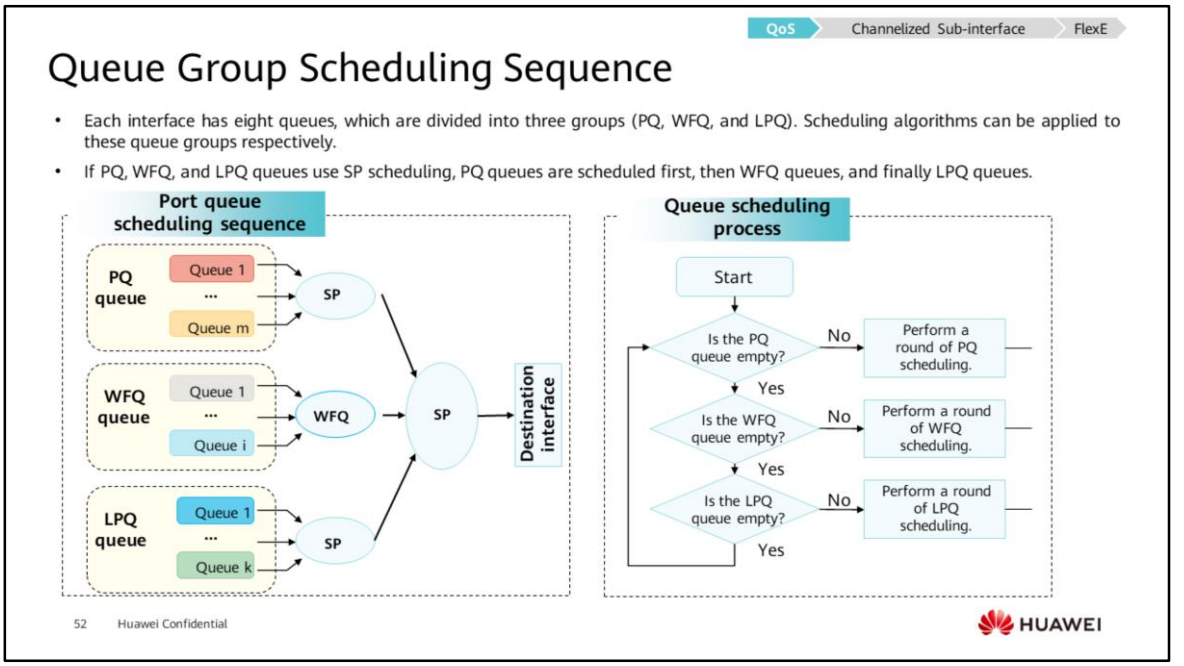
- Both traffic policing and traffic shaping are traffic rate limiting technologies. The former monitors the incoming traffic of a device and limits the incoming traffic rate to a permitted range. If the traffic rate is too high, excess packets are discarded or the packet priorities are re-set. Traffic shaping controls the rate of outgoing packets, so that packets can be sent at an even rate.
- Both traffic policing and traffic shaping use the token bucket algorithm to evaluate the traffic rate.



HUAWEI



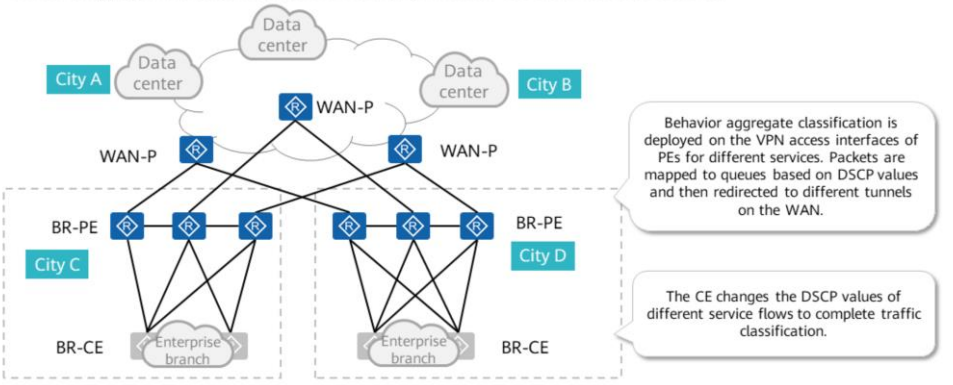
- Weighted Random Early Detection (WRED): The system discards packets based on the drop policies configured for data packets or queues with different priorities. WRED is a congestion avoidance mechanism used to discard packets to prevent queues from being congested. For details, see the product documentation for Huawei NetEngine products.
- FIFO: FIFO does not classify packets. FIFO allows packets to be queued and forwarded in the same order as they arrive at an interface.
- SP: Queues are scheduled strictly according to their priorities. Packets in queues with a low priority can be scheduled only after all packets in queues with a higher priority are scheduled.
- WFQ: The egress bandwidth is allocated to each flow according to the queue weight.
- Other scheduling algorithms, such as RR polling, WRR weighted polling, and DRR differential polling, are not described here.



- PQ queue
  - PQ queues use the SP scheduling algorithm. That is, the packets in the queue with the highest priority are scheduled first. In this way, an absolute priority can be provided for different service data, the delay of delay-sensitive applications such as VoIP can be guaranteed, and the use of bandwidth by high-priority services can be absolutely prioritized.
  - Disadvantage: If the bandwidth of high-priority packets is not limited, low-priority packets may fail to obtain bandwidth and be scheduled.
  - Generally, only delay-sensitive services enter PQ queues.
- WFQ queue
  - WFQ queues are scheduled based on weights. The WFQ scheduling algorithm can be used to allocate the remaining bandwidth based on weights.

WAN QoS Application: Flow Marking

- WAN QoS is mainly applied to WAN links. A WAN link has much lower bandwidth than a local link and is often a congestion point for traffic forwarding. The purpose of deploying QoS on the WAN is to ensure the SLAs of different services.
- It is recommended that CEs or the downstream devices of CEs mark DSCP priorities for service flows on the WAN. Then only behavior aggregate classification needs to be deployed on the VPN access interfaces of PEs.



HUAWEI

- If the CE or the downstream device of the CE does not have the traffic marking capability, deploy multi-field classification on the ingress PE on the bearer WAN to mark traffic for queuing. This, however, affects the forwarding performance of the bearer WAN.

WAN QoS Application: Flow Scheduling

- The WAN's forwarding bottleneck lies in the egress WAN link in each region. For example, remote sites are interconnected through a carrier's MSTP private line, which has limited bandwidth. To prevent service flows from being discarded due to rate limiting after entering WAN links on the carrier network, deploy traffic shaping on egress WAN links to ensure that the traffic entering the carrier's MSTP private line does not exceed the rate limit.
- Properly design QoS policies for enterprises' internal services. For example, deploy high-priority PQ scheduling for core production services or services with high QoS requirements and WFQ scheduling for services insensitive to packet loss and delay (only basic bandwidth for service continuity is needed in this case).

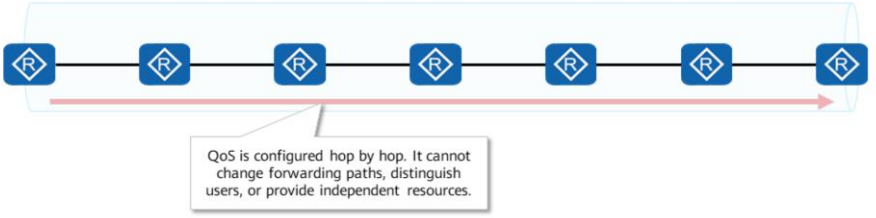
QoS design for an enterprise

Service Type	Importance	Priority	Scheduling Mode
Protocol packet	High	CS7/CS6	PQ
Core service	High	EF	PQ
Video service	High	AF4	PQ
External service	High	AF3	WFQ
Office and test services	Medium	AF2	WFQ
Others	Low	BE	WFQ

HUAWEI

QoS Limitations

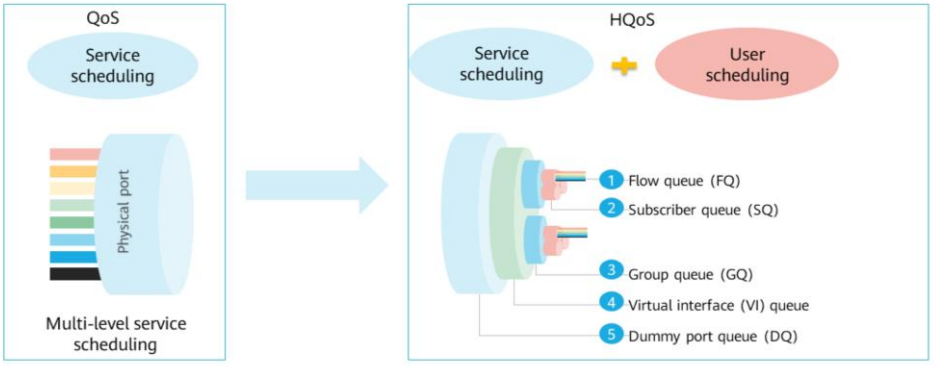
- QoS itself cannot solve the congestion problem or provide isolation and deterministic delay assurance for services:
  - QoS involves only single-hop behaviors and does not change the network topology.
  - QoS does not change service behaviors. If the bursty traffic of a single flow is too heavy, congestion still occurs.
  - The number of QoS queues is small, and SLA assurance cannot be provided for specific users. As a result, deterministic delay assurance cannot be provided.
  - The QoS mechanism is an experience system for resource management and cannot provide independent resources for users.



HUAWEI

HQoS: Providing Scheduling and Bandwidth Guarantee for Users

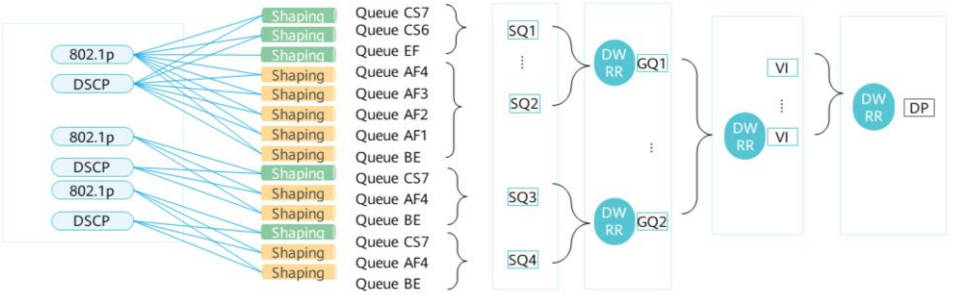
- On the basis of QoS, hierarchical quality of service (HQoS) provides finer-grained and more hierarchical scheduling and management of interface resources to implement fine-grained allocation and management of interface resources. Every scheduling queue can provide bandwidth guarantee, but the entire mechanism is based on QoS and cannot meet the deterministic delay and isolation requirements either.



HUAWEI

HQoS: Providing User-Level Bandwidth Guarantee When the CIR of Each Scheduler Does Not Exceed the Upper Limit

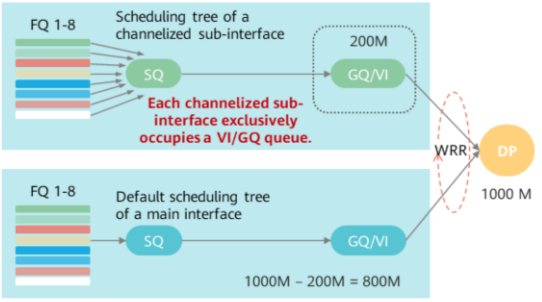
- HQoS adds hierarchical schedulers such as SQ, GQ, and VI. Each scheduler has two attributes (CIR and PIR) and uses SP scheduling for flows whose rates are between the CIR and PIR. The CIR is preferentially guaranteed.
- HQoS guarantees bandwidth through strict CIR deduction and shaping. The sum of the CIRs of all schedulers is less than the interface bandwidth.



HUAWEI

Channelized Sub-interfaces Providing Management Entities on the Basis of HQoS to Provide SLA Assurance

- Bandwidth guarantee: Channelized sub-interfaces use HQoS-based hierarchical scheduling to implement flexible and refined resource management. Each channelized scheduling tree has independent buffer resources and bandwidth resources, providing bandwidth guarantee.
- Delay guarantee: Because the resources for channelized sub-interface are strictly guaranteed, the delay can be guaranteed within a certain range.

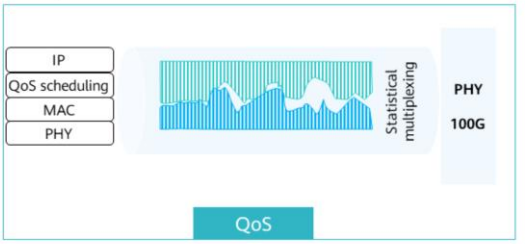


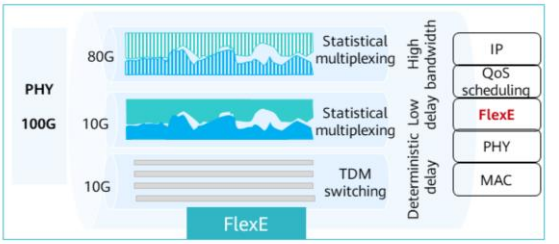
- Channelized sub-interfaces use the HQoS mechanism and exclusively occupy the HQoS VI/GQ scheduling trees and bandwidth to implement strict scheduling isolation.
- Remaining bandwidth of a main interface = Total interface bandwidth – Total bandwidth of all channelized sub-interfaces. Bandwidth is automatically deducted during the channelized sub-interface enabling process, simplifying the HQoS application model.
- Channelized sub-interfaces provide management entities and can work with the controller for resource management and E2E resource reservation, meeting the SLA assurance requirements of P2MP services.

HUAWEI

FlexE: Delivering Isolation and Deterministic Delay Assurance Based on Independent Resources

- Flexible Ethernet (FlexE) is a new technology introduced between the MAC and PHY layers. It is a lightweight enhanced technology for IP networks and is compatible with the existing Ethernet standard and QoS capabilities.
- It provides isolated FlexE interface links, enabling one network to carry different types of services.
- It provides an independent P2P time division multiplexing (TDM) tunnel for each service, meeting isolation and deterministic delay requirements.

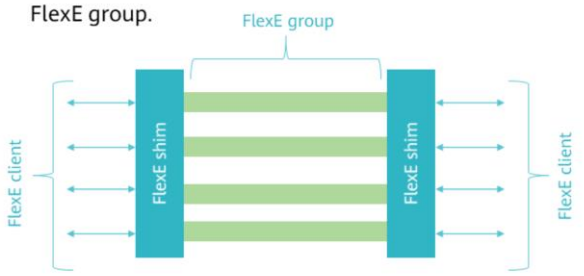




HUAWEI

Basic FlexE Architecture

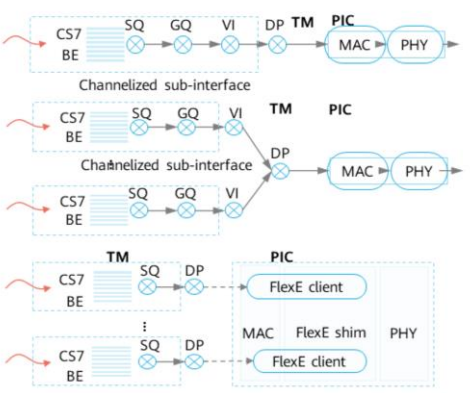
- FlexE decouples the MAC layer from the PHY layer by introducing the FlexE shim layer on the basis of IEEE 802.3, thereby providing flexible rates.
- FlexE adopts the client/group architecture and allows client interfaces at different rates to coexist in a FlexE group.



Client: corresponds to various user interfaces on a network. Each FlexE client can flexibly apply for bandwidth from the resource pool of a FlexE group, adjust the bandwidth, and transfer data flows to the FlexE shim layer as 64B-/66B encoded bit streams.
Group: consists of various Ethernet PHY layers defined in IEEE 802.3 and divides the PHY bandwidth into 1G timeslots (supported by Huawei devices) or 5G timeslots.
Shim: an extra logical layer inserted between the MAC and PHY layers of the traditional Ethernet architecture. It implements key FlexE functions through calendar timeslot distribution.

HUAWEI

Comparison of Slicing Technologies



QoS: All traffic shares eight queues. QoS schedules resources in a unified manner to maximize the statistical multiplexing capability. It cannot differentiate users, and so cannot provide independent resource reservation for different users.

Channelized sub-interface: Queue resources are isolated. Hierarchical scheduling is used to implement flexible and refined management of interface resources, provide bandwidth guarantee, and work with the controller to provide E2E resource reservation.

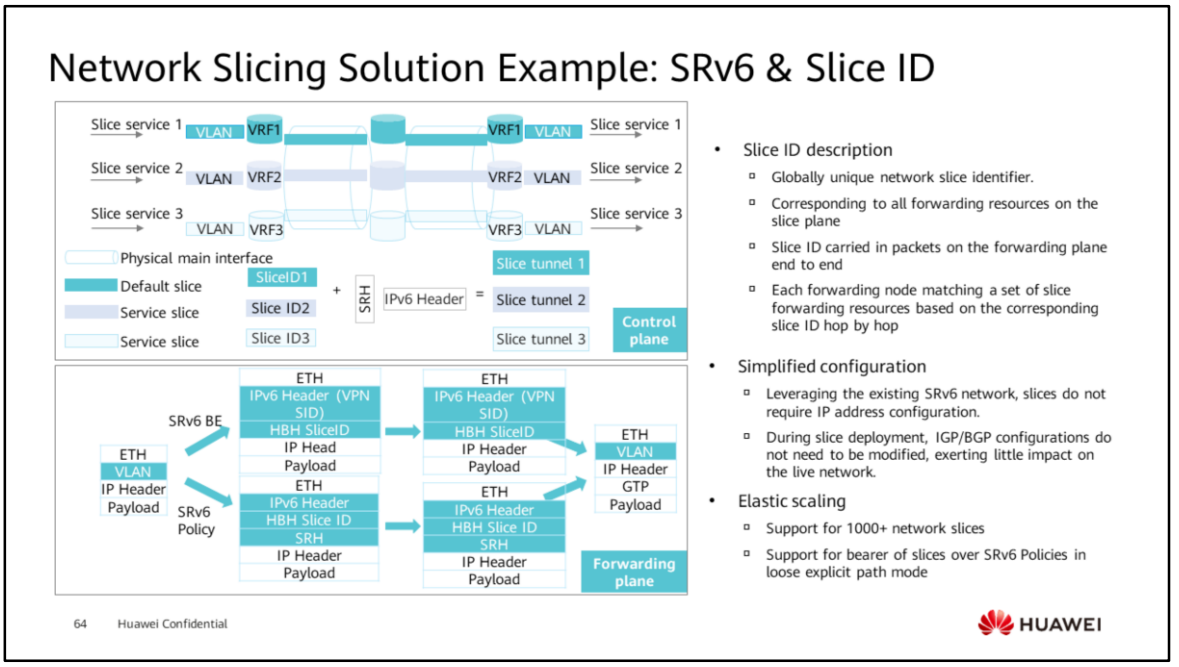
FlexE: Queue and interface resources are isolated. Every resource is divided by TDM timeslot. This meets the requirements for exclusive resource use and resource isolation and provides flexible and refined management of interface resources.

HUAWEI

Slicing Technology Implementation Modes

Interface Name	Usage Scenario and Feature Description
FlexE sub-interface (also called client interface)	A physical interface in standard Ethernet mode has fixed bandwidth. FlexE, however, can enable one or more physical interfaces to work in FlexE mode and add them to a group. The total bandwidth of this group can be allocated on demand to logical interfaces in the group. The group to which physical interfaces are added is referred to as a FlexE group. The logical interfaces that share bandwidth of the physical interfaces in the FlexE group are called FlexE interfaces (also referred to as FlexE service interfaces). FlexE interface bandwidth varies, which allows services to be isolated. Compared with traditional technologies, FlexE technology permits bit-level interface bundling, which solves uneven per-flow or per-packet hashing that challenges traditional trunk technology. In addition, each FlexE interface has a specific MAC address, and forwarding resources between interfaces are isolated. This prevents head-of-line (HOL) blocking that occurs when traditional logical interfaces such as VLAN sub-interfaces are used for forwarding. FlexE interface technology especially fits scenarios in which high-performance interfaces are required for converged bearer, such as mobile bearer, home broadband, and private line access. Services of different types are carried on different FlexE interfaces and are assigned bandwidth based on FlexE interfaces. In this way, FlexE achieves service-specific bandwidth control, meeting network slicing requirements in 5G scenarios.
VLAN channelized sub-interface	A channelized interface can strictly isolate interface bandwidth. A VLAN channelized sub-interface is a channelization-enabled sub-interface of an Ethernet physical interface. Different types of services are carried on different channelized sub-interfaces and assigned bandwidth based on channelized sub-interfaces. This implementation strictly isolates bandwidth among different channelized sub-interfaces on the same physical interface and achieves service-specific bandwidth control, preventing bandwidth preemption among different sub-interfaces.
Ethernet sub-interfaces	An Ethernet sub-interface is a virtual interface configured on a main interface and has Layer 3 features. You can configure an IP address for an Ethernet sub-interface to implement inter-VLAN communication. The main interface can be either a physical interface or a logical interface. The sub-interface inherits the physical layer parameters of the main interface but has its own link layer and network layer parameters. You can activate or deactivate the sub-interface, without affecting the performance of the main interface. The change of the main interface status, however, affects the sub-interface.

HUAWEI



- Different VLANs are used for service access. Logical interfaces correspond to VPN instances VRF1, VRF2, VR3... on the network slice.
- The ingress PE encapsulates the VPN SID and SRv6 Policy information into the service flow on the network slice with the slice ID being 2, and inserts an extension header with the Hop By Hop Slice ID being 2 between the IPv6 header and SRH of each packet.
- Each transit node queries the SRv6 SID in the SRH hop by hop to obtain the physical outbound interface, and then queries the specific "resource reservation" sub-interface of the physical outbound interface based on the slice ID. The Hop By Hop Slice ID remains unchanged throughout this process.
- The egress PE pops the Hop By Hop extension header and forwards the packet to the AC interface of the corresponding VPN instance based on the VPN SID.
- By default, the slice ID is 0, and the IPv6 Hop By Hop extension header does not need to be inserted. The packet format on the forwarding plane is the same as that of traditional L3VPN over SRv6 Policy.

Contents

1. Bearer WAN Architecture
2. Bearer WAN Basics
3. VPN Service
4. Network Traffic Optimization
5. SLA
6. **Network Reliability**
7. Network Management and O&M

HUAWEI

WAN Reliability Overview

- WAN reliability covers two parts: device reliability and network reliability.
- Device reliability: includes router reliability and controller reliability.
  - Controller reliability is implemented through cluster deployment and disaster recovery (DR) system deployment.
  - Router reliability can be implemented using device features such as non-stop routing (NSR). These features are beyond the scope of this course.
- Network reliability: reduces the impact of link and node faults on services through fast detection and convergence mechanisms at each layer.

HUAWEI

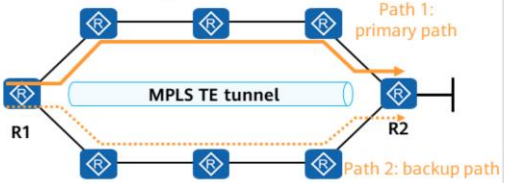
- Because different WAN VPN technologies use different terms, this section briefly describes various protection mechanisms, but does not describe specific protection technologies.

TE Tunnel Protection Technology Basics

- MPLS TE tunnel protection can be provided from two perspectives: local protection and E2E protection. TE tunnels, including MPLS TE tunnels, SR-MPLS TE tunnels, and SR-MPLS Policies can all be protected from the two perspectives, but their technical implementation is slightly different.

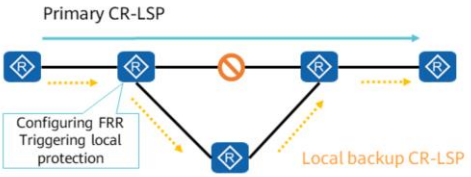
E2E protection

- E2E protection: establishes multiple CR-LSPs between the ingress and egress of a tunnel, with each CR-LSP traversing a different path. When detecting that the primary CR-LSP is faulty, the ingress switches traffic to the backup CR-LSP.
- In E2E protection, path switching is slow. Therefore, E2E protection needs to be used with a detection mechanism for fast path switching.



Local protection

- Fast reroute (FRR), also called local protection, is a temporary protection measure. If a transit node/link fails, local protection is triggered. A backup CR-LSP is then established locally for traffic switching. In addition, the tunnel ingress is instructed to recompute a path and switch traffic to the backup path in a timely manner.



HUAWEI

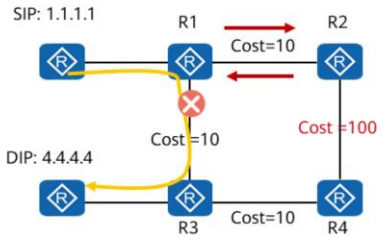
- MPLS TE E2E protection is classified into HSB and ordinary backup. In HSB protection, the backup path and primary path are created at the same time.
- Segment Routing adopts Topology Independent-Loop Free Alternate (TI-LFA), an enhancement of FRR, for local protection.
- Fast detection mechanism: Fast detection mechanisms represented by BFD support fast detection of communication faults between devices.

TI-LFA FRR

- TI-LFA FRR provides link and node protection for SR tunnels. If a link or node fails, traffic is rapidly switched to the backup path.

Limitations of the traditional LFA algorithm

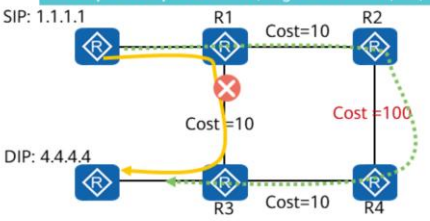
- The traditional LFA algorithm has topological limitations. As shown in the figure, SIP traffic is forwarded to the DIP through R1. If the R1-R3 link fails, R1 forwards the traffic to R2. However, no backup path can be formed before R2 detects the failure.



TI-LFA algorithm

- Using the source routing capability of SR, TI-LFA computes a backup path on each node to protect the failure point. When a node detects a failure, traffic is rapidly switched to the backup path.

Primary R1-R3 path: 4.4.4.4; segment list: R1, R3
Backup R1-R3 path: 4.4.4.4; segment list: R1, R2, R4, R3

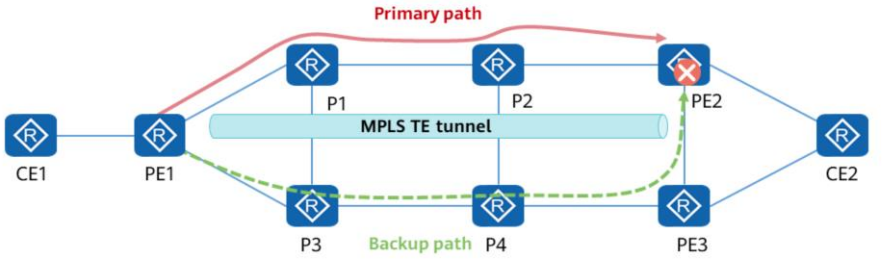


HUAWEI

- In a distributed network architecture, each device independently calculates paths, and there is no consensus on the shortest path when a fault occurs. As a result, traditional LFA cannot form a backup path.

Limitations of E2E Protection

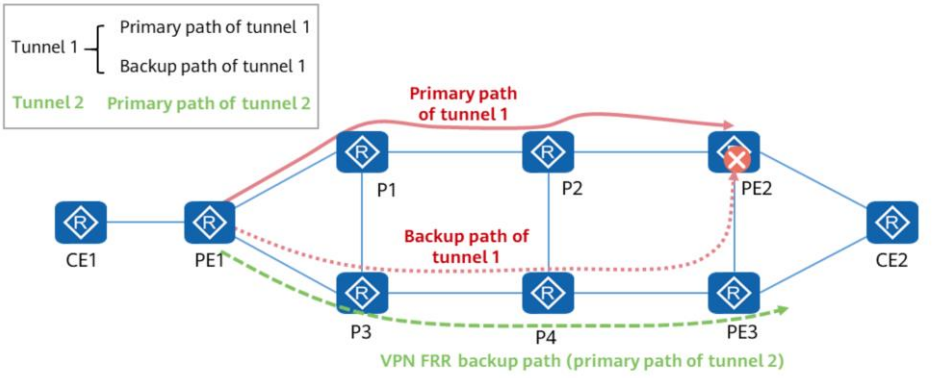
- Technologies such as HSB can protect E2E paths, but cannot rectify faults on PEs.
- In this example, a TE tunnel with both primary and backup paths is established between PE1 and PE2. If PE2 is faulty, the backup path cannot solve the problem.



HUAWEI

VPN FRR

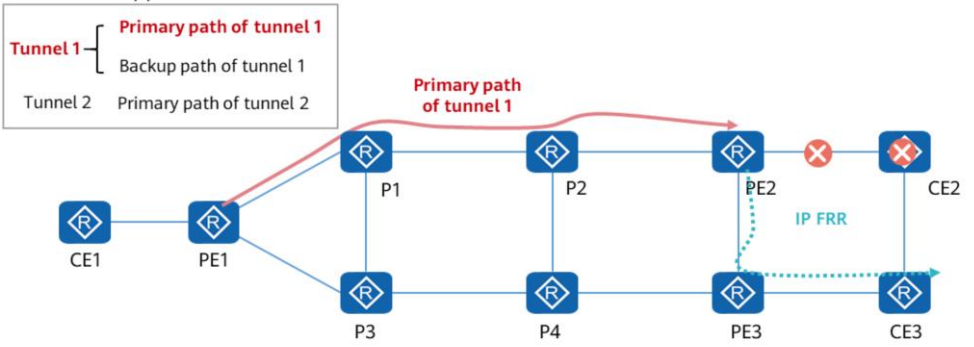
- VPN FRR sets the primary and backup forwarding paths pointing to the active and standby PEs on the remote PE in advance. It works with fast PE fault detection to accelerate fault-triggered E2E service convergence in scenarios where a CE is dual-homed to two PEs.



HUAWEI

IP FRR

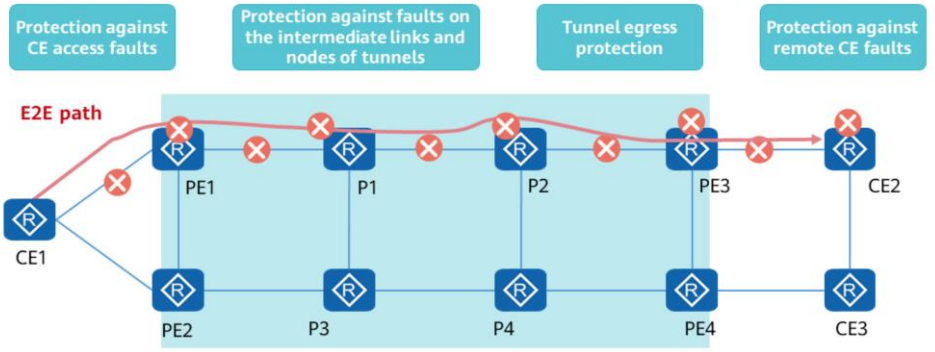
- If the link between PE2 and CE2 fails but PE2 still functions properly, the tunnel between PE1 and PE2 is still available. In this case, E2E tunnel switching is not required. PE2 selects PE3 as the backup next hop. If the link between PE2 and CE2 or CE2 fails, IP FRR is implemented to rapidly switch IP traffic.
- IP FRR is applicable to the IP network between CEs and PEs.



HUAWEI

Summary: Multi-Level Network Protection

- To sum up, different protection measures are taken to ensure tunnel reliability based on the locations of faults on E2E paths.
- From the perspective of E2E forwarding paths, it is recommended that multi-level protection be used:



HUAWEI

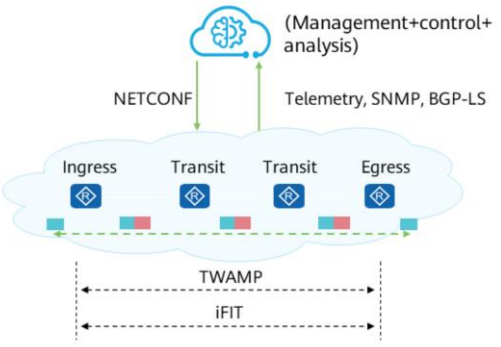
Contents

1. Bearer WAN Architecture
2. Bearer WAN Basics
3. VPN Service
4. Network Traffic Optimization
5. SLA
6. Network Reliability
7. **Network Management and O&M**

HUAWEI

WAN Management and O&M

- Among mainstream WAN solutions provided by different vendors, the SDN solution is preferred. The SDN controller centrally manages and delivers WAN services to forwarders.
- In Huawei's solution, the controller not only provides "control" functions, but also provides "management" and "analysis" functions.



Network analysis

- The AI algorithm is used to obtain network analysis results based on massive amounts of network performance and monitoring data.

Network collection/management

- Multiple protocols and channels are used to collect network configuration data, performance data, and monitoring data.
- Efficient network configuration management is provided.

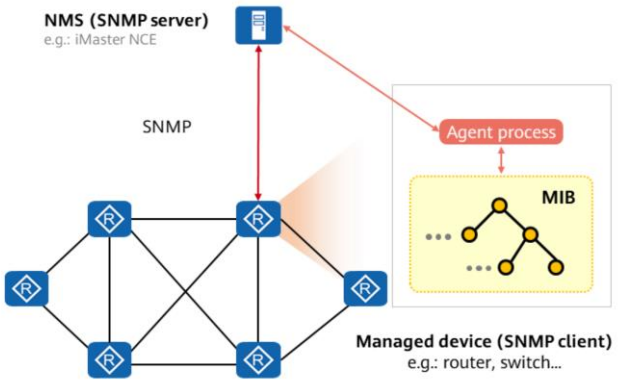
Network measurement

- Network performance indicators, such as the delay, jitter, and packet loss rate, are measured.

HUAWEI

- This figure does not show protocols related to tunneling and traffic statistics collection.

SNMP

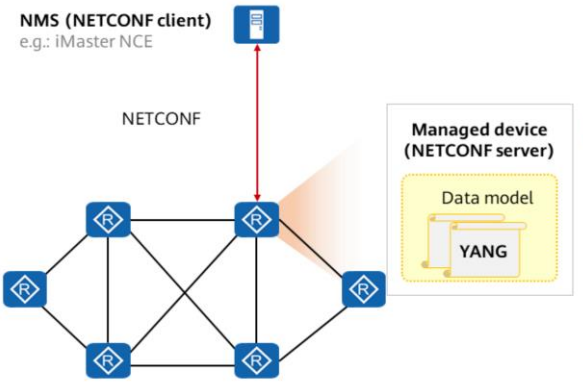


SNMP overview

- Simple Network Management Protocol (SNMP) is a network management standard widely used on TCP/IP networks.
- It provides a method for managing devices through a central computer that runs network management software — known as a network management station (NMS).
- By employing the "network management over networks" mode, SNMP implements efficient and batch network device management. In addition, SNMP enables unified management of network devices of different types and from different vendors.

HUAWEI

NETCONF



SNMP's drawbacks

SNMP is not a configuration-oriented protocol. On a large-sized network with a complex topology, SNMP cannot meet network management requirements, especially the configuration management requirements.

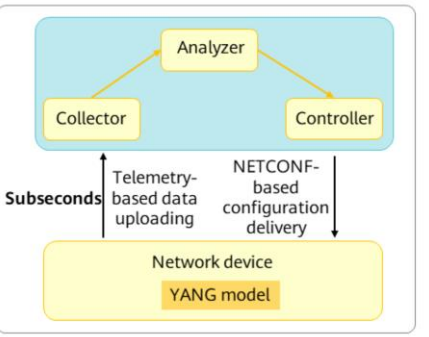
NETCONF overview

- The Network Configuration Protocol (NETCONF) provides a mechanism for the NMS to communicate with network devices.
- To be specific, the network administrator can use this mechanism to add, modify, and delete the configurations of network devices as well as obtain the configurations and status of network devices.
- NETCONF uses a YANG file to describe the data model of a device. A YANG file has a more hierarchical structure than a MIB file.

HUAWEI

Telemetry

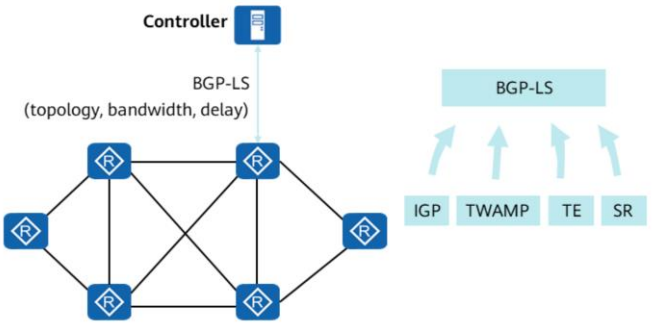
- Telemetry, also known as network telemetry, is mainly used to monitor networks, including packet check and analysis, intrusion and attack detection, intelligent data collection, and application performance management. Generally, it is used together with NETCONF. The analyzer analyzes the data collected by telemetry and then instructs the controller to automatically modify device configurations based on analysis results.
- Advantages of telemetry:
  - Is developed based on the YANG model.
  - 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.



HUAWEI

BGP-LS

- BGP-LS introduces new NLRI into BGP. The NLRI carries link, node, topology prefix, and other information, and is also referred to as the link state NLRI.
- BGP-LS can aggregate network-layer topology, bandwidth, delay, and other information and send the information to the controller for path computation.



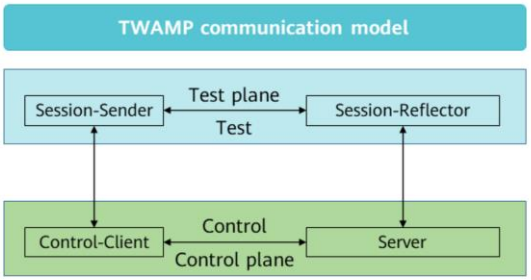
BGP-LS aggregates information collected by various network layer protocols, including:

- IGP: IGP information of each AS
- TWAMP: measurement information, such as interface delay
- TE: TE information, such as bandwidth
- SR: SR information, such as SR labels

HUAWEI

TWAMP

- Two-Way Active Measurement Protocol (TWAMP) measures the two-way delay, jitter, and packet loss rate between devices on an IP network. It performs negotiation over a TCP connection and uses UDP data packets as measurement packets.



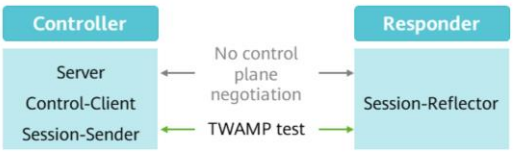
- Control-Client: establishes, starts, and stops a test session and also collects statistics.
- Server: responds to the Control-Client's request for establishing, starting, or stopping a test session
- Session-Sender: proactively sends probes for performance statistics after being notified by the Control-Client.
- Session-Reflector: replies to the probes sent by the Session-Sender with response probes after being notified by the Server.

HUAWEI

- TWAMP Light is a lightweight version of TWAMP defined based on standard protocols. Compared with TWAMP, TWAMP Light simplifies the control protocol used to establish performance measurement sessions.

TWAMP Light

- As the light version of TWAMP, TWAMP Light eliminates the necessity of the TWAMP-Control protocol, and moves the control plane from the Responder to the Controller so that TWAMP control modules can be centrally deployed on the Controller. Therefore, TWAMP Light greatly relaxes its requirements on the Responder performance, allowing the Responder to be rapidly deployed.
- The Control-Client, Server, and Session-Sender are deployed on the same host and function as the Controller. Therefore, the control session establishment process in the standard architecture can be ignored in the performance detection process. The Session-Reflector is deployed on another host as the Responder.



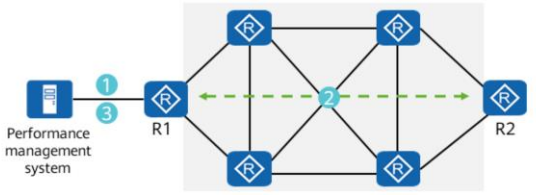
1. The Controller sends test packets to the Responder, and the Responder functions as the Session-Reflector to reflect the test packets.
2. The Session-Reflector does not need to learn Session status. After receiving the test packets, the Session-Reflector copies the necessary information, generates a test packet sequence number and timestamp, and returns the test packets to the Controller.
3. Upon receiving the reflected test packets, the Controller collects the indicators in both path directions.

HUAWEI

- The standard TWAMP version uses TCP for control plane negotiation, and test packets are based on UDP. The reflector needs to know the session status so that devices of different vendors can communicate with each other.
- TWAMP Light does not involve control plane negotiation, and test packets are also based on UDP. The implementation and configuration are simple, and the reflector does not need to know the session status.

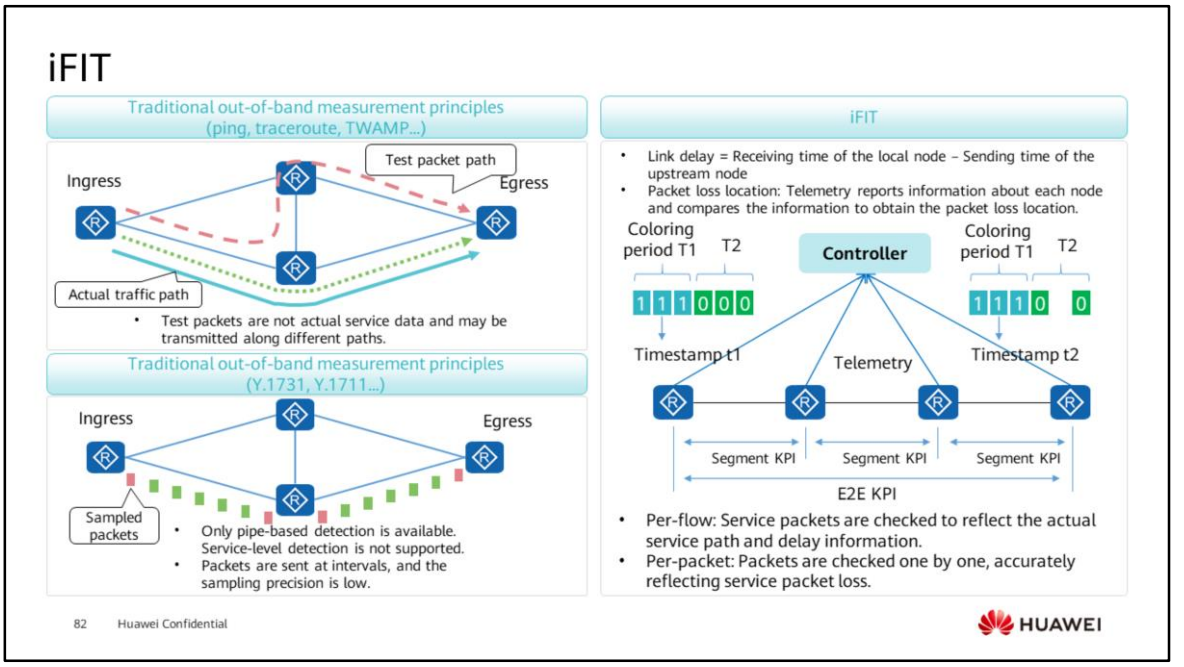
TWAMP Application Scenarios

- With TWAMP, NEs do not need to generate or maintain IP network performance statistics. The performance management system can easily obtain statistics about the entire network by managing only the TWAMP clients initiating statistics collection requests. In this way, IP performance statistics are collected quickly and flexibly.
- TWAMP is used on enterprise WANs to measure the delay, jitter, and packet loss rate between any two nodes, providing a reference for troubleshooting and traffic optimization.



1. The Controller instructs network devices to establish performance measurement sessions. For example, R1 functions as the Control-Client to initiate IP performance measurement.
2. R1 and R2 initiate a TWAMP measurement session.
3. R1 reports the measurement result to the Controller.

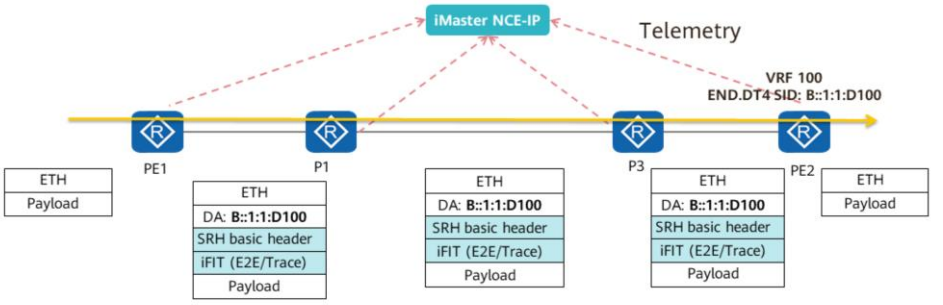
HUAWEI



- iFIT measures E2E service packets to obtain performance indicators of an IP network, such as the packet loss rate and delay. iFIT adds a color flag to the packet header in the service flow. Telemetry is used to periodically collect information. Features such as E2E delay measurement and packet loss measurement are supported.

iFIT for VPN over SRv6 Scenario

- Basic SRv6 scenario:
  - The ingress inserts the SRH between the IPv6 header and payload. The SRH carries the SRH basic header and iFIT extension header.
  - SRv6-capable nodes can report iFIT statistics in either E2E or hop-by-hop mode.
  - A node that does not support SRv6 but supports IPv6 forwarding can properly forward service packets carrying iFIT information.

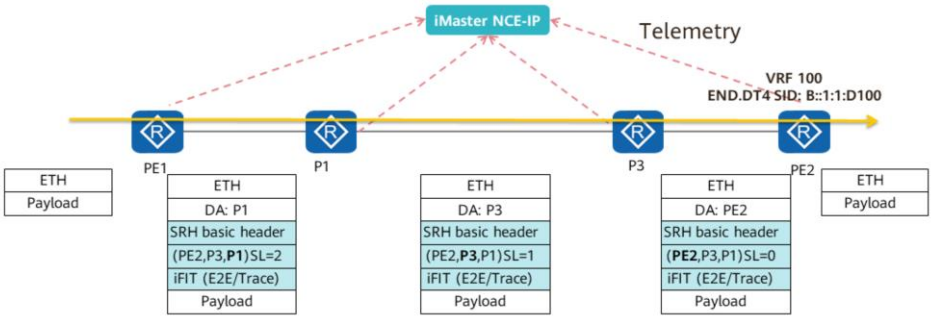


HUAWEI

- An End.DT4 SID (PE endpoint SID) identifies an IPv4 VPN instance on a network.
- For MPLS packets, the iFIT header is inserted between the MPLS label and MPLS payload.

iFIT for SRv6 Policy Scenarios

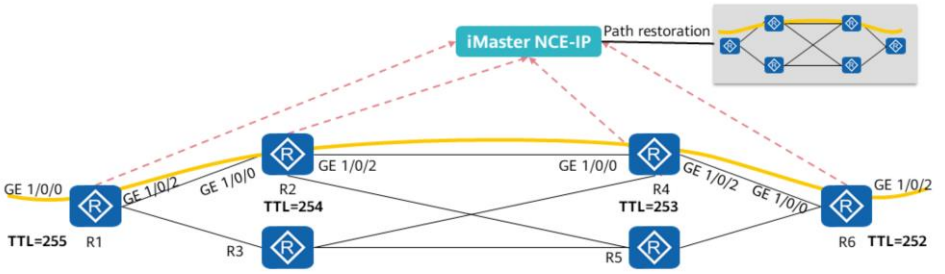
- In an SRv6 Policy scenario:
  - The iFIT extension header is encapsulated into the Optional TLV field of the SRH.
  - SRv6-capable nodes can report iFIT statistics in either E2E or hop-by-hop mode.
  - A node that does not support SRv6 but supports IPv6 forwarding can properly forward service packets carrying iFIT information.



HUAWEI

iFIT-based Service Path Display

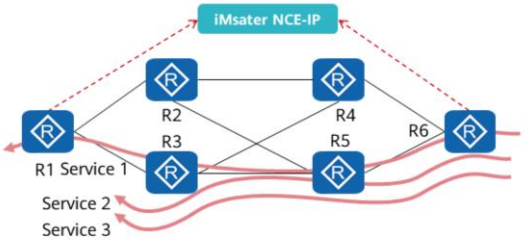
- In each measurement period, a device reports the flow direction, interface number, and TTL information when reporting packet statistics.
- iMaster NCE-IP restores the path information of the flow based on the information reported by each node.
- The implementation is independent of the tunnel type (SRv6/SRv6 Policy/SR-MPLS TE/SR-MPLS BE/MPLS...).



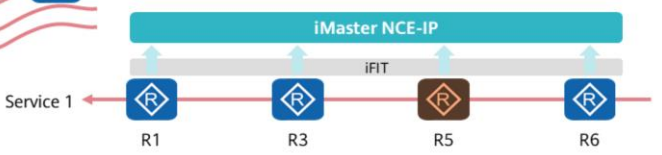
HUAWEI

iFIT-based Fault Locating

- Path aggregation is performed based on the physical topology, end nodes of poor-QoE services, and the built-in AI algorithm of iMaster NCE-IP to determine the minimum area that causes poor service quality, helping further locate and demarcate faults.



1. Find out poor-QoE services, such as services 1 to 3, on the entire network.
2. Intelligently compute the common paths of the three services.
3. Select a service for iFIT-based fault detection. Service 1 is used as an example here.
4. Determine the faulty device R5 based on per-hop iFIT measurement.



HUAWEI

Quiz

1. (True or False) An SR Policy is identified by <headend, color, endpoint> and contains multiple candidate paths. ( )
   A. True
   B. False
2. (Multiple-answer question) Which of the following are WAN bearer technologies? ( )
   A. SR-MPLS
   B. SRv6
   C. MPLS LDP
   D. MPLS TE

HUAWEI

1. A
2. ABCD

Summary

- This course introduces the concepts and principles of the enterprise bearer WAN's typical architecture, bearer technologies, VPN services, traffic optimization, SLA, reliability, and network management and analysis. To introduce these key aspects, this course uses a large enterprise with three data centers in two cities and multiple branches in different regions as an example.
- On a real production network, engineers need to determine the network architecture and technical applications based on the live network conditions and enterprises' services.

HUAWEI

Thank you.

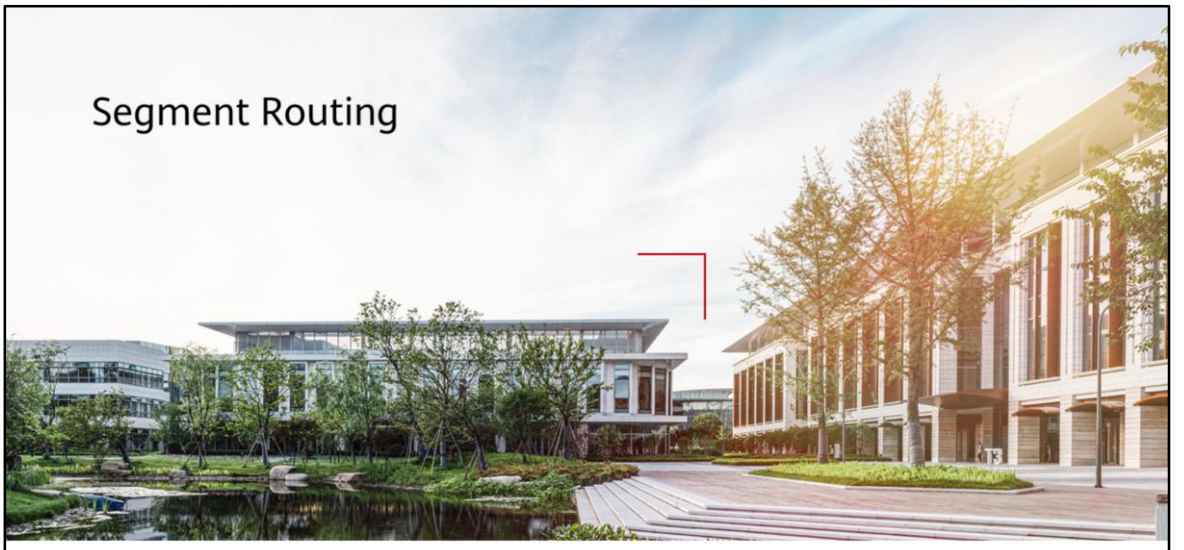
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Segment Routing



HUAWEI

Foreword

- Segment Routing (SR) is designed to forward data packets on a network using the source routing model.
- This document describes the source routing model of SR, segment definition, differences between SR-MPLS and SRv6, and scenario-specific SR-MPLS applications for Huawei NetEngine series routers.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the background of SR.
  - Describe the technical advantages of SR.
  - Describe the basic concepts involved in SR.
  - Describe the forwarding fundamentals of SR.
  - Master basic SR-MPLS configurations.

HUAWEI

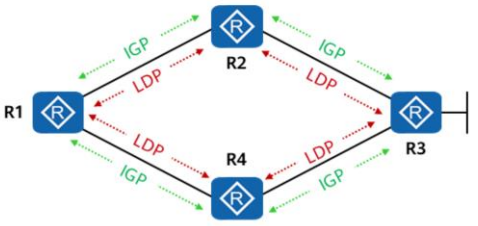
Contents

1. **Segment Routing Overview**
2. Segment Routing Fundamentals
3. Segment Routing Tunnel Protection and Detection Technologies
4. Typical Usage Scenarios of Segment Routing
5. Basic Configurations of Segment Routing

HUAWEI

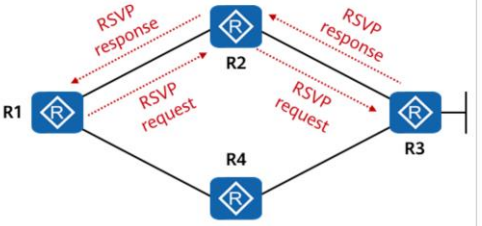
Problems in MPLS LDP and RSVP-TE

MPLS LDP



- LDP itself does not have the path computation capability and requires an IGP for path computation.
- Both the IGP and LDP need to be deployed for the control plane, and devices need to exchange a large number of packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- If LDP-IGP synchronization is not achieved, data forwarding may fail.

RSVP-TE



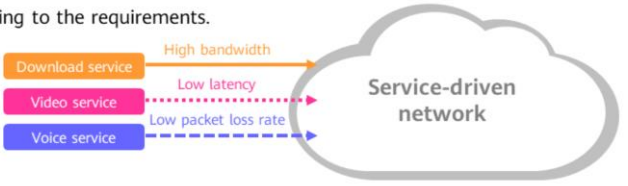
- RSVP-TE configuration is complex and load balancing is not supported.
- To implement TE, devices need to exchange a large number of RSVP packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- RSVP-TE uses a distributed architecture, so that each device only knows its own state and needs to exchange signaling packets with other devices.

HUAWEI

- In essence, MPLS is a tunneling technology used to guide data forwarding and has complete tunnel creation, management, and maintenance mechanisms. For the preceding mechanisms, networks are driven by network operation and management requirements, not by applications.

Service-Driven Network: Services Define the Network Architecture

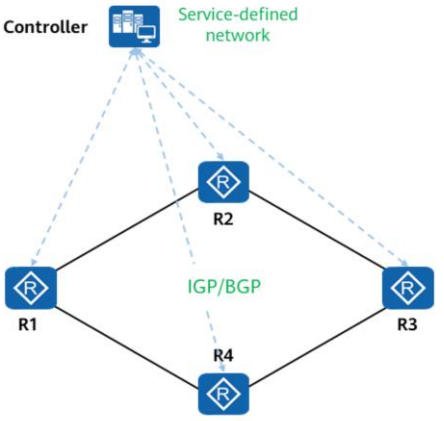
- The development of 5G and cloud services has changed the attributes and scope of network connections. More requirements are raised on connections, such as requiring better SLA guarantee, deterministic latency, or more information to be carried in packets.
- In this situation, the model that requires networks to adapt to services cannot keep up with rapid service development and even complicates network deployment and maintenance.
- To address this issue, the service-driven network model can be used, so that the network architecture is defined by services. Specifically, after an application raises requirements (e.g. latency, bandwidth, and packet loss rate), a controller is used to collect information (e.g. network topology, bandwidth usage, and latency) and compute an explicit path according to the requirements.



HUAWEI

- Traditionally, IP data packet forwarding is implemented based on IP addresses reachable to the destination over the shortest path. To meet the reliability requirements of services such as voice, online gaming, and video conferencing, the FRR technology is introduced. To meet the high bandwidth requirements of private line services such as group customer services, the TE technology is introduced. These technologies all represent network adaptation to services.
- The increasing types of services pose a variety of network requirements. For example, real-time Unified Communications and Collaboration (UC&C) applications usually prefer to paths with low latency and jitter, and big data applications prefer to high-bandwidth tunnels with a low packet loss rate. In this situation, the model that requires networks to adapt to services cannot keep up with rapid service development and even complicates network deployment and maintenance.
- The solution to this issue is to enable services to drive networks and define the network architecture. Specifically, after an application raises requirements (e.g. latency, bandwidth, and packet loss rate), a controller is used to collect information (e.g. network topology, bandwidth usage, and latency) and compute an explicit path according to the requirements.

SR Roadmap

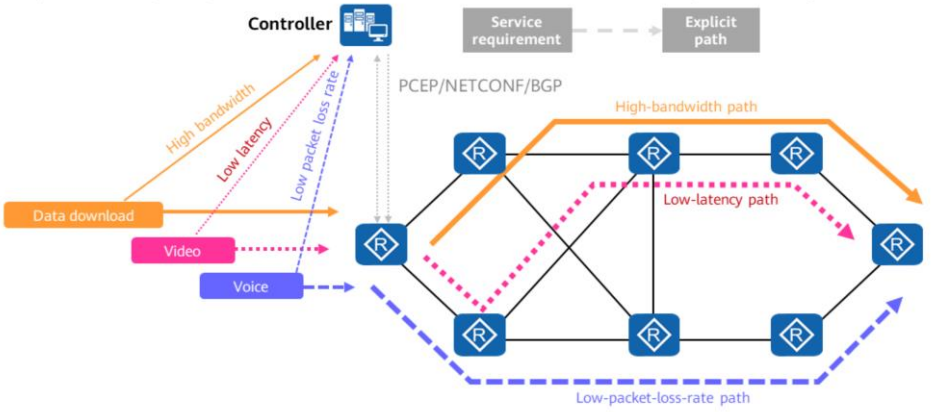


- Simplifies protocols and extends existing protocols.
  - The extended IGP/BGP supports label distribution. Therefore, LDP is not required on the network, achieving protocol simplification. In addition, devices require only software upgrades instead of hardware replacement, protecting the investment on the live network.
  - The source routing mechanism is introduced.
  - The specific forwarding policy is instantiated as a label list on the ingress to control the traffic forwarding path.
- Enables networks to be defined by services.
  - After an application raises requirements (e.g. latency, bandwidth, and packet loss rate), a controller is used to collect information (e.g. network topology, bandwidth usage, and latency) and compute an explicit path according to the requirements.

HUAWEI

SR Solution

- After services raise network requirements (e.g. latency, bandwidth, and packet loss rate), a controller computes an explicit path in a centralized manner and delivers an SR path to carry the services.



HUAWEI

SR Overview

- SR is designed to forward data packets on a network using the source routing model.
- SR divides a network path into several segments and assigns a segment ID (SID) to each segment and forwarding node. The segments and nodes are sequentially arranged into segment lists to form a forwarding path.
- SR encapsulates segment list information that identifies a forwarding path into the packet header for transmission. After a node receives the packet, it parses the segment list information. If the top SID in the segment list identifies the local node, the node removes the SID and executes the follow-up procedure. Otherwise, the node forwards the packet to the next hop in equal cost multiple path (ECMP) mode.
- SR has the following characteristics:
  - Extends existing protocols (e.g. IGP) to facilitate network evolution.
  - Supports both controller-based centralized control and forwarder-based distributed control, providing a balance between the two control modes.
  - Enables networks to quickly interact with upper-layer applications through the source routing technology.

HUAWEI

- https://datatracker.ietf.org/doc/rfc8402/

SR Advantages

Simplified control plane of the MPLS network	• SR uses a controller or IGP to uniformly compute paths and allocate labels, without the need to use tunneling protocols such as RSVP-TE and LDP. • SR can be directly used in the MPLS architecture, without requiring changes to the forwarding plane.
Efficient TI-LFA FRR protection against path failures	• SR works with remote loop-free alternate (RLFA) FRR to provide efficient topology-independent loop-free alternate (TI-LFA) FRR. • TI-LFA FRR offers node and link protection for all topologies, addressing the weakness in traditional tunnel protection technologies.
Enhanced network capacity expansion capability	• MPLS TE is a connection-oriented technology. To maintain connection states, devices need to exchange and process numerous keepalive packets, straining the control plane. • SR can control any service path by merely performing label operations for packets on the ingress. It does not require transit nodes to maintain path information, thereby freeing up the control plane. Moreover, the SR label quantity is the sum of the node quantity and local adjacency quantity on the entire network, meaning that it is related only to the network scale, rather than the tunnel quantity or service volume.
Smoother evolution to SDN networks	• As SR is designed based on the source routing model, the ingress controls packet forwarding paths. • SR can work with the centralized path computation module to flexibly and easily control and adjust paths. • SR supports both traditional networks and SDN networks and is compatible with existing devices, ensuring smooth evolution to SDN networks.

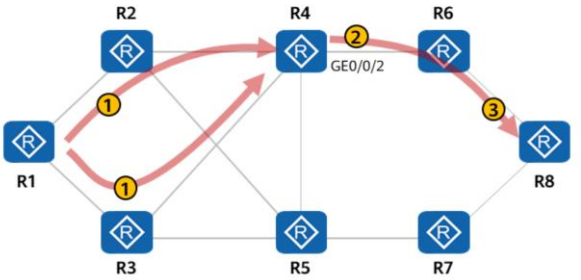
HUAWEI

Contents

1. Segment Routing Overview
2. **Segment Routing Fundamentals**
3. Segment Routing Tunnel Protection and Detection Technologies
4. Typical Usage Scenarios of Segment Routing
5. Basic Configurations of Segment Routing

HUAWEI

Basic Concept: Segment

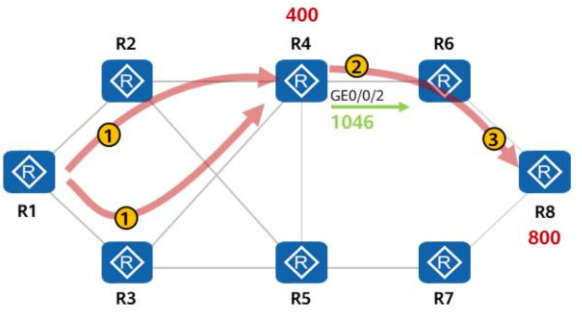


- A segment represents an instruction to be executed by a node for a received data packet, and the instruction is encapsulated in the packet header.
- For example:
  - Instruction 1: Forward the packet to R4 over the shortest path (ECMP supported).
  - Instruction 2: Forward the packet through GE0/0/2 of R4.
  - Instruction 3: Forward the packet to R8 over the shortest path.

HUAWEI

Basic Concept > Fundamentals > SR-MPLS BE > SR-MPLS TE > SR-MPLS Policy

Basic Concept: Segment ID

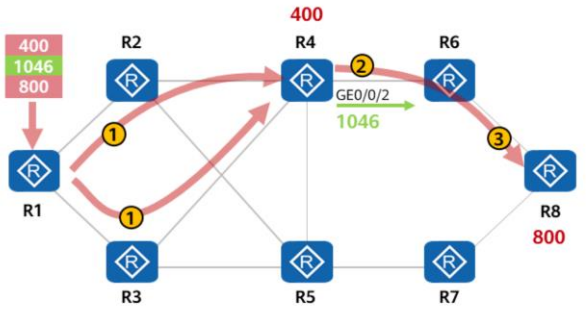


- Segment IDs (SIDs) identify segments. The SID format depends on the specific technical implementation. For example, SIDs can be MPLS labels, indexes in an MPLS label space, or IPv6 addresses.
- A segment list is an ordered list of one or more SIDs.
- For example:
  - Instruction 1 (400): Forward the packet to R4 over the shortest path (ECMP supported).
  - Instruction 2 (1046): Forward the packet through GE0/0/2 of R4.
  - Instruction 3 (800): Forward the packet to R8 over the shortest path (ECMP supported).

HUAWEI

- The label values used in this course are only examples. For details about the label allocation scope, see the corresponding product documentation.

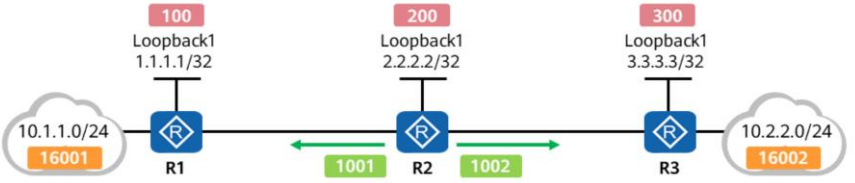
Basic Concept: Source Routing



- Source routing: The source node selects a forwarding path and encapsulates an ordered segment list into a packet. After receiving the packet, other nodes forward it based on the segment list information.

HUAWEI

Basic Concept: Segment Classification

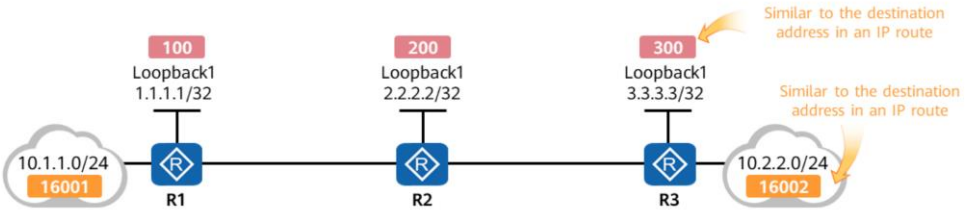


Category	Description
Prefix segment	Identifies the prefix of a destination address on a network. Generation mode: manual configuration Prefix segments are propagated to other devices through an IGP. They are visible to and effective on all the devices. **Node segments are special prefix segments.**
Adjacency segment	Identifies an adjacency on a network. Generation mode: dynamic allocation by the ingress through a protocol Adjacency segments are propagated to other devices through an IGP. They are visible to all the devices but effective only on the local device.

Prefix SID | Node SID | Adjacency SID Note: SIDs are identified in the same way in the following parts.

HUAWEI

Basic Concept: Prefix Segment



Prefix Segment

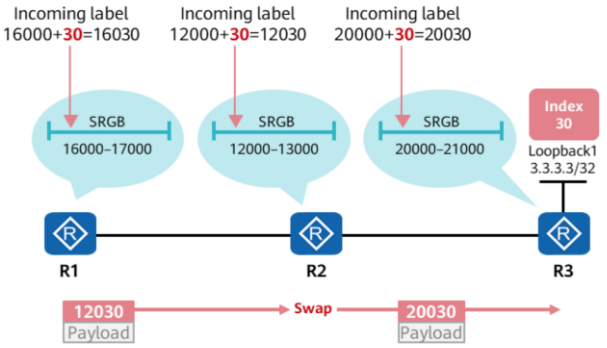
- Identifies the prefix of a destination address on a network. Prefix segments are propagated to other devices through an IGP. They are visible to and effective on all the devices.

- Prefix segments are identified using prefix SIDs.
- A prefix SID is an offset value within the Segment Routing global block (SRGB) range advertised by the advertising end. The receiving end calculates the actual label value based on its own SRGB to generate an MPLS forwarding entry.

- Node segments are special prefix segments used to identify specific nodes.
- When an IP address is configured as a prefix for a node's loopback interface, the prefix SID of the node is the node SID.

HUAWEI

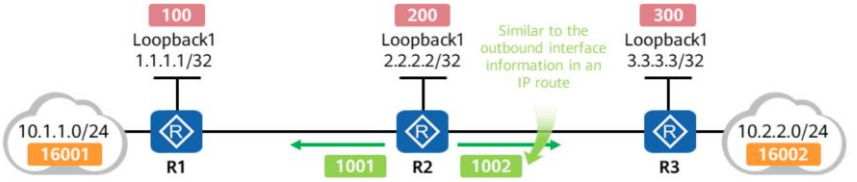
Basic Concept: SRGB



- Segment Routing global block (SRGB): a set of user-specified global labels reserved for SR-MPLS.
- Each device advertises its SRGB through an extended routing protocol.
- After a node advertises the prefix SID index through an extended routing protocol, each device receiving the index calculates the incoming and outgoing SIDs based on the SRGB.
- In actual deployment, it is recommended that devices use the same SRGB.
- Why is SRGB required?
  - SR requires prefix SIDs to be globally valid.
  - In MPLS, some label space of a device may be occupied by other protocols, such as LDP. Therefore, a specific space must be specified for global SR labels.

HUAWEI

Basic Concept: Adjacency Segment



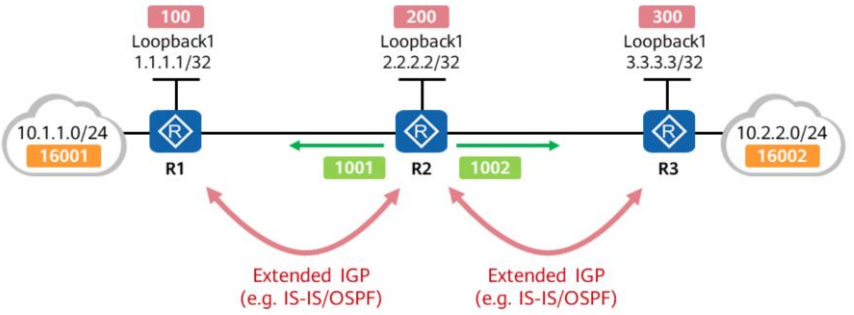
Adjacency Segment
Identifies an adjacency on a network. Adjacency segments are propagated to other devices through an IGP. They are visible to all the devices but effective only on the local device.

- Adjacency segments are identified using adjacency SIDs.
- Adjacency SIDs are local SIDs that are not in the SRGB range.

HUAWEI

Intra-AS Propagation of Node SIDs and Adjacency SIDs

- SR-MPLS uses an IGP to advertise topology, prefix, SRGB, and label information. This is achieved by extending the TLVs of protocol packets for the IGP.



HUAWEI

OSPF for SR-MPLS

Name	Function	Carried In
SR-Algorithm TLV	Advertises the algorithm that is used.	Type 10 Opaque LSA
SID/Label Range TLV	Advertises the SR-MPLS SID or MPLS label range.	Type 10 Opaque LSA
SRMS Preference TLV	Advertises the priority of an NE functioning as an SR mapping server.	Type 10 Opaque LSA
SID/Label Sub-TLV	Advertises SR-MPLS SIDs or MPLS labels.	SID/Label Range TLV
		OSPFv2 Extended Prefix TLV and OSPF Extended Prefix Range TLV in OSPFv2 Extended Prefix Opaque LSA
		OSPFv2 Extended Link TLV in OSPFv2 Extended Link Opaque LSA
Prefix SID Sub-TLV	Advertises SR-MPLS prefix SIDs.	OSPFv2 Extended Prefix TLV and OSPF Extended Prefix Range TLV in OSPFv2 Extended Prefix Opaque LSA
Adj-SID Sub-TLV	Advertises SR-MPLS adjacency SIDs on a P2P network.	OSPFv2 Extended Link TLV in OSPFv2 Extended Link Opaque LSA
LAN Adj-SID Sub-TLV	Advertises SR-MPLS adjacency SIDs on a LAN.	OSPFv2 Extended Link TLV in OSPFv2 Extended Link Opaque LSA

HUAWEI

IS-IS for SR-MPLS

Name	Function	Carried In
Prefix-SID Sub-TLV	Advertises SR-MPLS prefix SIDs.	IS-IS Extended IPv4 Reachability TLV-135 IS-IS Multitopology IPv4 Reachability TLV-235 IS-IS IPv6 IP Reachability TLV-236 IS-IS Multitopology IPv6 IP Reachability TLV-237 SID/Label Binding TLV
Adj-SID Sub-TLV	Advertises SR-MPLS adjacency SIDs on a P2P network.	IS-IS Extended IS reachability TLV-22 IS-IS IS Neighbor Attribute TLV-23 IS-IS inter-AS reachability information TLV-141 IS-IS Multitopology IS TLV-222 IS-IS Multitopology IS Neighbor Attribute TLV-223
LAN-Adj-SID Sub-TLV	Advertises SR-MPLS adjacency SIDs on a LAN.	IS-IS Extended IS reachability TLV-22 IS-IS IS Neighbor Attribute TLV-23 IS-IS inter-AS reachability information TLV-141 IS-IS Multitopology IS TLV-222 IS-IS Multitopology IS Neighbor Attribute TLV-223
SID/Label Sub-TLV	Advertises SR-MPLS SIDs or MPLS labels.	SR-Capabilities Sub-TLV and SR Local Block Sub-TLV
SID/Label Binding TLV	Advertises the mapping between prefixes and SIDs.	IS-IS LSP
SR-Capabilities Sub-TLV	Advertises SR-MPLS capabilities.	IS-IS Router Capability TLV-242
SR Local Block Sub-TLV	Advertises the range of labels reserved for local SIDs.	IS-IS Router Capability TLV-242

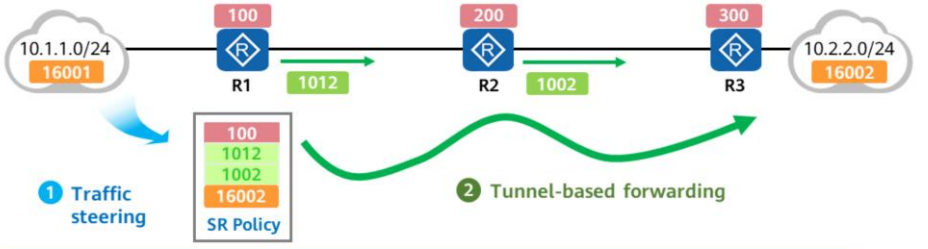
HUAWEI

Basic Concept: SR Policy

- According to RFC 8402, an SR Policy is an ordered list of segments. In addition, it defines a framework for SR technologies used to calculate/generate/maintain the segment list and steer traffic. Currently, SR Policy is the mainstream SR implementation mode.
- Traffic is steered into an SR Policy by the headend. The involved segment list is accurately encapsulated as a label stack to guide traffic forwarding. It is calculated based on a series of optimization objectives and constraints, such as latency, affinity, and SRLG. The calculation can be performed locally or by a controller and then applied to the network.

HUAWEI

SR Policy Example



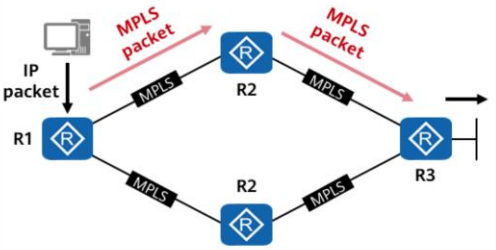
SR Policy:

- Can be generated using different modes, such as CLI, NETCONF, PCEP, and BGP SR Policy.
- Contains segment lists to guide traffic steering and forwarding.

HUAWEI

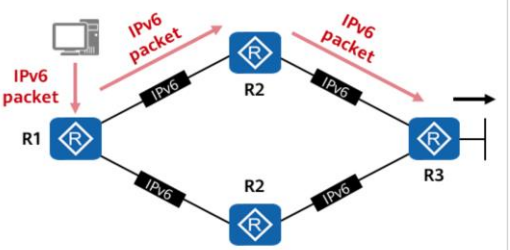
Basic Concept: SR-MPLS and SRv6

SR-MPLS



- Data forwarding plane: based on MPLS
- MPLS labels are used as SIDs.
- Segment list information is encoded as a label stack. The segment to be processed is at the stack top. Once a segment is processed, the corresponding label is removed from the label stack.

SRv6

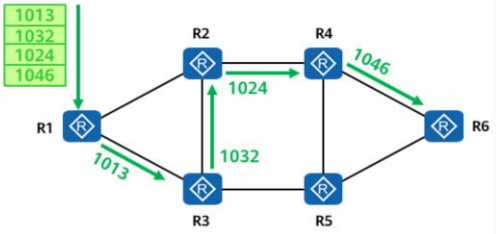


- Data forwarding plane: based on IPv6
- IPv6 addresses are used as SIDs.
- Segment list information is encoded as a label stack and carried using the IPv6 Segment Routing header (SRH).

HUAWEI

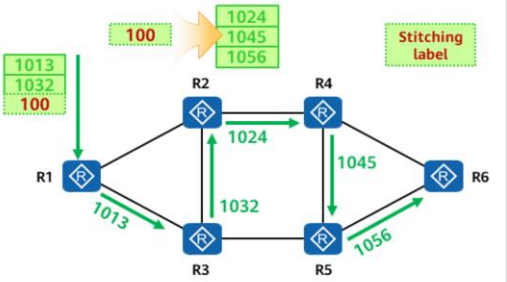
Label Stack, Stitching Label, and Stitching Node

Label Stack



- A label stack is an ordered set of labels used to identify a complete LSP.
- Each adjacency label in the label stack identifies an adjacency, and the entire label stack identifies all adjacencies along the LSP.
- During packet forwarding, a node searches for the corresponding adjacency according to each adjacency label in the label stack, removes the label, and then forwards the packet. After all the adjacency labels in the label stack are removed, the packet traverses the entire LSP and reaches the tunnel destination.

Stitching Label and Stitching Node



- If the label stack depth exceeds the maximum depth supported by forwarders, the controller needs to allocate multiple label stacks to the forwarders and a special label to an appropriate node to stitch these label stacks, thereby implementing segment-by-segment forwarding.
- This special label is called a stitching label, and this appropriate node is called a stitching node. The controller allocates a stitching label to the stitching node and pushes it to the bottom of the label stack.

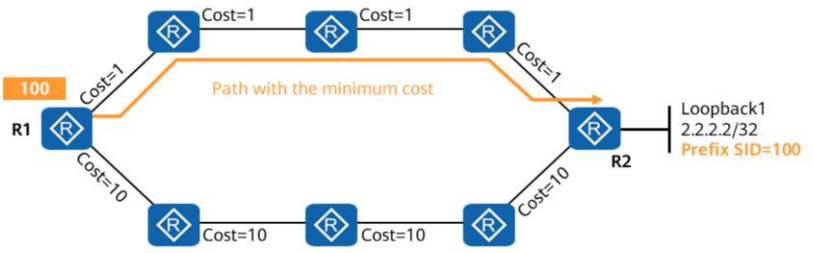
HUAWEI

How Are SIDs Used

- Combining prefix (node) and adjacency SIDs in sequence can construct any network path.
- Every hop on a path identifies the next hop based on the top SID in the label stack.
- SID information is stacked in sequence at the top of the data header.
- If the top SID identifies another node, the receive node forwards the data packet to that node in ECMP mode.
- If the top SID identifies the local node, the receive node removes the top SID and proceeds with the follow-up procedure.
- In real-world applications, prefix segments and adjacency segments can be used separately or together.

HUAWEI

Scenario 1: Prefix Segment-based Forwarding Path



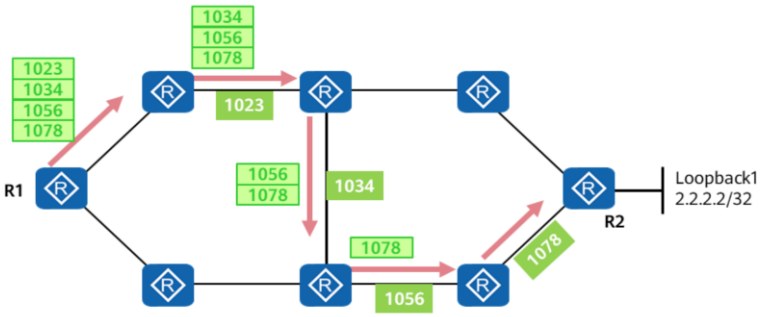
A prefix segment-based forwarding path is computed by an IGP using the SPF algorithm.

1. After the prefix SID (100) of R2 is propagated using an IGP, all devices in the IGP domain learn the SID.
2. R1 is used as an example (the implementation for other devices is similar to this). It runs SPF to compute the shortest path to R2.

Prefix segment-based forwarding paths are not fixed, and the ingress cannot control the entire packet forwarding path.

HUAWEI

Scenario 2: Adjacency Segment-based Forwarding Path

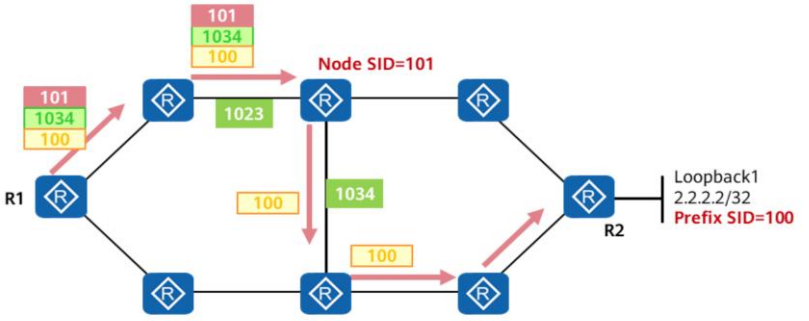


An adjacency segment is allocated to each adjacency on the network, and a segment list containing multiple adjacency segments is defined on the ingress.

This method can be used to specify any strict explicit path, facilitating SDN implementation.

HUAWEI

Scenario 3: Adjacency Segment+Node Segment-based Forwarding Path

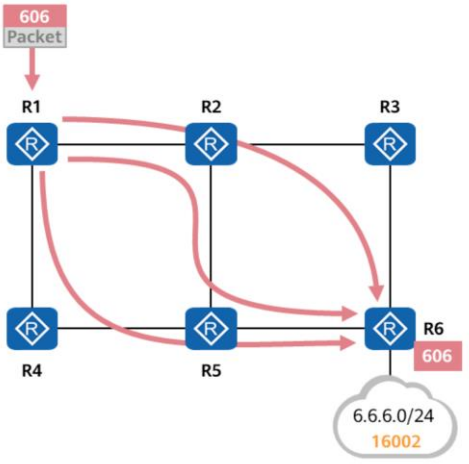


Adjacency and node segments can be used together. An adjacency segment can be specified to force a path to traverse an adjacency. The node corresponding to a node segment can run SPF to compute the shortest path that supports ECMP.

Paths established in this mode are not strictly fixed, and therefore, they are also called loose explicit paths.

HUAWEI

SR-MPLS BE

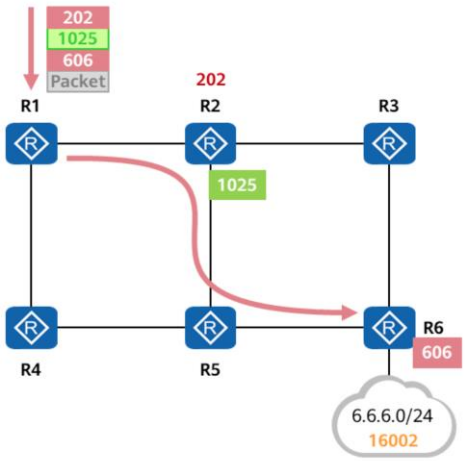


SR-MPLS BE

- In SR-MPLS best effort (BE) mode, SIDs are used to guide data forwarding over the shortest path.
- In this example, node SID 606 of R6 is used to instruct data to be forwarded over the shortest path to R6. The shortest path is computed through a routing protocol and supports ECMP.
- SR-MPLS BE is a new solution that replaces the LDP+IGP solution.

HUAWEI

SR-MPLS TE



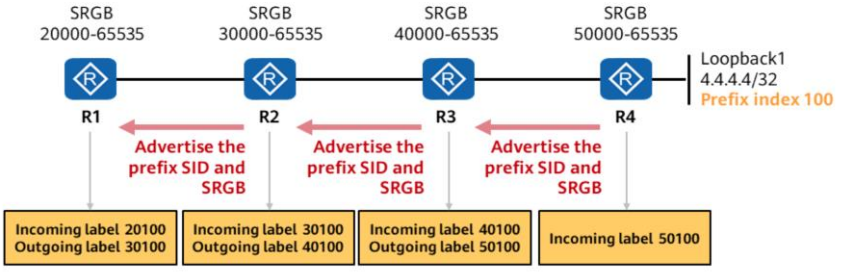
SR-MPLS TE

- In SR-MPLS TE mode, multiple SIDs are combined to guide data forwarding based on constraints, thereby meeting traffic engineering requirements.
- Methods of combining SIDs:
  - Combine multiple node SIDs.
  - Combine multiple adjacency SIDs.
  - Combine node and adjacency SIDs, as shown in the figure.

HUAWEI

SR-MPLS BE LSP

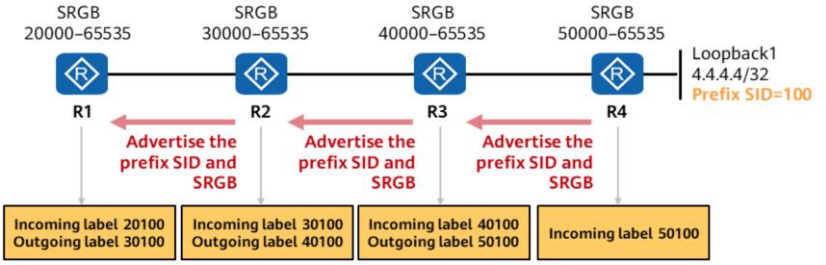
- An SR-MPLS BE LSP is a label forwarding path established using the SR technology. It uses a prefix or node segment to guide packet forwarding.
- An SR-MPLS BE LSP is the optimal SR LSP computed by an IGP using the SPF algorithm.
- The creation and data forwarding of SR-MPLS BE LSPs are similar to those of LDP LSPs. SR-MPLS BE LSPs do not have tunnel interfaces.



HUAWEI

SR-MPLS BE LSP Creation

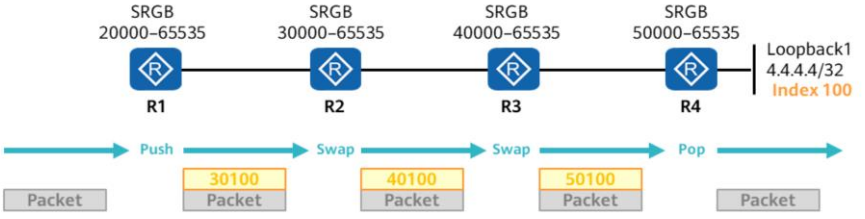
- LSP creation involves the following operations:
  - Network topology reporting (required only in controller-based LSP creation) and label allocation
  - Path computation
- SR-MPLS BE LSPs are created primarily based on prefix labels. Specifically, the destination node runs an IGP to advertise a prefix SID. After receiving the packet carrying the SID, forwarders parse the packet to obtain the SID and compute label values based on their own SRGBs. Then, using the IGP-collected topology information, each node runs the SPF algorithm to compute a label forwarding path, and delivers the computed next hop and outgoing label (OuterLabel) information to the forwarding table to guide data packet forwarding.



HUAWEI

Data Forwarding Process

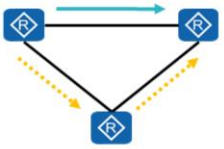
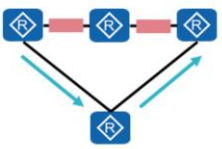
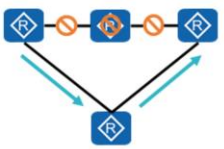
- Push: When a packet enters an LSP, the ingress adds a label between the Layer 2 and IP headers of the packet or adds a new label on top of the existing label stack.
- Swap: After receiving a packet forwarded within the SR domain, a node uses the label allocated by the next hop to replace the top label according to the label forwarding table.
- Pop: When a packet leaves the SR domain, the egress searches for the outbound interface according to the top label in the packet and then removes the top label.



HUAWEI

Traffic Engineering

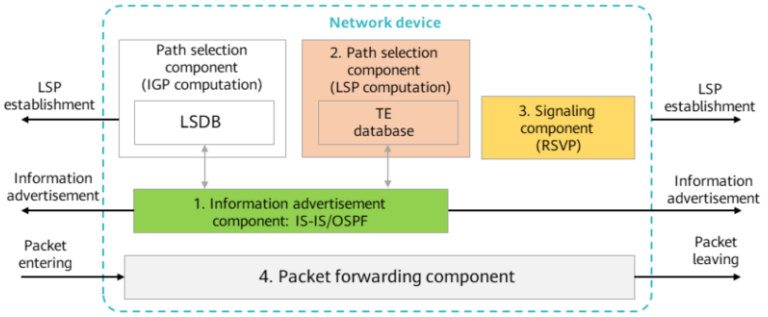
- Traffic engineering (TE) is one of the most important network services. The traditionally popular TE technology is based on MPLS and therefore is called MPLS TE. It can accurately control the path through which traffic passes, maximizing bandwidth utilization.

Path Planning	Traffic Optimization	Fault Protection
		
• Different paths are planned for different services.	• When traffic is unbalanced due to major events, traffic is evenly distributed to idle links.	• A fast protection switching is performed in the case of a device or link fault.

HUAWEI

Traditional Distributed MPLS TE Architecture

- MPLS TE uses the distributed architecture, in which the ingress computes paths according to constraints and uses RSVP-TE signaling to establish constraint-based LSPs.
- MPLS nodes are used to maintain a complete TE architecture through four components: information advertisement component, path computation component, path establishment component (or signaling component), and packet forwarding component.

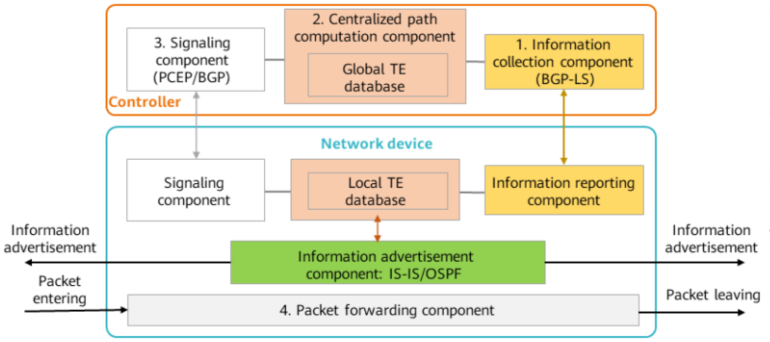


1. The extended IS-IS/OSPF carries TE information, advertises IGP and TE information in the domain, and generates a TEDB.
2. The CSPF algorithm is used to compute a path that meets constraints based on the TEDB.
3. RSVP-TE is used to establish LSPs.
4. Data is forwarded based on MPLS labels.

HUAWEI

Centralized SR-MPLS TE Architecture

- Segment Routing-MPLS Traffic Engineering (SR-MPLS TE) is a new TE tunneling technology that uses SR as the control protocol. SR-MPLS TE supports the centralized architecture, in which the controller collects global network topology and TE information, computes paths in a centralized manner, and delivers path computation results to network devices.
- SR-MPLS TE also supports manual configuration.



Centralized SR-MPLS TE:

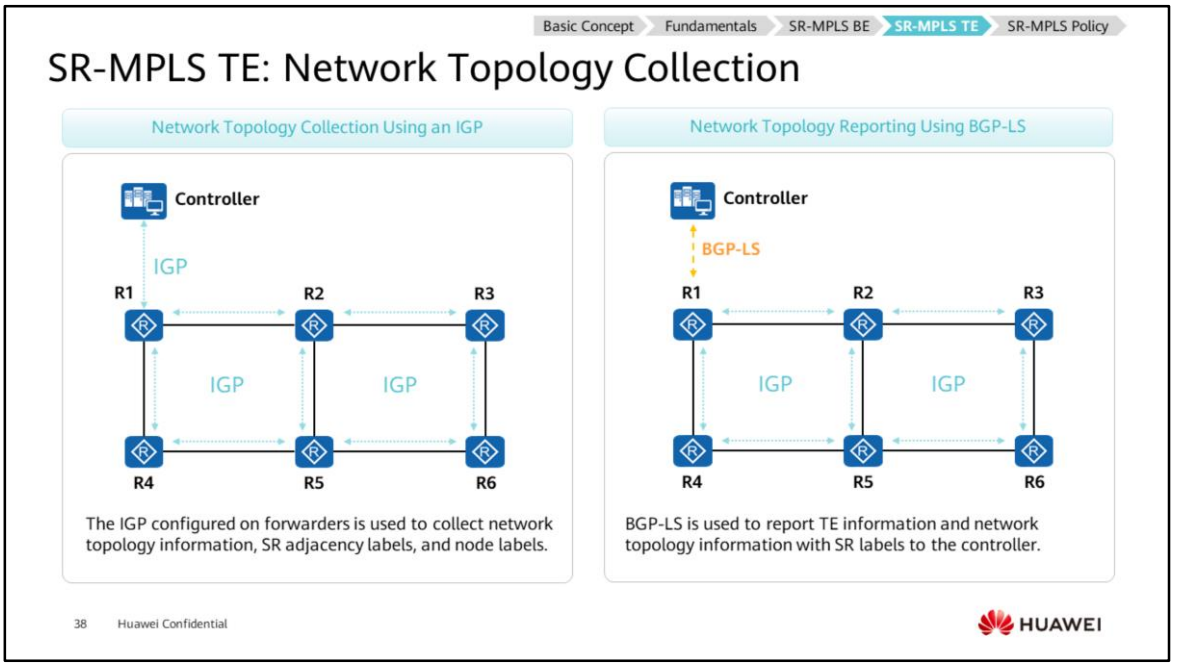
1. The extended IS-IS/OSPF carries TE information, advertises IGP and TE information in the domain, and generates a TEDB.
2. BGP-LS is used to collect network information and establish a global TE database.
3. The controller globally computes paths based on constraints.
4. PCEP or BGP SR Policy is used to deliver path computation results to devices.

HUAWEI

Comparison Between SR-MPLS TE and RSVP-TE

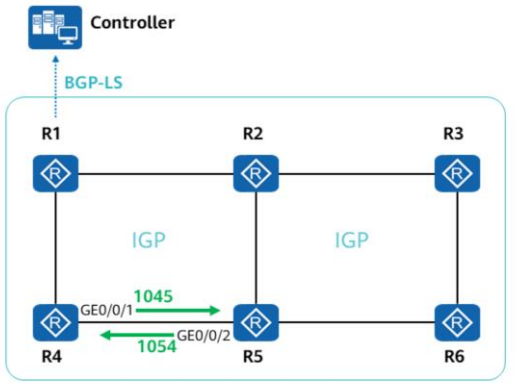
Item	SR-MPLS TE	RSVP-TE
Label allocation	Labels are allocated and propagated using IGP extensions. Each link is allocated with only one label. All LSPs traversing a link share the label of this link, reducing label resource consumption and the workload in label forwarding table maintenance.	Labels are allocated and propagated using RSVP extensions. Each LSP is allocated with a label. When there are multiple LSPs, multiple labels need to be allocated to the same link, occupying a large number of label resources and increasing the workload of maintaining the label forwarding table.
Control plane	IGP extensions are used for signaling control, reducing the number of required protocols.	RSVP-TE needs to be used as the MPLS control protocol, complicating the control plane.
Scalability	As transit nodes are unaware of tunnels and use packets to carry tunnel information, they only need to maintain forwarding entries instead of tunnel state information, enhancing scalability.	Tunnel state information and forwarding entries need to be maintained, resulting in poor scalability.
Path adjustment and control	Transit nodes are unaware of tunnels. The service path can be controlled only by performing label operations on the packet sent from the ingress, eliminating the need of hop-by-hop configuration delivery. If a node in the path fails, the controller re-computes a path and updates the label stack of the ingress to complete path adjustment.	Configurations need to be delivered node by node regardless of whether the path is adjusted in normal or fault scenarios.

HUAWEI



- For SR-capable IGP instances, all IGP-enabled outbound interfaces are allocated with SR adjacency labels, which are propagated to the entire network through an IGP.
- In Huawei's early solutions, an IGP can also be used to collect network topology information. Due to IGP area-related restrictions, BGP-LS is mainly used at present.

SR-MPLS TE: Label Allocation

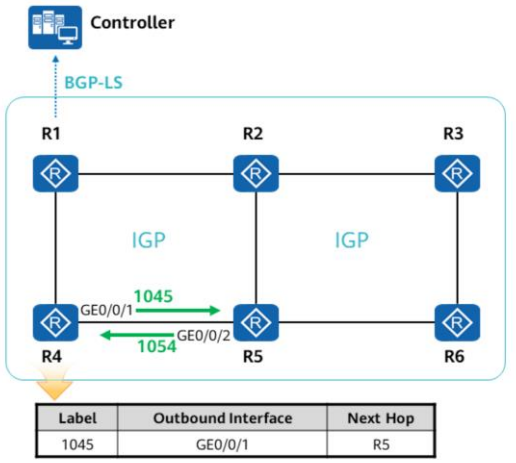


In SR-MPLS TE, labels are allocated through the IGP configured on forwarders and reported to a controller through BGP-LS.

- SR-MPLS TE mainly uses adjacency labels and can also use node labels.
- Adjacency labels are allocated by the ingress, and are valid locally and unidirectional.

HUAWEI

Label Allocation Example



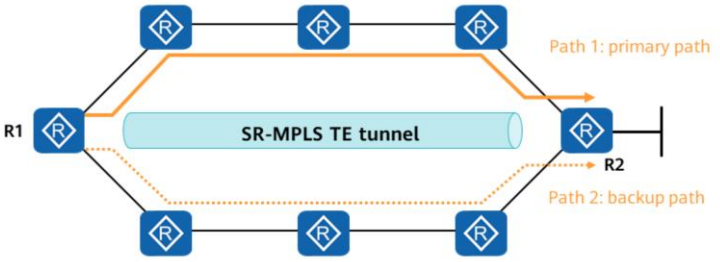
Label	Outbound Interface	Next Hop
1045	GE0/0/1	R5

- IGP SR is enabled on each device. For SR-capable IGP instances, all IGP-enabled outbound interfaces are allocated with SR adjacency labels.
- Adjacency labels are propagated to the entire network through an IGP SR extension.
- Taking R4 as an example, the process of label allocation through an IGP is as follows:
  1. R4 allocates a local dynamic label to an adjacency through an IGP. For example, adjacency label 1045 is allocated to the R4->R5 adjacency.
  2. R4 propagates the adjacency label to the entire network through the IGP.
  3. R4 generates a label forwarding entry corresponding to the adjacency label.
  4. Other nodes learn the R4-propagated adjacency label through the IGP but do not generate label forwarding entries.
- Other devices allocate and propagate adjacency labels in the same way as R4 and generate label forwarding entries. BGP-LS is used to report TE information and network topology information with SR labels to the controller.

HUAWEI

SR-MPLS TE LSP Creation

- SR-MPLS TE tunnels are created using the SR protocol based on TE constraints. The figure shows two LSPs working in primary/backup mode. The two LSPs correspond to the same SR-MPLS TE tunnel with a specified ID.



- SR-MPLS TE tunnel creation involves tunnel attribute configuration and tunnel establishment.

HUAWEI

- Before SR-MPLS TE tunnel creation, IS-IS/OSPF neighbor relationships must be established between forwarders to implement network layer connectivity, allocate labels, and collect network topology information. In addition, the forwarders need to report label and network topology information to a controller for path computation. If no controller is available, CSPF can be enabled on the ingress of the SR-MPLS TE tunnel so that forwarders can compute paths using CSPF.

SR-MPLS TE Tunnel Attribute Configuration

- SR-MPLS TE tunnel attributes must be configured before tunnel establishment. An SR-MPLS TE tunnel can be configured on a controller or forwarder.
- Tunnel configuration on a controller: After an SR-MPLS TE tunnel is configured on a controller, the controller uses NETCONF to deliver tunnel attributes to a forwarder, which then uses PCEP to delegate the tunnel to the controller for management.
- Tunnel configuration on a forwarder: After an SR-MPLS TE tunnel is configured on a forwarder, the forwarder delegates the tunnel to the controller for management.

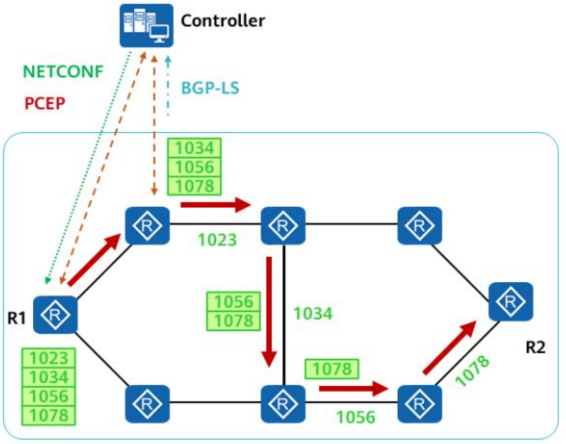
Manual Configuration of a Tunnel with an Explicit Path	NETCONF-based Tunnel Configuration Delivery by a Controller
[R1] interface tunnel1 [R1-Tunnel1] ip address unnumbered interface LoopBack0 [R1-Tunnel1] tunnel-protocol mpls te [R1-Tunnel1] destination 3.3.3.3 [R1-Tunnel1] mpls te tunnel-id 1 [R1-Tunnel1] mpls te signal-protocol segment-routing [R1-Tunnel1] mpls te path explicit-path p1 # A path is manually specified.	[R1] interface tunnel1 [R1-Tunnel1] ip address unnumbered interface LoopBack0 [R1-Tunnel1] tunnel-protocol mpls te [R1-Tunnel1] destination 3.3.3.3 [R1-Tunnel1] mpls te tunnel-id 1 [R1-Tunnel1] mpls te signal-protocol segment-routing [R1-Tunnel1] mpls te pce delegate # The tunnel is delegated to the PCE server.

SR-MPLS TE tunnels are established and managed using tunnel interfaces. As such, you need to configure a tunnel interface on the ingress of each SR-MPLS TE tunnel.

HUAWEI

- For SR-MPLS TE tunnel configuration on a forwarder, in addition to manually specifying an explicit path, you can also use the function of path computation by the ingress.

SR-MPLS TE Tunnel Establishment (Path Computation by the Controller)



If a configured service (e.g. VPN service) needs to be bound to an SR-MPLS TE tunnel, the tunnel can be established as follows:

1. Based on SR-MPLS TE tunnel constraints, the controller uses the path computation element (PCE) to compute a path similar to a common TE tunnel and generates a label stack (path computation result).
2. The controller uses NETCONF and PCEP to deliver tunnel configurations and the tunnel stack, respectively, to forwarders.
3. The forwarders establish an SR-MPLS TE tunnel with a specific LSP based on the tunnel configurations and label stack delivered by the controller.

BGP-LS: used to report labels and network topology information by forwarders.

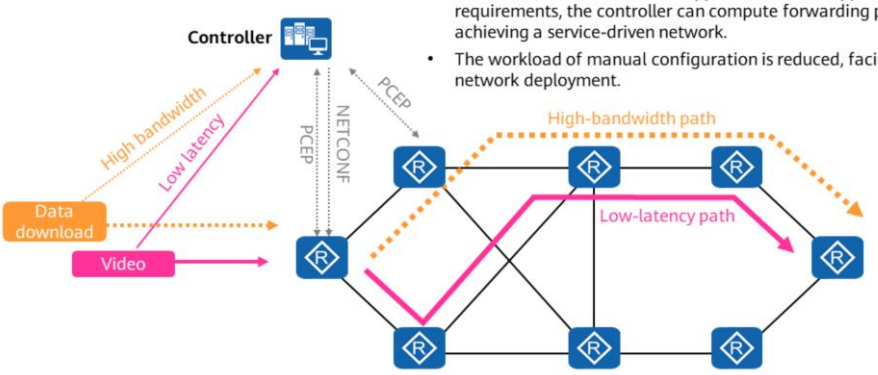
PCEP: used to deliver a label stack by a controller and report LSP states by forwarders.

NETCONF: used to deliver tunnel configurations by a controller.

HUAWEI

Advantages of Controller-based SR-MPLS TE Tunnel Establishment

- Bandwidth calculation and resource reservation are supported.
- The optimal path can be computed from a global perspective.
- The controller can work with applications. After applications raise network requirements, the controller can compute forwarding paths as required, achieving a service-driven network.
- The workload of manual configuration is reduced, facilitating large-scale network deployment.



HUAWEI

SR-MPLS TE Data Forwarding

- Forwarders perform label operations on packets according to the label stacks corresponding to a specific SR-MPLS TE tunnel's LSP and search for outbound interfaces hop by hop according to the top label to guide packet forwarding to the destination. Data can be forwarded based on adjacency labels or a combination of node and adjacency labels.
- Forwarding based on adjacency labels
  - Forwarding based on adjacency labels is also called strict-path forwarding. The label stack strictly determines the forwarding path and does not support load balancing.
- Forwarding based on a combination of node and adjacency labels
  - Forwarding based on a combination of node and adjacency labels is also called loose-path forwarding. When processing node labels, a device can forward packets along the shortest path or perform load balancing because the path is not strictly fixed in this case.

HUAWEI

- Currently, the mainstream solution is strict-path forwarding based on adjacency labels.

SR-MPLS BE and SR-MPLS TE Traffic Steering

- Traffic steering: After an SR tunnel is established, service traffic needs to be steered to it.
- SR-MPLS BE (without tunnel interfaces) traffic steering
  - Tunnel policy: Use a tunnel type prioritizing policy to select an SR-BE tunnel.
  - Static route: Specify the next hop of a static route as the destination address of an SR-BE tunnel and recurse traffic to the tunnel based on the next hop.
  - Recursion based on the next hop of a route: Recurse a public network route (e.g. BGP route) to an SR-BE tunnel based on the route's next hop.
- SR-MPLS TE (with tunnel interfaces) traffic steering
  - Tunnel policy: Use a tunnel type prioritizing policy to select an SR-TE tunnel.
  - Static route: When configuring a static route, specify the outbound interface of the route as an SR-TE tunnel interface.
  - Auto route: Use an SR-TE tunnel as a logical link in IGP route calculation.
  - Policy-based routing (PBR): Specify an SR-TE tunnel interface as an outbound interface in the involved clause.

HUAWEI

SR-MPLS TE Disadvantages in the Early Stage

- SR-MPLS TE in the early stage inherits the tunnel interface concept of RSVP-TE and uses tunnel interfaces to implement SR.

```
[R1] interface tunnel1
[R1-Tunnel1] ip address unnumbered interface LoopBack0
[R1-Tunnel1] tunnel-protocol mpls te
[R1-Tunnel1] destination 3.3.3.3
[R1-Tunnel1] mpls te tunnel-id 1
[R1-Tunnel1] mpls te signal-protocol segment-routing
...
```

- Using tunnel interfaces to implement SR is simple and easy to understand, but has the following disadvantages:
  - Tunnel interfaces and traffic steering are implemented separately, leading to complex traffic steering and low performance.
  - Tunnels need to be configured and deployed in advance, imposing a restriction in scenarios where the tunnel destination cannot be determined.
  - The application scenarios of tunnel interface-based ECMP are limited.

HUAWEI

SR Policy Overview

- An SR Policy uses a segment list to specify a forwarding path, without the need to use tunnel interfaces.
- SR Policies are classified into SR-MPLS Policies and SRv6 Policies based on segments. This document focuses on SR-MPLS Policies.
- The controller computes paths based on the color attribute that represents SLAs and delivers the computation results to forwarders to form SR-MPLS Policies. (In this example, the forwarder's tunnel information is different from SR-TE tunnel information.) According to the color attribute and next hop of the involved service route, the headend recurses the route to the corresponding SR-MPLS Policy for service forwarding.

```
<PE1>display tunnel-info all
Tunnel ID                Type          Destination          Status
--------------------------------------------------------------------------------
0x0000000001004c4c04     ldp           1.0.0.12             UP
0x000000002900000004     srbe-lsp      1.0.0.12             UP
0x000000000300002001     sr-te         1.0.0.12             UP
0x000000000320000c001    srtepolicy    1.0.0.12             UP
0x000000003400002001     srv6tepolicy  FC01::12             UP
```


HUAWEI

- https://datatracker.ietf.org/doc/draft-ietf-spring-segment-routing-policy/
- An SR Policy is a framework that enables instantiation of an ordered list of segments on a node for implementing a source routing policy with a specific intent for traffic steering from that node.

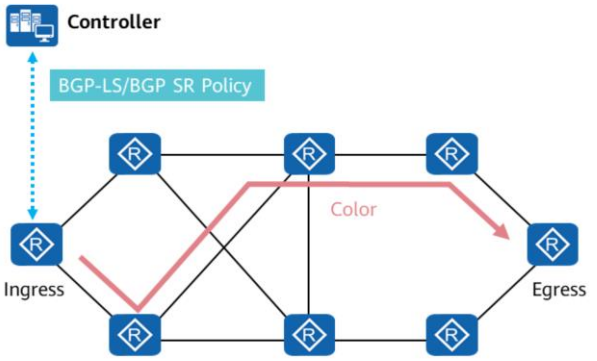
SR-MPLS Policy Tuple

- An SR-MPLS Policy is identified by the tuple <headend, color, endpoint>.
- For an SR-MPLS Policy with a specified node, it is identified only using <color, endpoint>.
  - Headend: node where an SR-MPLS Policy is originated. Generally, it is a globally unique IP address.
  - Color: 32-bit extended community attribute. It is used to identify a service intent (e.g. low latency).
  - Endpoint: destination address of an SR-MPLS Policy. Generally, it is a globally unique IP address.
- Color and endpoint are used to identify a forwarding path on the specific headend of an SR-MPLS Policy.

HUAWEI

SR-MPLS Policy Standards

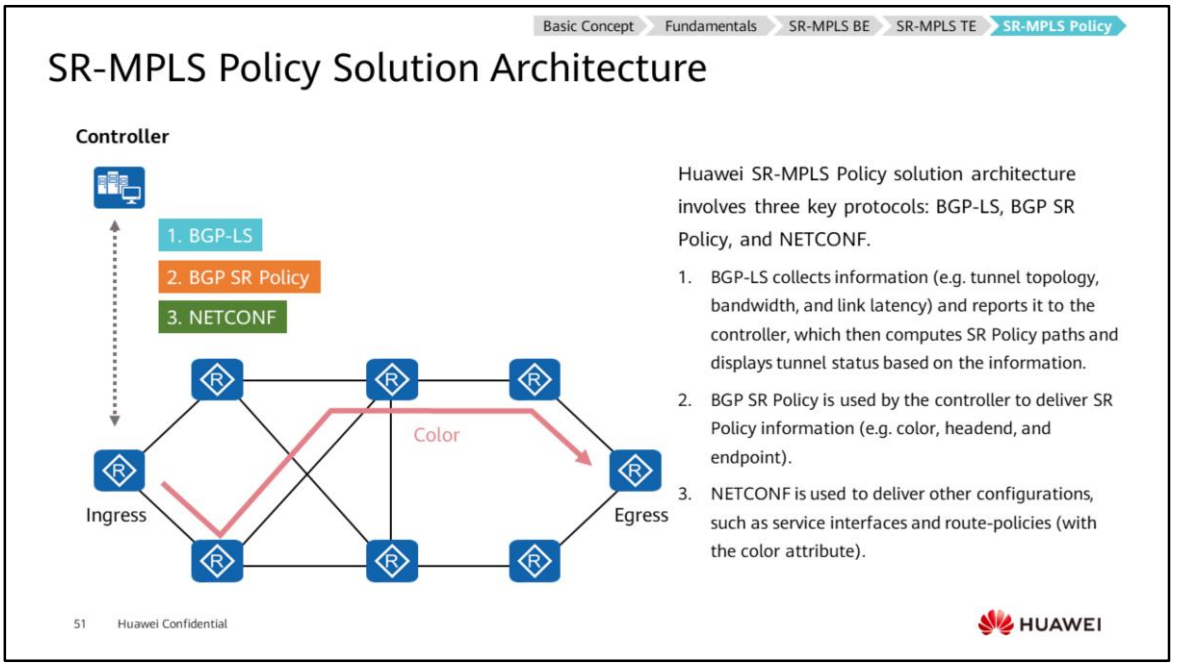
- According to RFC draft-ietf-spring-segment-routing-policy, BGP multi-protocol extension supports the BGP SR Policy (SAFI = 73) address family for delivering SR-MPLS Policies:



- The controller uses BGP to deliver a combination of SR SIDs to the ingress. A TE tunnel carrying the policy color and destined for the egress is then created on the ingress.
- If the tunnel needs to be referenced, you can locate the tunnel based on the policy color.

HUAWEI

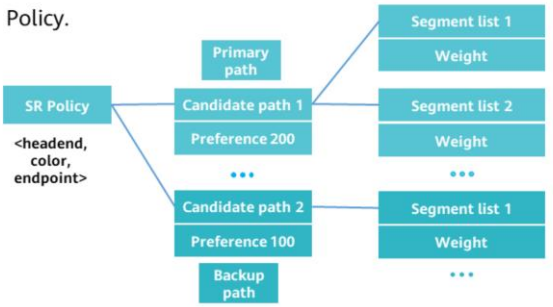
- There are three mainstream methods for SR Policy implementation.
  - BGP: BGP-LS is used to collect topology information, so that no new interface protocol needs to be introduced for customer-developed controllers. BGP SR Policy is used to deliver route information.
  - PCEP: PCEP is a mature southbound protocol used in SR-MPLS TE scenarios. However, the tunnel implementation models of vendors are different and cannot interwork, and the interaction process of PCEP is more complex than that of BGP. As such, BGP extension is recommended.
  - NETCONF/YANG: delivers tunnel paths to forwarders as configurations. This method is not recommended because it delivers configurations in essence and offers the poorest performance. In a comprehensive solution, NETCONF is used to deliver configurations other than tunnel configurations.
- For details about SR Policy, see [I-D.ietf-spring-segment-routing-policy]. (https://datatracker.ietf.org/doc/draft-ietf-idr-segment-routing-te-policy/)



- BGP-LS connection:
  - Collects tunnel topology information for SR Policy path computation.
  - BGP-LS supports the collection of SR Policy status information, based on which the controller displays tunnel status. https://datatracker.ietf.org/doc/draft-ietf-idr-te-lsp-distribution/
  - BGP-LS supports SRLB information encapsulation and decapsulation, so that the controller can obtain the SRLB information for binding SID allocation. (The backup path of each SR Policy corresponds to a binding SID.)
- BGP SR Policy connection:
  - The controller delivers SR Policy information to forwarders to generate SR Policies.
  - BGP routes delivered by the controller carry the color community attribute, and this attribute can be transmitted. The ingress finds a matching BGP route and recurses it to an SR Policy based on the color and endpoint information.
  - In the SR Policy solution, path computation constraints of each application need to be planned in a unified manner on the controller based on SLAs, different colors are used to identify SR Policies. An SR Policy is uniquely identified by <headend, color, endpoint>. The BGP route of services to be steered into an SR Policy needs to carry the corresponding color attribute.
- Huawei SR-MPLS Policy solution also uses PCEP for tunnel status query.

SR-MPLS Policy Model

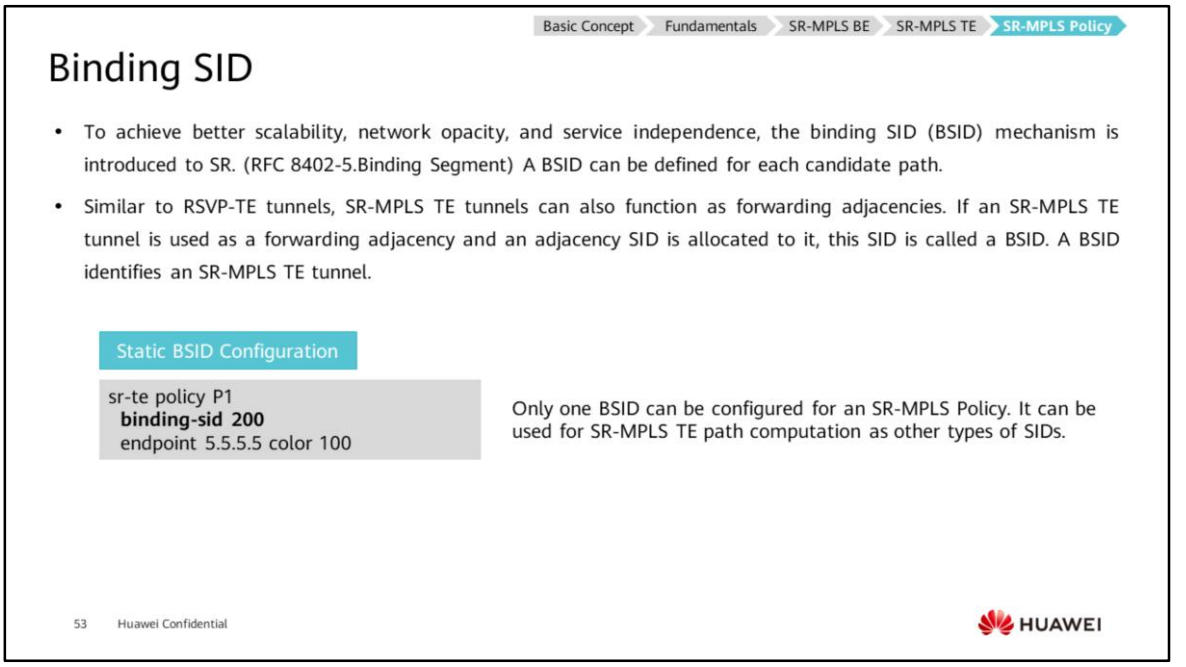
- An SR-MPLS Policy can contain multiple candidate paths with the preference attribute. The valid candidate path with the highest preference functions as the primary path of the SR-MPLS Policy, and the valid candidate path with the second highest preference functions as the backup path.
- A candidate path is an SR-MPLS Policy's segment list sent to the headend through PCEP or BGP SR Policy.



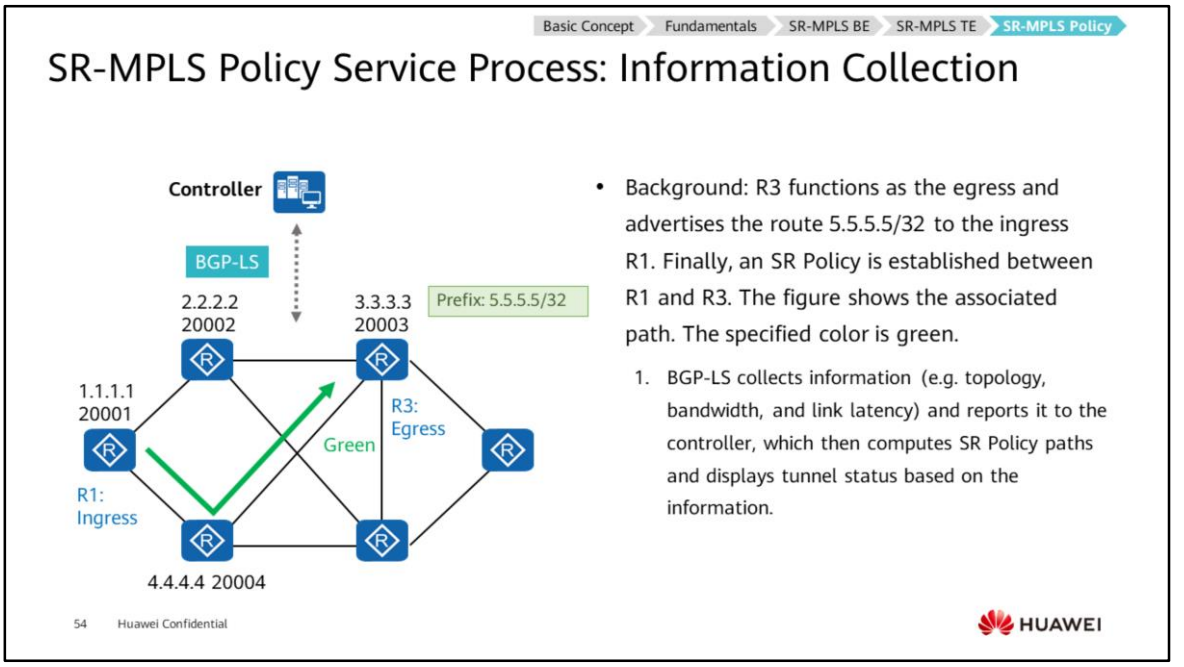
SR policy P1 <headend, color, endpoint>
Candidate-path CP1 <protocol, origin, discriminator>
Preference 200
Weight W1, SID-List1 <SID11...SID1i>
Weight W2, SID-List2 <SID21...SID2j>
Candidate-path CP2 <protocol, origin, discriminator>
Preference 100
Weight W3, SID-List3 <SID31...SID3i>
Weight W4, SID-List4 <SID41...SID4j>

HUAWEI

- An SR Policy can contain multiple candidate paths (e.g. CP1 and CP2). Each of the paths is uniquely determined by the triplet <protocol, origin, discriminator>.
- CP1 is the primary path because it is valid and has the highest preference. The two SID lists of CP1 are delivered to the forwarder, and traffic is balanced between the two paths based on weights. For SID-List <SID11...SID1i>, traffic is balanced according to W1/(W1+W2). In the current mainstream implementation, a candidate path has only one segment list.

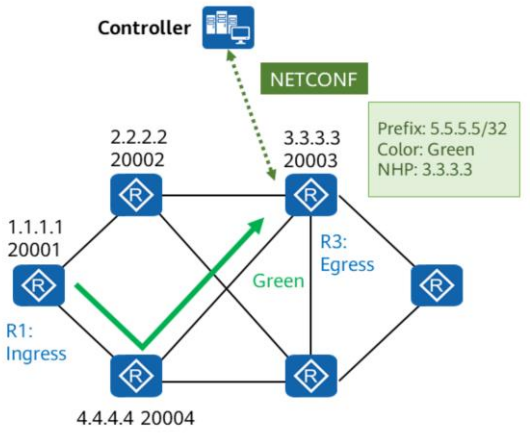


- Source of BSIDs: SRLB or SRGB
- Each candidate path of an SR Policy has a BSID. The BSIDs of different candidate paths of the same SR Policy are generally the same. The BSIDs of different SR Policies must be different. Generally, the BSID range needs to be planned and cannot be shared with other services.
- The headend of an SR Policy forwards packets over the SR Policy based on the BSID. For example, when the headend receives a packet carrying a BSID, it uses the corresponding SR Policy to forward the packet.
- BSIDs are used in label-based traffic steering scenarios, especially label stitching scenarios and tunnel protocol interworking scenarios, such as LDP over SR.
- For details, see draft-ietf-spring-segment-routing-policy-6.Binding SID.



- Preparations:
    1. Controller planning: You can plan the color attribute and the mapping between the color attribute and SR tunnels' SLA requirements (path computation constraints) on the controller based on the SLA requirements of services.
    2. Enable SR on involved devices.
    3. Create a BGP session between the ingress and egress to advertise BGP VPN route information.
    4. Check that the BGP peer relationship is established successfully and a reachable route carrying the color attribute exists between the ingress and egress.

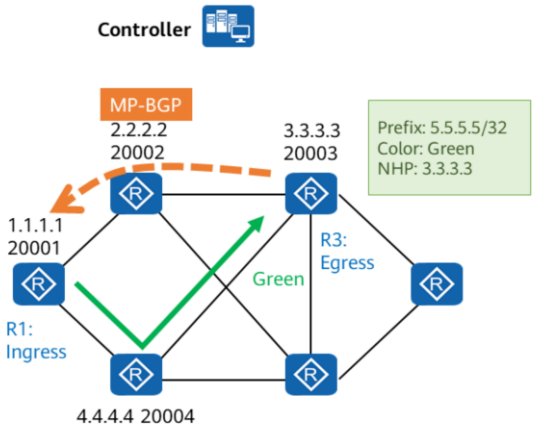
SR Policy Service Process: Route Coloring



- Background: R3 functions as the egress and advertises the route 5.5.5.5/32 to the ingress R1. Finally, an SR Policy is established between R1 and R3. The figure shows the associated path. The specified color is green.
  2. The controller uses NETCONF to deliver a VPN or BGP export route-policy to the egress. The color attribute (green) is set for the route prefix 5.5.5.5/32, and the next hop of the route is R3 address 3.3.3.3.

HUAWEI

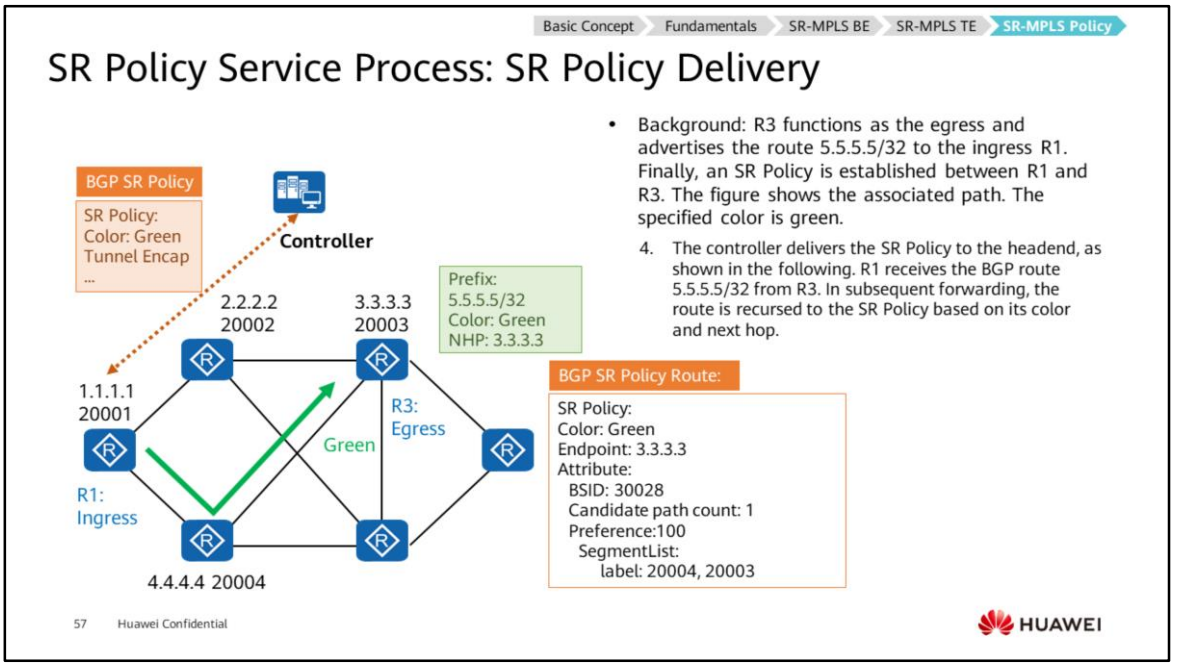
SR Policy Service Process: Route Advertisement



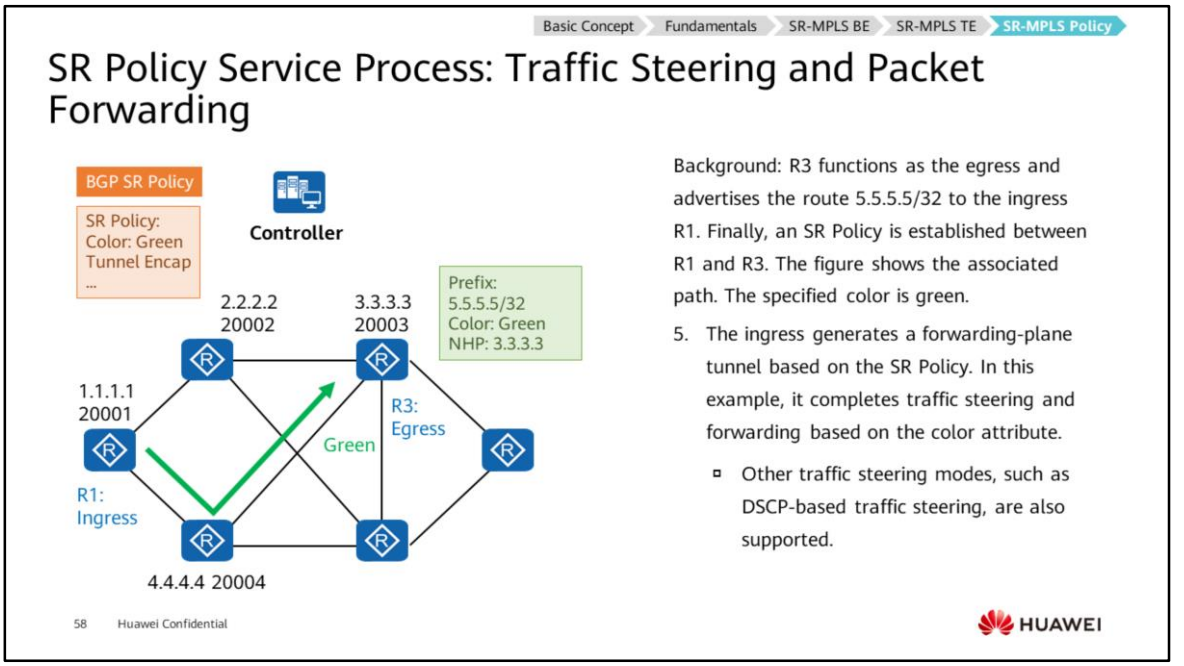
- Background: R3 functions as the egress and advertises the route 5.5.5.5/32 to the ingress R1. Finally, an SR Policy is established between R1 and R3. The figure shows the associated path. The specified color is green.

3. The egress advertises the colored route 5.5.5.5/32 to the ingress through MP-BGP.

HUAWEI



- The steps in this document do not represent the actual configuration sequence. They are only used to help you understand the implementation process. In real-world situations, the controller may deliver SR Policies and use NETCONF to deliver configurations at the same time.



- DSCP-based traffic steering does not support color-based route recursion. Instead, it recurses a route to an SR-MPLS Policy based on the next-hop address in the route. Specifically, it searches for the SR-MPLS Policy group matching specific endpoint information and then finds the corresponding SR-MPLS Policy based on the DSCP value of packets. For details, see the corresponding product documentation.

Summary: SR-MPLS Path Generation Modes

- SR is a technology that allows route selection on the ingress without depending on hop-by-hop signaling exchange (LDP/RSVP-TE). SR-MPLS paths are composed of segments advertised through an IGP. SR-MPLS paths support the following generation modes:
  - Forwarder-based path computation (SPF/CSPF)
  - Static explicit path configuration (CLI/NETCONF)
  - Controller-based path computation (PCEP/BGP SR Policy)
- Currently, BGP SR Policy is the mainstream path delivery mode.

HUAWEI

- Path Computation Element Communication Protocol (PCEP) Extensions for SR: https://datatracker.ietf.org/doc/rfc8664/?include_text=1

Contents

1. Segment Routing Overview
2. Segment Routing Fundamentals
3. **Segment Routing Tunnel Protection and Detection Technologies**
4. Typical Usage Scenarios of Segment Routing
5. Basic Configurations of Segment Routing

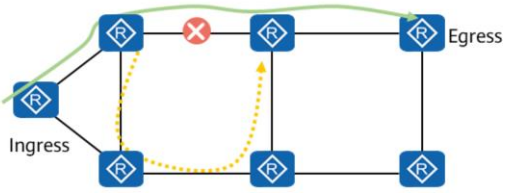
HUAWEI

Overview of SR-MPLS Protection Technologies

- TE tunnel protection is classified into local protection and E2E protection. These protection mechanisms are inherited and also enhanced for SR-MPLS TE.

Local protection

- Fast switching
- Only links and nodes protected

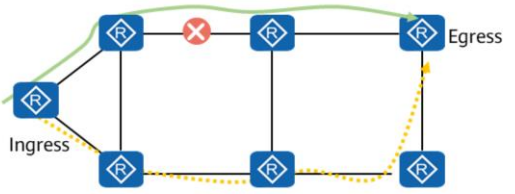


TI-LFA FRR

Anycast FRR

E2E protection

- Detection-dependent fast switching
- E2E paths protected



Hot Standby

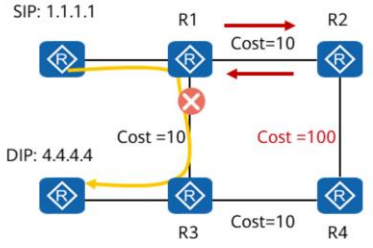
HUAWEI

TI-LFA FRR

- Topology-independent loop-free alternate (TI-LFA) FRR provides link and node protection for SR tunnels. If a link or node fails, traffic is rapidly switched to the backup path.

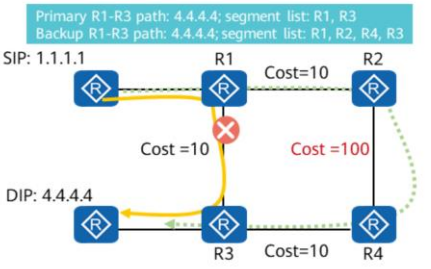
Limitations of the Traditional LFA Algorithm

- The traditional LFA algorithm has topological limitations. As shown in the figure, SIP traffic is forwarded to the DIP through R1. If the R1-R3 link fails, R1 forwards the traffic to R2. However, no backup path can be formed before R2 detects the failure.



TI-LFA Algorithm

- Using the source routing capability of SR, TI-LFA computes a backup path on each node to protect the failure point. When a node detects a failure, traffic is rapidly switched to the backup path.

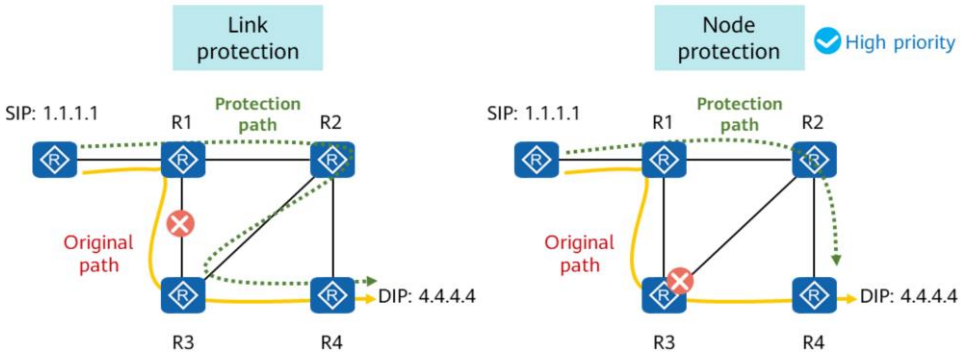


HUAWEI

- In a distributed network architecture, each device independently computes a path, and there is no consensus on the shortest path when a fault occurs. As a result, a backup path cannot be formed using traditional LFA.

TI-LFA FRR Protection Path Computation

- TI-LFA FRR protects services against both link and node failures. TI-LFA preferentially computes a node protection path because this path can definitely protect services against a link failure.



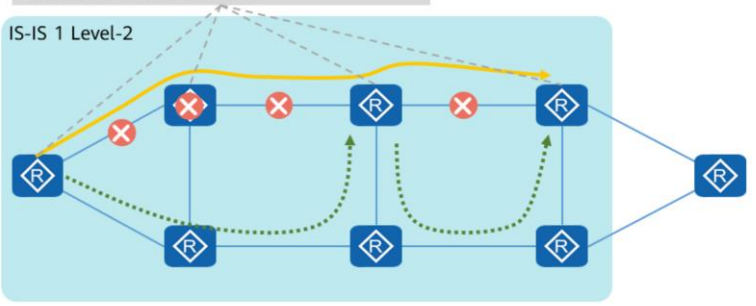
HUAWEI

- For details about TI-LFA FRR, see the "TI-LFA FRR" section in NE series product documentation.

TI-LFA FRR Usage Scenarios and Configuration

- To protect the entire path, you need to enable TI-LFA FRR local protection for the IGP processes of multiple nodes.

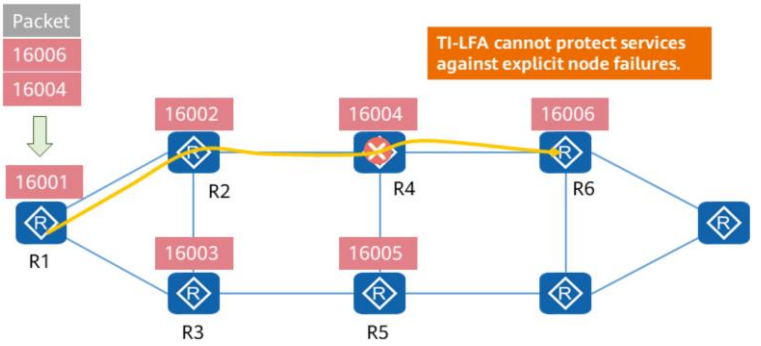
```
[Router] isis 1
[Router-isis-1] frr
[Router-isis-1-frr] loop-free-alternate level-2
[Router-isis-1-frr] ti-lfa level-2
```



HUAWEI

Limitations of TI-LFA FRR

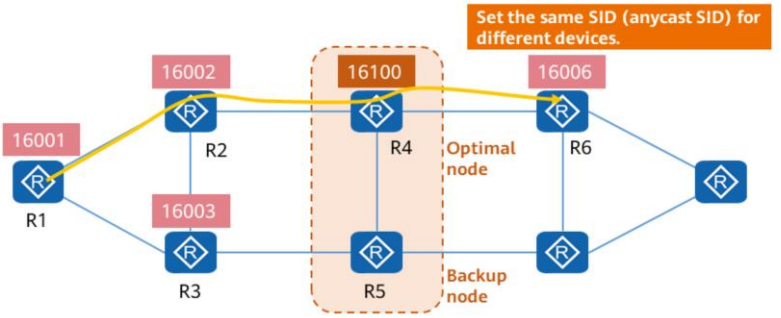
- TI-LFA cannot provide protection if a specified explicit node (ingress, egress, or constraint node) along an SR tunnel fails. For example, on the SR path shown in the following figure, TI-LFA cannot generate protection paths for explicit nodes R1, R4, and R6.



HUAWEI

Anycast FRR

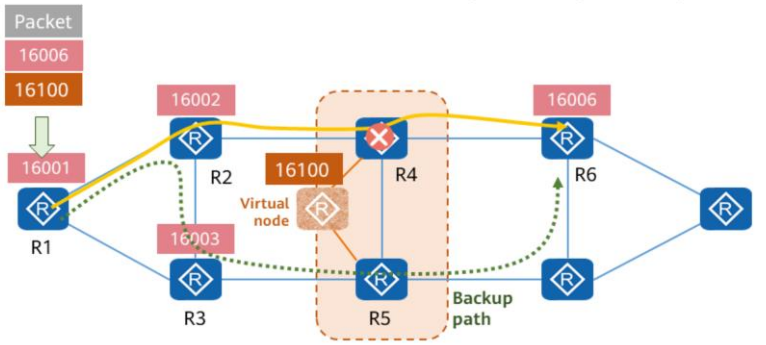
- Anycast FRR can protect services against failures of specified nodes.
- Assume that R4 and R5 advertise the same SID. This SID is called an anycast SID. The anycast SID is advertised in the IGP, with the next hop pointing to the nearest node on the path, such as R4. In this case, R4 is the optimal node of the anycast SID, and R5 is the backup node.



HUAWEI

Anycast FRR Protection

- Anycast FRR constructs a virtual node for SID advertisement and uses the TI-LFA algorithm to compute the backup next hop of the virtual node.
- If R4 fails, TI-LFA continues to forward traffic through R5 along the computed backup path.

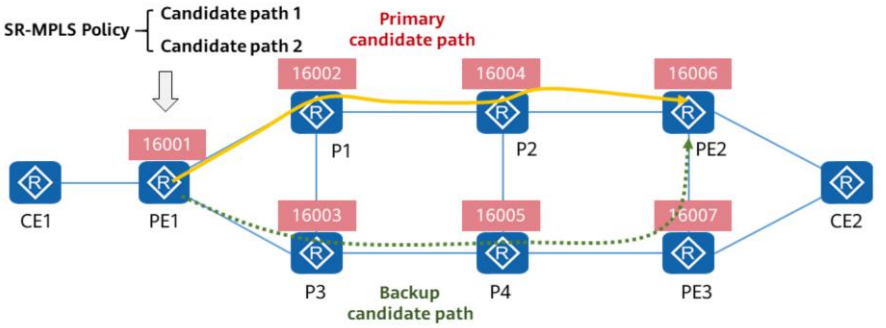


HUAWEI

- The cost values of the links from R4 and R5 to the virtual node are both 0. However, the cost values of the links from the virtual node to R4 and R5 are infinite.

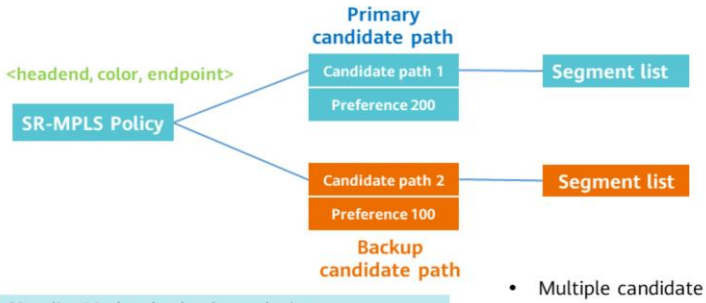
Hot Standby

- SR hot standby enables the controller to compute a backup path that is different from the primary path to implement E2E path protection.
- For SR-MPLS Policies, the primary and backup candidate paths implement hot standby protection. The primary and backup candidate paths belong to the same SR-MPLS Policy.



HUAWEI

Hot Standby Implementation for SR-MPLS Policy



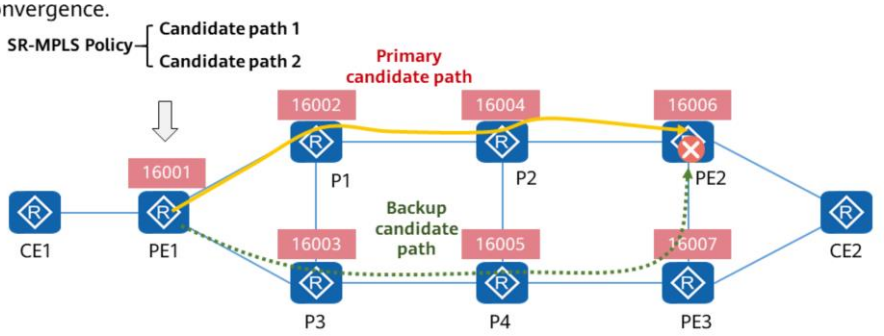

```
SR policy P1 <headend, color, endpoint>
  Candidate-path CP1 <protocol, origin, discriminator>
    Preference 200
      SID-List <SID11...SID1i>
  Candidate-path CP2 <protocol, origin, discriminator>
    Preference 100
      SID-List <SID21...SID2i>
```

- Multiple candidate paths of an SR-MPLS Policy implement hot standby protection. If a segment list fails, a failover is triggered.
- SR-MPLS Policy fault detection depends on detection mechanisms such as BFD or SBFD.

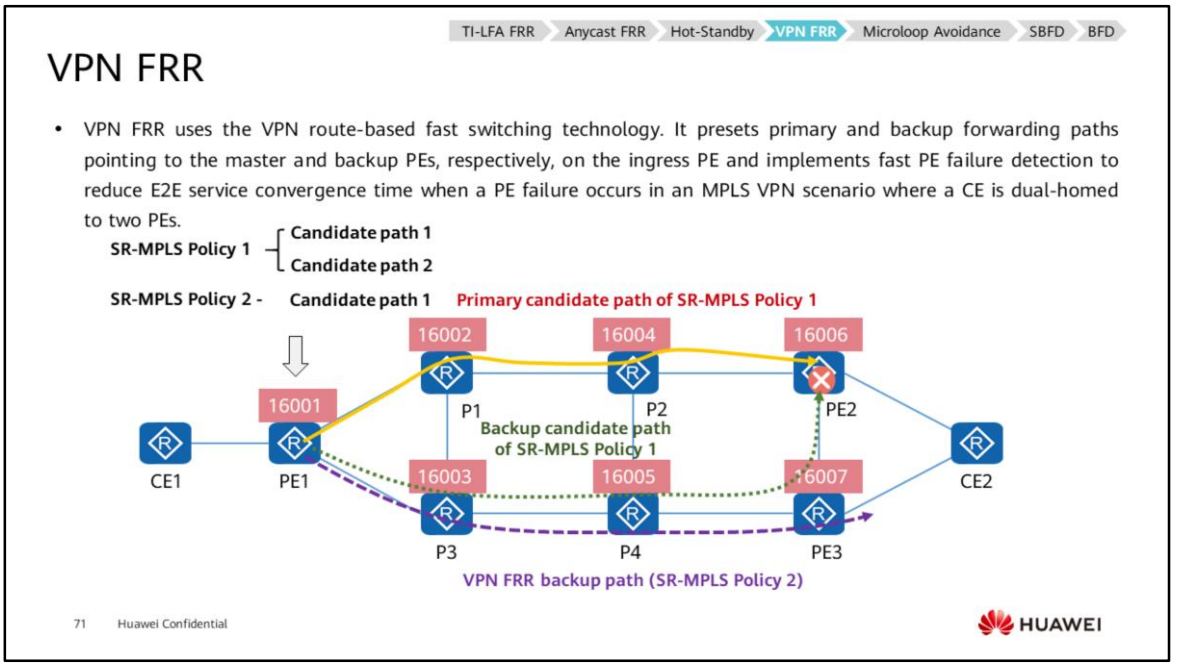
HUAWEI

Limitations of Hot Standby

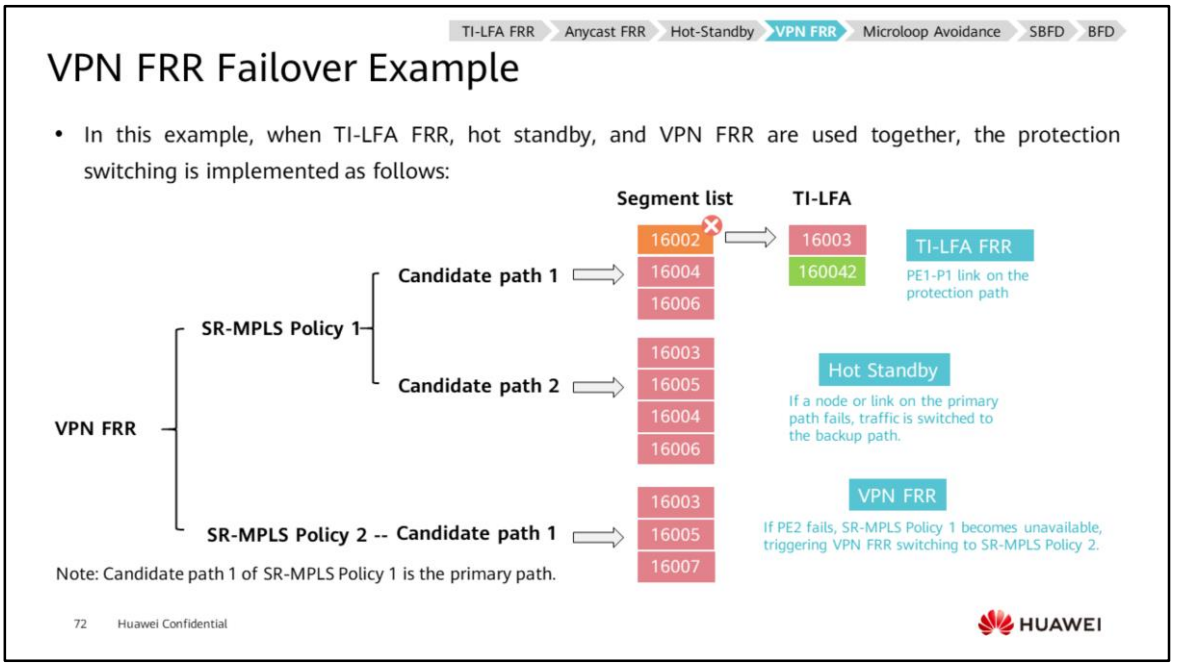
- Hot standby can protect E2E paths but does not apply to scenarios where the egress PE of a tunnel fails. In this example, PE1 receives the routes advertised by PE2 and PE3 at the same time and preferentially selects the route advertised by PE2. If PE2 fails, services can recover only through route convergence.



HUAWEI



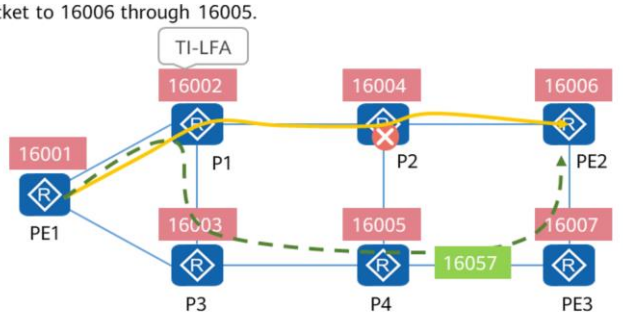
- In traditional TE tunnel protection technologies, if a PE fails, services can recover only through E2E route convergence and LSP convergence. The service convergence time is closely related to the number of internal MPLS VPN routes and the number of hops on the bearer network. The greater the number of VPN routes, the longer the service convergence time.
- In VPN FRR, service convergence time depends on only the time required to detect remote PE failures and change tunnel status, making service convergence time irrelevant to the number of VPN routes on the bearer network.
- In this example, VPN FRR primary and backup paths exist from PE1 to PE3. They are not all displayed in the figure.



- Fault detection in hot standby and VPN FRR scenarios depends on detection mechanisms such as BFD or SBFD.

SR Microloop Avoidance Overview

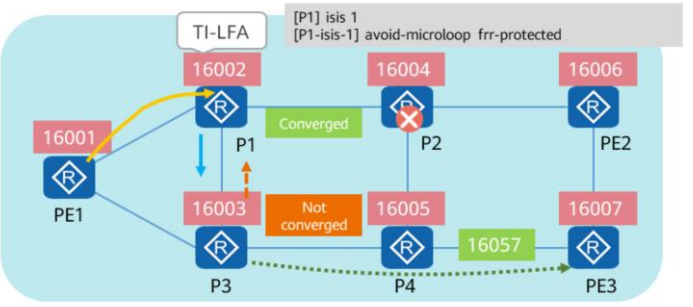
- Each node independently calculates the IGP LSDB, which may lead to a loop during unordered convergence. This may in turn result in microloops, a kind of transient loop that disappears after all the nodes on the forwarding path have converged.
- On the network shown in the following figure, TI-LFA FRR is working properly. After detecting that P2 fails, P1 enters the TI-LFA FRR switching process. Specifically, it inserts the repair list <16005, 16057> into the packet to forward the packet to 16006 through 16005.



HUAWEI

SR Local Microloop Avoidance in a Traffic Switchover Scenario

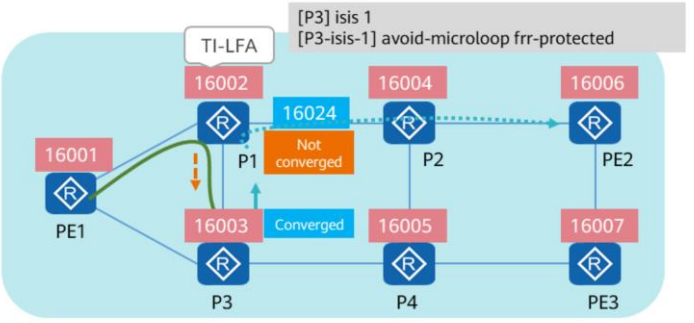
- Devices converge at different time points, leading to a microloop. For example, the route of P1 does not carry a repair list after convergence. In this case, the next hop of the route to 16006 is P3. If P3 has not converged, the next hop pointing to 16006 is still P1, causing a local microloop in a traffic switchover scenario.
- After microloop avoidance is enabled, P1 starts the T1 timer during which the packet is still forwarded according to the TI-LFA policy <16005, 16057> and waits for other nodes to converge.



HUAWEI

SR Local Microloop Avoidance in a Traffic Switchback Scenario

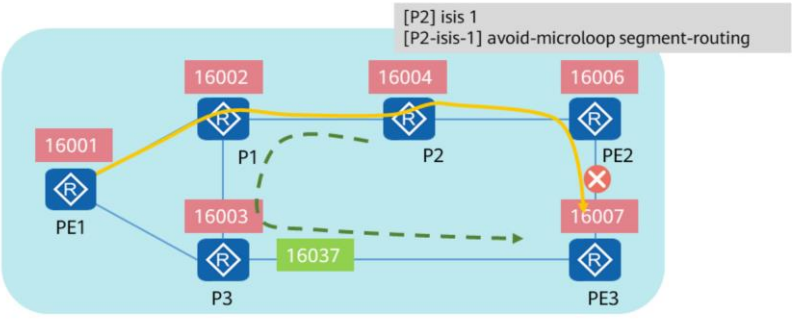
- A microloop may also occur during traffic switchback implemented after fault rectification. Assume that P2 recovers. If P1 has not converged and forwards traffic to P3 that has converged, traffic will be forwarded back to P1, resulting in a local microloop.
- With microloop avoidance enabled, after P3 converges, it computes the microloop avoidance segment list <16002, 16024>. PE1 forwards the packet to P1. As P1 has not converged, it forwards the packet to P3. P3 inserts the segment list into the packet and forwards the packet to P2 through P1 and finally to PE2.



HUAWEI

SR Remote Microloop Avoidance

- Traffic switching may cause not only a local microloop but also a microloop between remote nodes (that is, a remote microloop).
- As shown in the figure, the link between PE2 and PE3 fails. If P2 has converged but P1 has not, a loop occurs between P1 and P2.
- With remote microloop avoidance enabled, after P2 converges, it computes the microloop avoidance segment list <16003,16037> for traffic accessing PE3. In this case, P1 still forwards traffic from P3 to PE3 even if P1 has not converged.



HUAWEI

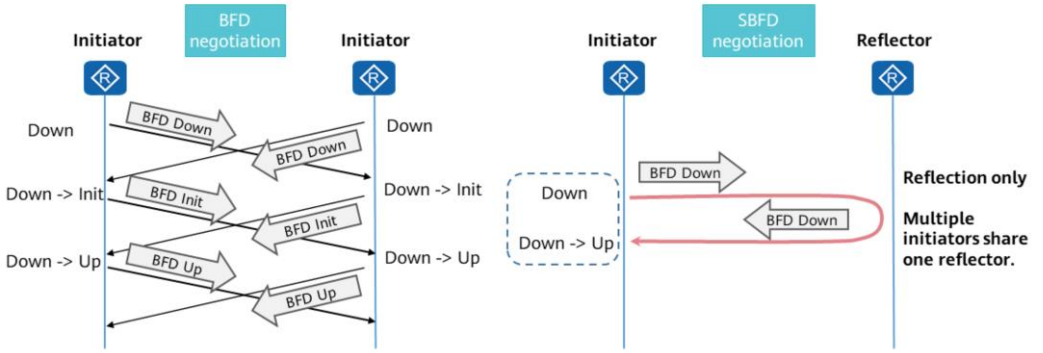
Summary: Comparison Between TI-LFA and Microloop Avoidance

TI-LFA	Microloop Avoidance
• Purpose: to locally compute a backup path for the destination address. • Trigger condition: link or node failure on the primary path.	• Purpose: to prevent temporary loops during the update of the primary path. • Trigger condition: primary path update.

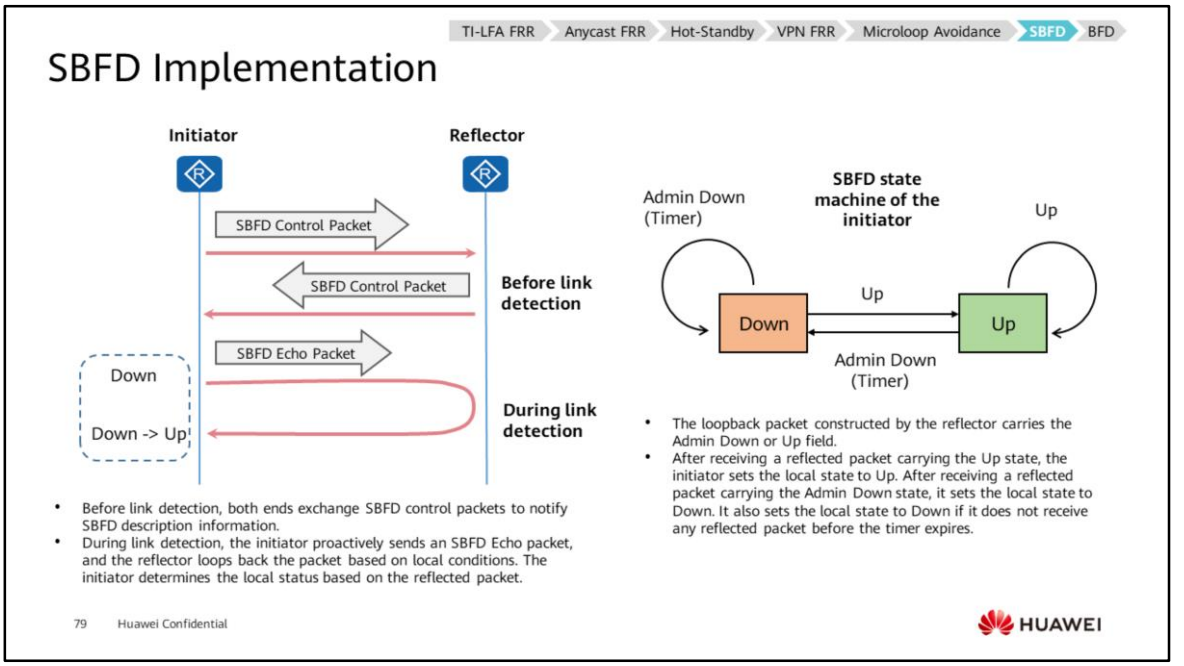
HUAWEI

SBFD Overview

- If BFD detects a large number of links, the negotiation time of the state machine is prolonged, which is not suitable for SR. To address this issue, seamless bidirectional forwarding detection (SBFD), which is a simplified BFD mechanism, is introduced to detect SR tunnels. With a simplified BFD state machine, SBFD shortens the negotiation time and improves network-wide flexibility.



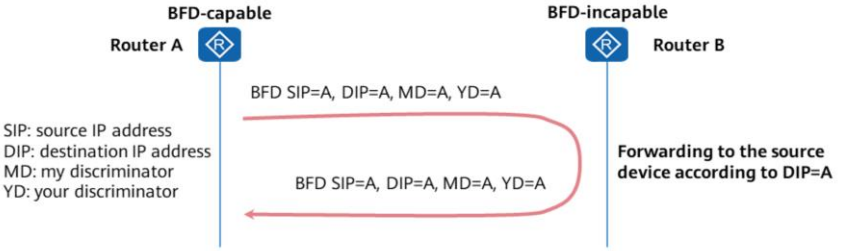
HUAWEI



- Because the state machine has only Up and Down states, the initiator can send packets carrying only the Up or Down state and receive packets carrying only the Up or Admin Down state. The initiator starts by sending an SBFD packet carrying the Down state to the reflector. The destination and source port numbers of the packet are 7784 and 4784, respectively; the destination IP address is a user-configured address on the 127 network segment; the source IP address is the locally configured LSR ID.
- The reflector runs no SBFD state machine or detection mechanism. For this reason, it does not proactively send SBFD Echo packets. Instead, it only reflects back received SBFD packets. The destination and source port numbers in the looped-back SBFD packet are 4784 and 7784, respectively; the source IP address is the locally configured LSR ID; the destination IP address is the source IP address of the initiator.

One-Arm BFD

- BFD/SBFD requires that devices at both ends support this function. If a Huawei device needs to communicate with a BFD-incapable device, you can configure one-arm BFD (also called one-arm BFD echo) for the Huawei device. A one-arm BFD Echo session can be established on the BFD-capable device. After receiving a BFD Echo packet, the BFD-incapable device immediately loops back the packet for quick link detection.
- One-arm BFD Echo does not require Echo negotiation capabilities at both ends; that is, BFD can be configured on only one end. The device with one-arm Echo enabled sends special BFD packets (source and destination IP addresses in the IP header are the IP address of the local device, and the local and remote discriminators in the BFD packet are the same). After receiving the packets, the peer device directly loops them back to the local device to check whether the link is normal. This function equips Huawei devices with a stronger adaptability to low-end devices.



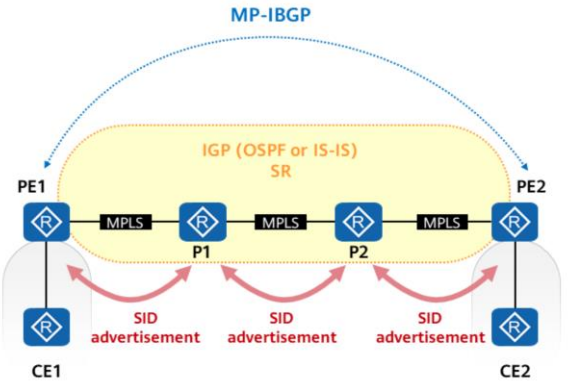
HUAWEI

Contents

1. Segment Routing Overview
2. Segment Routing Fundamentals
3. Segment Routing Tunnel Protection and Detection Technologies
4. **Typical Usage Scenarios of Segment Routing**
5. Basic Configurations of Segment Routing

HUAWEI

Intra-AS SR-MPLS BE

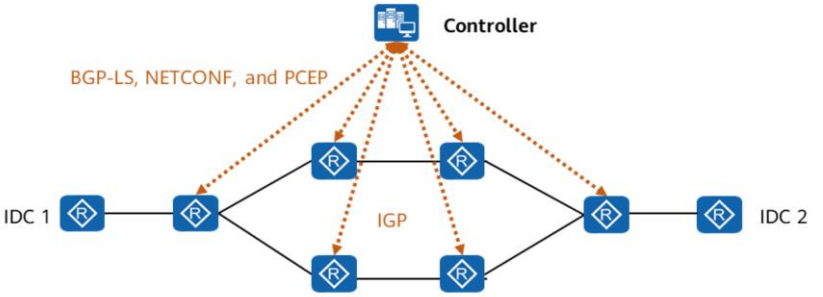


- SR-MPLS BE applies to services that do not have strict SLA requirements or require path planning.
- Downstream routers allocate SIDs to upstream routers to form SR-MPLS forwarding paths.
- MP-BGP is used on the control plane to advertise VPN labels.
- SR-MPLS BE can be used as a backup solution for SR-MPLS TE services on a production network.

HUAWEI

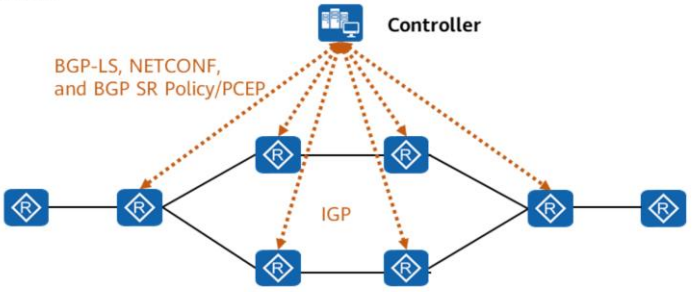
Intra-AS SR-MPLS TE

- SR-MPLS TE applies to scenarios that have strict SLA requirements and require path planning, such as DCI scenarios.
- SR labels are advertised by an IGP. The controller uses BGP-LS to collect information (e.g. network topology, bandwidth, latency, and label information).
- The controller computes qualified forwarding paths based on constraints and delivers path computation results to forwarders through PCEP or NETCONF. Engineers can also manually configure strict forwarding paths and delegate the paths to the controller through PCEP.



Intra-AS SR-MPLS Policy

- SR-MPLS Policy applies to scenarios that have strict SLA requirements and require path planning.
- SR labels are advertised by an IGP. The controller uses BGP-LS to collect information (e.g. network topology, bandwidth, latency, and label information).
- The controller computes qualified forwarding paths based on constraints and delivers path computation results to forwarders through BGP SR Policy or PCEP. Engineers can also manually configure strict forwarding paths and delegate the paths to the controller through PCEP.

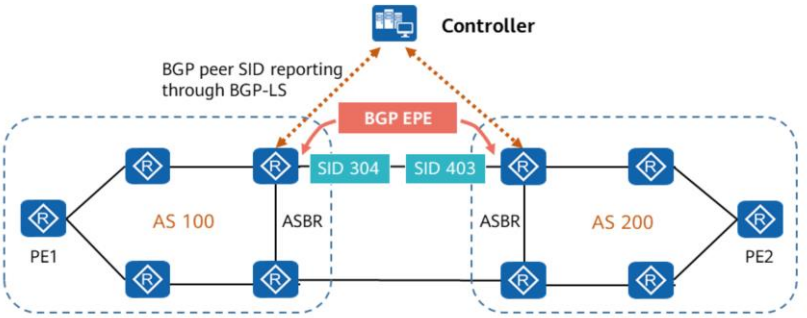


HUAWEI

- PCEP was first proposed in the optical transport field. It is seldom deployed on enterprises' production networks due to its few applications on IP networks, difficult interoperability between vendors, and poor performance. Therefore, BGP SR-Policy is recommended on an SR-MPLS network.

Inter-AS E2E SR-MPLS TE (1)

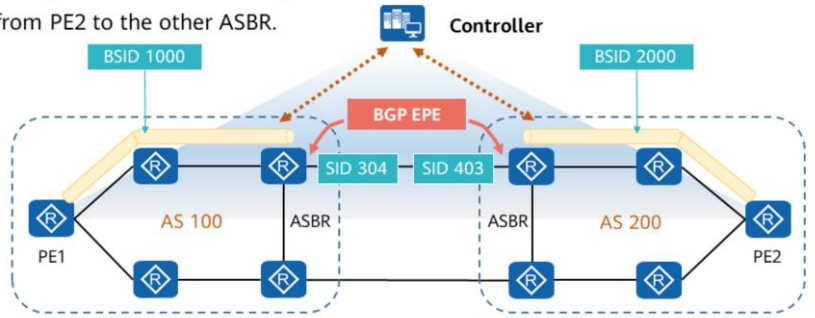
- In inter-AS access scenarios, it is recommended that the controller perform centralized computation and deliver E2E SR-MPLS TE paths.
- BGP egress peer engineering (EPE) is configured on ASBRs for them to allocate a BGP peer SID to each other.
- The ASBRs then use BGP-LS to report the BGP EPE-generated labels and network topology information.



HUAWEI

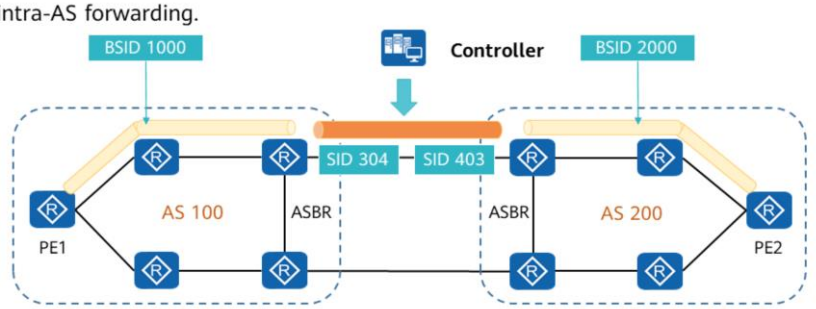
Inter-AS E2E SR-MPLS TE (2)

- Before an E2E SR-MPLS TE tunnel is created, the controller needs to create intra-AS SR-MPLS TE tunnels.
- To reduce the label stack depth, you can configure a BSID for each intra-AS tunnel.
- In this example, BSID 1000 is configured for the tunnel from PE1 to one ASBR, and BSID 2000 for the tunnel from PE2 to the other ASBR.



Inter-AS E2E SR-MPLS TE (3)

- The controller performs global computation, integrates path labels into a label stack, and then delivers it to forwarders.
- In this example, the label stack for the path from PE1 to PE2 is <1000, 304, 2000>.
- In the label stack, 1000 and 2000 are BSIDs, which will be replaced with corresponding SR label stacks during intra-AS forwarding.



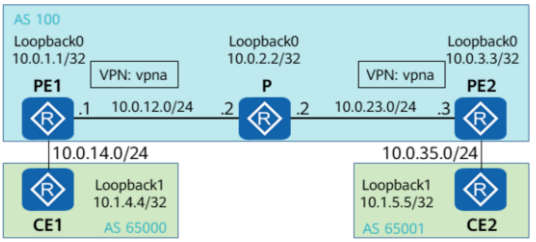
HUAWEI

Contents

1. Segment Routing Overview
2. Segment Routing Fundamentals
3. Segment Routing Tunnel Protection and Detection Technologies
4. Typical Usage Scenarios of Segment Routing
5. **Basic Configurations of Segment Routing**
   - SR-MPLS BE
   - SR-MPLS TE
   - SR-MPLS Policy

HUAWEI

L3VPN over SR-MPLS BE (1)



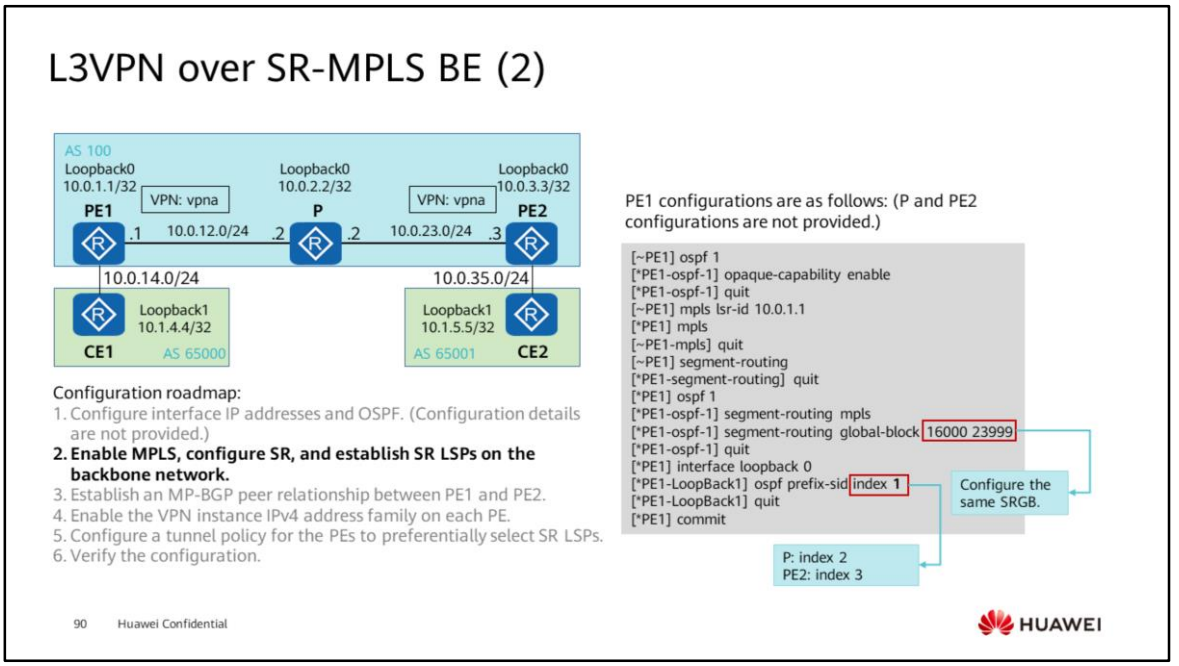
Networking requirements:

1. Connect PE1 and PE2 to different CEs that belong to VPN instance vpna.
2. Deploy L3VPN service recursion to SR-MPLS BE tunnel on the backbone network so that CE1 and CE2 can communicate through Loopback1.

Configuration roadmap:

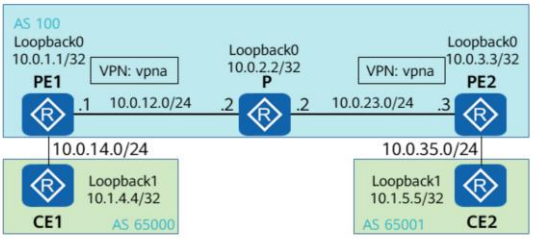
1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Configure a tunnel policy for the PEs to preferentially select SR LSPs.
6. Verify the configuration.

HUAWEI



- Before configuring an SR-MPLS BE tunnel, you need to enable MPLS on each device in the SR-MPLS domain. The configuration procedure is as follows:
  - Run the **system-view** command to enter the system view.
  - Run the **mpls lsr-id** *lsr-id* command to configure an LSR ID for the local device.
    - Note the following during LSR ID configuration:
      - Configuring LSR IDs is the prerequisite for all MPLS configurations.
      - LSRs do not have default LSR IDs, and such IDs must be manually configured.
      - Using the address of a loopback interface as the LSR ID is recommended for an LSR.
  - Run the **mpls** command to enable MPLS.
- Basic SR-MPLS BE function configurations mainly involve enabling SR globally, specifying an SRGB, and configuring an SR prefix SID.
  - Enable SR globally.
    - Run the **system-view** command to enter the system view.
    - Run the **segment-routing** command to enter the Segment Routing view.
    - Run the **commit** command to commit the configuration.
    - Run the **quit** command to return to the system view.

L3VPN over SR-MPLS BE (3)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR LSPs on the backbone network.
3. **Establish an MP-BGP peer relationship between PE1 and PE2.**
4. **Enable the VPN instance IPv4 address family on each PE.**
5. Configure a tunnel policy for the PEs to preferentially select SR LSPs.
6. Verify the configuration.

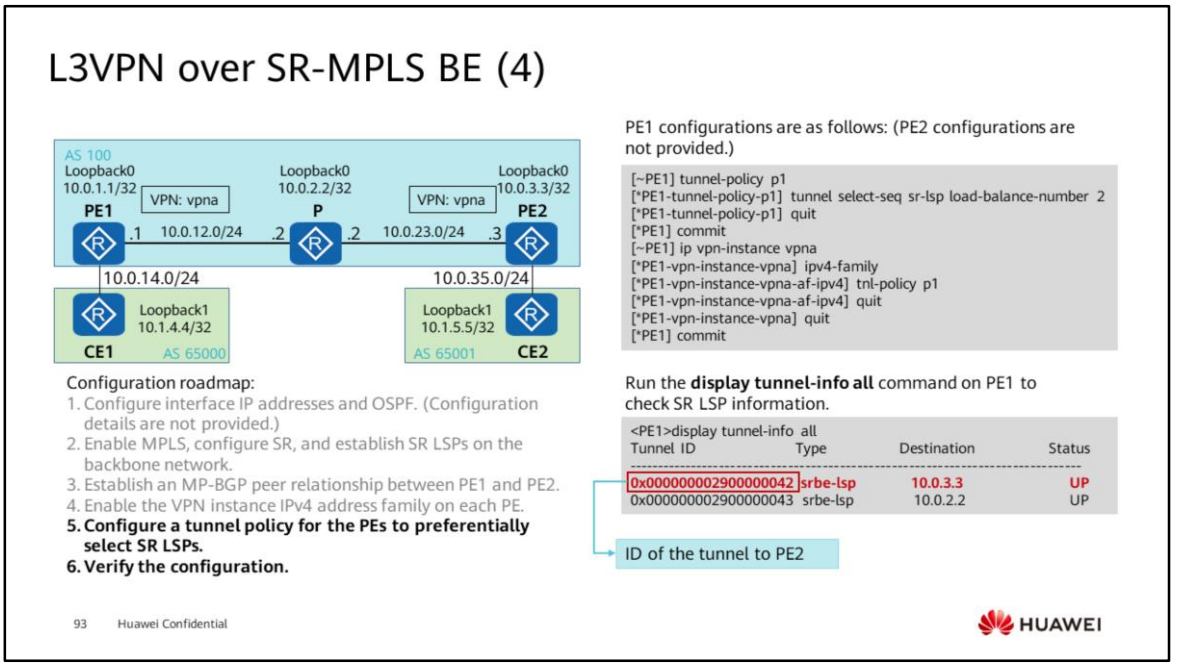
PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~PE1] bgp 100
[~PE1-bgp] peer 10.0.3.3 as-number 100
[*PE1-bgp] peer 10.0.3.3 connect-interface loopback 0
[*PE1-bgp] ipv4-family vpnv4
[*PE1-bgp-af-vpnv4] peer 10.0.3.3 enable
[*PE1-bgp-af-vpnv4] commit
[~PE1-bgp-af-vpnv4] quit
[~PE1-bgp] quit
```

PE1 configurations are as follows: (PE2 configurations are not provided.)

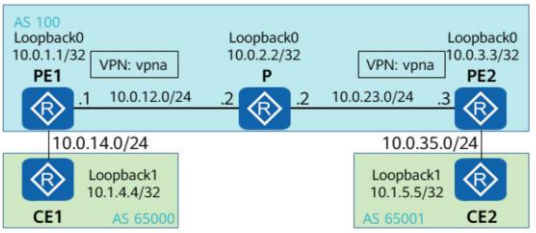
```
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] route-distinguisher 100:1
[*PE1-vpn-instance-vpna-af-ipv4] vpn-target 111:1 both
[*PE1-vpn-instance-vpna-af-ipv4] quit
[*PE1-vpn-instance-vpna] quit
[*PE1]bgp 100
[*PE1-bgp]ipv4-family vpn-instance vpna
[*PE1-bgp-vpna]peer 10.0.14.4 as-number 65000
```


HUAWEI



- Configure a tunnel policy and tunnel selection sequence.
  - Run the **system-view** command to enter the system view.
  - Run the **tunnel-policy** *policy-name* command to create a tunnel policy and enter the tunnel policy view.
  - Run the **tunnel select-seq** *sr-lsp* **load-balance-number** *load-balance-number* [**unmix**] command to configure a tunnel selection sequence and the number of tunnels for load balancing.
  - Run the **commit** command to commit the configuration.
  - Run the **quit** command to return to the system view.
- Configure BGP L3VPN service recursion to SR-MPLS BE tunnels.
  - Run the **ip vpn-instance** *vpn-instance-name* command to enter the VPN instance view.
  - Run the **ipv4-family** command to enter the VPN instance IPv4 address family view.
  - Run the **tnl-policy** *policy-name* command to apply a tunnel policy to the VPN instance IPv4 address family.
  - Run the **commit** command to commit the configuration.

L3VPN over SR-MPLS BE (5)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. **Configure a tunnel policy for the PEs to preferentially select SR LSPs.**
6. **Verify the configuration.**

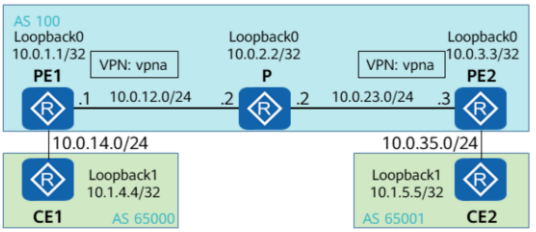
Check VPNv4 routing information on PE1.

```
<PE1>display bgp vpnv4 all routing-table 10.1.5.5

 BGP local router ID : 10.0.1.1
 Local AS number : 100

 Total routes of Route Distinguisher(100:1): 1
 BGP routing table entry information of 10.1.5.5/32:
 Label information (Received/Applied): 48122/NULL
 From: 10.0.3.3 (10.0.3.3)
 Route Duration: 0d00h39m18s
 Relay IP Nexthop: 10.0.12.2
 Relay IP Out-Interface: GigabitEthernet0/3/1
 Relay Tunnel Out-Interface: GigabitEthernet0/3/1
 Original nexthop: 10.0.3.3
 Qos information : 0x0
 Ext-Community: RT <111 : 1>
 AS-path 65001, origin incomplete, MED 0, localpref 100, pref-val 0,
valid, internal, best, select, pre 255, IGP cost 2
 Not advertised to any peer yet
```

Label allocated by PE2 to 10.1.5.5/32

HUAWEI

L3VPN over SR-MPLS BE (6)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. **Configure a tunnel policy for the PEs to preferentially select SR LSPs.**
6. **Verify the configuration.**

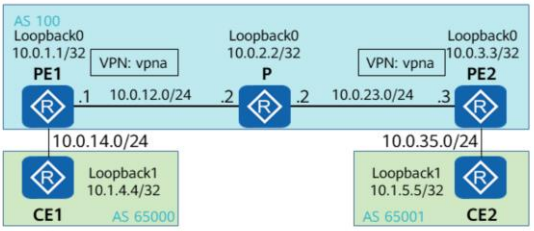
Check vpna's routing information on PE1.

```
<PE1>display ip routing-table vpn-instance vpna 10.1.5.5 verbose
Route Flags: R - relay, D - download to fib, T - to vpn-instance
------------------------------------------------------------------------------
Routing Table : vpna
Summary Count : 1

Destination: 10.1.5.5/32
    Protocol: IBGP              Process ID: 0
  Preference: 255                     Cost: 0
     NextHop: 10.0.3.3          Neighbour: 10.0.3.3
       State: Active Adv Relied       Age: 00h35m03s
         Tag: 0                Priority: low
       Label: 48122            QoSInfo: 0x0
  IndirectID: 0x100013A              Instance:
RelayNextHop: 10.0.12.2          Interface: GigabitEthernet0/3/1
    TunnelID: 0x00000002900000042  Flags: RD
```

The VPNv4 label and SR LSP are combined to guide packet forwarding.

HUAWEI

L3VPN over SR-MPLS BE (7)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. **Configure a tunnel policy for the PEs to preferentially select SR LSPs.**
6. **Verify the configuration.**

Tracert the SR LSP on PE1.

```
<PE1>tracert lsp  segment-routing  ip 10.0.3.3 32
  LSP Trace Route FEC: SEGMENT ROUTING IPV4 PREFIX 10.0.3.3/32 ,
press CTRL_C to break.
  TTL    Replier          Time    Type       Downstream
  0                               Ingress    10.0.12.2/[16003 ]
  1      10.0.12.2        8 ms    Transit    10.0.23.3/[3 ]
  2      10.0.3.3         9 ms    Egress
```

Question: How are the labels computed?

Verify the configuration on CE1.

```
<CE1>ping -a 10.1.4.4 10.1.5.5
  PING 10.1.5.5: 56  data bytes, press CTRL_C to break
    Reply from 10.1.5.5: bytes=56 Sequence=1 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=2 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=3 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=4 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=5 ttl=254 time=1 ms
```

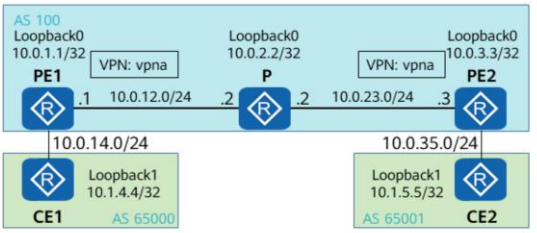

HUAWEI

Contents

1. Segment Routing Overview
2. Segment Routing Fundamentals
3. Segment Routing Tunnel Protection and Detection Technologies
4. Typical Usage Scenarios of Segment Routing
5. **Basic Configurations of Segment Routing**
   - SR-MPLS BE
   - SR-MPLS TE
   - SR-MPLS Policy

HUAWEI

L3VPN over SR-MPLS TE (1)



Networking requirements:

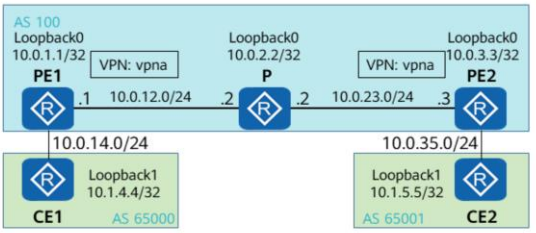
1. Connect PE1 and PE2 to different CEs that belong to VPN instance vpna.
2. Deploy L3VPN service recursion to SR-MPLS TE tunnel on the backbone network so that CE1 and CE2 can communicate through Loopback1.

Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR-MPLS TE LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs.
6. Configure a tunnel policy for the PEs to preferentially select SR-MPLS TE LSPs.
7. Verify the configuration.

HUAWEI

L3VPN over SR-MPLS TE (2)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. **Enable MPLS, configure SR, and establish SR-MPLS TE LSPs on the backbone network.**
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs.
6. Configure a tunnel policy for the PEs to preferentially select SR-MPLS TE LSPs.
7. Verify the configuration.

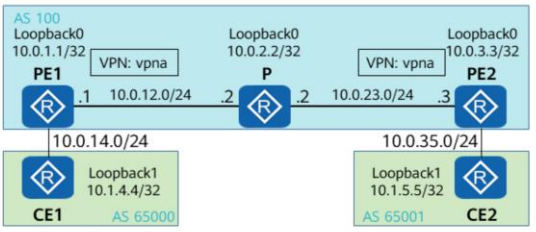
Configure basic SR-MPLS TE functions. PE1 configurations are as follows: (P and PE2 configurations are not provided.)

```
[~PE1] mpls lsr-id 10.0.1.1
[*PE1] mpls
[*PE1-mpls] mpls te
[*PE1-mpls] quit
[~PE1] segment-routing
[*PE1-segment-routing] quit
[~PE1] ospf 1
[*PE1-ospf-1] opaque-capability enable
[*PE1-ospf-1] segment-routing mpls
[*PE1-ospf-1] segment-routing global-block 16000 23999
[*PE1-ospf-1] area 0
[*PE1-ospf-1-area-0.0.0.0] mpls-te enable
[*PE1-ospf-1-area-0.0.0.0] quit
[*PE1] interface loopback 0
[*PE1-LoopBack1] ospf prefix-sid index 1
[*PE1-LoopBack1] quit
```


HUAWEI

- Before configuring an SR-MPLS TE tunnel, you need to enable MPLS TE on each device in the SR-MPLS domain.
  - Run the **system-view** command to enter the system view.
    - Run the **mpls lsr-id** *lsr-id* command to configure an LSR ID for the local device.
    - Run the **mpls** command to enter the MPLS view.
    - Run the **mpls te** command to enable MPLS TE globally on the local device.
  - (Optional) Enable interface-specific MPLS TE. In a scenario where the controller or ingress performs path computation, interface-specific MPLS TE must be enabled. In a static explicit path scenario, this step can be ignored.
    - Run the **quit** command to return to the system view.
    - Run the **interface** *interface-type interface-number* command to enter the view of an interface on an MPLS TE link.
    - Run the **mpls** command to enable MPLS on the interface.
    - Run the **mpls te** command to enable MPLS TE on the interface.
  - Run the commit command to commit the configuration.

L3VPN over SR-MPLS TE (3)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. **Enable MPLS, configure SR, and establish SR-MPLS TE LSPs on the backbone network.**
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs.
6. Configure a tunnel policy for the PEs to preferentially select SR-MPLS TE LSPs.
7. Verify the configuration.

Configure an SR-MPLS TE explicit path. PE1 configurations are as follows: (P and PE2 configurations are not provided.)

```
[~PE1]explicit-path te1
[*PE1-explicit-path-te1]next sid label 16002 type prefix
[*PE1-explicit-path-te1]next sid label 16003 type prefix
[*PE1-explicit-path-te1]commit
```

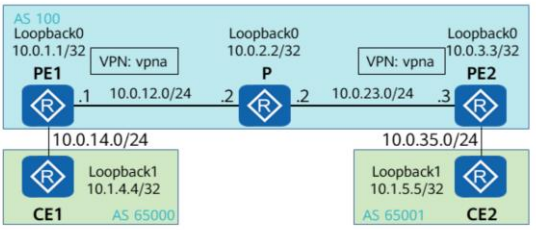
Configure an SR-MPLS TE tunnel interface. PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[*PE1] interface tunnel1
[*PE1-Tunnel1] ip address unnumbered interface LoopBack1
[*PE1-Tunnel1] tunnel-protocol mpls te
[*PE1-Tunnel1] destination 10.0.3.3
[*PE1-Tunnel1] mpls te tunnel-id 1
[*PE1-Tunnel1] mpls te signal-protocol segment-routing
[*PE1-Tunnel1] mpls te path explicit-path te1
[*PE1-Tunnel1] commit
[~PE1-Tunnel1] quit
```


HUAWEI

- In this example, an explicit path is established by specifying prefix SIDs.
- An explicit path is a vector path comprised of a series of nodes that are arranged in the configuration sequence. The path through which an SR-MPLS TE LSP passes can be planned by specifying next-hop labels or next-hop IP addresses on an explicit path. Generally, the IP addresses involved in an explicit path are interface IP addresses. An explicit path that is in use can be updated. To configure an explicit path, perform the following steps:
  - Run the **system-view** command to enter the system view.
  - Run the **explicit-path** *path-name* command to create an explicit path and enter the explicit path view.
  - Run the **next sid label** *label-value* **type** { **adjacency** | **prefix** | **binding-sid** } command to specify a next-hop SID for the explicit path.
  - Run the **commit** command to commit the configuration.

L3VPN over SR-MPLS TE (4)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR-MPLS TE LSPs on the backbone network.
3. **Establish an MP-BGP peer relationship between PE1 and PE2.**
4. **Enable the VPN instance IPv4 address family on each PE.**
5. **Establish an MP-IBGP peer relationship between the PEs.**
6. Configure a tunnel policy for the PEs to preferentially select SR-MPLS TE LSPs.
7. Verify the configuration.

PE1 configurations are as follows: (PE2 configurations are not provided.)

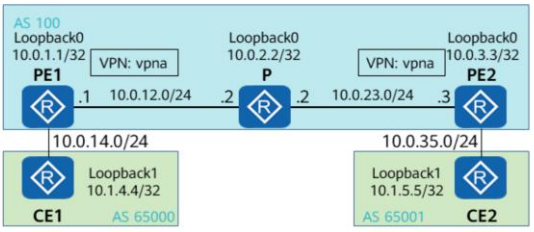
```
[~PE1] bgp 100
[~PE1-bgp] peer 10.0.3.3 as-number 100
[*PE1-bgp] peer 10.0.3.3 connect-interface loopback 0
[*PE1-bgp] ipv4-family vpnv4
[*PE1-bgp-af-vpnv4] peer 10.0.3.3 enable
[*PE1-bgp-af-vpnv4] commit
[~PE1-bgp-af-vpnv4] quit
[~PE1-bgp] quit
```

PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] route-distinguisher 100:1
[*PE1-vpn-instance-vpna-af-ipv4] vpn-target 111:1 both
[*PE1-vpn-instance-vpna-af-ipv4] quit
[*PE1-vpn-instance-vpna] quit
[*PE1]bgp 100
[*PE1-bgp]ipv4-family vpn-instance vpna
[*PE1-bgp-vpna]peer 10.0.14.4 as-number 65000
```


HUAWEI

L3VPN over SR-MPLS TE (5)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR-MPLS TE LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs.
6. **Configure a tunnel policy for the PEs to preferentially select SR-MPLS TE LSPs.**
7. **Verify the configuration.**

PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~PE1] tunnel-policy p2
[*PE1-tunnel-policy-p2] tunnel select-seq sr-te load-balance-number 1
[*PE1-tunnel-policy-p2] quit
[*PE1] commit
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] tnl-policy p2
[*PE1-vpn-instance-vpna-af-ipv4] quit
[*PE1-vpn-instance-vpna] quit
[*PE1] commit
```

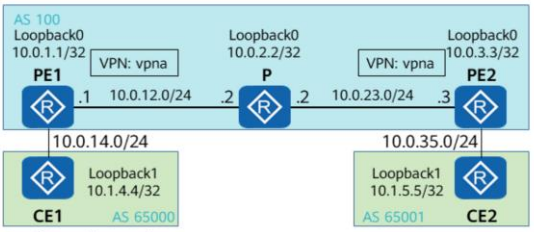
Run the **display tunnel-info all** command on PE1 to check SR LSP information.

Tunnel ID	Type	Destination	Status
0x000000000300000001	sr-te	10.0.3.3	UP
0x000000002900000042	srbe-lsp	10.0.3.3	UP
0x000000002900000043	srbe-lsp	10.0.2.2	UP

ID of the SR-TE tunnel to PE2

HUAWEI

L3VPN over SR-MPLS TE (6)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR-MPLS TE LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs.
6. Configure a tunnel policy for the PEs to preferentially select SR-MPLS TE LSPs.
7. **Verify the configuration.**

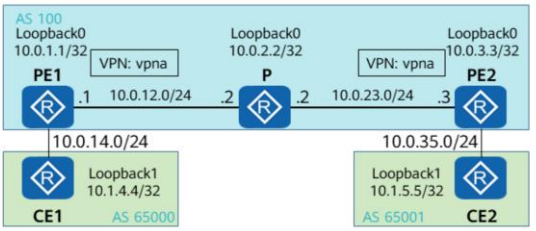
Check vpna's routing information on PE1.

```
[~PE1]display ip routing-table vpn-instance vpna 10.1.5.5 verbose
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B -
black hole route
------------------------------------------------------------------------------
Routing Table : vpna
Summary Count : 1

Destination: 10.1.5.5/32
    Protocol: IBGP                Process ID: 0
  Preference: 255                     Cost: 0
     NextHop: 10.0.3.3            Neighbour: 10.0.3.3
       State: Active Adv Relied         Age: 00h04m18s
         Tag: 0                  Priority: low
       Label: 48122              QoSInfo: 0x0
  IndirectID: 0x100013D               Instance:
RelayNextHop: 0.0.0.0             Interface: Tunnel1
    TunnelID: 0x00000000300000001 Flags: RD
```

The VPNv4 label and SR TE LSP are combined to guide packet forwarding.

HUAWEI

L3VPN over SR-MPLS TE (7)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR-MPLS TE LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs.
6. Configure a tunnel policy for the PEs to preferentially select SR-MPLS TE LSPs.
7. **Verify the configuration.**

Tracert the SR LSP on PE1.

```
<PE1>tracert lsp segment-routing  te Tunnel 1
  LSP Trace Route FEC: SEGMENT ROUTING TE TUNNEL IPV4 SESSION
QUERY Tunnel1 , press CTRL_C to break.
  TTL    Replier          Time    Type        Downstream
  0                               Ingress     10.0.12.2/[16003 ]
  1      10.0.12.2        21 ms   Transit     10.0.23.3/[3 ]
  2      10.0.3.3         9 ms    Egress
```

Question: How are the labels computed?

Verify the configuration on CE1.

```
<CE1>ping -a 10.1.4.4 10.1.5.5
  PING 10.1.5.5: 56  data bytes, press CTRL_C to break
    Reply from 10.1.5.5: bytes=56 Sequence=1 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=2 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=3 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=4 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=5 ttl=254 time=1 ms
```

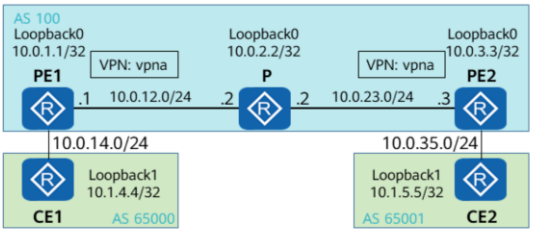

HUAWEI

Contents

1. Segment Routing Overview
2. Segment Routing Fundamentals
3. Segment Routing Tunnel Protection and Detection Technologies
4. Typical Usage Scenarios of Segment Routing
5. **Basic Configurations of Segment Routing**
   - SR-MPLS BE
   - SR-MPLS TE
   - SR-MPLS Policy

HUAWEI

L3VPN over Static SR-MPLS Policy (1)



Networking requirements:

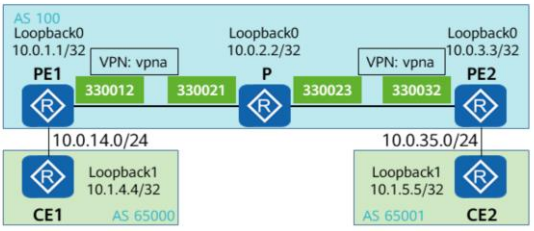
1. Connect PE1 and PE2 to different CEs that belong to VPN instance vpna.
2. Deploy L3VPN service recursion to static SR-MPLS Policy on the backbone network so that CE1 and CE2 can communicate through Loopback1.

Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS and configure an SR-MPLS Policy on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Configure the color attribute for routes on the PEs and enable the PEs to exchange routing information.
6. Configure a tunnel policy on the PEs.
7. Verify the configuration.

HUAWEI

L3VPN over Static SR-MPLS Policy (2)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. **Enable MPLS and configure an SR-MPLS Policy on the backbone network.**
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Configure the color attribute for routes on the PEs and enable the PEs to exchange routing information.
6. Configure a tunnel policy on the PEs.
7. Verify the configuration.

Configure basic SR-MPLS functions. PE1 configurations are as follows: (P and PE2 configurations are not provided.)

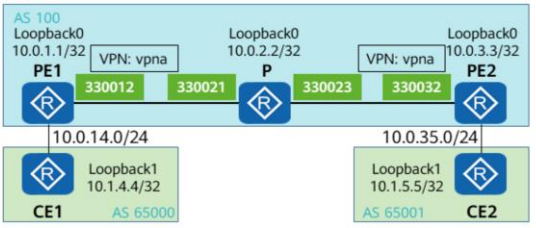
```
[~PE1] mpls lsr-id 10.0.1.1
[*PE1] mpls
[*PE1-mpls] mpls te
[*PE1-mpls] quit
[~PE1] segment-routing
[*PE1-segment-routing] ipv4 adjacency local-ip-addr 10.0.12.1 remote-ip-addr 10.0.12.2 sid 330012
[*PE1-segment-routing] quit
[~PE1] ospf 1
[*PE1-ospf-1] opaque-capability enable
[*PE1-ospf-1] segment-routing mpls
[*PE1-ospf-1] segment-routing global-block 16000 23999
[*PE1-ospf-1-area-0.0.0.0] quit
[*PE1] interface loopback 0
[*PE1-LoopBack1] ospf prefix-sid index 1
[*PE1-LoopBack1] quit
```

In scenarios where SR-MPLS Policies are statically configured, you are advised to use statically configured adjacency SIDs.

HUAWEI

- In this example, adjacency SIDs are configured statically. The values of adjacency SIDs are shown in the figure.

L3VPN over Static SR-MPLS Policy (3)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. **Enable MPLS and configure an SR-MPLS Policy on the backbone network.**
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Configure the color attribute for routes on the PEs and enable the PEs to exchange routing information.
6. Configure a tunnel policy on the PEs.
7. Verify the configuration.

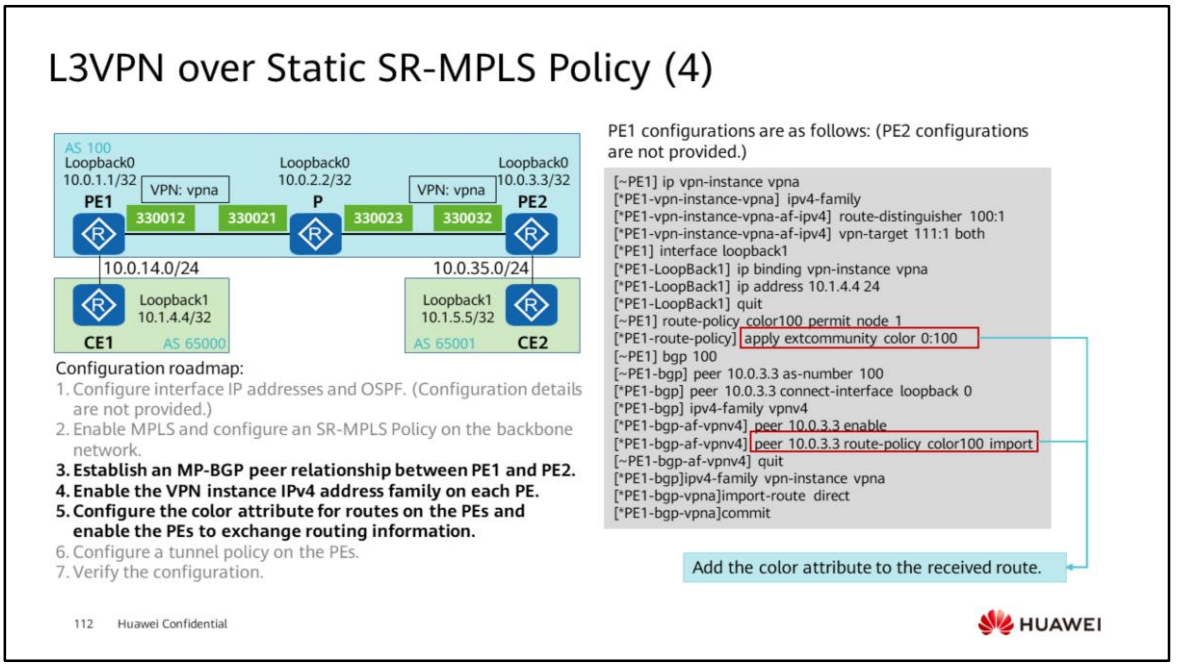
Configure an SR-MPLS Policy. PE1 configurations are as follows: (P and PE2 configurations are not provided.)

```
[~PE1] segment-routing
[~PE1-segment-routing] segment-list pe1
[*PE1-segment-routing-segment-list-pe1] index 10 sid label 330012
[*PE1-segment-routing-segment-list-pe1] index 20 sid label 330023
[*PE1-segment-routing-segment-list-pe1] quit
[*PE1-segment-routing] sr-te policy policy100 endpoint 10.0.3.3 color 100
[*PE1-segment-routing-te-policy-policy100] binding-sid 115
[*PE1-segment-routing-te-policy-policy100] mtu 1000
[*PE1-segment-routing-te-policy-policy100] candidate-path preference 200
[*PE1-segment-routing-te-policy-policy100-path] segment-list pe1
[*PE1-segment-routing-te-policy-policy100-path] quit
[*PE1-segment-routing-te-policy-policy100] quit
[*PE1-segment-routing] quit
[*PE1] commit
```

Configure a destination address and color for the SR-MPLS Policy.

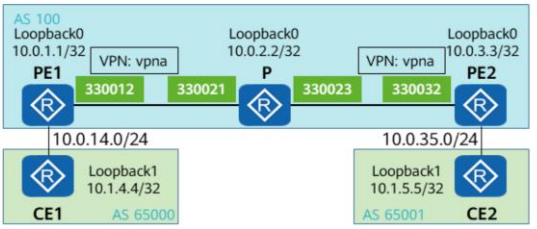
HUAWEI

- SR-MPLS Policies are used to direct traffic to traverse an SR-MPLS TE network. Each SR-MPLS Policy can have multiple candidate paths with different preferences. A valid candidate path with the highest preference is selected as the primary path, and a valid candidate path with the second highest preference as the backup path. The SR-MPLS Policy configuration procedure is as follows:
  - Configure a segment list.
    - Run the **system-view** command to enter the system view.
    - Run the **segment-routing** command to enable SR globally and enter the Segment Routing view.
    - Run the **segment-list** (Segment Routing view) list-name command to configure a segment list for an SR-MPLS TE candidate path and enter the segment list view.
    - Run the **index** *index* **sid label** *label* command to specify a next-hop SID for the segment list.
      - You can run the command multiple times. The system generates a label stack for the segment list by index in ascending order. If a candidate path in an SR-MPLS Policy is preferentially selected, traffic is forwarded using the segment list of the candidate path. A maximum of 10 SIDs can be configured for each segment list.



- The color attribute is added to a route through a route-policy. This enables the route to recurse to an SR-MPLS Policy based on the color value and next-hop address in the route.
    - Configure a route-policy.
        - Run the **system-view** command to enter the system view.
        - Run the **route-policy** *route-policy-name* { **deny** | **permit** } node node command to create a route-policy and enter the route-policy view.
        - (Optional) Configure if-match clauses for the route-policy. The community attributes of routes can be added or modified only if the routes match specified if-match clauses.
        - Run the **apply extcommunity color** *color* command to configure a BGP extended community, that is, the color attribute.
        - Run the **commit** command to commit the configuration.

L3VPN over Static SR-MPLS Policy (5)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS and configure an SR-MPLS Policy on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Configure the color attribute for routes on the PEs and enable the PEs to exchange routing information.
6. **Configure a tunnel policy on the PEs.**
7. **Verify the configuration.**

PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~PE1] tunnel-policy p3
[*PE1-tunnel-policy-p3] tunnel select-seq sr-te-policy load-balance-number 1 unmix
[*PE1-tunnel-policy-p3] quit
[*PE1] commit
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] tnl-policy p3
[*PE1-vpn-instance-vpna-af-ipv4] quit
[*PE1-vpn-instance-vpna] quit
[*PE1] commit
```

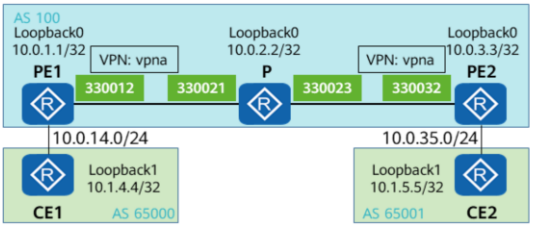
Run the **display tunnel-info all** command on PE1 to check SR LSP information.

```
<PE1>display tunnel-info all
Tunnel ID                  Type       Destination     Status
-------------------------------------------------------------
0x000000000300000001 sr-te          10.0.3.3        UP
0x000000002900000042 srbe-lsp       10.0.3.3        UP
0x000000002900000043 srbe-lsp       10.0.2.2        UP
0x000000003200000001 srtepolicy     10.0.3.3        UP
```

Tunnel ID of the SR-TE Policy destined for PE2

HUAWEI

L3VPN over Static SR-MPLS Policy (6)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS and configure an SR-MPLS Policy on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Configure the color attribute for routes on the PEs and enable the PEs to exchange routing information.
6. Configure a tunnel policy on the PEs.
7. **Verify the configuration.**

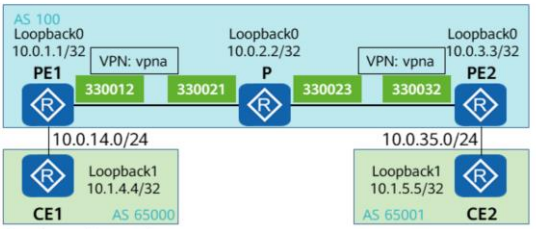
Check vpna's routing information on PE1.

```
[~PE1]display ip routing-table vpn-instance vpna 10.1.5.5 verbose
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B -
black hole route
------------------------------------------------------------------------------
Routing Table : vpna
Summary Count : 1

Destination: 10.1.5.5/32
    Protocol: IBGP              Process ID: 0
  Preference: 255                     Cost: 0
     NextHop: 10.0.3.3          Neighbour: 10.0.3.3
       State: Active Adv Relied       Age: 00h01m04s
         Tag: 0                Priority: low
       Label: 48122            QoSInfo: 0x0
  IndirectID: 0x100013F              Instance:
RelayNextHop: 0.0.0.0           Interface: policy100
    TunnelID: 0x00000003200000001  Flags: RD
```

The VPNv4 label and SR-TE Policy LSP are combined to guide packet forwarding.

HUAWEI

L3VPN over Static SR-MPLS Policy (7)



Configuration roadmap:

1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS and configure an SR-MPLS Policy on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Configure the color attribute for routes on the PEs and enable the PEs to exchange routing information.
6. Configure a tunnel policy on the PEs.
7. **Verify the configuration.**

Tracert the SR LSP on PE1.

```
<PE1>tracert lsp sr-te policy endpoint-ip 10.0.3.3 color 100
sr-te policy's segment list:
  Preference: 200; Path Type: primary; Protocol-Origin: local; Originator:
0, 0.0.0.0; Discriminator: 200; Segment-List ID: 65; Xcindex: 2000065
 TTL   Replier          Time    Type      Downstream
 0                              Ingress   10.0.12.2/[330023 ]
 1     10.0.12.2        24 ms   Transit   10.0.23.3/[3 ]
 2     10.0.3.3         113 ms  Egress
```

Question: How are the labels computed?

Verify the configuration on CE1.

```
<CE1>ping -a 10.1.4.4 10.1.5.5
  PING 10.1.5.5: 56  data bytes, press CTRL_C to break
    Reply from 10.1.5.5: bytes=56 Sequence=1 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=2 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=3 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=4 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=5 ttl=254 time=1 ms
```


HUAWEI

Quiz

1. (Single-answer question) Which of the following types of LSAs is used by OSPF to carry node IDs? ( )

   A. Type 1

   B. Type 2

   C. Type 7

   D. Type 10

HUAWEI

1. D

Quiz

2. (Multiple-answer question) Which of the following ports are used by SBFD packets by default? ( )

   A. 4784

   B. 3784

   C. 6784

   D. 7784

HUAWEI

2. AD

Summary

- SR is designed to forward data packets on a network using the source routing model. Compared with LDP and RSVP-TE, SR-MPLS simplifies the control plane of an MPLS network, enabling information such as labels to be carried only through IGP extensions. It provides higher scalability, freeing transit nodes from maintaining path information. The packet forwarding path can be controlled only by using the ingress. In addition, SR-MPLS can work with the centralized path computation module to flexibly and easily control and adjust paths, achieving smoother evolution to SDN.
- SR-MPLS supports three types of LSPs: SR-MPLS BE, SR-MPLS TE, and SR-MPLS Policy. SR-MPLS provides multiple detection and protection mechanisms for these different LSPs, such as TI-LFA FRR, anycast FRR, hot standby, VPN FRR, microloop avoidance, BFD, and SBFD.
- SR-MPLS supports both traditional and SDN networks, is compatible with existing devices, and supports multiple scenarios such as inter-AS interconnection. To facilitate understanding, this course provides examples for configuring SR-MPLS using commands. In the following courses, we will introduce how to use the controller to configure SR-MPLS.

HUAWEI

Thank you.

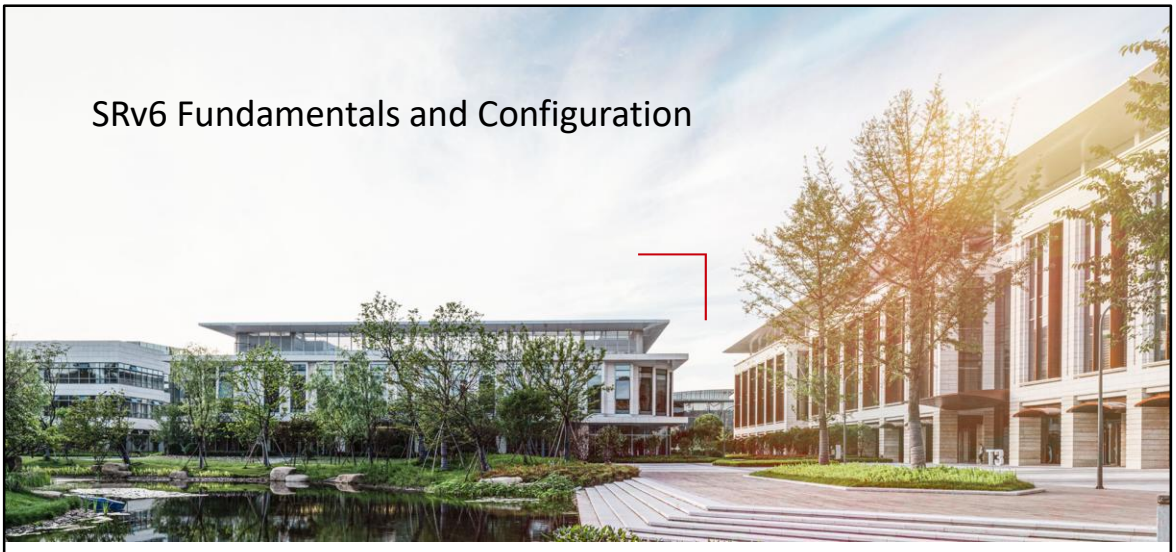
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

SRv6 Fundamentals and Configuration



HUAWEI

Foreword

- At the beginning of the Segment Routing (SR) architecture design, two implementation modes were designed for the data plane. One is Segment Routing-Multiprotocol Label Switching (SR-MPLS), which reuses the MPLS data plane and can be incrementally deployed on the existing IP/MPLS network. The other is Segment Routing IPv6 (SRv6), which uses the IPv6 data plane and implements extension based on the IPv6 Routing header.
- This document describes the concepts and fundamentals of SRv6 and its applications for Huawei NetEngine series routers.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the background of SRv6.
  - Describe the technical advantages of SRv6.
  - Describe the basic concepts and fundamentals of SRv6.
  - Describe the concepts and fundamentals of SRv6 high reliability.
  - Describe how to configure SRv6 BE and static SRv6 Policies.
  - Describe how to deploy SRv6 using iMaster NCE-IP.

HUAWEI

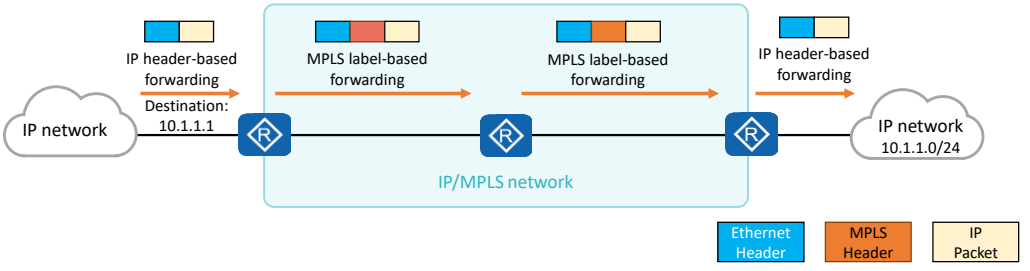
Contents

1. **SRv6 Overview**
2. SRv6 Fundamentals
3. SRv6 High Reliability
4. Basic SRv6 Configuration
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

IP/MPLS Network Introduction

- As a Layer 2.5 technology that runs between Layer 2 and Layer 3, MPLS adds connection-oriented attributes to connectionless IP networks. Traditional MPLS label-based forwarding improves the forwarding efficiency of IP networks. However, as hardware capabilities continue to improve, MPLS no longer features distinct advantages in forwarding efficiency. Nevertheless, MPLS provides good QoS guarantee for IP networks through connection-oriented label forwarding and also supports TE, VPN, and FRR.
- IP/MPLS networks have gradually replaced dedicated networks, such as ATM, frame relay (FR), and X.25. Ultimately, MPLS is applied to various networks, including IP backbone, metro, and mobile transport, to support multi-service transport and implement the Internet's all-IP transformation.



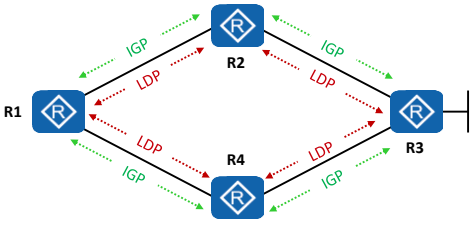
HUAWEI

- In the initial stage of network development, multiple types of networks, such as X.25, FR, ATM, and IP, co-existed to meet different service requirements. These networks could not interwork with each other, and on top of that, also competed, with mainly ATM and IP networks taking center stage. ATM is a transmission mode that uses fixed-length cell switching. It establishes paths in connection-oriented mode, and can provide better QoS capabilities than IP. The design philosophy of ATM involves centering on networks and providing reliable transmission, and its design concepts reflect the reliability and manageability requirements of telecommunications networks. This is the reason why ATM was widely deployed on early telecommunications networks. The design concepts of IP differ greatly from those of ATM. To be more precise, IP is a connectionless communication mechanism that provides the best-effort forwarding capability, and the packet length is not fixed. On top of that, IP networks mainly rely on the transport-layer protocols (e.g., TCP) to ensure transmission reliability, and the requirement for the network layer involves ease of use. The design concept of IP networks embodies the "terminal-centric and best-effort" notion of the computer network, enabling IP to meet the computer network's service requirements. The competition between the two can essentially be represented as a competition between telecommunications and computer networks. As the network scale expanded and network services increased in number, ATM networks became more complex than IP networks, while also bearing higher management costs. Within the context of costs versus benefits for telecom carriers, ATM networks were gradually replaced by IP networks.

- Although IP is more suitable for the development of computer networks than ATM, computer networks require a certain level of QoS guarantee. To compensate for the IP network's insufficient QoS capabilities, numerous technologies integrating IP and ATM, such as local area network emulation (LANE) and IP over ATM (IPoA), have been proposed. However, these technologies only addressed part of the issue, until 1996 when MPLS technology was proposed to provide a better solution to this issue.

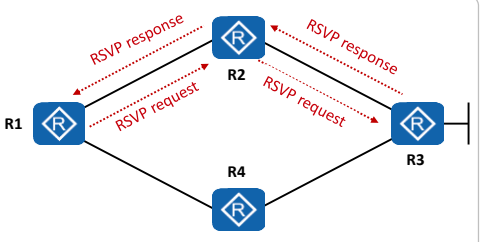
Issues with MPLS LDP and RSVP-TE

MPLS LDP



- LDP itself does not have the path computation capability and requires an IGP for path computation.
- Both the IGP and LDP need to be deployed for the control plane, and devices need to exchange a large number of packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- If LDP-IGP synchronization is not achieved, data forwarding may fail.

RSVP-TE

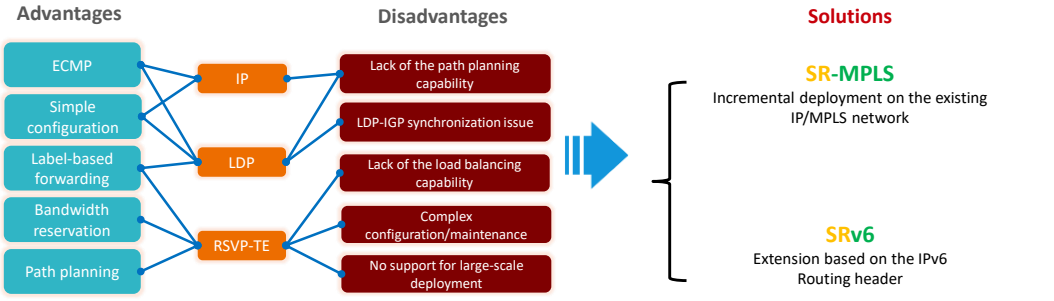


- RSVP-TE configuration is complex and load balancing is not supported.
- To implement TE, devices need to exchange a large number of RSVP packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- RSVP-TE uses a distributed architecture, so that each device only knows its own state and needs to exchange signaling packets with other devices.

HUAWEI

SR Origin and Solution

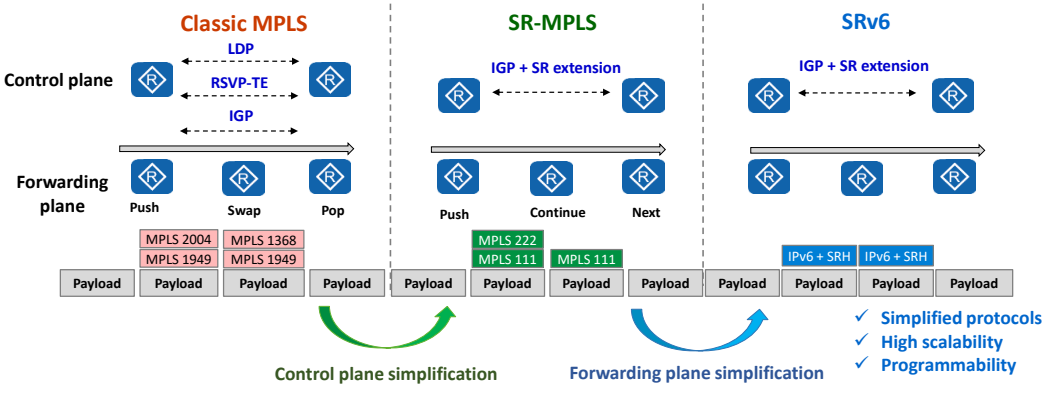
- The SDN concept has a great impact on the network industry, and many protocols used for SDN implementation emerge in the industry, including OpenFlow, Protocol Oblivious Forwarding (POF), Programming Protocol-independent Packet Processors (P4), and SR. Compared with revolutionary protocols, SR considers compatibility with the existing network and smooth evolution, and also provides programmability. It is a de facto SDN standard.



- SR resolves many issues on IP/MPLS networks through two solutions: SR-MPLS (based on MPLS forwarding) and SRv6 (based on IPv6 forwarding).

From MPLS to SRv6

- MPLS causes isolated network islands. SRv6 provides a unified forwarding plane and has advantages such as simplified protocols, high scalability, and programmability.

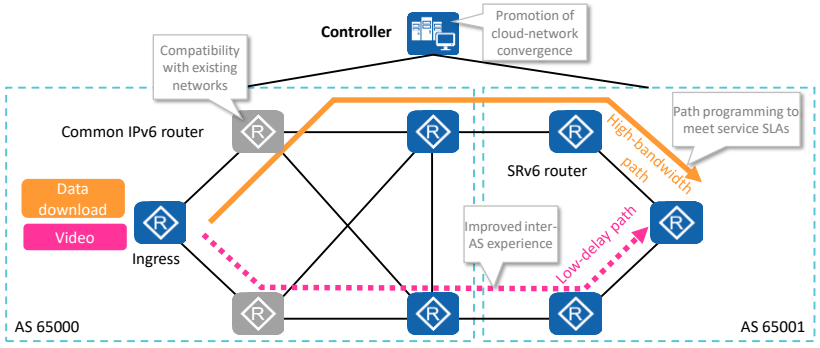


HUAWEI

- Although MPLS plays an important role in the all-IP transformation of networks, it causes isolated network islands. On the one hand, it increases the complexity of cross-domain network interconnection. For example, solutions such as the MPLS VPN Option A/B/C solution are complex to deploy and involve difficult E2E service deployment. On the other hand, as the Internet and cloud computing develop, more and more cloud data centers are built. To meet tenants' networking requirements, multiple overlay technologies have been proposed, among which VXLAN is a typical example. In the past, quite a few attempts were made to provide VPN services by introducing MPLS to data centers. However, these attempts all wound up in failure due to multiple factors, including numerous network boundaries, complex management, and insufficient scalability. As such, the traffic from an end user to a service in a data center may typically need to pass through the VLAN, IP network, IP/MPLS network, and VXLAN network.
- The combination of MPLS and SR is intended to provide programmability for networks as a practice of SDN implementation. However, this cannot satisfy services (such as SFC and IOAM) that need to carry metadata, as MPLS encapsulation has poor scalability. Nowadays, the IPv4 address space is almost exhausted. IPv6 and SR are combined, promoting the advent of SRv6.

Technical Benefits of SRv6

- SRv6 simplifies existing network protocols and network management. In addition, SRv6 supports native IPv6 and provides the network programming capability, making SRv6 more advantageous.
  - Thanks to the native IPv6 attribute, SRv6 can better promote cloud-network convergence, be compatible with existing networks, and improve inter-AS experience.
  - Thanks to the network programming capability, SRv6 can not only better implement path programming to meet service SLAs but also connect networks and applications to build intelligent cloud-networks.



HUAWEI

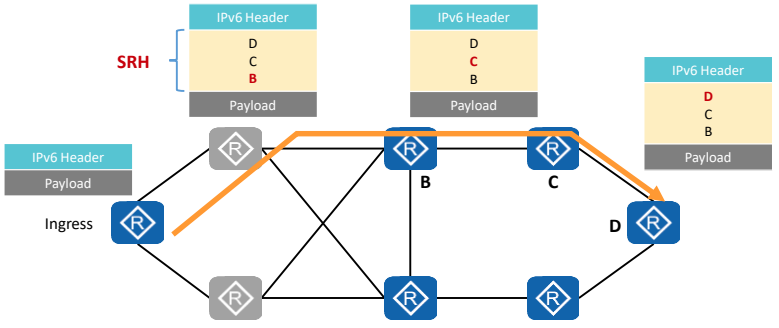
Contents

1. SRv6 Overview
2. **SRv6 Fundamentals**
   - Basic Concepts of SRv6
   - SRv6 Policy Path Establishment and Traffic Steering
   - Typical SRv6 Applications
3. SRv6 High Reliability
4. Basic SRv6 Configuration
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

SRv6 Fundamentals

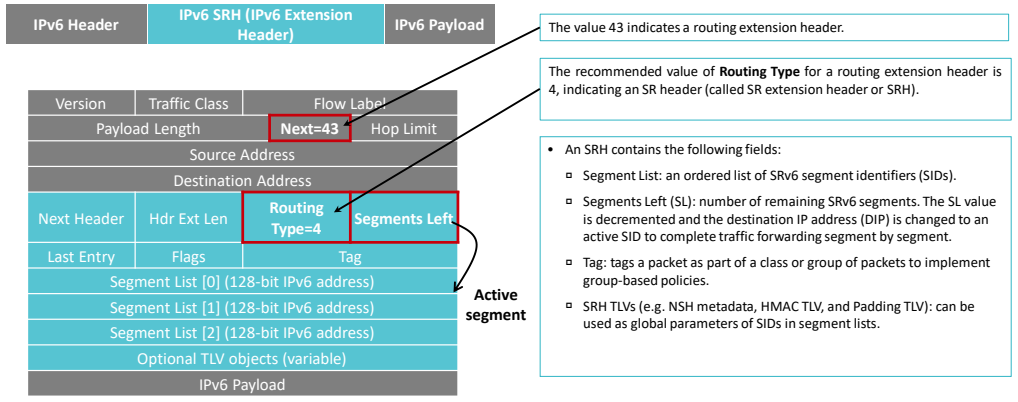
- SRv6 adds a Segment Routing header (SRH) to data on the ingress to guide data forwarding.
- Without changing the original encapsulation format of IPv6 packets, SRv6 packets are still IPv6 ones and can be identified by common IPv6 devices. As such, SRv6 is a native IPv6 technology.
- SRv6's native IPv6 attribute enables SRv6 devices to interwork with common IPv6 devices, offering excellent compatibility on an existing network.



HUAWEI

IPv6 SRH

- RFC 8754 defines the IPv6 SRH added to IPv6 packets. The SRH format is as follows:

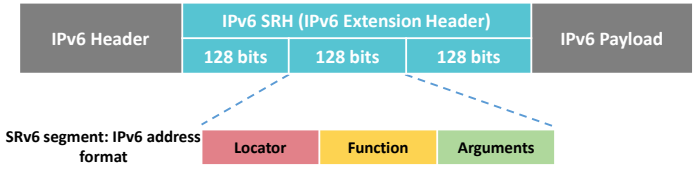


HUAWEI

- The biggest difference between SRv6 and SR-MPLS lies in the IPv6 SRH. SRv6 uses IPv6 extension headers to implement Segment Routing.
- For details, see https://datatracker.ietf.org/doc/rfc8754.

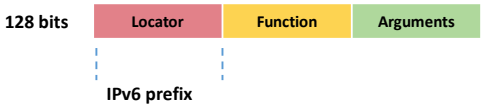
SRv6 Segment

- SRv6 segments are expressed using IPv6 addresses and usually called SRv6 SIDs.
- As shown in the figure, an SRv6 SID usually consists of three fields: Locator, Function, and Arguments. They are expressed in the Locator:Function:Arguments format. Note that the total length (Locator + Function + Arguments) is less than or equal to 128 bits. If the total length is less than 128 bits, the reserved bits are padded with 0.
- If the Arguments field does not exist, the format is Locator:Function. The Locator field occupies the most significant bits of an IPv6 address, and the Function field occupies the remaining part of the IPv6 address.



HUAWEI

SRv6 Segment: Locator



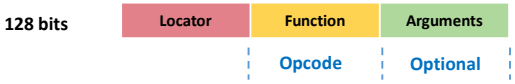
- The Locator field identifies the location of a network node, and is used for other nodes to route and forward packets to this identified node so as to implement network instruction addressing.
- A locator has two important characteristics: routable and aggregatable. After a locator is configured for a node, the system generates a locator route and propagates the route throughout the SR domain using an IGP, allowing other nodes to locate the node based on the received locator route information. In addition, all SRv6 SIDs advertised by the node are reachable through the route.
- In the following example, a locator with the 64-bit prefix 2001:DB8:ABCD:: is configured for a Huawei device.

```
[Huawei] segment-routing ipv6
[Huawei-segment-routing-ipv6] locator srv6_locator1 ipv6-prefix 2001:DB8:ABCD:: 64
```


HUAWEI

- The locator is routable and therefore usually unique in an SR domain. In some scenarios, such as an anycast protection scenario, multiple devices may be configured with the same locator.

SRv6 Segment: Function & Arguments



- The Function field identifies the forwarding behavior to be performed. In SRv6 network programming, forwarding behaviors are identified using different functions. For example, RFC defines End, End.X, End.DX4, and End.DX6 behaviors.
- An End.X SID is similar to an adjacency SID in SR-MPLS and is used to identify a link. A configuration example is as follows:

```
[Huawei-segment-routing-ipv6] locator srv6_locator1 ipv6-prefix 2001:DB8:ABCD:: 64
[Huawei-segment-routing-ipv6] opcode ::1 end-x interface G3/0/0 next-hop 2001:DB8:200::1
```

- The opcode corresponding to the function is ::1. In this example, the Arguments field is not carried, and the SRv6 SID is 2001:db8:abcd::1.
- This function guides packet forwarding from the specified interface (G3/0/0) to the corresponding neighbor (2001:DB8:200::1).

HUAWEI

- In some scenarios, an SRv6 endpoint behavior may require additional actions. In this case, the Arguments field must be encapsulated. For example, in an EVPN VPLS scenario where CE multi-homing is deployed for BUM traffic forwarding, the Function field is set to End.DT2M, and the Arguments field is used to provide local ESI mapping to implement split horizon.
- The Function and Arguments fields can both be defined by engineers, resulting in an SRv6 SID structure that improves network programmability. In most scenarios, the Arguments field is not configured.

SRv6 Segment Types

Category	Function Description	Protocol	Type
End	Indicates an endpoint SID that identifies a destination node. The corresponding function is to update the IPv6 DA and then search the IPv6 forwarding information base (FIB) for packet forwarding.	IGP	Path SID
End.X	Indicates a Layer 3 cross-connect endpoint SID that identifies a link. The corresponding function is to update the IPv6 DA and then forward packets through the outbound interface bound to the End.X SID.	IGP	Path SID
End.DT4	Indicates a PE-specific endpoint SID that identifies an IPv4 VPN instance. The corresponding function is to decapsulate packets and then search the routing table of the involved IPv4 VPN instance for packet forwarding. This SID is equivalent to an IPv4 VPN label and used in L3VPNv4 scenarios.	BGP	Service SID
End.DT6	Indicates a PE-specific endpoint SID that identifies an IPv6 VPN instance. The corresponding function is to decapsulate packets and then search the routing table of the involved IPv6 VPN instance for packet forwarding. This SID is equivalent to an IPv6 VPN label and used in L3VPNv6 scenarios.	BGP	Service SID
End.DX4	Indicates a PE-specific Layer 3 cross-connect endpoint SID that identifies an IPv4 CE. The corresponding function is to decapsulate packets and then forward the resulting IPv4 packets through the Layer 3 interface bound to the SID. This SID is equivalent to a label identifying an adjacency to a CE and used in L3VPNv4 scenarios.	BGP	Service SID
End.DX6	Indicates a PE-specific Layer 3 cross-connect endpoint SID that identifies an IPv6 CE. The corresponding function is to decapsulate packets and then forward the resulting IPv6 packets through the Layer 3 interface bound to the SID. This SID is equivalent to a label identifying an adjacency to a CE and used in L3VPNv6 scenarios.	BGP	Service SID

HUAWEI

- In addition to L3VPN services, SRv6 can carry L2VPN services. L2VPN-related SIDs are as follows:
  - End.DX2: Indicates a Layer 2 cross-connect endpoint SID that identifies an endpoint. The corresponding function is to decapsulate packets, remove the IPv6 header (along with all its extension headers), and then forward the remaining packet data to the outbound interface associated with the SID. This SID can be used in EVPN VPWS scenarios. If a bypass tunnel exists on the network, an End.DX2L SID is generated automatically.
  - End.DT2U: Indicates a Layer 2 cross-connect endpoint SID that requires unicast MAC table lookup and identifies an endpoint. If a bypass tunnel exists on the network, an End.DT2UL SID is generated automatically. This SID can be used to guide unicast traffic forwarding over the bypass tunnel when a CE is dual-homed to PEs. The corresponding function is to remove the IPv6 header (along with all its extension headers), search the MAC address table for a MAC entry based on the exposed destination MAC address, and then forward the remaining packet data to the corresponding outbound interface based on the entry. This SID can be used in EVPN VPLS unicast scenarios.

    - End.DT2M: Indicates a Layer 2 cross-connect endpoint SID that requires broadcast-based flooding and identifies an endpoint. The corresponding function is to remove the IPv6 header (along with all its extension headers) and then broadcast the remaining packet data in the Bridge Domain (BD). This SID can be used in EVPN VPLS BUM scenarios.

- SRv6 SID mainly used for SRv6 O&M:
    - End.OP (OAM Endpoint with Punt): Indicates an OAM SID. The corresponding function is to send OAM packets to the OAM process. This SID is mainly used in ping/tracert scenarios.

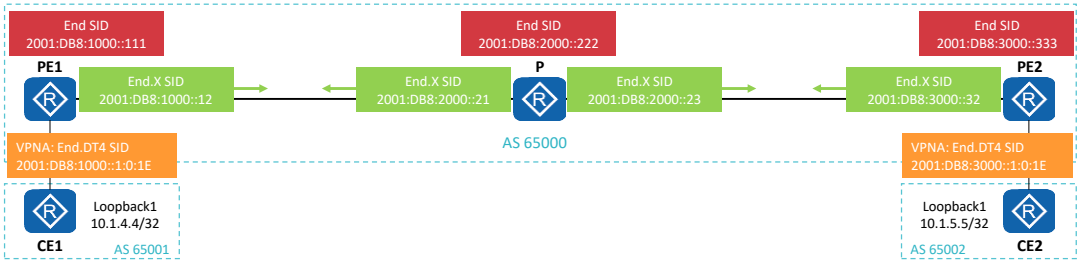
Naming Rules for SRv6 Segments

- SRv6 segments are named according to certain rules. You can quickly determine the corresponding instruction function based on the naming rule combination.
  - End: the most basic instruction executed by a segment endpoint node, directing the node to terminate the current instruction and start the next instruction. The corresponding forwarding behavior is to decrement the SL field by 1 and copy the SID pointed by the SL field to the DA field in the IPv6 header.
  - X: forwards packets through one or a group of Layer 3 outbound interfaces.
  - T: searches a specified routing table and forwards packets.
  - D: decapsulates packets by removing the IPv6 header and related extension headers.
  - V: searches a specified table for packet forwarding based on virtual local area network (VLAN) information.
  - U: searches a specified table for packet forwarding based on unicast MAC address information.
  - M: searches a Layer 2 forwarding table for multicast forwarding.
  - B6: applies a specified SRv6 Policy.
  - BM: applies a specified SR-MPLS Policy.

HUAWEI

SRv6 Segment Examples

- An End SID identifies a destination node. It is similar to a node SID in SR-MPLS. After an End SID is generated on a node, the node propagates the SID to all the other nodes in the SRv6 domain through an IGP. All nodes in the SRv6 domain know how to implement the instruction bound to the SID.
- An End.X SID is a Layer 3 cross-connect endpoint SID that identifies a link. It is similar to an adjacency SID in SR-MPLS. After an End.X SID is generated on a node, the node propagates the SID to all the other nodes in the SRv6 domain through an IGP. Although the other nodes can all obtain the SID, only the node generating the SID knows how to implement the instruction bound to the SID.
- An End.DT4 SID is a PE-specific endpoint SID that identifies an IPv4 VPN instance. The instruction bound to the End.DT4 SID is to decapsulate packets and search the routing table of the corresponding IPv4 VPN instance for packet forwarding. The End.DT4 SID is equivalent to an IPv4 VPN label and used in L3VPNv4 scenarios. It can be either manually configured or automatically allocated by BGP within the dynamic SID range of the specified locator.



HUAWEI

SRv6 Flavors

- Flavors are additional behaviors defined for SRv6 segment enhancement. These behaviors are optional and used to enhance SRv6 segment-based actions in order to meet diverse service requirements.
- SRv6-Network-Programming defines the following additional behaviors: penultimate segment pop of the SRH (PSP), ultimate segment pop of the SRH (USP), and ultimate segment decapsulation (USD).

Flavor	Function Description	Attached End Instruction
PSP	Removes the SRH on the penultimate endpoint node.	End, End.X, End.DT2, End.DT4, and End.DT6
USP	Removes the SRH on the ultimate endpoint node.	End, End.X, End.DT2, End.DT4, and End.DT6
USD	Decapsulates the outer IPv6 header on the ultimate endpoint node.	End, End.X, End.DT2, End.DT4, and End.DT6

HUAWEI

- Different flavors can be combined. For example, if an End SID carries PSP and USP flavors, the PSP action is performed on the penultimate node, and the USD action is performed on the ultimate node.

SRv6 Locator Configuration Commands

1. Enable SRv6 and enter the SRv6 view.

   [Huawei] **segment-routing ipv6**

   After running the **segment-routing ipv6** command, you can configure a locator and SRv6 SID in the SRv6 view so that an SRv6 local SID forwarding entry can be generated.

2. Configure an SRv6 SID locator.

   [Huawei-segment-routing-ipv6] **locator** *locator-name* [**ipv6-prefix** *ipv6-address prefix-length* [[**static** *static-length*] | [**args** *args-length*]] *]

   An SRv6 SID is a 128-bit IPv6 address expressed in the *Locator:Function:Arguments* format.
   - The Locator field corresponds to the **ipv6-prefix** *ipv6-address* parameter and its length is determined by the **prefix-length** parameter.
   - The Function field is also called opcode, which can be dynamically allocated using an IGP or be configured using the **opcode** command. When configuring a locator, you can use the **static** *static-length* parameter to specify the static segment length, which determines the number of static opcodes that can be configured in the locator. In dynamic opcode allocation, the IGP allocates opcodes outside the range of the static segment, so that no SRv6 SID conflict occurs.
   - The Args field is determined by the **args** *args-length* parameter. It is optional in SRv6 SIDs and depends on command configurations.

HUAWEI

- **static** *static-length*: specifies the static segment length in the Function field. This length determines the number of static opcodes that can be configured in the specified locator.
- **args** *args-length*: specifies the length of the Arguments field. The Arguments field is located at the end of a SID. If **args** *args-length* is configured, the Arguments field is reserved and will not be occupied by configured static SIDs or generated dynamic SIDs.

SRv6 SID Configuration Commands

1. Configure a static End SID opcode.

   [Huawei-segment-routing-ipv6-locator] **opcode** *func-opcode* **end [no-flavor | psp | psp-usp-usd]**

   An End SID identifies an SRv6 node.

2. Configure a static End.X SID opcode.

   [Huawei-segment-routing-ipv6-locator] **opcode** *func-opcode* **end-x interface** {*interface-name* | *interface-type interface-number*} **nexthop** *nexthop-address* **[no-flavor | psp | psp-usp-usd]**

   An End.X SID identifies a Layer 3 adjacency of an SRv6 node. Therefore, you need to specify an interface and the next hop address of the interface during the configuration.

HUAWEI

SRv6 Locator Configuration Example

- SRv6 SIDs can be statically configured or dynamically allocated. In dynamic allocation mode, only the locator command needs to be run, and the required opcode is dynamically allocated by an IGP. In static configuration mode, you need to manually configure an opcode for the SIDs of the corresponding type.
- The relationship of parameters in the locator command is as follows:

  |--Locator--|--Dynamic Opcode--|--Static Opcode--|--Args--|

  ```
  [Huawei-segment-routing-ipv6] locator srv6_locator1 ipv6-prefix 2001:DB8:ABCD:: 64 static 32
  ```

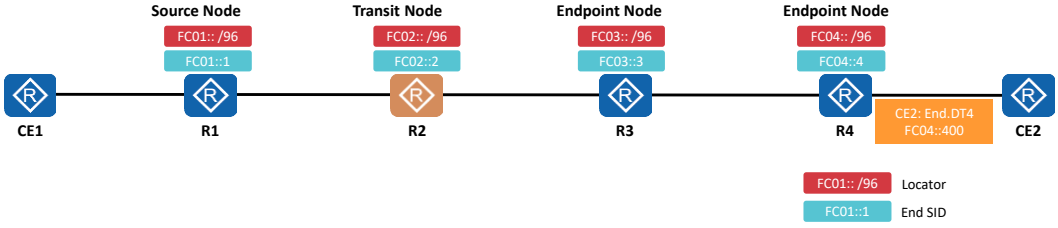
- In static configuration mode, SIDs occupy only the static segment with values starting from 1, and the dynamic segment is set to 0. In dynamic allocation mode, SIDs occupy both the dynamic segment and static segment. The values in the dynamic segment start from 1, and those in the static segment start from 0.
- In this example, the locator 2001:DB8:ABCD:: is configured, and its length is 64 bits. The static segment occupies 32 bits, the dynamic segment 32 bits, and the Args field 0 bits. The value range is as follows:
  - Static segment: The start value is 2001:DB8:ABCD:0000:0000:0000:0000:0001, and the end value is 2001:DB8:ABCD:0000:0000:0000:FFFF:FFFF.
  - Dynamic segment: The start value is 2001:DB8:ABCD:0000:0000:0001:0000:0000, and the end value is 2001:DB8:ABCD:0000:FFFF:FFFF:FFFF:FFFF.

  Statically configuring End and End.X SIDs is recommended. Dynamically allocated SIDs will change after a device restart, adversely affecting maintenance.

HUAWEI

SRv6 Node

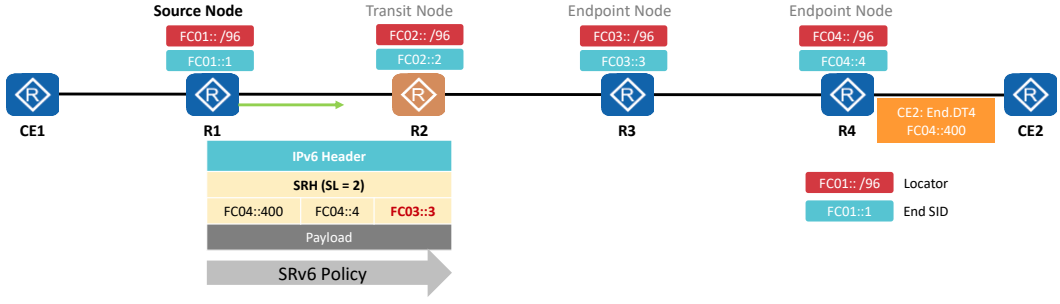
- RFC 8754 defines three types of SR nodes:
  - SR source node: a source node that encapsulates packets with SRv6 headers.
  - Transit node: an IPv6 node that forwards SRv6 packets but does not perform SRv6 processing.
  - SRv6 segment endpoint node: a node that receives and processes SRv6 packets in which the destination IPv6 address is a local SID or local interface address of the node.



HUAWEI

SRv6 Source Node

- An SRv6 source node steers a packet using an SRv6 segment list. If the SRv6 segment list contains only one SID, and no Type Length Value (TLV) or other information needs to be added to the packet, the DA field of the packet is set to this SID.
- An SRv6 source node can be either an SRv6-capable host where IPv6 packets originate or an edge device in an SRv6 domain.

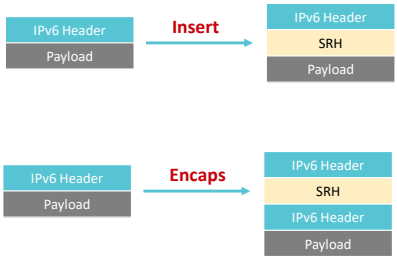


HUAWEI

Source Node Behaviors

- An SRv6 source node steers packets into an SRv6 Policy and, if possible, encapsulates SRHs into the packets. The following table lists the behaviors of an SRv6 source node.

Source Node Behavior	Function Description
H.Insert	Inserts an SRH into a received IPv6 packet and searches the corresponding routing table for packet forwarding.
H.Insert.Red	Inserts a reduced SRH into a received IPv6 packet and searches the corresponding routing table for packet forwarding.
H.Encaps	Encapsulates an outer IPv6 header and SRH for a received IP packet, and searches the corresponding routing table for packet forwarding.
H.Encaps.Red	Encapsulates an outer IPv6 header and reduced SRH for a received IP packet, and searches the corresponding routing table for packet forwarding.
H.Encaps.L2	Encapsulates an outer IPv6 header and SRH for a received Layer 2 frame, and searches the corresponding routing table for forwarding.
H.Encaps.L2.Red	Encapsulates an outer IPv6 header and reduced SRH for a received Layer 2 frame, and searches the corresponding routing table for forwarding.

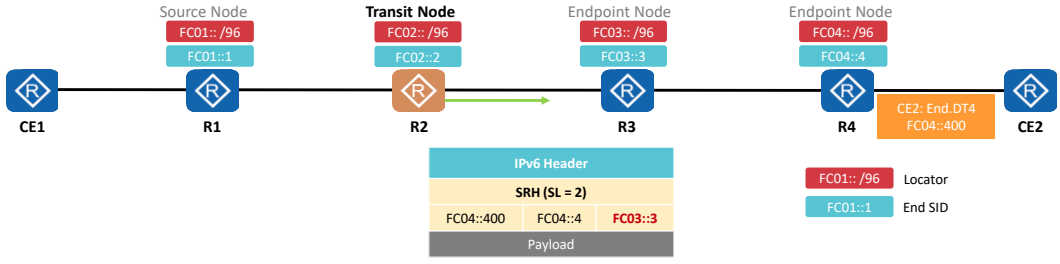


HUAWEI

- The difference between a reduced SRH and an SRH is that the segment list in a reduced SRH does not contain the first segment in an existing IPv6 DA.

Transit Node

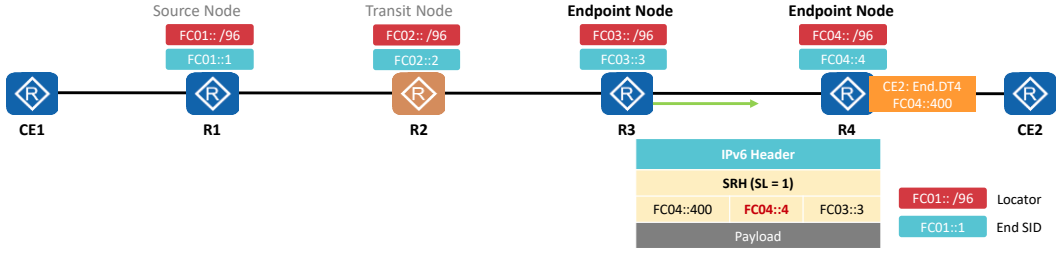
- A transit node is an IPv6 node that does not participate in SRv6 processing on the SRv6 packet forwarding path. That is, the transit node just performs ordinary IPv6 packet forwarding.
- After receiving an SRv6 packet, the node parses the IPv6 DA field in the packet. If the IPv6 DA is neither a locally configured SRv6 SID nor a local interface address, the node considers the SRv6 packet as an ordinary IPv6 packet and searches the routing table for packet forwarding without processing the SRH.
- A transit node can be either an ordinary IPv6 node or an SRv6-capable node.



HUAWEI

Endpoint Node

- An endpoint node is a node that receives an SRv6 packet destined for itself (a packet of which the IPv6 destination address is a local SID).
- For example, R3 searches its local SID table based on the IPv6 DA FC03::3 of the packet and finds a matching End SID. Then, R3 decrements the SL value by 1, uses the SID whose SL value is 1 as the destination IPv6 address, searches the routing table, and forwards the packet.
- There may be multiple endpoint nodes on the data forwarding path. Each endpoint node provides services such as packet forwarding, encapsulation, and decapsulation.



HUAWEI

- Each SRv6 node maintains a local SID table that contains all SRv6 SIDs generated on the node, and an SRv6 FIB can be generated based on the table. The local SID table provides the following functions:
  - Defines locally generated SIDs, such as End.X SIDs.
  - Specifies instructions bound to the SIDs.
  - Stores forwarding information related to the instructions, such as outbound interface and next hop information.

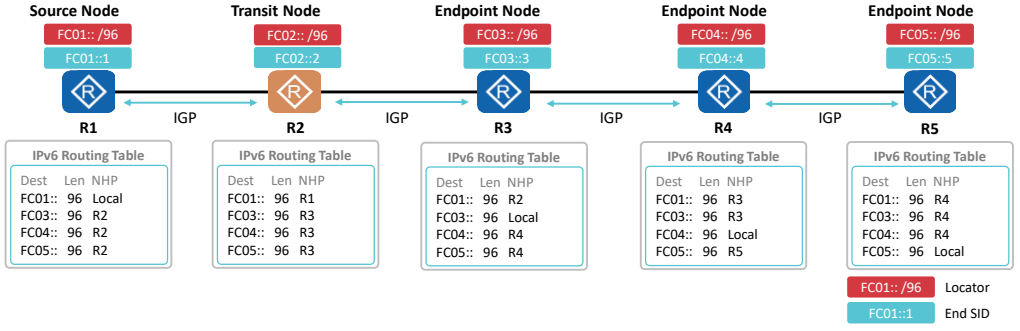
SRv6 Forwarding Modes

- SRv6 supports two forwarding modes: SRv6 Policy and SRv6 BE.
  - In addition to implementing traffic engineering, SRv6 Policy can work with a controller to meet differentiated service requirements more effectively, achieving a service-driven network.
  - SRv6 BE is a simplified implementation of SRv6. Typically, it can provide only best-effort forwarding and does not involve SRHs.
- In the initial phase of SRv6 deployment, SRv6 BE can be used to quickly provision services based on IPv6 route reachability, offering unparalleled advantages. During future evolution, transit nodes can be upgraded on demand and SRv6 Policy can be deployed to meet the requirements of high-value services.

HUAWEI

SRv6 Locator Information Propagation

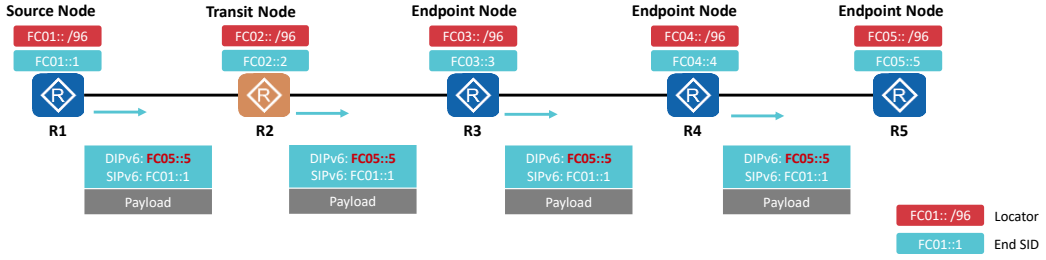
- No matter whether traffic is forwarded in SRv6 BE or SRv6 Policy mode, a router can forward SRv6 packets only after obtaining SRv6 locator-related routing information.
- SRv6 nodes usually use an extended IGP (extended OSPFv3 or IS-IS) to propagate locator-related routing information to network nodes, including the source, transit, and endpoint nodes.



HUAWEI

Forwarding in SRv6 BE Mode

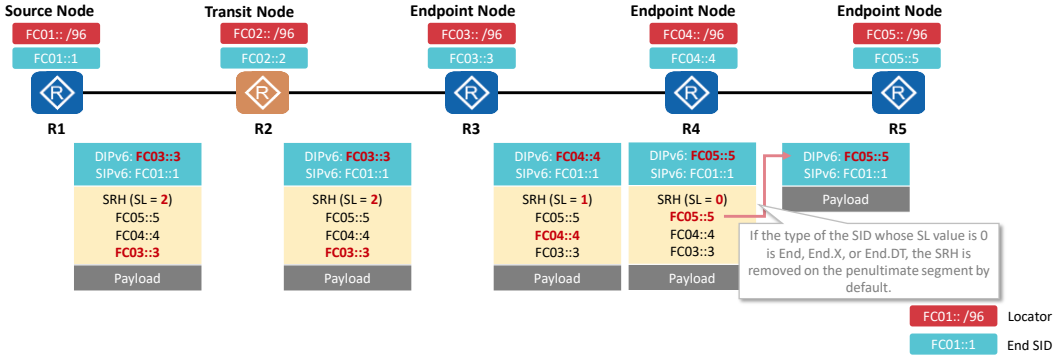
- Traditional MPLS involves two control protocols: LDP and RSVP-TE. The former uses IGP path computation results to establish LDP LSPs and guide traffic forwarding, but it does not support traffic engineering. Similar to LDP, SRv6 BE uses only one service SID to guide packet forwarding on an IP network in best-effort mode.
- In SRv6 BE mode, the forwarding path is computed based on the IGP cost.



HUAWEI

Forwarding in SRv6 Policy Mode

- In SRv6 forwarding, each time a packet passes through an SRv6 endpoint node, the SL field is decremented by 1 and the IPv6 DA in the IPv6 header changes. The SL and Segment List fields are both used to determine an IPv6 DA.



- Different from SR-MPLS label processing, SRv6 SRH processing is implemented from the bottom up, and segments in the SRv6 SRH are not popped after being processed by a node. Therefore, the SRv6 header can be used for path backtracking.

HUAWEI

- In MPLS, different removal options are defined using the Implicit-Null and Non-null options. Penultimate hop popping (PHP) in the MPLS data plane refers to the process in which the outermost label of the MPLS label stack is removed by an LSR before the packet reaches the adjacent label edge router (LER). If PHP is not enabled on the MPLS network, the LER is responsible for removing the label.
- These behaviors are defined as two functions in SRv6: PSP and USP.

Contents

1. SRv6 Overview
2. **SRv6 Fundamentals**
   - Basic Concepts of SRv6
   - SRv6 Policy Path Establishment and Traffic Steering
   - Typical SRv6 Applications
3. SRv6 High Reliability
4. Basic SRv6 Configuration
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

SRv6 Policy Overview

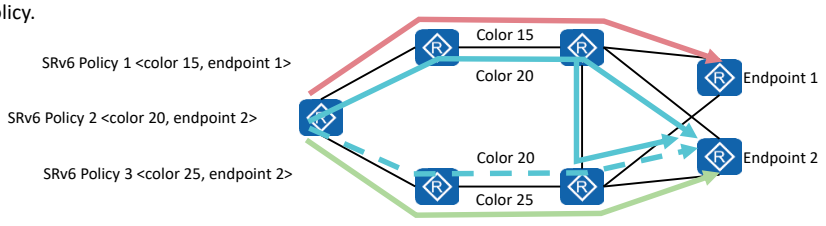
- SRv6 Policy is a new traffic steering technology developed based on SRv6.
- An SRv6 Policy is a set of candidate paths consisting of one or more segment lists, that is, SID lists. Each SID list identifies an E2E path from the source to the destination, instructing a device to forward traffic through the path rather than the shortest path computed using an IGP.
- If a packet is steered into an SRv6 Policy, the headend adds a SID list into the packet, and other devices receiving the packet execute the instructions encapsulated into the list.

HUAWEI

- https://datatracker.ietf.org/doc/draft-ietf-spring-segment-routing-policy/

SRv6 Policy Identification

- An SRv6 Policy is identified by the tuple <headend, color, endpoint>.
- For an SRv6 Policy with a specified headend, it is identified only using <color, endpoint>.
  - Headend: node where an SRv6 Policy is originated. Generally, it is a globally unique IP address.
  - Color: 32-bit extended community attribute. It is used to identify a type of service intent (e.g. low delay).
  - Endpoint: destination address of an SRv6 Policy. Generally, it is a globally unique IPv6 address.
- On the specified headend, the color and endpoint are used to identify the forwarding path of the corresponding SRv6 Policy.

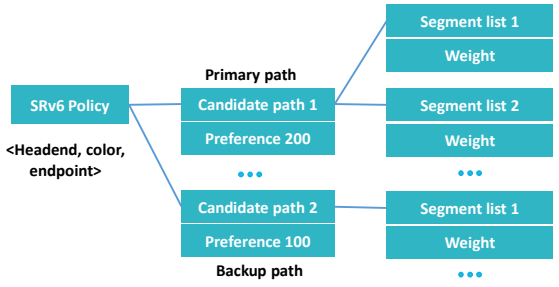


HUAWEI

- An endpoint in an SRv6 Policy is different from an endpoint node in SRv6.
  - The endpoint node in SRv6 refers to the type of the device that processes the SRH.
  - The endpoint in an SRv6 Policy refers to the policy's egress, which is generally expressed using an IPv6 address.

SRv6 Policy Path Model

- One SRv6 Policy may contain multiple candidate paths with the preference attribute. The valid candidate path with the highest preference functions as the primary path of the SRv6 Policy, and the valid candidate path with the second highest preference functions as a backup path.
- A candidate path is an SRv6 Policy's basic unit that is manually configured or is sent to the headend through BGP IPv6 SR Policy.
- Weights can be configured for segment lists to control load balancing among SRv6 paths.



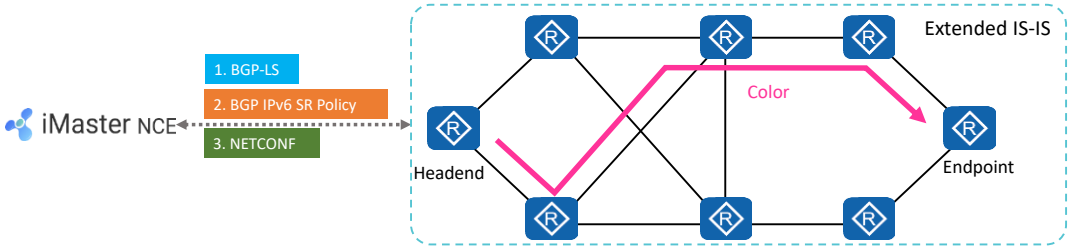
SR Policy P1 <headend, color, endpoint>
Candidate-path CP1 <protocol, origin, discriminator>
Preference 200
Weight W1, SID-List1 <SID11...SID1i>
Weight W2, SID-List2 <SID21...SID2j>
Candidate-path CP2 <protocol, origin, discriminator>
Preference 100
Weight W3, SID-List3 <SID31...SID3i>
Weight W4, SID-List4 <SID41...SID4j>

HUAWEI

- SR Policy P1 is uniquely determined by the triplet <headend, color, endpoint>.
- An SR Policy can contain multiple candidate paths (e.g. CP1 and CP2). Each of the paths is uniquely determined by the triplet <protocol, origin, discriminator>.
- CP1 is the primary path because it is valid and has the highest preference. The two SID lists of CP1 are delivered to the forwarder, and traffic is balanced between the two paths based on weights. For the SID list <SID11...SID1i>, traffic is balanced according to W1/(W1+W2).

Overview of SRv6 Policy Path Establishment

- In Huawei's SRv6 Policy solution architecture, the controller (iMaster NCE-IP) is used to establish SRv6 Policy paths.
  - The controller uses BGP-LS to collect topology information collected through an extended IGP to compute SRv6 Policy paths and display state information.
  - The controller uses BGP IPv6 SR Policy to deliver SRv6 Policy information (e.g. headend, color, and endpoint) to the headend.
- Huawei's SRv6 Policy solution also uses NETCONF to deliver other configurations, such as service interfaces and route-policies (with the color attribute).
- In addition to delivering SRv6 Policies through iMaster NCE-IP, you can also manually deploy SRv6 Policies.

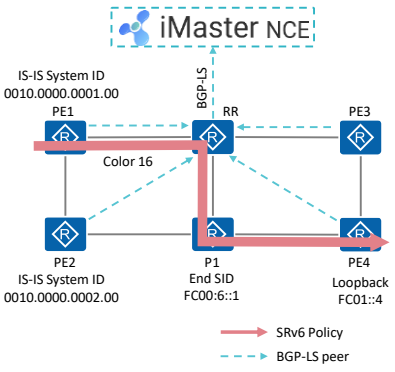


HUAWEI

- Mainstream methods for SRv6 Policy implementation:
  - BGP: BGP-LS is used to collect topology information, so that no new interface protocol needs to be introduced for customer-developed controllers. BGP IPv6 SR Policy is used to deliver route information.
  - PCEP: is a mature southbound protocol used in SR-MPLS TE scenarios. However, the tunnel implementation models of vendors are different and cannot interwork, and the interaction process of PCEP is more complex than that of BGP. As such, BGP extension is recommended.
  - NETCONF/YANG: delivers tunnel paths to forwarders as configurations. This method is not recommended because it delivers configurations in essence and offers the poorest performance. In a comprehensive solution, NETCONF is used to deliver configurations other than tunnel configurations.
- Huawei devices mainly use BGP to deploy SRv6 Policies.

Information Collection Through BGP-LS

- In Huawei's SRv6 Policy solution architecture, BGP-LS mainly provides the following functions:
  - Advertises the topology, prefix, SRv6 locator, and SID information collected by IS-IS/OSPFv3 to iMaster NCE-IP through BGP-LS routes.
  - Advertises the SRv6 Policy status through BGP-LS routes.



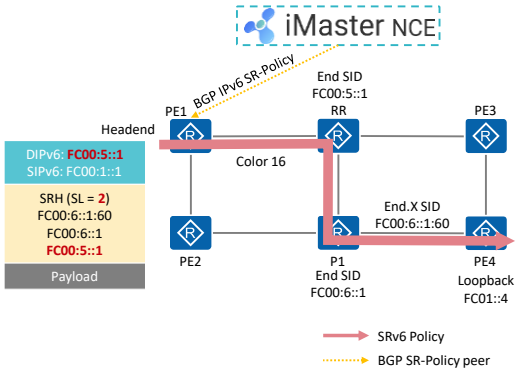

```
[~PE1]display bgp link-state unicast routing-table

Total Number of Node Routes: 36
 *>    Network  : [NODE][ISIS-LEVEL-2][IDENTIFIER0][LOCAL[as65001][bgp-ls-identifier1.0.0.1][ospf-
area-id0.0.0.0][igp-router-id0010.0000.0001.00]] //Topology node information

 *>    Network  : [LINK][ISIS-LEVEL-2][IDENTIFIER0][LOCAL[as65001][bgp-ls-identifier1.0.0.1][ospf-
area-id0.0.0.0][igp-router-id0010.0000.0001.00]][REMOTE[as65001][bgp-ls-identifier1.0.0.1][ospf-
area-id0.0.0.0][igp-router-id0010.0000.0002.00]][LINK[if-address10.0.0.26][peer-address10.0.0.25][if-
address::][peer-address::]] //Topology link information

 *>    Network  : [SRV6-SID][ISIS-LEVEL-2][IDENTIFIER0][LOCAL[as65001][bgp-ls-identifier1.0.0.1][ospf-
area-id0.0.0.0][igp-router-id0010.0000.0006.00]][SID[mt-id2][sidFC00:6::1]] //SRv6 SID information

 *>    Network  : [TEPOLICY][SEGMENT-ROUTING][IDENTIFIER0][LOCAL[as65001][bgp-ls-
identifier1.0.0.1][bgp-router-id1.0.0.1][ipv4-router-id0.0.0.0][ipv6-router-idFC01::1]][TE[protocol-
origin2][Flag128][endpointFC01::4][color16][originator-as65001][originator-
address172.21.17.102][discriminator49]] //SRv6 Policy status
```


HUAWEI

- BGP-LS connection:
  - Collects tunnel topology information for SR Policy path computation.
  - BGP-LS supports the collection of SR Policy status information, based on which the controller displays tunnel status. https://datatracker.ietf.org/doc/draft-ietf-idr-te-lsp-distribution/
  - BGP-LS supports SRLB information encapsulation and decapsulation, so that the controller can obtain the SRLB information for binding SID allocation. (The backup path of each SR Policy corresponds to a binding SID.)
  - BGP-LS also needs to be deployed on the headend to advertise the SRv6 Policy status.

Path Delivery Through BGP IPv6 SR Policy

- iMaster NCE-IP can deliver an SRv6 Policy to the specified headend through the BGP IPv6 SR Policy peer relationship.
- An SRv6 Policy is represented as [distinguisher][color][endpoint] in the BGP routing table.



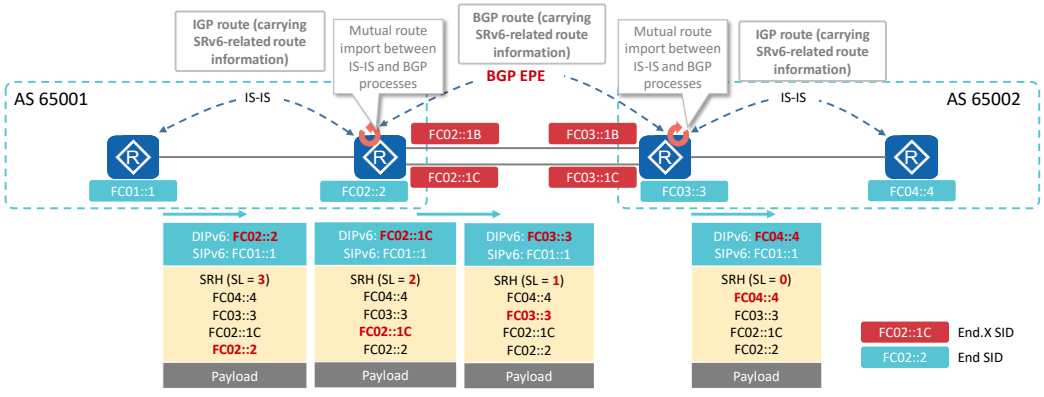

```
[~PE1]display bgp sr-policy ipv6 routing-table [49][16][FC01::4]

BGP local router ID : 1.0.0.1
Local AS number : 65001
Paths:   2 available, 1 best, 1 select, 0 best-external, 0 add-path
BGP routing table entry information of [49][16][FC01::4]:
...
Tunnel Encaps Attribute (23):
Tunnel Type: SR Policy (15)
 Preference: 10
 Binding SID:  FC00:1::1:1, s-flag(0), i-flag(0)
 Segment List
   Weight: 1
   Path MTU: 9600
   Segment: type:2, SID: FC00:5::1
   Segment: type:2, SID: FC00:6::1
   Segment: type:2, SID: FC00:6::1:60
 Template ID: 4294967278
Not advertised to any peer yet
```


HUAWEI

- BGP IPv6 SR Policy connection:
  - The controller delivers SRv6 Policy information to forwarders for SRv6 Policy generation.
  - BGP routes delivered by the controller carry the color community attribute, which can be transmitted. The headend finds a matching BGP route and recurses it to the corresponding SRv6 Policy based on the color and endpoint information.
  - In Huawei's SRv6 Policy solution, path computation constraints of each application need to be uniformly planned on the controller based on SLAs, and different colors are used to identify SRv6 Policies. An SRv6 Policy is uniquely identified by <headend, color, endpoint>. The BGP route of services to be steered into an SRv6 Policy needs to carry the corresponding color attribute.
- Binding SIDs are mainly used for path stitching.

Inter-AS SRv6 Policy Path Establishment

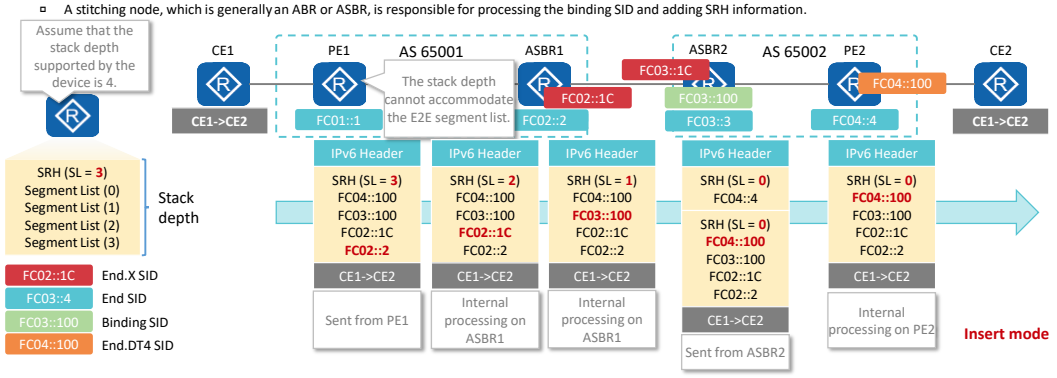
- SRv6 devices need to obtain network-wide SIDs to implement E2E data forwarding.
- In an AS, devices can use an extended IGP (extended OSPFv3 or IS-IS) to obtain intra-AS SID information. In inter-AS scenarios, however, BGP egress peer engineering (EPE) needs to be used to transmit SID information.



- BGP EPE can allocate BGP peer SIDs to inter-AS paths. Peer SIDs are classified into the following types:
  - A Peer-Node SID identifies a peer node. The peers at both ends of each BGP session are allocated with Peer-Node SIDs. An EBGP peer relationship established based on loopback interfaces may traverse multiple physical links. In this case, the Peer-Node SID of a peer is mapped to multiple outbound interfaces. Peer-Node SIDs are End SIDs.
  - A Peer-Adj SID identifies an adjacency to a peer. An EBGP peer relationship established based on loopback interfaces may traverse multiple physical links. In this case, each adjacency is allocated with a Peer-Adj SID. Only the specified link (mapped to the specified outbound interface) can be used for forwarding. Peer-Adj SIDs are End.X SIDs.
- BGP EPE allocates SIDs only to BGP peers and links, but cannot be used to construct a forwarding path. BGP peer SIDs must be used with IGP SIDs to form an E2E path.

SRv6 Policy Path Stitching (1)

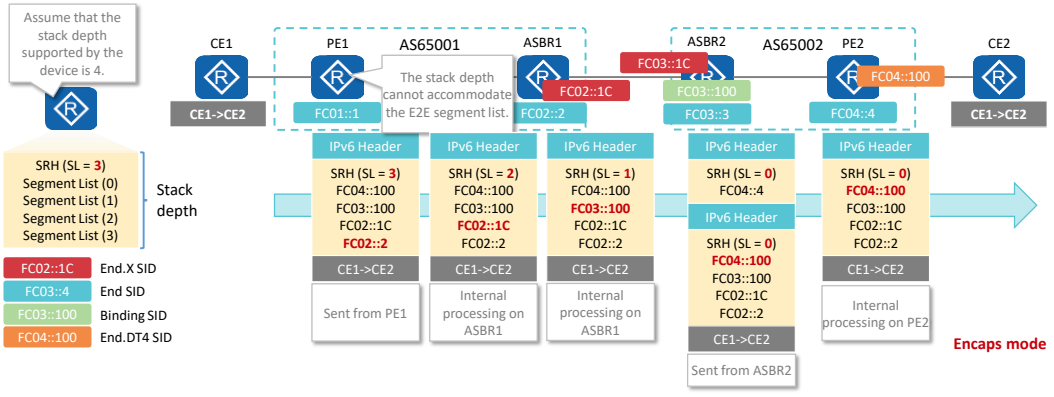
- Due to the restrictions on the stack depth in the SRv6 SRH, SRv6 path stitching can be deployed on large networks.
- Path stitching is mainly implemented using stitching SIDs and nodes.
  - A stitching SID (also called binding SID) can be used to represent an SRv6 Policy's forwarding path.
  - A stitching node, which is generally an ABR or ASBR, is responsible for processing the binding SID and adding SRH information.



- On a large network, the SRv6 SRH may be of a large size. Considering device limitations and forwarding efficiency, the number of SIDs in the SRH must be limited.
- Generally, there are two methods for reducing the SRH size:
  - SRv6 header compression
    - Huawei mainly uses the G-SRv6 solution for SRv6 header compression, reducing the SRH size and improving the forwarding efficiency without sacrificing SID information.
    - This course will not cover relevant header compression technologies.
  - SRv6 path stitching
    - Binding SIDs are used to stitch different SRv6 paths together, so that the SRH of each SRv6 path is not too large.
- The SRv6 stack depth is generally determined by device capabilities.
- SRv6 mainly supports the following four types of binding SIDs:
  - End.B6.Insert
  - End.B6.Insert.Red
  - End.B6.Encaps
  - End.B6.Encaps.Red

SRv6 Policy Path Stitching (2)

- Main encapsulation modes supported by SRv6 stitching labels:
  - Insert mode: inserts an SRH after the original IPv6 header so that the packet is forwarded based on the original IPv6 header and the new SRH.
  - Encaps mode: inserts an outer IPv6 header and SRH before the original IPv6 header so that the packet is forwarded based on the outer IPv6 header and SRH.

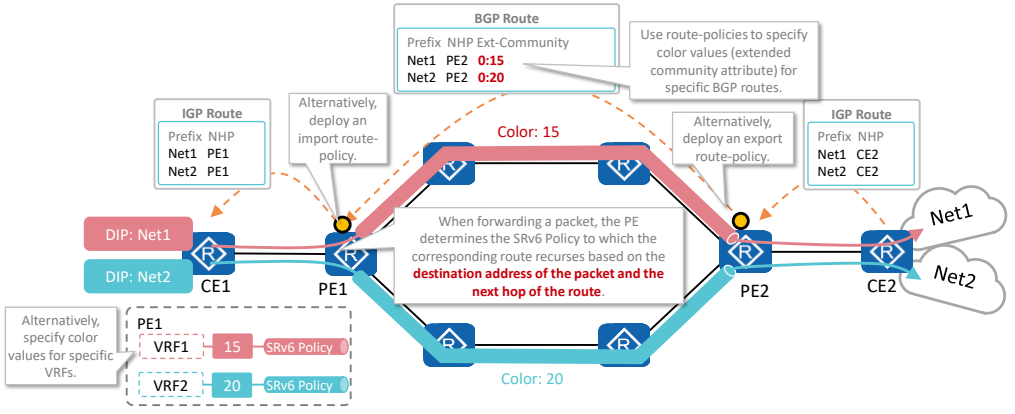


HUAWEI

- The End.B6.Encaps instruction used in Encaps mode can be disassembled into End + B6 + Encaps, where B6 indicates the application of an SRv6 Policy and Encaps indicates the encapsulation of an outer IPv6 header and SRH. This instruction includes the following operations: decrements the SL value of the inner SRH by 1, copies the SID to which the SL field is pointing to the DA field of the inner IPv6 header, encapsulates an IPv6 header and SRH (including segment lists), sets the source address to the address of the current node and the destination address to the first SID of the involved SRv6 Policy, sets other fields in the outer IPv6 header, looks up the corresponding table, and forwards the new IPv6 packet accordingly.

Color-based Traffic Steering to SRv6 Policies

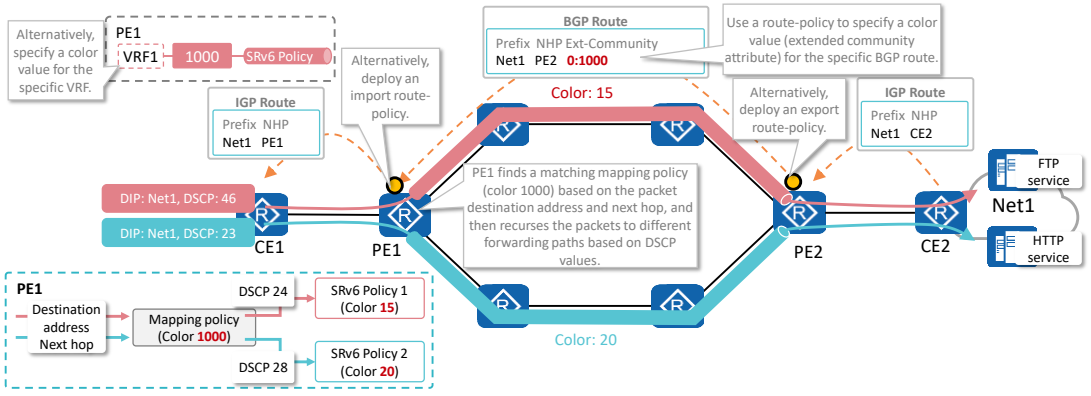
- SRv6 Policies support color-based traffic steering, enabling traffic to be steered to an SRv6 Policy through route recursion that is implemented based on the color value and destination address in the route.



- Configure a tunnel policy on PE1. After receiving the BGP routes (Net1 and Net2), PE1 recurses the routes to different SRv6 Policies based on the color values (0:15 and 0:20) and the next hop (PE2). Before forwarding packets to specified subnets (Net1 and Net2), PE1 adds specific SRv6 SID stacks to the packets.
- The color attribute in route entries can be modified before the local router (for example, PE2) sends routes or after the peer router (for example, PE1) receives routes.
- You can also directly configure the color attribute for the VPN instance of the originating router (for example, PE1), so that all traffic of the VPN instance is forwarded over the specified SRv6 Policy.

DSCP-based Traffic Steering to SRv6 Policies

- Color-based traffic steering causes different traffic (such as HTTP and FTP traffic) destined for the same address to recurse to the same forwarding path, leading to a failure to implement refined traffic control. In this case, DSCP-based traffic steering can be performed to recurse traffic to different forwarding paths.



HUAWEI

- In DSCP-based traffic steering, the color attribute in route entries is mainly used to find a matching mapping policy.
- The color attribute in route entries can be modified in the outbound direction of the originating router (for example, PE2) or in the inbound direction of the receiving router (for example, PE1).
- You can also directly configure the color attribute for the VPN instance of the originating router (for example, PE1), so that all traffic of the VPN instance is forwarded over the specified SRv6 Policy.

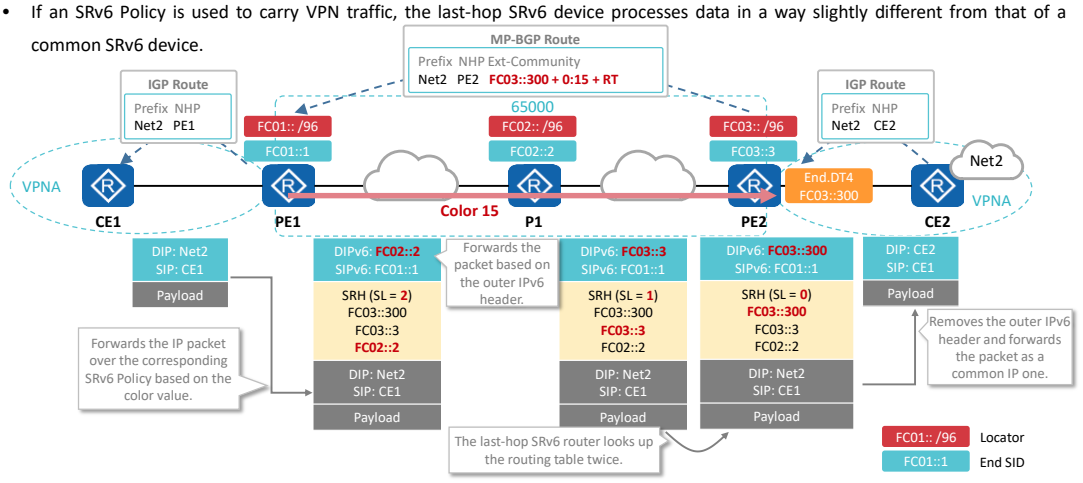
Contents

1. SRv6 Overview
2. **SRv6 Fundamentals**
   - Basic Concepts of SRv6
   - SRv6 Policy Path Establishment and Traffic Steering
   - Typical SRv6 Applications
3. SRv6 High Reliability
4. Basic SRv6 Configuration
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

L3VPN over SRv6 Policy

- If an SRv6 Policy is used to carry VPN traffic, the last-hop SRv6 device processes data in a way slightly different from that of a common SRv6 device.



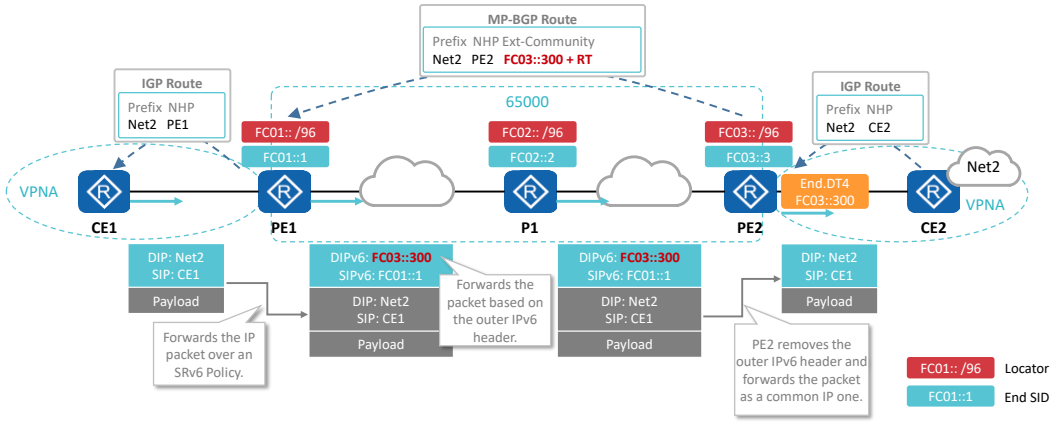
HUAWEI

- Route advertisement process:
  - SRv6 and SRv6 VPN are configured on each PE, and IPv6 or SRv6 is enabled on the transit node.
  - PE2 advertises an SRv6 locator route to PE1.
  - CE-to-PE route advertisement: CE2 advertises its route to PE2. Either a static route or a routing protocol (RIP, OSPFv3, IS-IS, or BGP) can be deployed between the CE and PE.
  - After learning the route advertised by CE2, PE2 installs it in the routing table of the corresponding VPN instance and converts the route into an MP-BGP one.
  - Inter-PE route advertisement: Configure a BGP or VPN export policy on PE2 (or a BGP or VPN import policy on PE1), and set a color value for the route (next hop: PE2). Then, configure PE2 to send routing information to PE1 through a BGP peer relationship using update messages with RT and SRv6 VPN SID attributes.
  - After PE1 receives the VPN route, if the next hop in the route is reachable and the route matches the BGP import policy, PE1 performs a series of actions to determine whether to install the route in the routing table of the corresponding VPN instance. These actions include VPN route leaking, route recursion to an SRv6 path, and route selection. In this example, PE1 installs the VPN route, which is associated with the SRv6 VPN SID.
  - PE-to-CE route advertisement: CE1 can learn the VPN route from PE1 through a static route or a routing protocol (RIP, OSPFv3, IS-IS, or BGP). The route advertisement process is similar to that from CE2 to PE2.

- Packet forwarding process:
  - After receiving a common unicast packet from CE1, PE1 searches the routing table of the corresponding VPN instance and finds that the outbound interface of the route is an SRv6 Policy interface. PE1 then inserts an SRH carrying the SID list of the SRv6 Policy and encapsulates an IPv6 header into the packet. After completing these operations, PE1 forwards the packet to P1.
  - The transit node P1 forwards the packet hop by hop based on SRH information.
  - After receiving the packet, the endpoint PE2 searches the My Local SID Table and finds an End SID that matches the IPv6 DA FC03::3 in the packet. According to the instruction bound to the SID, PE2 decrements the SL value of the packet by 1 and updates the IPv6 DA to the VPN SID FC03::300.
  - Based on the VPN SID FC03::300, PE2 searches the My Local SID Table and finds a matching End.DT4 SID. According to the instruction bound to the SID, PE2 decapsulates the packet, removes the SRH and IPv6 header, searches the routing table of the VPN instance corresponding to the VPN SID FC03::300 according to the DA in the inner packet, and forwards the packet to CE2.

L3VPN over SRv6 BE

- When SRv6 BE is used to carry VPN traffic, the P between PEs does not function as an endpoint.



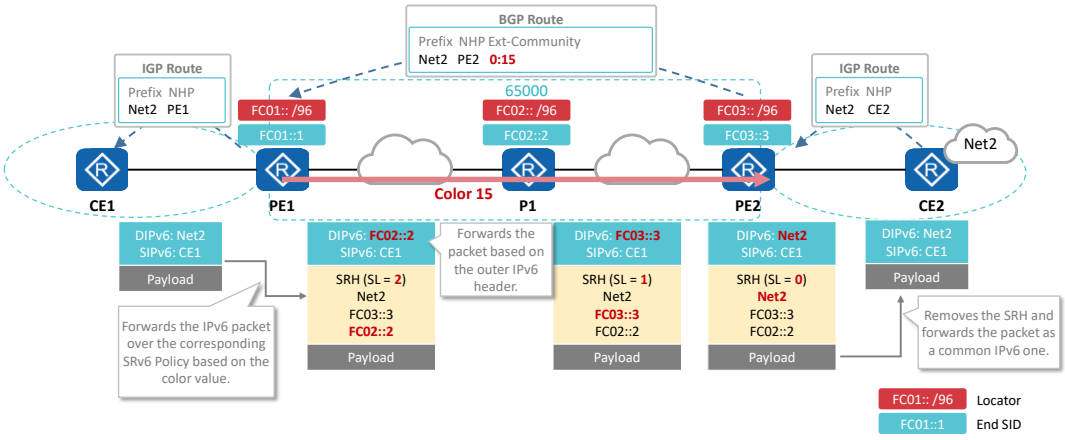
HUAWEI

- Route advertisement process:
  - SRv6 and SRv6 VPN are configured on each PE, and IPv6 is enabled on the transit node.
  - PE2 advertises an SRv6 locator route to PE1.
  - CE-to-PE route advertisement: CE2 advertises its route to PE2. Either a static route or a routing protocol (RIP, OSPFv3, IS-IS, or BGP) can be deployed between the CE and PE.
  - After learning the route advertised by CE2, PE2 installs it in the routing table of the corresponding VPN instance and converts the route into an MP-BGP one.
  - Inter-PE route advertisement: PE2 advertises the VPN route to egress node PE1 through MP-BGP using update messages with RT and SRv6 VPN SID attributes.
  - After PE1 receives the VPN route, if the next hop in the route is reachable and the route matches the BGP import policy, PE1 performs a series of actions to determine whether to install the route in the routing table of the corresponding VPN instance. These actions include VPN route leaking, route recursion to an SRv6 BE path, and route selection. In this example, PE1 installs the VPN route, which is associated with the SRv6 VPN SID.
  - PE-to-CE route advertisement: CE1 can learn the VPN route from PE1 through a static route or a routing protocol (RIP, OSPFv3, IS-IS, or BGP). The route advertisement process is similar to that from CE2 to PE2.

- Packet forwarding process:
    - After receiving a common unicast packet from CE1, PE1 searches the routing table of the corresponding VPN instance and finds that the outbound interface of the route is an SRv6 Policy interface. PE1 then inserts an SRH carrying the SID list of the SRv6 Policy and encapsulates an IPv6 header into the packet. After completing these operations, PE1 forwards the packet to P1.
    - The transit node P1 forwards the packet hop by hop based on SRH information.
    - After receiving the packet, the endpoint PE2 searches the My Local SID Table and finds an End SID that matches the IPv6 DA FC03::3 in the packet. According to the instruction bound to the SID, PE2 decrements the SL value of the packet by 1 and updates the IPv6 DA to the VPN SID FC03::300.
    - Based on the VPN SID FC03::300, PE2 searches the My Local SID Table and finds a matching End.DT4 SID. According to the instruction bound to the SID, PE2 decapsulates the packet, removes the SRH and IPv6 header, searches the routing table of the VPN instance corresponding to the VPN SID FC03::300 according to the DA in the inner packet, and forwards the packet to CE2.

Native IPv6 over SRv6 Policy

- Common IPv6 data can also be carried using SRv6.



HUAWEI

- Route advertisement process:
  - SRv6 and SRv6 VPN are configured on each PE, and IPv6 or SRv6 is enabled on the transit node.
  - PE2 advertises an SRv6 locator route to PE1.
  - CE-to-PE route advertisement: CE2 advertises its route to PE2. Either a static route or a routing protocol (RIP, OSPFv3, IS-IS, or BGP) can be deployed between the CE and PE.
  - Inter-PE route advertisement: Configure a BGP export policy on PE2 (or a BGP import policy on PE1), and set a color value for the route (next hop: PE2). Then, configure PE2 to send routing information to PE1 through a BGP peer relationship.
  - After PE1 receives the IPv6 route, if the next hop in the route is reachable and the route matches the BGP import policy, PE1 performs a series of actions, including route recursion to an SRv6 path and route selection.
  - PE-to-CE route advertisement: CE1 can learn the IPv6 route from PE1 through a static route or a routing protocol (RIP, OSPFv3, IS-IS, or BGP). The route advertisement process is similar to that from CE2 to PE2.

- Packet forwarding process:
    - After receiving a unicast IPv6 packet from CE1, PE1 searches the IPv6 routing table and finds that the outbound interface of the route is an SRv6 Policy interface. PE1 then inserts an SRH carrying the SID list of the SRv6 Policy and encapsulates an IPv6 header into the packet. After completing these operations, PE1 forwards the packet to P1.
    - The transit node P1 forwards the packet hop by hop based on SRH information.
    - After receiving the packet, the endpoint PE2 searches the My Local SID Table and finds an End SID that matches the IPv6 DA FC03::3 in the packet. According to the instruction bound to the SID, PE2 decrements the SL value of the packet by 1 and updates the IPv6 DA to the End SID Net2.
    - Based on the End SID Net2, PE2 searches the My Local SID table, finds a matching End SID, removes the SRH and IPv6 header, and forwards the packet to CE2.

Contents

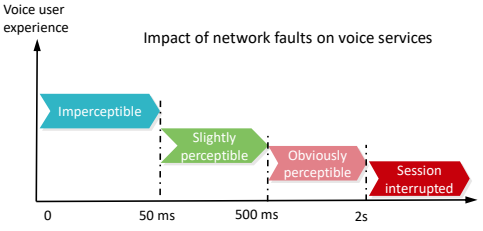
1. SRv6 Overview
2. SRv6 Fundamentals
3. **SRv6 High Reliability**
   - Overview of SRv6 High Reliability
   - SRv6 Tunnel Status Detection and Tunnel Protection
   - SRv6 Egress Protection and Access Protection
4. Basic SRv6 Configuration
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

WAN Bearer Networks' Requirements for High Reliability

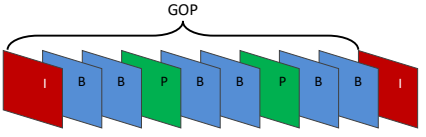
- Fault recovery within 50 ms has become a basic requirement of WAN bearer networks.
  - Voice services: have high requirements for real-time performance. If fault recovery is completed within milliseconds, the fault is imperceptible or only slightly perceptible to users, without much impacts on services. However, if the fault recovery time reaches seconds, the corresponding session is interrupted.
  - IPTV services: If fault recovery is completed within milliseconds, IPTV services may encounter transient pixelation.

Impact of network faults on voice services



Reference standard:
YD/T 1071-2000 <IP Telephone Gateway technical specification>

Impact of network faults on IPTV services

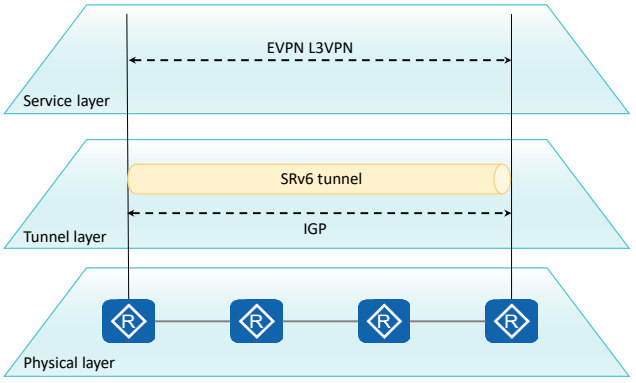


I-frame damage caused by packet loss is the key cause of erratic display.

HUAWEI

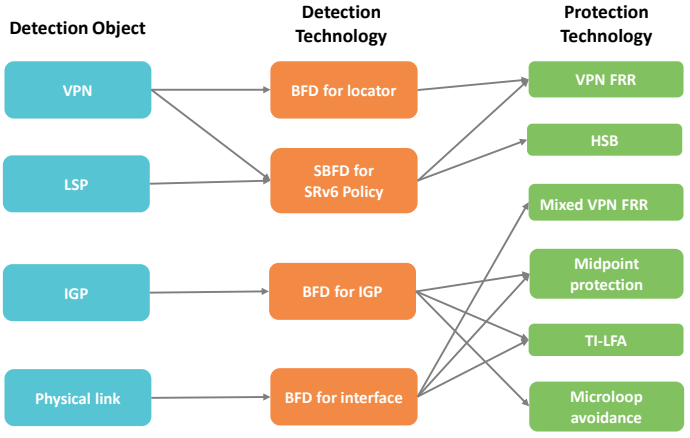
Overview of Multi-Layer Reliability Solutions

- WAN bearer networks require high reliability to be provided at device, network, and service layers to achieve E2E high availability of 99.999% and fast protection switching of all services within 50 ms.



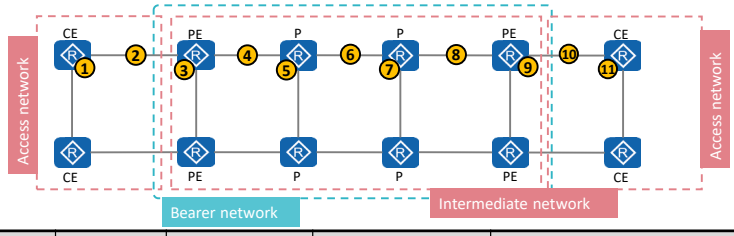
- VPN FRR
- TI-LFA, midpoint protection (TE FRR), and candidate path protection (HSB)
- Highly reliable hardware architecture: Key components, such as main control boards, SFUs, interface boards, power supply modules, and fans, adopt redundancy design.
- Highly reliable networking architecture: effective redundant paths, reliable networking mode (e.g. ring, spine-leaf, or dual-plane networking), and sufficient protection bandwidth reserved.

Overview of Reliability Technologies for Multi-Layer Networks



HUAWEI

Usage Scenarios of Reliability Technologies for Multi-Layer Networks



Service Category	Protection Type	Tunnel Type	Failure Point	Detection Technology	Protection Technology	Specification
Common services	Local protection, E2E protection	SRv6 BE, SRv6 Policy	4 to 8	BFD for interface	1. TI-LFA (SRv6 BE + SRv6 Policy) 2. Midpoint TI-LFA (SRv6 Policy)	50 ms
			3 and 9	BFD for locator	VPN FRR	50 ms
			1, 2, 10, and 11	BFD for interface	VPN mixed FRR, IP FRR	50 ms
Services with high SLA requirements	Local protection, E2E protection	SRv6 Policy	4 to 8	SBFD for SRv6 Policy	HSB	50 ms
			3 and 9	SBFD for SRv6 Policy	VPN FRR	200 ms/50 ms
			1, 2, 10, and 11	BFD for interface	VPN mixed FRR, IP FRR	50 ms

HUAWEI

Contents

1. SRv6 Overview
2. SRv6 Fundamentals
3. **SRv6 High Reliability**
   - Overview of SRv6 High Reliability
   - SRv6 Tunnel Status Detection and Tunnel Protection
   - SRv6 Egress Protection and Access Protection
4. Basic SRv6 Configuration
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

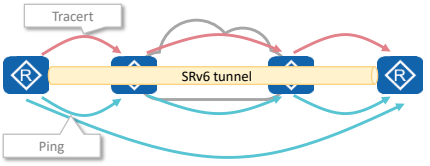
SRv6 Tunnel Detection Technologies

- SRv6 tunnel connectivity can be detected using multiple technologies, such as SBFD, ping, and tracert.

SBFD

- SBFD can be used to detect tunnel connectivity in an E2E manner.
- However, SBFD cannot detect the specific fault location on the network. As such, it is usually used with HSB or VPN FRR.

Ping/Tracert

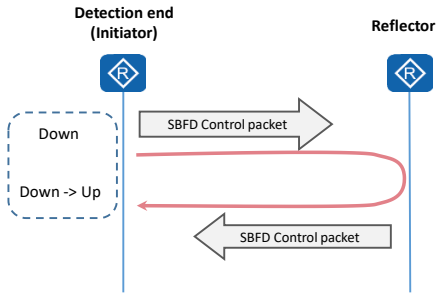


- SRv6 SID ping is mainly used to check network connectivity and host reachability.
  - Ping tests are classified into segment-by-segment tests and non-segment-by-segment tests.
- In addition to checking network connectivity and host reachability, SRv6 SID tracert can be used to analyze the specific fault location on the network.

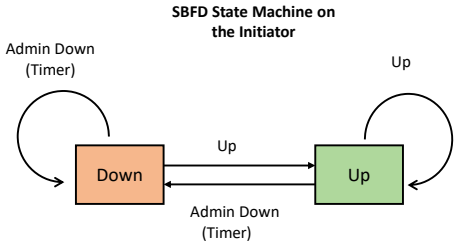
HUAWEI

- Seamless bidirectional forwarding detection (SBFD) is a simplified BFD mechanism. With a simplified BFD state machine, SBFD shortens the negotiation time and improves network-wide flexibility.
  - The IPv6 address of the SBFD reflector must be the same as the endpoint of the corresponding SRv6 Policy.
- As SRv6 simply adds a new type of routing extension header to implement forwarding based on the IPv6 data plane, ICMPv6 ping and tracert can be directly used on an SRv6 network for connectivity check based on common IPv6 addresses, without requiring any changes to hardware or software. ICMPv6 ping and tracert both support packet forwarding to a destination address over the shortest path, thereby checking the reachability to the destination. If the destination address is an SRv6 SID, the check can be performed through either ICMPv6 ping & tracert or SRv6 OAM extensions. Currently, SRv6 OAM can be extended in either of the following ways:
  - Set the O-bit (OAM bit) in the SRH.
  - Insert an End.OP SID into the SRH.

SBFD Implementation



- Before link detection, an SBFD initiator and an SBFD reflector exchange SBFD Control packets to notify each other of SBFD parameters (for example, discriminators). In link detection, the initiator proactively sends an SBFD packet, and the reflector loops this packet back. The initiator then determines the local status based on the looped packet.

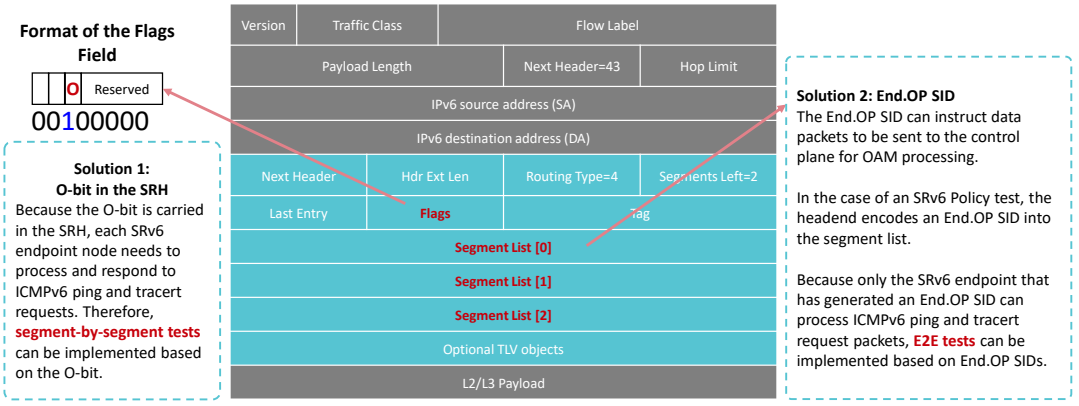


- The loopback packet constructed by the reflector carries the Admin Down or Up field.
- After receiving a reflected packet carrying the Up state, the initiator sets the local state to Up. After receiving a reflected packet carrying the Admin Down state, it sets the local state to Down. It also sets the local state to Down if it does not receive any reflected packet before the timer expires.

HUAWEI

- Because the state machine has only Up and Down states, the initiator can send packets carrying only the Up or Down state and receive packets carrying only the Up or Admin Down state. The initiator starts by sending an SBFD packet carrying the Down state to the reflector. The destination and source port numbers of the packet are 7784 and 4784, respectively; the destination IP address is a user-configured address on the 127 network segment; the source IP address is the locally configured LSR ID.
- The reflector does not have any SBFD state machine or detection mechanism. For this reason, it does not proactively send SBFD Echo packets, but rather, it only reflects SBFD packets. The destination and source port numbers in the looped-back packet are 4784 and 7784, respectively; the source IP address is the locally configured LSR ID; the destination IP address is the source IP address of the initiator.

Introduction to SRv6 Ping and Tracert

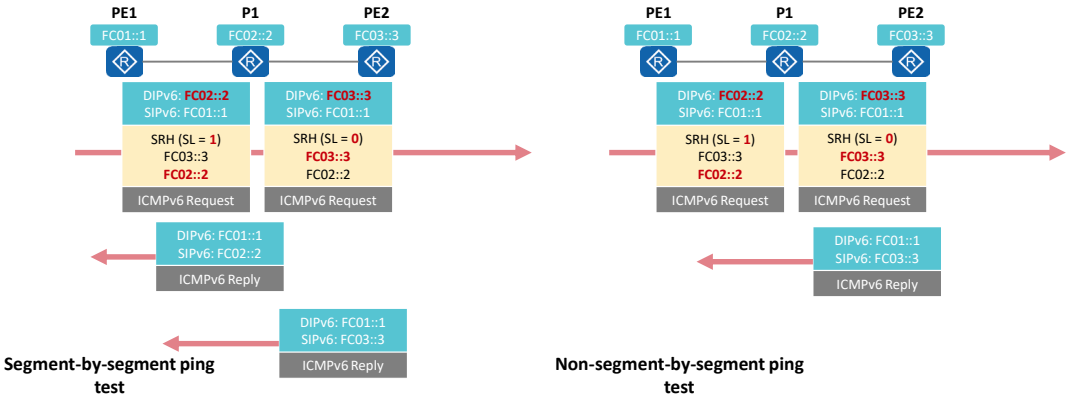


HUAWEI

- Currently, SRv6 ping and tracert can be implemented using the following two methods:
  - One method is to use the O-bit (OAM bit) in the SRH. Because the O-bit is carried in the SRH, each SRv6 endpoint node needs to process and respond to ICMPv6 ping and tracert requests. Therefore, segment-by-segment tests can be implemented based on the O-bit. You can run the **ping ipv6-sid** and **tracert ipv6-sid** commands to initiate tests based on one or more SIDs.
  - The other method is to introduce End.OP SIDs, which instruct data packets to be sent to the control plane for OAM processing. In the case of an SRv6 Policy test, the headend encodes an End.OP SID into the segment list. Because only the SRv6 endpoint that has generated an End.OP SID can process ICMPv6 ping and tracert request packets, E2E tests can be implemented based on End.OP SIDs.
    - For SID stack-based tests, specify one or more End.OP SIDs in the **ping ipv6-sid** and **tracert ipv6-sid** commands.
    - For SRv6 Policy-based tests, specify the **end-op** parameter in the **ping srv6-te policy** and **tracert srv6-te policy** commands.

SRv6 Ping Implementation

- SRv6 ping can be classified into segment-by-segment ping and non-segment-by-segment ping.

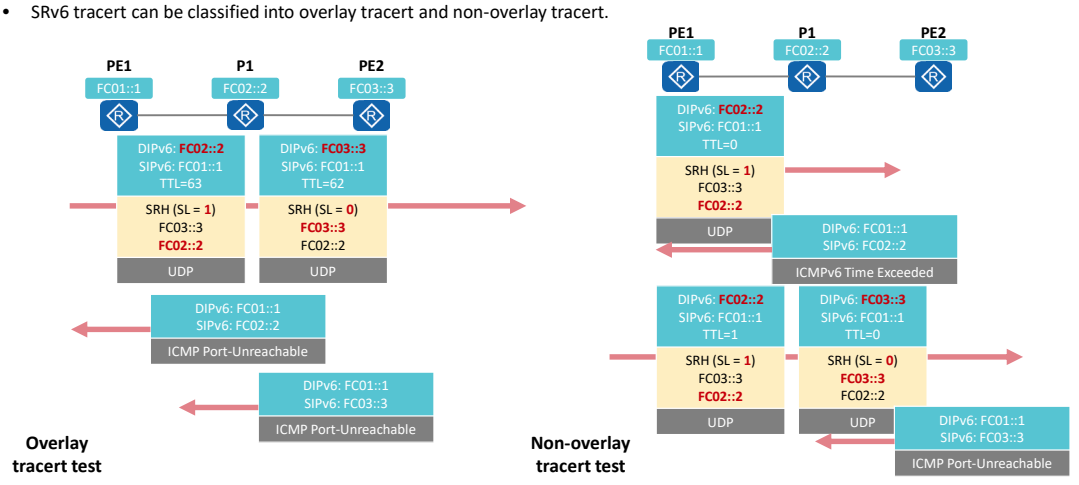


HUAWEI

- For a segment-by-segment test:
  - PE1 initiates a ping operation to PE2. Specifically, it constructs an ICMPv6 Request packet carrying the SRv6 SIDs (End and End.X SIDs) of P1 and the End SID of PE2 and then forwards the packet.
  - After receiving the ICMPv6 Request packet, P1 sends an ICMPv6 Reply packet to PE1 and forwards the ICMPv6 Request packet to PE2.
  - After receiving the ICMPv6 Request packet, PE2 sends an ICMPv6 Reply packet to PE1.
- For a non-segment-by-segment test:
  - PE1 initiates a ping operation to PE2. Specifically, it constructs an ICMPv6 Request packet carrying the End SID of PE2 and then forwards the packet.
  - After receiving the ICMPv6 Request packet, P1 forwards it to PE2.
  - After receiving the ICMPv6 Request packet, PE2 sends an ICMPv6 Reply packet to PE1. In this case, you can view detailed information about the ping operation on PE1.

SRv6 Tracert Implementation

- SRv6 tracert can be classified into overlay tracert and non-overlay tracert.



HUAWEI

- For an overlay test:
  - PE1 initiates a tracert operation to PE2. Specifically, it constructs a UDP packet carrying the SRv6 SIDs (End and End.X SIDs) of P1 and the End SID of PE2, encapsulates an SRH into the packet, and then forwards the packet. In this case, the value of the Hop Limit field (equal to the TTL field in IPv4) in the IPv6 header is set to 64 and decrements by 1 each time the packet passes through a device. When the value of the Hop Limit field is 0, the packet is discarded, and then an ICMPv6 Time Exceeded message is sent to PE1.
  - After receiving the UDP packet, P1 changes the value of the Hop Limit field to 63, sends an ICMPv6 Port Unreachable message to PE1, and forwards the UDP packet to PE2.
  - After receiving the UDP packet, PE2 changes the value of the Hop Limit field to 62 and sends an ICMPv6 Port Unreachable message to PE1.

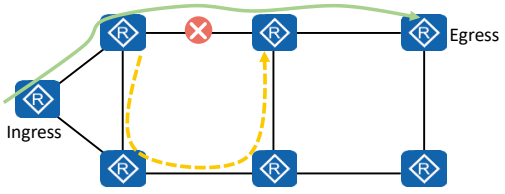
- For a non-overlay test:
    - PE1 initiates a tracert operation to PE2. Specifically, it constructs a UDP packet carrying the SRv6 SID of P1 and the End SID of PE2, encapsulates an SRH into the packet, and then forwards the packet. In this case, the value of the Hop Limit field in the IPv6 header is set to 1 and decrements by 1 each time the packet passes through a device. When the value of the Hop Limit field is 0, the packet is discarded, and then an ICMPv6 Time Exceeded message is sent to PE1.
    - After receiving the UDP packet, P1 changes the value of the Hop Limit field to 0 and sends an ICMPv6 Time Exceeded message to PE1.
    - After receiving the ICMPv6 Time Exceeded message from P1, PE1 increments the value of the Hop Limit field by 1 (the value now becomes 2) and continues to send the UDP packet.
    - After receiving the UDP packet, P1 changes the value of the Hop Limit field to 1 and forwards the packet to PE2.
    - After receiving the UDP packet, PE2 changes the value of the Hop Limit field to 0, determines that the SID type is End SID, and checks whether the upper-layer protocol header is a UDP or an ICMPv6 header.

Overview of Tunnel Protection Technologies

- SRv6 tunnel protection can be classified into local protection and E2E protection.

Local Protection

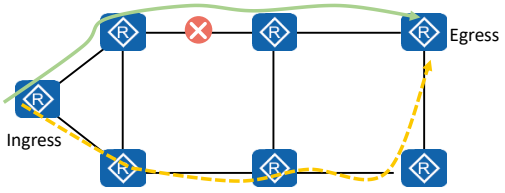
- Fast switching
- Only links and nodes protected



- TI-LFA FRR
- Midpoint TI-LFA FRR
- Microloop avoidance

E2E protection

- Detection-dependent fast switching
- E2E paths protected

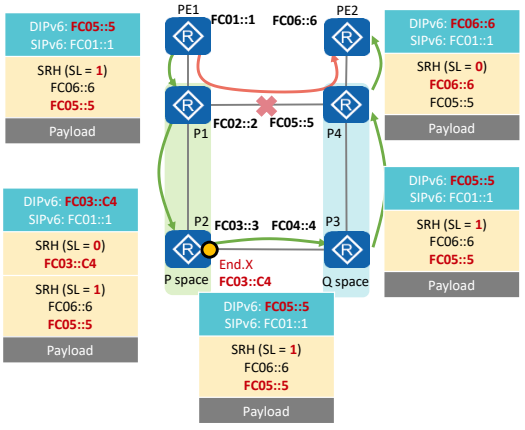


- HSB
- ECMP
- Best-effort path

HUAWEI

TI-LFA FRR

- TI-LFA FRR provides link and node protection for SRv6 tunnels. It enables traffic to be rapidly switched to the backup path if a link or node failure occurs.



- As shown in the figure, the shortest path from PE1 to PE2 is PE1 -> P1 -> P4 -> PE2, which is the primary path. P1 needs to compute a TI-LFA backup path to PE2 through the following operations:
  1. Excludes the primary next hop (link P1 -> P4) and computes the post-convergence shortest path: P1 -> P2 -> P3 -> P4 -> PE2.
  2. Computes the P space and Q space, which are (P1, P2) and (P3, P4, PE2), respectively.
  3. Computes the TI-LFA backup path. In this case, any path can be represented as a multi-segment path (source node <-> P <-> Q <-> destination node). Both the segments from the source node to the P node and from the Q node to the destination node are loop-free. The P-to-Q path is expressed using a strict explicit path (End.X SID), ensuring that the entire TI-LFA strict explicit path is loop-free. To simplify repair path computation, P2 (which is farthest from the source node and resides in the P space), P3 (which is farthest from the destination node and resides in the Q space), and a link between the P and Q spaces are selected.
  4. After detecting that the P1-to-PE2 link goes down, P1 uses backup forwarding entries and encapsulates a new SRH into the packet, with the segment list being <FC03::C4, FC06::6>. In addition, the node changes the IPv6 destination address to FC03::C4 and then forwards the packet to the backup outbound interface in the B-to-C direction.

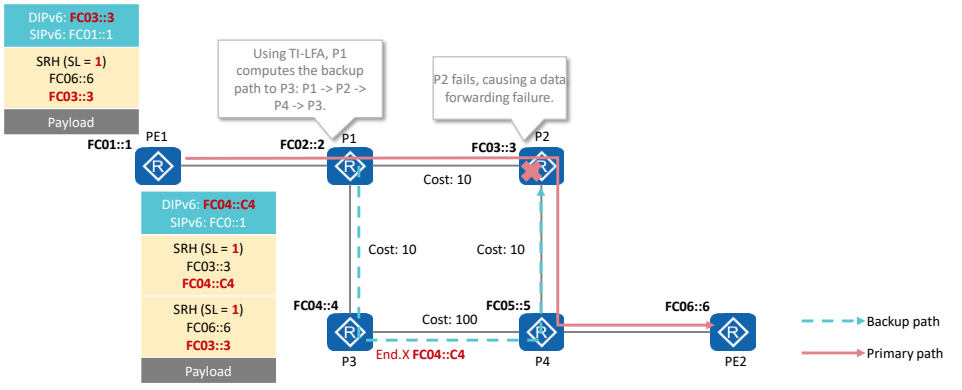
HUAWEI

- P space and Q space:
  - P space: is a set of nodes reachable from the source node of a protected link using the shortest path tree (SPT) rooted at the source node without traversing the protected link.
  - Extended P space: is a set of nodes reachable from all the neighbors of a protected link's source node using SPTs rooted at the neighbors without traversing the protected link.
  - Q space: is a set of nodes reachable from the destination node of a protected link using the reverse SPT rooted at the destination node without traversing the protected link.
- Advantages of TI-LFA:
  - Provides protection for all topologies, preventing cost planning from affecting protection path computation.
  - Simplifies deployments. Backup paths are computed based on an IGP, eliminating the need to deploy additional protocols for reliability purposes.
  - Preferentially uses the post-convergence path as the backup path, reducing the number of path switchovers and facilitating bandwidth planning.

- TI-LFA computation rules
  - Priority: SRLG disjoint > Node protection > Link protection > Minimum cost
  - TI-LFA computes a backup path that meets both the SRLG disjoint and node protection conditions. If multiple backup paths meet the two conditions, TI-LFA selects the path with the minimum cost.
  - If no qualified backup path is available, TI-LFA computes a backup path that meets both the SRLG disjoint and link protection conditions. If multiple backup paths meet the two conditions, TI-LFA selects the path with the minimum cost.
  - If no qualified backup path is available, TI-LFA computes a backup path that meets the node protection condition with the minimum cost.
  - If no qualified backup path is available, TI-LFA computes a backup path that meets the link protection condition with the minimum cost.

Limitations of TI-LFA FRR

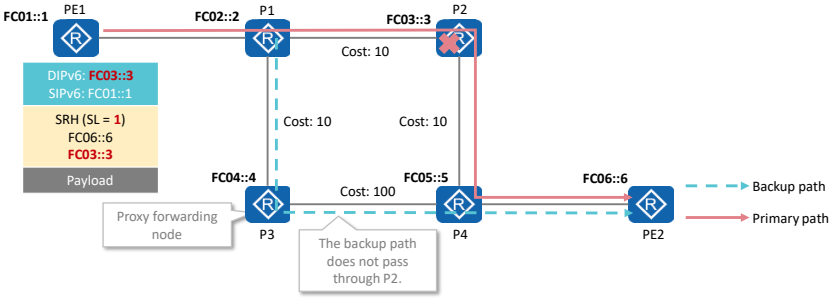
- In SRv6 Policy scenarios, the forwarding path of data packets usually needs to be constrained by specifying the nodes or links they need to traverse. However, if some nodes or links that must be traversed are faulty, TI-LFA FRR cannot provide protection.



HUAWEI

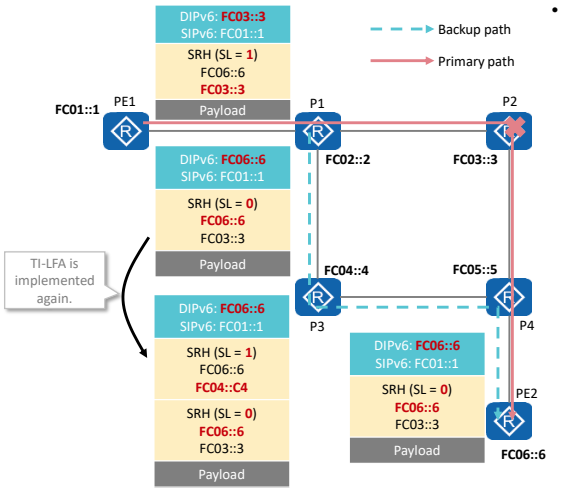
Overview of Midpoint Protection

- In SRv6 Policy scenarios, strict node constraints may result in a TI-LFA FRR protection failure. To resolve this issue, a proxy forwarding node (a node upstream to the failed midpoint) takes over from the failed midpoint to complete the forwarding.
- Specifically, after detecting that the next-hop interface of the packet fails, the next-hop address is the destination address of the packet, and the SL value is greater than 0, the proxy forwarding node performs the End behavior on behalf of the midpoint. The behavior involves decrementing the SL value by 1, copying the next SID to the DA field in the outer IPv6 header, and then forwarding the packet according to the instruction bound to the SID. In this way, the failed midpoint is bypassed, achieving SRv6 midpoint protection.



HUAWEI

Midpoint Protection Implementation

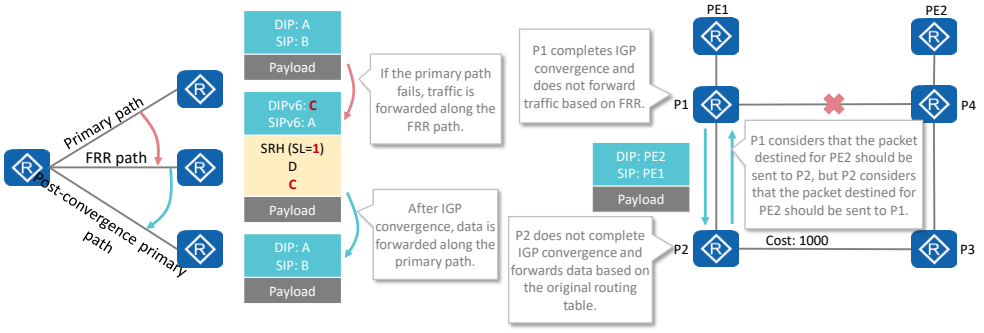


- Midpoint protection is implemented as follows:
  - Typically, P1 computes the shortest IGP path and TI-LFA backup path to each node, such as PE2.
  - After midpoint TI-LFA is configured for P1 and P2 fails, P1 detects that the outbound interface of P2 is faulty. The destination address of the packet is P2, which is a neighbor directly connected to P1. In addition, the SL value in the SRH is greater than 0. In this case, the midpoint TI-LFA action is triggered.
  - P1 performs proxy forwarding. Specifically, it decrements the SL value by 1, copies the next SID FC06::6 to the IPv6 destination address, and forwards the traffic to PE2 based on the destination address. In this way, the traffic can successfully reach PE2 through the P1 -> P3 -> P4 path.
  - If P1 finds that the next hop to PE2 is still P2 after performing proxy forwarding, P1 forwards the traffic to the TI-LFA backup path. In other words, after implementing midpoint TI-LFA, P1 implements basic TI-LFA processing to enable the traffic to be switched to the correct backup path.
  - After completing IGP convergence, PE1 deletes the FIB entries destined for P2. In this case, if iMaster NCE-IP delivers the segment list of a new policy, PE2 also performs proxy forwarding: decrementing the SL value by 1, using the next SID FC06::6 as the IPv6 destination address, and forwarding the traffic to PE2.

HUAWEI

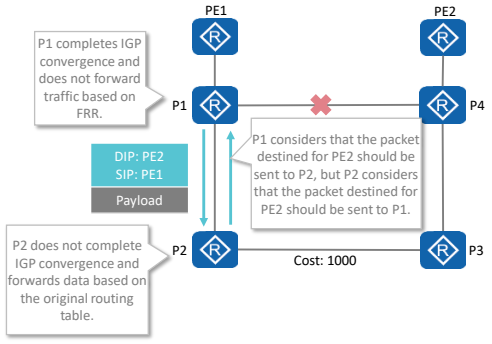
Microloop Introduction

- TI-LFA FRR and midpoint protection can maintain data forwarding for a short time before IGP convergence is complete. After IGP convergence, however, data is forwarded through IGP routes instead of FRR (tunnel mode).
- However, the convergence speed of devices on the live network may be different. As a result, a temporary loop, which is called a microloop, may be generated. The loop disappears only after all routers on the forwarding path complete convergence.



HUAWEI

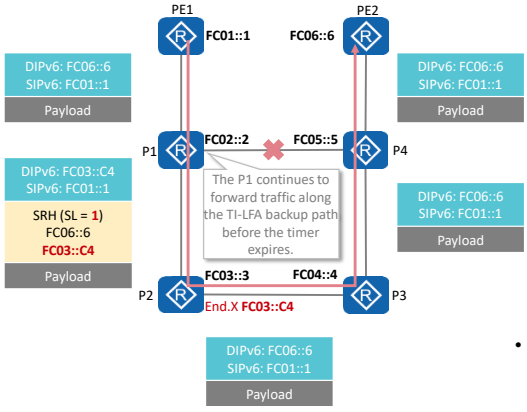
Local Microloop in a Traffic Switchover Scenario



- In a traffic switchover scenario, a local microloop occurs when a node adjacent to the failed node converges earlier than the other nodes on the network.
  - When the link between P1 and P4 fails, traffic is first forwarded along the TI-LFA backup path.
  - After P1 completes convergence, the next hop becomes P2, so that P1 forwards the traffic to P2 along the post-convergence path.
  - Because P2 and other nodes on the network have not converged, the next hop for the traffic from P2 to PE2 is still P1. In this case, a loop is formed between P1 and P2 until P2 completes convergence.

HUAWEI

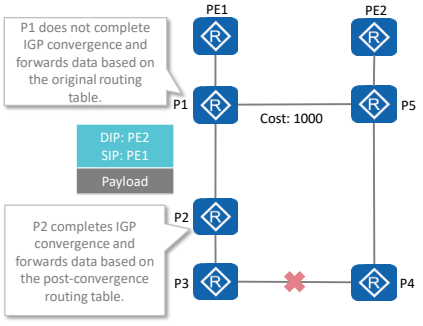
Local Microloop Avoidance in a Traffic Switchover Scenario



- The microloop avoidance measure is to deliver the corresponding forwarding path after the involved node has already completed convergence for a period of time. This prevents the loop caused by IGP convergence on the node adjacent to the failed node. The detailed process is as follows:
  - After P1 detects the fault of the link connected to P4, it enters the TI-LFA process and forwards the packet to PE2 along the TI-LFA backup path.
  - P1 starts a timer. Before the timer expires, the forwarding table retains unchanged, and the TI-LFA backup path continues to be used for packet forwarding.
  - When the timer expires, other nodes on the network have completed convergence. P1 can now perform convergence and then forward the packet along the post-convergence path.
- Because each TI-LFA backup path is loop-free, the packet can be forwarded along a TI-LFA backup path for a period of time, and then the TI-LFA process can exit after the other nodes complete convergence.

HUAWEI

Remote Microloop in a Traffic Switchover Scenario

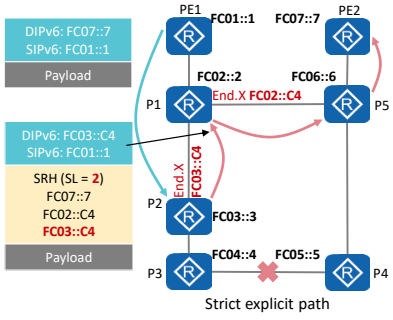


- If a remote fault occurs, a loop may also be formed between nodes.
- A loop may occur if the node closer to the failure point converges earlier than the node farther from the failure point on the packet forwarding path.
  - After detecting a fault on the P3 -> P4 link, a node (for example, P3) performs delayed convergence and forwards traffic along a TI-LFA backup path. In this case, only P3 performs delayed convergence, and the other nodes on the network still perform normal IGP convergence.
  - P2 converges first and computes the P2 -> P1 -> P5 -> PE2 path as the new path to PE2. As such, after arriving at P2, the traffic is sent back to P1 instead of being forwarded to P3.
  - Because P1 has not completed convergence, it still forwards the traffic along the P1 -> P2 -> P3 -> P4 -> P5 -> PE2 path. As a result, the traffic is sent back to P2, forming a loop.

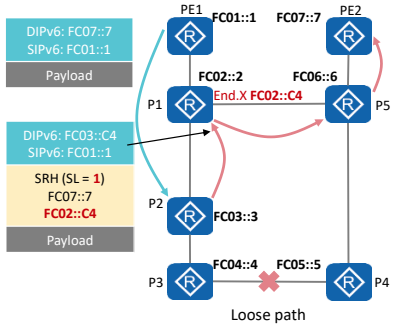
HUAWEI

Remote Microloop Avoidance in a Traffic Switchover Scenario

- A network node can pre-compute a loop-free backup path only when a directly connected link or node fails. That is, no loop-free path can be pre-computed against any other potential fault on the network. As such, a loop-free path needs to be computed after node convergence is completed, thereby resolving the microloop issue.
- As shown in the following figures, the loop-free TI-LFA path P2 -> P1 -> P5 -> PE2 is computed after P2 converges. The loop-free path computed by P3 can be either a strict explicit one or a loose one.



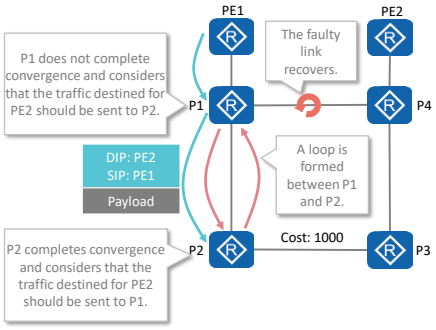
Strict explicit path



Loose path

HUAWEI

Microloop in a Traffic Switchback Scenario

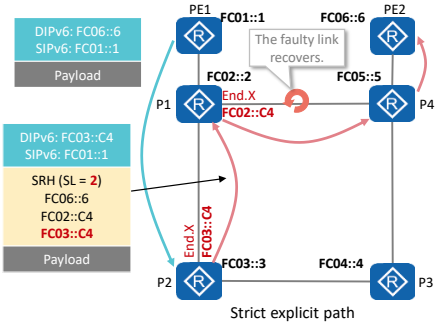


- Due to the lack of a convergence order in distributed computing, a microloop may occur in a traffic switchback scenario.
  - Assume that the link between P1 and P4 fails, and the path from PE1 to PE2 is PE1 -> P1 -> P2 -> P3 -> P4 -> PE2.
  - If the link between P1 and P4 recovers and P2 completes convergence, the path from P2 to PE2 is P2 -> P1 -> P4 -> PE2. In this case, P2 forwards traffic to P1.
  - If P1 has not completed IGP convergence due to some reasons, for example, a large number of services are running or CPU usage is high, the node still forwards the traffic back to P2 through the original path, forming a loop.

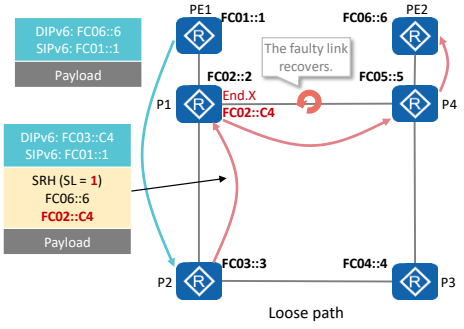
HUAWEI

Microloop Avoidance in a Traffic Switchback Scenario

- The method of microloop avoidance in a traffic switchback scenario is similar to that of remote microloop avoidance in a traffic switchover scenario.
- As shown in the following figures, after P2 completes convergence, it encapsulates a new SRH into the packet, enabling traffic to be forwarded along the post-convergence path (which is a strict explicit path) to the destination address. After P1 converges, it forwards the traffic to PE2 along the post-convergence path.



Strict explicit path



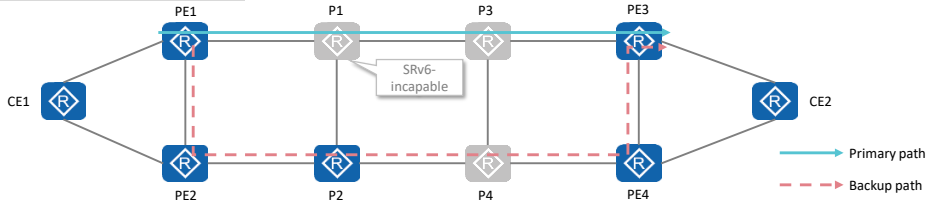
Loose path

HUAWEI

SRv6 Policy HSB

- E2E SRv6 Policy protection can be implemented by configuring two candidate paths with different preferences for an SRv6 Policy. The path with a higher preference is the primary and that with a lower preference is the backup. SBFD is enabled for the two paths.
- When SBFD on the primary path detects a fault, traffic is switched from the primary path to the backup path.
- If some devices on an SRv6 network do not support SRv6, local protection cannot be implemented. In this case, HSB can be used to provide high reliability.

```
SRv6 policy PE1toPE3
endpoint PE3 color 100
candidate-path preference 200 //Primary
  segment-list <PE1.End, PE3.End>
candidate-path preference 100 //Backup
  segment-list <PE2.End, P2.End, PE4.End, PE3.End >
```

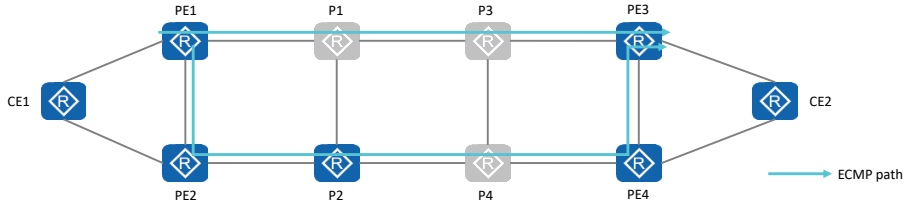


HUAWEI

SRv6 Policy ECMP

- During SRv6 Policy deployment, two segment lists can be configured in a candidate path to implement SRv6 Policy load balancing. In this way, data is load balanced between two paths.
- If one of the paths fails, data can still be forwarded over the other path.

```
SRv6 policy PE1toPE3
endpoint PE3 color 100
candidate-path preference 200
  segment-list <PE1.End, PE3.End> //ECMP
  segment-list <PE2.End, P2.End, PE4.End , PE3.End > //ECMP
```

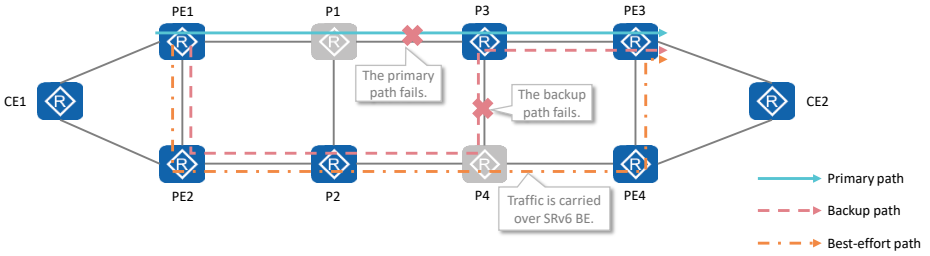


HUAWEI

Best-Effort Paths for SRv6 Policies

- If SBFD on PE1 detects that both the primary and backup paths fail, PE1 forwards traffic over an SRv6 BE tunnel.

```
BGP 100
  ipv6-family vpn-instance vpn1
    segment-routing ipv6 traffic-engineer best-effort
```



HUAWEI

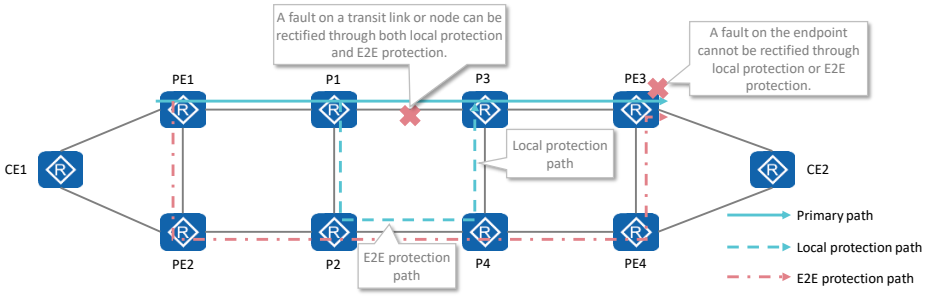
Contents

1. SRv6 Overview
2. SRv6 Fundamentals
3. **SRv6 High Reliability**
   - Overview of SRv6 High Reliability
   - SRv6 Tunnel Status Detection and Tunnel Protection
   - SRv6 Egress Protection and Access Protection
4. Basic SRv6 Configuration
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

Overview of SRv6 Egress Protection

- The tunnels of an SRv6 Policy have the same headend. In this case, the specific tunnel is determined using <Color, Endpoint>. If the endpoint of the tunnel fails, data cannot be sent to the destination network segment.
- As such, local protection and E2E protection technologies commonly used for SRv6 tunnels can protect only the source PE and transit nodes (Ps), but not the SRv6 endpoint.
- If a fault occurs on the endpoint, it is mainly rectified through VPN FRR. In addition, it can also be rectified through anycast FRR or mirror protection.

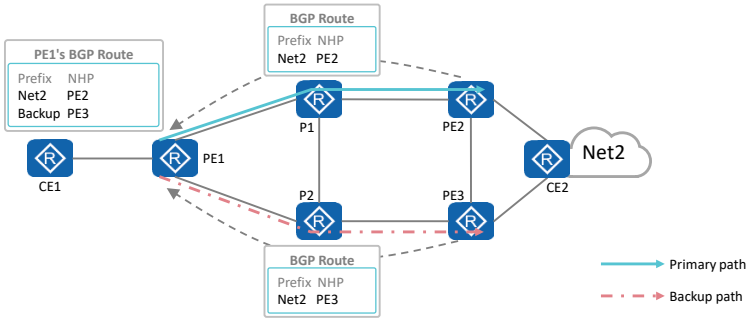


HUAWEI

- Anycast FRR and mirror protection technologies are complex and therefore rarely used on live networks.

VPN FRR

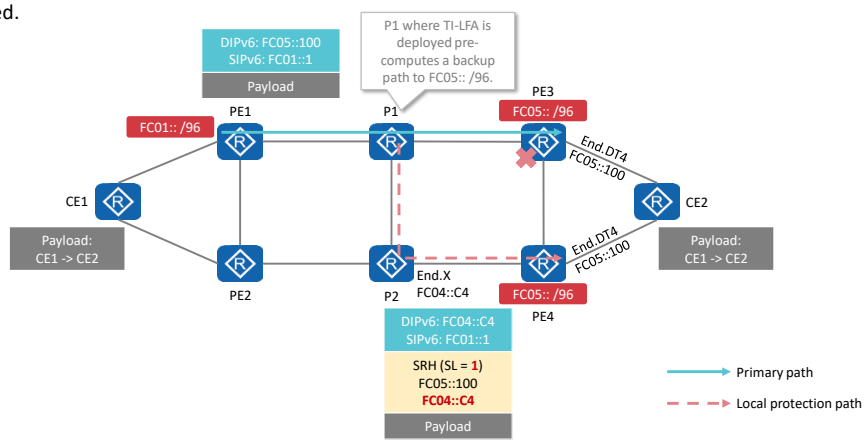
- VPN FRR helps rectify endpoint faults by directly forming VPN backup routes. It is implemented as follows:
  - The source PE pre-computes primary and backup routes based on the two learned VPN routes with different next-hop PEs and then delivers the computed routes to the FIB table. In addition, after detecting a remote PE fault through BFD, the source PE switches VPN traffic to the backup path before VPN route convergence.



HUAWEI

Anycast FRR

- Anycast FRR implements SRv6 egress protection by deploying the same locator and VPN SID on the PEs to which a CE is dual-homed.

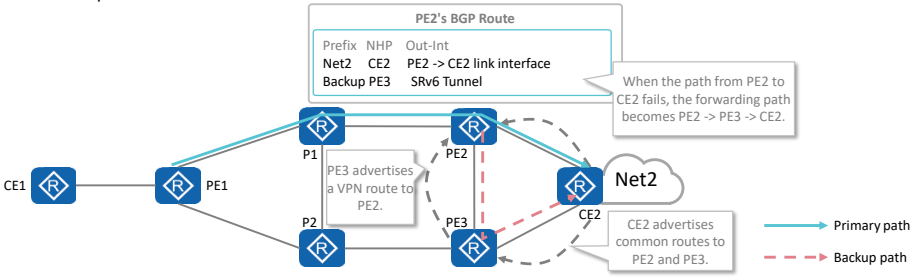


HUAWEI

- Anycast FRR can be used in both egress protection and local protection scenarios.
- Although anycast FRR can provide protection against PE failures, it has the following drawbacks:
  - VPN SIDs must be manually configured to ensure that the two PEs configured with the same VPN instance have the same VPN SID.
  - Only IGP route selection (not VPN route selection) can be performed. For example, if VPN services need to be load-balanced between PE3 and PE4 or the route advertised by PE3 needs to be preferentially selected, VPN route selection cannot be performed if the route advertised by PE4 is preferentially selected through an IGP on the path to FC05::.
  - If there is a PE-CE link interface failure, such as a failure on the link between PE3 and CE2, traffic is still forwarded to PE3 and then to PE4, resulting in a traffic loop that cannot be eliminated.

SRv6 Access Protection

- When a CE is dual-homed to PEs, if the link between the CE and endpoint PE fails, traffic may be lost. In this case, mixed FRR can be used to resolve this problem.



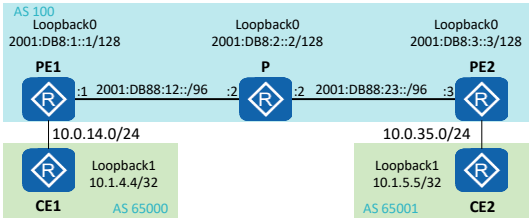
- PE2 receives a VPN route from CE2 and another VPN route from PE3, forming FRR protection. When the link between PE2 and CE2 fails, PE2 detects the fault and steers all relevant traffic to the backup path to PE3. In this case, the next hop of the primary path is an access interface and the backup path is an SRv6 tunnel, forming mixed FRR protection.

Contents

1. SRv6 Overview
2. SRv6 Fundamentals
3. SRv6 High Reliability
4. **Basic SRv6 Configuration**
   - SRv6 BE
   - SRv6 Policy
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

L3VPNv4 over SRv6 BE (1)



Networking requirements:

1. Connect PE1 and PE2 to different CEs that belong to VPN instance vpna.
2. Deploy L3VPN service recursion to SRv6 BE paths on the backbone network to enable CE1 and CE2 to communicate through Loopback1.

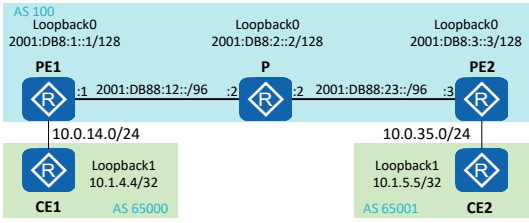
Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
6. Verify the configuration.

HUAWEI

- The experiment configuration is based on the NE20E-S2F (software version: NE20E V800R012C10SPC300). The configuration roadmaps for different devices are similar. For details, see the corresponding product documentation.

L3VPNv4 over SRv6 BE (2)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. **Establish an MP-BGP peer relationship between PE1 and PE2.**
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
6. Verify the configuration.

Establish an MP-IBGP peer relationship between the PEs. The configuration on PE1 is used as an example.

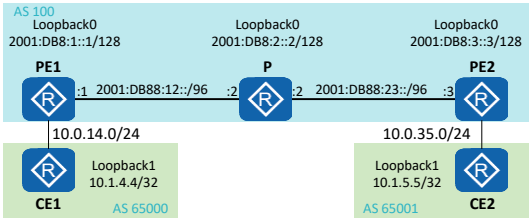
```
[~PE1] bgp 100
[~PE1-bgp] peer 2001:DB8:3::3 as-number 100
[*PE1-bgp] peer 2001:DB8:3::3 connect-interface loopback 0
[*PE1-bgp] ipv4-family vpnv4
[*PE1-bgp-af-vpnv4] peer 2001:DB8:3::3 enable
[*PE1-bgp-af-vpnv4] commit
[~PE1-bgp-af-vpnv4] quit
[~PE1-bgp] quit
```

Check the VPNv4 peer relationship on PE1.

```
<PE1>display bgp vpnv4 all 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  Up/Down      State
  2001:DB8:3::3   4    100        3        4     00:00:04  Established
```


HUAWEI

L3VPNv4 over SRv6 BE (3)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. **Enable SR and establish an SRv6 BE path on the backbone network.**
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
6. Verify the configuration.

Establish an SRv6 BE path between the PEs. PE1 configurations are as follows: (PE2 configurations are not provided here, and the P does not require such configurations.)

```
[~PE1] segment-routing ipv6
[*PE1-segment-routing-ipv6] encapsulation source-address 2001:DB8:1::1
[*PE1-segment-routing-ipv6] locator as100 ipv6-prefix 2001:DB8:100:: 64 static
32
[*PE1-segment-routing-ipv6-locator] quit
[*PE1-segment-routing-ipv6] quit
[*PE1] bgp 100
[*PE1-bgp] ipv4-family vpnv4
[*PE1-bgp-af-vpnv4] peer 2001:DB8:3::3 prefix-sid
[*PE1-bgp-af-vpnv4] quit
[~PE1-bgp] quit
[~PE1] isis 1
[~PE1-isis-1] segment-routing ipv6 locator as100
[*PE1-isis-1] commit
[~PE1-isis-1] quit
```

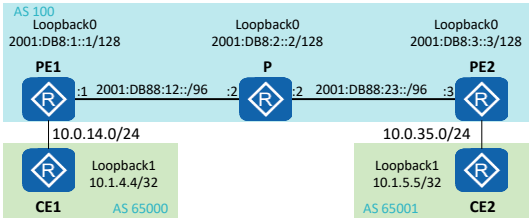

HUAWEI

- Configure basic SRv6 functions as follows:
  1. Run the **segment-routing ipv6** command to enable SRv6 and enter the SRv6 view.
  2. Run the **encapsulation source-address** *ipv6-address* [**ip-ttl** *ttl-value*] command to configure a source address for SRv6 VPN encapsulation.
     - When traffic enters an SRv6 VPN tunnel, the address configured using this command functions as the source address in the IPv6 header. The source address must be an existing interface address on the device.
  3. Run the **locator** *locator-name* [**ipv6-prefix** *ipv6-address prefix-length* [**static** *static-length* | **args** *args-length*] *] command to configure an SRv6 locator.
     - The Locator field corresponds to the **ipv6-prefix** *ipv6-address* parameter and its length is determined by the *prefix-length* parameter. A locator identifies an IPv6 subnet on which all IPv6 addresses can be allocated as SRv6 SIDs. After a locator is configured for a node, the system generates a locator route through which other nodes can locate this node. In addition, all SIDs advertised by the node are reachable through the route.
     - The Function field is also called opcode, which can be dynamically allocated using an IGP or be configured using the opcode command. When configuring a locator, you can use the **static** *static-length* parameter to specify the static segment length, which determines the number of static opcodes that can be configured in the locator. In dynamic opcode allocation, the IGP allocates opcodes outside the range of the static segment, so that no SRv6 SID conflict occurs.

4. (Optional) Run the **opcode** *func-opcode* **end-dt4 vpn-instance** *vpn-instance-name* command to configure a static SID opcode.
    - An End.DT4 SID can be dynamically allocated by BGP or be manually configured. If you want to use the **segment-routing ipv6 locator** *locator-name* command to enable dynamic End.DT4 SID allocation by BGP in the future, skip this step.
5. Run the **quit** command to exit the SRv6 locator view.
6. Run the **quit** command to exit the SRv6 view.

- Enable IS-IS SRv6.
    1. Run the **isis** [*process-id*] command to enter the IS-IS view.
    2. Run the **ipv6 enable topology ipv6** command to enable the IPv6 capability for the IS-IS process in the IPv6 topology.
    3. Run the **segment-routing ipv6 locator** *locator-name* [**auto-sid-disable**] command to enable IS-IS SRv6.
        - **segment-routing ipv6 locator** *locator-name*: enables IS-IS SRv6 and dynamic SID allocation. SIDs must be allocated within the locator range.
        - **segment-routing ipv6 locator** *locator-name* **auto-sid-disable**: enables IS-IS SRv6 and disables dynamic SID allocation so that SIDs are statically allocated within the locator range.
        - During SRv6 SID allocation, if a static opcode is configured, the static opcode is preferentially used to form a static SID. If no static opcode is configured, a SID is dynamically allocated. The process of dynamic SRv6 SID allocation using IS-IS is as follows:
        - Configure a locator in the SRv6 view, and run the **segment-routing ipv6 locator** command in the IS-IS process view to enable SRv6 and reference the locator. An IS-IS process can reference multiple locators. However, only one locator can be used to advertise End/End.X SIDs.
        - IS-IS allocates an End SID to each locator in the range of the dynamic segment based on locator configurations. Both PSP and non-PSP SIDs can be allocated.
        - Configure IPv6 addresses for interfaces and enable IS-IS IPv6. IS-IS allocates both PSP and non-PSP End.X SIDs to the interfaces in the range of the dynamic segment based on locator configurations.
    4. Run the **quit** command to exit the IS-IS view.
- Configure the device to add SIDs to VPN routes.
    - Run the **peer** *ipv6-address* **prefix-sid** command to enable the device to exchange IPv4 prefix SID information with the specified IPv6 peer.

L3VPNv4 over SRv6 BE (4)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. **Enable the VPN instance IPv4 address family on each PE.**
5. **Establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.**
6. Verify the configuration.

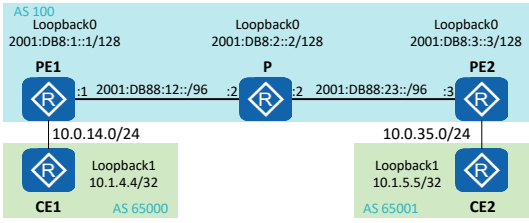
Enable the VPN instance IPv4 address family on each PE. PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] route-distinguisher 100:1
[*PE1-vpn-instance-vpna-af-ipv4] vpn-target 111:1 both
[*PE1-vpn-instance-vpna-af-ipv4] quit
[*PE1-vpn-instance-vpna] quit
[~PE1] bgp 100
[*PE1-bgp] ipv4-family vpn-instance vpna
[*PE1-bgp-vpna] peer 10.0.14.4 as-number 65000
[*PE1-bgp-vpna] segment-routing ipv6 best-effort
[*PE1-bgp-vpna] segment-routing ipv6 locator as100
[*PE1-bgp-vpna] commit
[~PE1-bgp-vpna] quit
[~PE1-bgp] quit
```


HUAWEI

- Configure VPN routes to recurse to SRv6 BE paths based on the carried SIDs.
  1. Run the **bgp** { *as-number-plain* | *as-number-dot* } command to enter the BGP view.
  2. Run the **ipv4-family vpn-instance** *vpn-instance-name* command to enter the BGP-VPN instance IPv4 address family view.
  3. Run the **segment-routing ipv6 best-effort** command to enable VPN route recursion based on the SIDs carried by routes.
  4. Run the **segment-routing ipv6 locator** *locator-name* [**auto-sid-disable**] command to enable the device to add SIDs to VPN routes.
  5. If **auto-sid-disable** is not specified, dynamic SID allocation is supported. If there are static SIDs in the range of the locator specified using *locator-name*, the static SIDs are used. Otherwise, dynamically allocated SIDs are used. If **auto-sid-disable** is specified, BGP does not dynamically allocate SIDs.
  6. Run the **commit** command to commit the configuration.

L3VPNv4 over SRv6 BE (5)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
6. **Verify the configuration.**

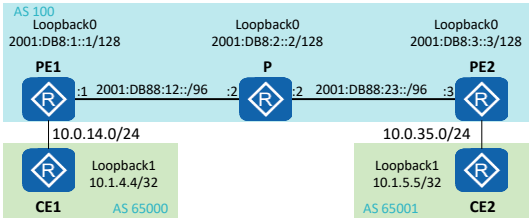
Check the local SID table containing all types of SRv6 SIDs on PE2.

```
<PE2>display segment-routing ipv6 local-sid forwarding

                My Local-SID Forwarding Table
                -------------------------------------

SID      : 2001:DB8:300::1:0:0/128              FuncType : End
LocatorName: as100                              LocatorID: 2

SID      : 2001:DB8:300::1:0:1/128              FuncType : End
LocatorName: as100                              LocatorID: 2

SID      : 2001:DB8:300::1:0:2/128              FuncType : End.X
LocatorName: as100                              LocatorID: 2

SID      : 2001:DB8:300::1:0:3/128              FuncType : End.X
LocatorName: as100                              LocatorID: 2

SID      : 2001:DB8:300::1:0:20/128             FuncType : End.DT4
LocatorName: as100
```

PE2 locally generates an End.DT4 SID and advertises the SID to PE1.

HUAWEI

L3VPNv4 over SRv6 BE (6)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
6. **Verify the configuration.**

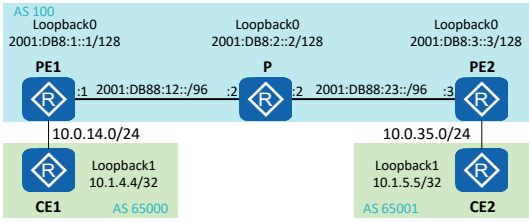
Check VPNv4 routing information on PE1.

```
<PE1>display bgp vpnv4 al routing-table 10.1.5.5
 BGP local router ID : 10.0.1.1
 Local AS number : 100
 Total routes of Route Distinguisher(100:1): 1
 BGP routing table entry information of 10.1.5.5/32:
 Label information (Received/Applied): 3/NULL
 From: 2001:DB8:3::3 (10.0.3.3)
 Route Duration: 0d00h15m54s
 Relay IP Nexthop: FE80::DE99:14FF:FE7A:C301
 Relay IP Out-Interface: GigabitEthernet0/3/0.12
 Relay Tunnel Out-Interface:
 Original nexthop: 2001:DB8:3::3
 Qos information : 0x0
 Ext-Community: RT <111 : 1>
 Prefix-sid: 2001:DB8:300::1:0:20
 AS-path Nil, origin incomplete, MED 0, localpref 100, pref-val 0, valid, internal,
best, select, pre 255, IGP cost 20
 Not advertised to any peer yet
```

IPv6 address of the peer; **SID corresponding to 10.1.5.5 (the same as that locally allocated on PE2)**

HUAWEI

L3VPNv4 over SRv6 BE (7)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
6. **Verify the configuration.**

Check vpna's routing information on PE1.

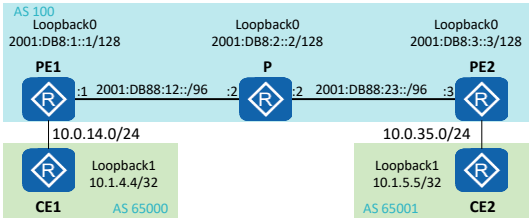
```
<PE1> display ip routing-table vpn-instance vpna 10.1.5.5 verbose
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - black hole
route
------------------------------------------------------------------------------
Routing Table : vpna
Summary Count : 1

Destination: 10.1.5.5/32
    Protocol: IBGP              Process ID: 0
  Preference: 255                     Cost: 0
     NextHop: 2001:DB8:300::1:0:20 Neighbour: 2001:DB8:3::3
       State: Active Adv Relied       Age: 00h17m40s
         Tag: 0                 Priority: low
       Label: 3                 QoSInfo: 0x0
  IndirectID: 0x1000177         Instance:
RelayNextHop: 2001:DB8:300::1:0:20 Interface: SRv6 BE
    TunnelID: 0x0               Flags: RD
```

Interface type: SRv6 BE

PE1 uses this IPv6 address as the next-hop address to forward packets destined for 10.1.5.5.

HUAWEI

L3VPNv4 over SRv6 BE (8)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
6. **Verify the configuration.**

Verify the configuration on CE1.

```
<CE1>ping -a 10.1.4.4 10.1.5.5
  PING 10.1.5.5: 56  data bytes, press CTRL_C to break
    Reply from 10.1.5.5: bytes=56 Sequence=1 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=2 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=3 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=4 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=5 ttl=254 time=1 ms
```

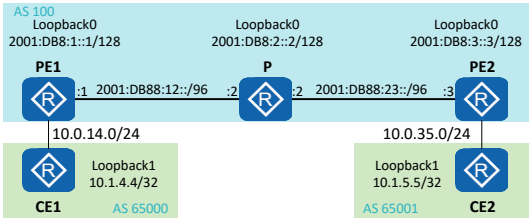

HUAWEI

Contents

1. SRv6 Overview
2. SRv6 Fundamentals
3. SRv6 High Reliability
4. **Basic SRv6 Configuration**
   - SRv6 BE
   - SRv6 Policy
5. SRv6 Deployment Using iMaster NCE-IP

HUAWEI

L3VPNv4 over SRv6 Policy (1)



Networking requirements:

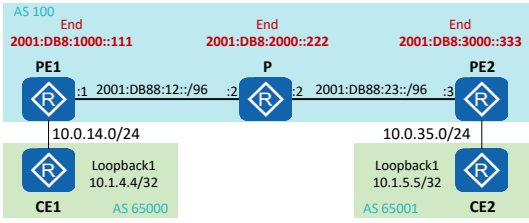
1. Connect PE1 and PE2 to different CEs that belong to VPN instance vpna.
2. Deploy L3VPN service recursion to SRv6 Policies on the backbone network to enable CE1 and CE2 to communicate through Loopback1.

Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 Policy on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
5. Configure a tunnel policy and import VPN traffic.
6. Verify the configuration.

HUAWEI

L3VPNv4 over SRv6 Policy (2)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. **Establish an MP-BGP peer relationship between PE1 and PE2.**
3. Enable SR and establish an SRv6 Policy on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
5. Configure a tunnel policy and import VPN traffic.
6. Verify the configuration.

Establish an MP-IBGP peer relationship between the PEs.

```
[~PE1] bgp 100
[~PE1-bgp] peer 2001:DB8:3::3 as-number 100
[*PE1-bgp] peer 2001:DB8:3::3 connect-interface loopback 0
[*PE1-bgp] ipv4-family vpnv4
[*PE1-bgp-af-vpnv4] peer 2001:DB8:3::3 enable
[*PE1-bgp-af-vpnv4] commit
[~PE1-bgp-af-vpnv4] quit
[~PE1-bgp] quit
```

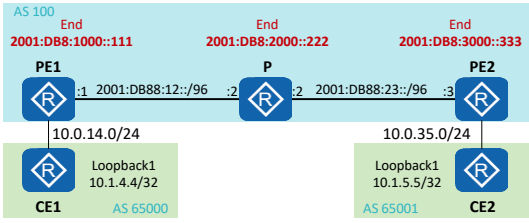
Check the VPNv4 peer relationship on PE1.

```
<PE1>display bgp vpnv4 all 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  Up/Down     State
 2001:DB8:3::3   4   100        3        4    00:00:04  Established
```


HUAWEI

- The SIDs of PE1, the P, and PE2 are 2001:DB8:1000::111, 2001:DB8:2000::222, and 2001:DB8:3000::333, respectively.
- In this experiment, the SRv6 Policy is established based on specified End SIDs.

L3VPNv4 over SRv6 Policy (3)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. **Enable SR and establish an SRv6 Policy on the backbone network.**
4. Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
5. Configure a tunnel policy and import VPN traffic.
6. Verify the configuration.

Configure an SRv6 SID. PE1 configurations are as follows: (P and PE2 configurations are not provided.)

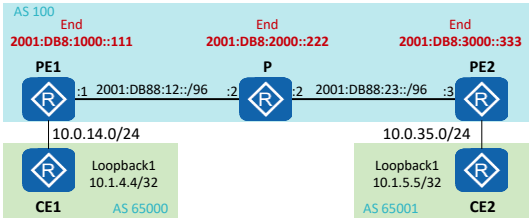
```
[~PE1] segment-routing ipv6
[*PE1-segment-routing-ipv6] encapsulation source-address 2001:DB8:1::1
[*PE1-segment-routing-ipv6] locator as1000 ipv6-prefix 2001:DB8:1000:: 64
static 32
[*PE1-segment-routing-ipv6-locator] opcode ::111 end
[*PE1-segment-routing-ipv6-locator] quit
[*PE1-segment-routing-ipv6] quit
[*PE1] bgp 100
[*PE1-bgp] ipv4-family vpnv4
[*PE1-bgp-af-vpnv4] peer 2001:DB8:3::3 prefix-sid
[*PE1-bgp-af-vpnv4] quit
[~PE1-bgp] quit
[~PE1] isis 1
[~PE1-isis-1] segment-routing ipv6 locator as1000 auto-sid-disable
[*PE1-isis-1] commit
[~PE1-isis-1] quit
```

Manually configure an SRv6 End SID.

HUAWEI

- SRv6 paths are established using SIDs. Static SRv6 SIDs are recommended. The configuration procedure is as follows:
  1. Run the **locator** *locator-name* [**ipv6-prefix** *ipv6-address prefix-length* [**static** *static-length* | **args** *args-length*] *] command to configure an SRv6 locator.
  2. Run the **opcode** *func-opcode* **end** command to configure a static End SID opcode.
  3. Run the **opcode** *func-opcode* **end-x interface** *interface-name* **nexthop** *nexthop-address* [**no-psp**] command to configure a static End.X SID opcode.
  4. Run the **quit** command to exit the SRv6 locator view.

L3VPNv4 over SRv6 Policy (4)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. **Enable SR and establish an SRv6 Policy on the backbone network.**
4. Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
5. Configure a tunnel policy and import VPN traffic.
6. Verify the configuration.

Configure an SRv6 Policy. PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~PE1] segment-routing ipv6
[~PE1-segment-routing-ipv6] segment-list list1
[*PE1-segment-routing-ipv6-segment-list-list1] index 5 sid ipv6
2001:DB8:2000::222
[*PE1-segment-routing-ipv6-segment-list-list1] index 10 sid ipv6
2001:DB8:3000::333
[*PE1-segment-routing-ipv6-segment-list-list1] commit
[~PE1-segment-routing-ipv6-segment-list-list1] quit
[~PE1-segment-routing-ipv6] srv6-te-policy locator as1000
[*PE1-segment-routing-ipv6] srv6-te policy policy1 endpoint 2001:DB8:3::3 color
101
[*PE1-segment-routing-ipv6-policy-policy1] binding-sid 2001:DB8:1000::100
[*PE1-segment-routing-ipv6-policy-policy1] candidate-path preference 100
[*PE1-segment-routing-ipv6-policy-policy1-path] segment-list list1
[*PE1-segment-routing-ipv6-policy-policy1-path] commit
[~PE1-segment-routing-ipv6-policy-policy1-path] quit
[~PE1-segment-routing-ipv6-policy-policy1] quit
[~PE1-segment-routing-ipv6] quit
```

Specify SRv6 End SIDs for the P and PE2 in sequence.

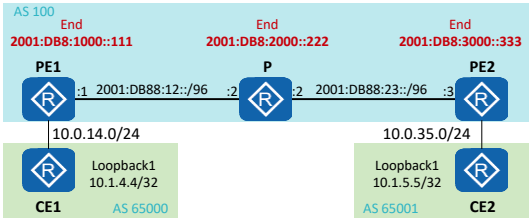
Specify the color and endpoint.

HUAWEI

- Configure a segment list.
  1. Run the **system-view** command to enter the system view.
  2. Run the **segment-routing ipv6** command to enable SRv6 and enter the SRv6 view.
  3. Run the **segment-list** *list-name* command to configure a segment list (an explicit path) for an SRv6 Policy candidate path and enter the segment list view.
  4. Run the **index** *index* **sid ipv6** *ipv6address* command to specify a next-hop SID for the segment list.
     - You can run the command multiple times. The system generates a SID stack for the segment list by **index** *index* in ascending order. If a candidate path in the SRv6 Policy is preferentially selected, traffic is forwarded using the segment lists of the candidate path. A maximum of 10 SIDs can be configured for each segment list.
  5. Run the **commit** command to commit the configuration.

- Configure an SRv6 Policy.
    1. Run the **system-view** command to enter the system view.
    2. Run the **segment-routing ipv6** command to enable SRv6 and enter the SRv6 view.
    3. Run the **srv6-te-policy locator** *locator-name* command to associate a locator with the SRv6 Policy to be created. This configuration allows you to specify a binding SID for the SRv6 Policy in the locator range.
    4. Run the **srv6-te policy** *name-value* **endpoint** *endpoint-ip* **color** *color-value* command to create an SRv6 Policy and enter the SRv6 Policy view.
    5. (Optional) Run the **binding-sid** *binding-sid* command to configure a binding SID for the SRv6 Policy.
        - The value of *binding-sid* must be within the range of the static segment specified using the **locator** *locator-name* [**ipv6-prefix** *ipv6-address prefix-length* [**static** *static-length* | **args** *args-length*] *] command.
    6. Run the **candidate-path preference** *preference* command to configure a candidate path and its preference for the SRv6 Policy.
        - Each SRv6 Policy supports multiple candidate paths. A larger preference value indicates a higher candidate path preference. If multiple candidate paths are configured, the one with the highest preference takes effect.
    7. Run the **segment-list** *list-name* [**weight** *weight-value* | **path-mtu** *mtu-value*] * command to configure a segment list for the candidate path of the SRv6 Policy.
        - The segment list must have been created using the **segment-list** (SRv6 view) command.
    8. Run the **commit** command to commit the configuration.

L3VPNv4 over SRv6 Policy (5)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 Policy on the backbone network.
4. **Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.**
5. Configure a tunnel policy and import VPN traffic.
6. Verify the configuration.

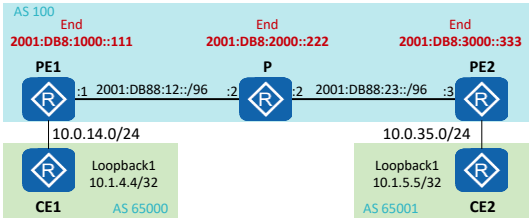
Enable the VPN instance IPv4 address family on each PE. PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] route-distinguisher 100:1
[*PE1-vpn-instance-vpna-af-ipv4] vpn-target 111:1 both
[*PE1-vpn-instance-vpna-af-ipv4] quit
[*PE1-vpn-instance-vpna] quit
[*PE1-bgp] ipv4-family vpn-instance vpna
[*PE1-bgp-vpna] segment-routing ipv6 traffic-engineer best-effort
[*PE1-bgp-vpna] segment-routing ipv6 locator as1000
[*PE1-bgp-vpna] commit
[~PE1-bgp-vpna] quit
[~PE1-bgp] quit
```


HUAWEI

- **segment-routing ipv6 traffic-engineer best-effort** //If an SRv6 BE path exists on the network, you can set the **best-effort** parameter, allowing the SRv6 BE path to function as a best-effort path in the case of an SRv6 Policy fault.

L3VPNv4 over SRv6 Policy (6)



Configuration roadmap:

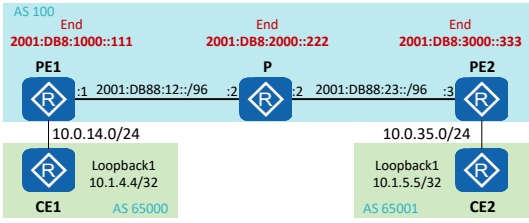
1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 Policy on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
5. **Configure a tunnel policy and import VPN traffic.**
6. Verify the configuration.

Configure a tunnel policy and import VPN traffic. PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~PE1] route-policy p1 permit node 10
[*PE1-route-policy] apply extcommunity color 0:101
[*PE1-route-policy] quit
[*PE1] bgp 100
[*PE1-bgp] ipv4-family vpnv4
[*PE1-bgp-af-vpnv4] peer 2001:DB8:3::3 route-policy p1 import
[*PE1-bgp-af-vpnv4] quit
[*PE1-bgp] quit
[*PE1] tunnel-policy p1
[*PE1-tunnel-policy-p1] tunnel select-seq ipv6 srv6-te-policy load-balance-number 1
[*PE1-tunnel-policy-p1] quit
[*PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] tnl-policy p1
[*PE1-vpn-instance-vpna-af-ipv4] commit
[~PE1-vpn-instance-vpna-af-ipv4] quit
[~PE1-vpn-instance-vpna] quit
```


HUAWEI

L3VPNv4 over SRv6 Policy (7)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 Policy on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
5. Configure a tunnel policy and import VPN traffic.
6. **Verify the configuration.**

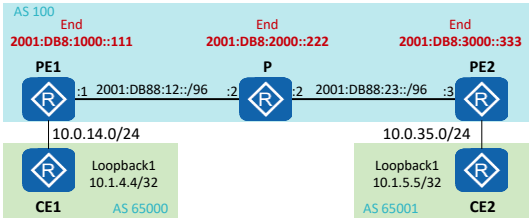
Check SRv6 Policy information on PE1.

```
<PE1>display srv6-te policy
PolicyName : policy1
Color                : 101          Endpoint          : 2001:DB8:3::3
TunnelId             : 1            Binding SID       : 2001:DB8:1000::100
TunnelType           : SRv6-TE Policy                 DelayTimerRemain    :
Policy State         : Up
Admin State          : UP                             Traffic Statistics  : Disable
Candidate-path Count : 1
 Candidate-path Preference : 100
 Path State          : Active                         Path Type           : Primary
 Protocol-Origin     : Configuration(30)              Originator          : 0, 0.0.0.0
 Discriminator       : 100                            Binding SID  : 2001:DB8:1000::100
 GroupId             : 1                              Policy Name         : policy1
 DelayTimerRemain    : -                              Segment-List Count  : 1
 Segment-List        : list1
  Segment-List ID    : 1                              XcIndex             : 1
  List State         : Up                             DelayTimerRemain    : -
  Weight             : 1                              BFD State           : -
  SID :
      2001:DB8:2000::222
      2001:DB8:3000::333
```

Color of the specified SRv6 Policy

HUAWEI

L3VPNv4 over SRv6 Policy (8)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 Policy on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
5. Configure a tunnel policy and import VPN traffic.
6. **Verify the configuration.**

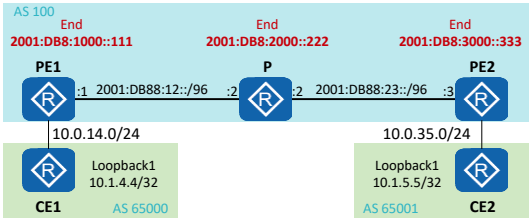
Check VPNv4 routing information on PE1.

```
<PE1> display bgp vpnv4 all routing-table 10.1.5.5
 BGP local router ID : 10.0.1.1
 Local AS number : 100
 Total routes of Route Distinguisher(100:1): 1
 BGP routing table entry information of 10.1.5.5/32:
 Label information (Received/Applied): 3/NULL
 From: 2001:DB8:3::3 (10.0.13.3)
 Route Duration: 0d00h03m30s
 Relay IP Nexthop: FE80::DE99:14FF:FE7A:C301
 Relay IP Out-Interface: GigabitEthernet0/3/0.12
 Relay Tunnel Out-Interface:
 Original nexthop: 2001:DB8:3::3
 Qos information : 0x0
 Ext-Community: RT <111 : 1>, Color <0 : 101>
 Prefix-sid: 2001:DB8:3000::1:0:1E
 AS-path 65000, origin incomplete, MED 0, localpref 100, pref-val 0, valid,
internal, best, select, pre 255, IGP cost 20
 Not advertised to any peer yet
```

The route recurses to the corresponding SRv6 Policy based on the color attribute.

HUAWEI

L3VPNv4 over SRv6 Policy (9)



Configuration roadmap:

1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 Policy on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE and establish an MP-IBGP peer relationship between the PEs for them to exchange routing information.
5. Configure a tunnel policy and import VPN traffic.
6. **Verify the configuration.**

Check vpna's routing information on PE1.

```
<PE1>display ip routing-table vpn-instance vpna 10.1.5.5 verbose
Routing Table : vpna
Summary Count : 1
Destination: 10.1.5.5/32
    Protocol: IBGP              Process ID: 0
  Preference: 255               Cost: 0
     NextHop: 2001:DB8:3::3     Neighbour: 2001:DB8:3::3
       State: Active Adv Relied      Age: 00h08m38s
         Tag: 0                 Priority: low
       Label: 3                 QoSInfo: 0x0
  IndirectID: 0x1000174         Instance:
RelayNextHop: ::                Interface: policy1
    TunnelID: 0x000000003400000001    Flags: RD
```

The outbound interface is an SRv6 Policy interface.

Verify the configuration on CE1.

```
<CE1>ping -a 10.1.4.4 10.1.5.5
  PING 10.1.5.5: 56  data bytes, press CTRL_C to break
    Reply from 10.1.5.5: bytes=56 Sequence=1 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=2 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=3 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=4 ttl=254 time=1 ms
    Reply from 10.1.5.5: bytes=56 Sequence=5 ttl=254 time=1 ms
```


HUAWEI

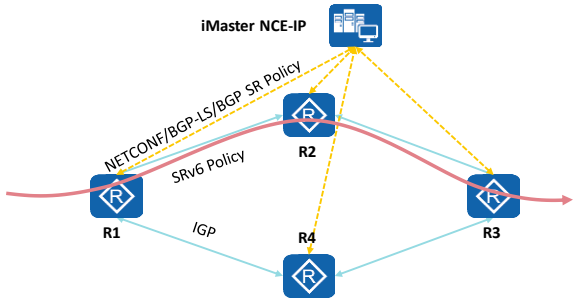
Contents

1. SRv6 Overview
2. SRv6 Fundamentals
3. SRv6 High Reliability
4. Basic SRv6 Configuration
5. **SRv6 Deployment Using iMaster NCE-IP**

HUAWEI

SRv6 Deployment Using iMaster NCE-IP

- Static SRv6 deployment on the live network poses great challenges to network O&M. For example, static SRv6 deployment takes a long period and does not support global path planning or network quality-based SRv6 forwarding path adjustment.
- Deploying SRv6 through iMaster NCE-IP can avoid the problems faced by static SRv6 deployment.
- iMaster NCE-IP uses the following protocols to deploy SRv6 tunnels and VPN services:
  - BGP-LS
  - BGP IPv6 SR Policy
  - NETCONF

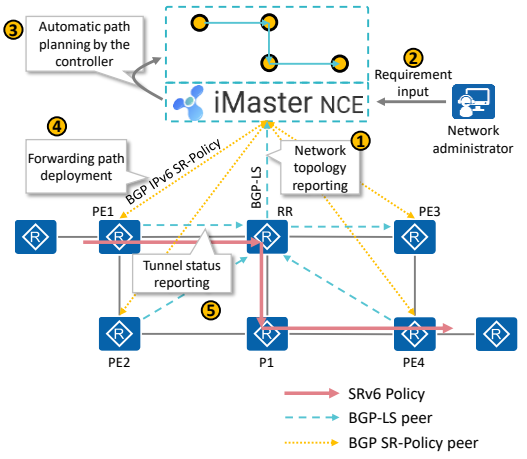


HUAWEI

- IGP: generates network topology information, such as bandwidth, delay, and SID information, on a router.
- BGP-LS: collects topology information and reports collected information to the controller. If an RR exists on the network, you only need to deploy BGP-LS on the RR and establish a BGP-LS peer relationship between the RR and controller.
- BGP IPv6 SR Policy: Such a peer relationship is established between the controller and forwarder, so that the controller can deliver an SRv6 Policy to the forwarder through the peer relationship to direct traffic forwarding. To reduce the number of peer relationships, you can deploy an RR and configure PEs and the controller to function as RR clients.
- NETCONF: delivers service configurations from the controller to forwarders. This document does not describe service delivery or NETCONF-related configuration.

SRv6 Policy Advertisement Process

- To facilitate configuration, the controller provides the following functions:
  - Directly creates a bidirectional tunnel between the ingress and egress. In other words, a tunnel from the egress to the ingress is automatically created when a tunnel from the ingress to the egress is created.
  - Allows you to configure tunnel and color templates to simplify the configuration of some parameters.
- In Huawei's CloudWAN solution:
  - The SRv6 Policy configurations are delivered by BGP IPv6 SR Policy, and the tunnel status is reported by BGP-LS.

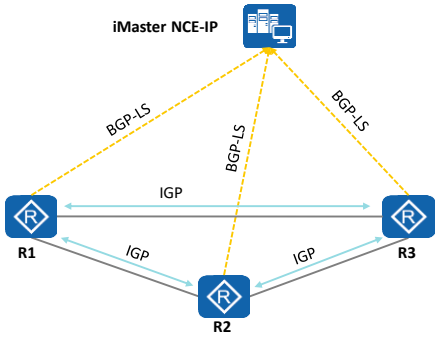


HUAWEI

- The process of planning and deploying forwarding paths through iMaster NCE-IP is as follows:
  - Devices use BGP-LS to report network topology information to the controller, which then generates forwarding paths based on requirements.
  - The controller delivers the computed paths to the devices through BGP IPv6 SR Policy.
  - Target traffic travels along the delivered paths.

Key Information Required for SRv6 Policy Path Computation

- iMaster NCE-IP uses an IGP to collect the following key information for SRv6 Policy path computation:
  - SID information: End and End.X SIDs.
  - Locator information: After a locator is configured for a node, the system generates a locator route and propagates the route throughout the SR domain using an IGP. Other nodes on the network can locate this node through the locator route.
  - Interface bandwidth information: physical bandwidth and reservable bandwidth.
  - Interface delay and packet loss rate.
  - Network topology information.
- BGP-LS advertises the information collected by an IGP to iMaster NCE-IP.



HUAWEI

Basic IGP Configurations

- In Huawei's SRv6 solution, it is recommended that extended IS-IS be used as an IGP for the collection of basic network information.
- Basic IGP configurations are as follows:
  - Global IS-IS configurations

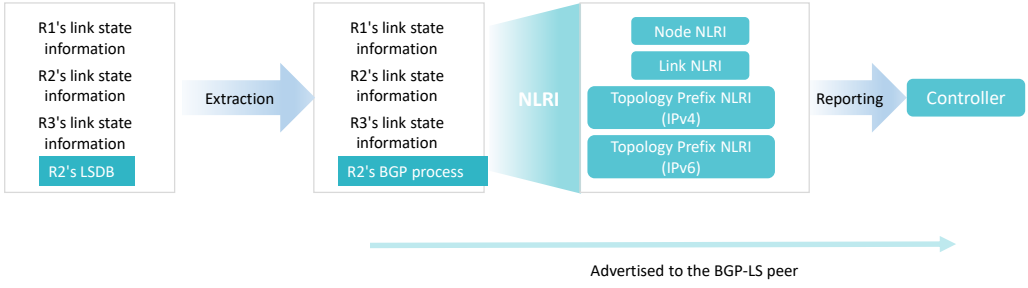
```
[P1]isis 1
[P1-isis-1] is-level level-2
[P1-isis-1] cost-style wide //TE information (such as bandwidth information) required in TE scenarios cannot be carried in narrow mode. Therefore, the wide type needs to be set.
[P1-isis-1] network-entity 49.0001.0010.0000.0005.00
[P1-isis-1] is-name P1
[P1-isis-1] ipv6 enable topology ipv6
[P1-isis-1] ipv6 bgp-ls enable level-2 //The device is enabled to send topology information collected by IS-IS to the controller through BGP-LS. This function only needs to be configured on the RR. That is, only one device in the IGP domain needs to send topology information to the controller through BGP-LS.
[P1-isis-1] ipv6 advertise link attributes //The device is enabled to carry link attribute-related TLVs in LSPs. TLV information includes the IPv6 addresses and indexes of interfaces.
[P1-isis-1] ipv6 metric-delay advertisement enable level-1-2 //The device is enabled to advertise IPv6 delay information. Intra-domain IPv4 link delay information is collected and flooded through IS-IS and then reported to the controller through BGP-LS. Based on the delay information, the controller computes the optimal path on the P2P network.
[P1-isis-1] ipv6 traffic-eng level-2 //IS-IS TE is enabled so that link bandwidth information is reported to the TE module.
```

  - Interface-specific IS-IS configurations

```
[P1]interface GigabitEthernet0/3/0
[P1-GigabitEthernet0/3/0] isis ipv6 enable 1
[P1-GigabitEthernet0/3/0] isis circuit-type p2p //The IS-IS interface's network type must be set to P2P. Otherwise, the required network topology cannot be formed on the controller.
```

Topology Information Reporting to the Controller Through BGP-LS

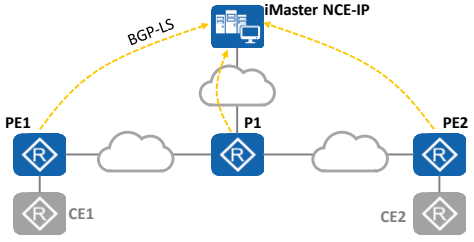
- Each router maintains one or more LSDBs. Each LSDB contains multiple link attributes, such as the interface IP address, link metric, TE metric, link bandwidth, and reservable bandwidth. The BGP process of a router obtains information from these LSDBs and carries the information in the extended NLRI attribute.



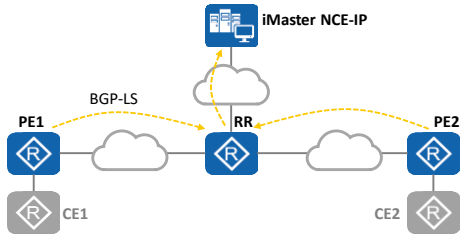
HUAWEI

BGP-LS Deployment

- BGP-LS is mainly used to upload network topology and tunnel information. As such, iMaster NCE-IP needs to establish BGP-LS peer relationships with at least two types of devices:
  - Establishes a BGP-LS peer relationship with an intra-AS network node, which is usually a BGP RR, in order to obtain network topology information.
  - Establishes a BGP-LS peer relationship with the ingress of the involved tunnel in order to obtain SRv6 Policy status.
- BGP-LS deployment solutions:



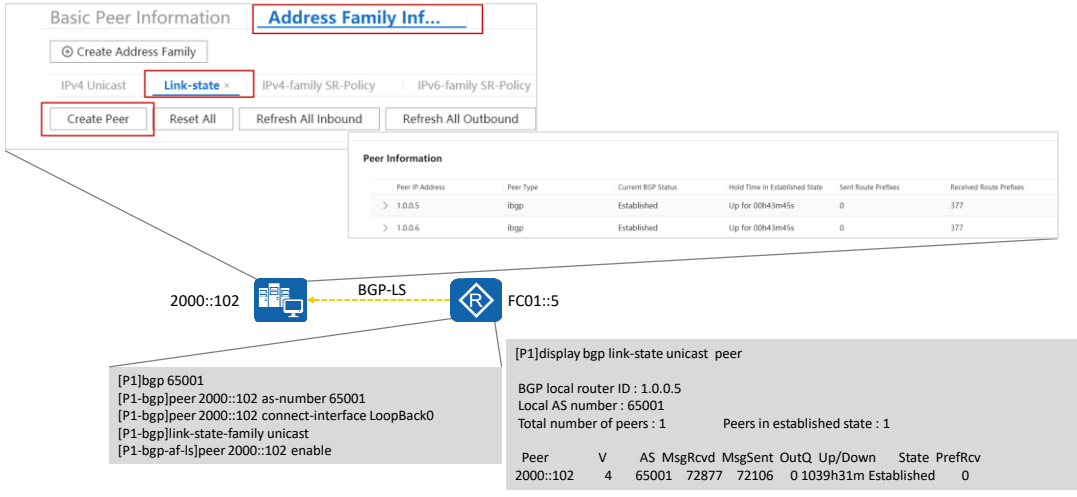
- Solution 1: Establish BGP-LS peer relationships between iMaster NCE-IP and all PEs and between iMaster NCE-IP and all Ps.



- Solution 2: Establish BGP-LS peer relationships between iMaster NCE-IP and RRs and between the RRs and other devices.
- Solution 2 is recommended to reduce the number of BGP peers maintained by iMaster NCE-IP.

HUAWEI

BGP-LS Peer Relationship Establishment

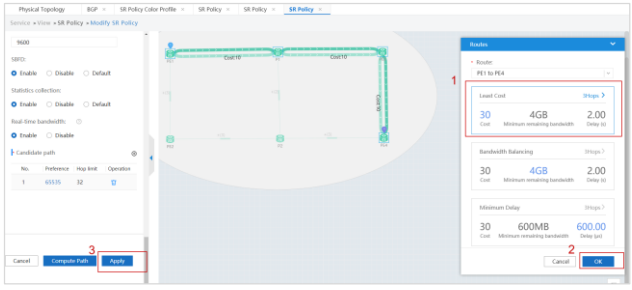


HUAWEI

- BGP-LS peer relationships can be established using IPv4 or IPv6 addresses. This course uses IPv6 addresses to establish such relationships.

SRv6 Policy Path Computation and Deployment

- With the SRv6 Policy path computation algorithm, the controller can provide the following path computation results if specified constraints are met:
  - Minimum cost: path with the minimum cost among all qualified paths
  - Minimum delay: path with the minimum delay among all qualified paths
  - Bandwidth balancing: path with the most remaining bandwidth among all qualified paths that have the same cost
- During SRv6 Policy creation, you need to specify a color value for each SRv6 Policy.

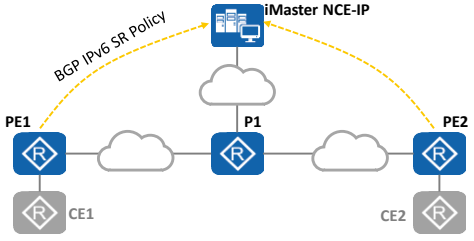


HUAWEI

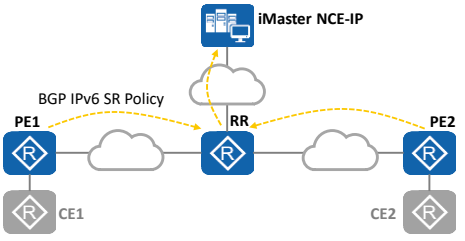
- Optional constraints:
  - Bandwidth constraint: ensures that the bandwidth configured for a service does not exceed the remaining bandwidth of the link that the service traverses.
  - PIR constraint: ensures that the peak bandwidth does not exceed the BC0 bandwidth of the link that the service traverses. PIR refers to the peak bandwidth of a service.
  - Delay limit constraint: ensures that the path delay of a service does not exceed the configured delay limit.
  - Hop limit constraint: ensures that the number of links that a service traverses does not exceed the configured hop limit.
  - Affinity constraint: determines which types of links are allowed and which types of links are not allowed for services.

BGP IPv6 SR Policy Deployment

- BGP IPv6 SR Policy is mainly used to deliver SRv6 tunnel information. As such, iMaster NCE-IP needs to establish a BGP IPv6 SR Policy peer relationship with the ingress of the involved tunnel.
- BGP IPv6 SR Policy deployment solutions:



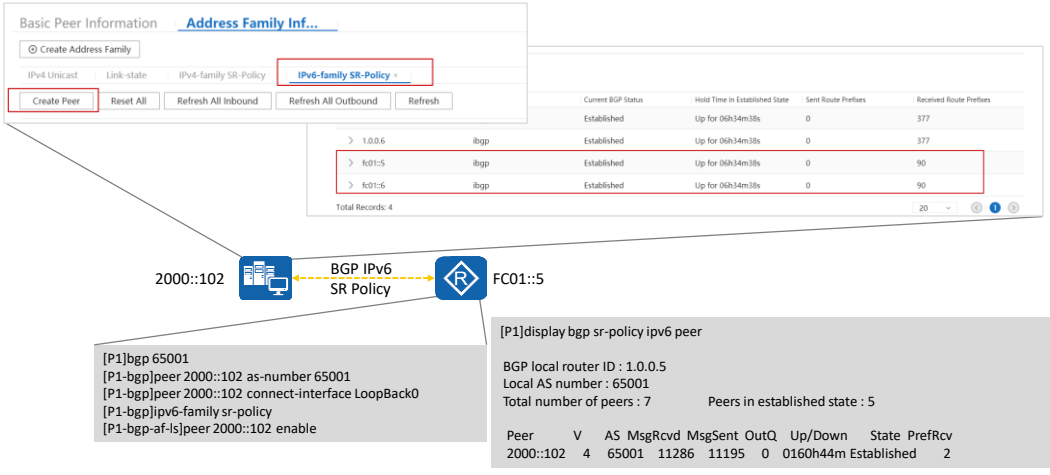
- Solution 1: Establish BGP IPv6 SR Policy peer relationships between iMaster NCE-IP and all PEs.



- Solution 2: Establish BGP IPv6 SR Policy peer relationships between iMaster NCE-IP and RRs and between the RRs and other devices.
- Solution 2 is recommended to reduce the number of BGP peers maintained by iMaster NCE-IP.

HUAWEI

BGP IPv6 SR Policy Peer Relationship Establishment




```
[P1]bgp 65001
[P1-bgp]peer 2000::102 as-number 65001
[P1-bgp]peer 2000::102 connect-interface LoopBack0
[P1-bgp]ipv6-family sr-policy
[P1-bgp-af-ls]peer 2000::102 enable
```

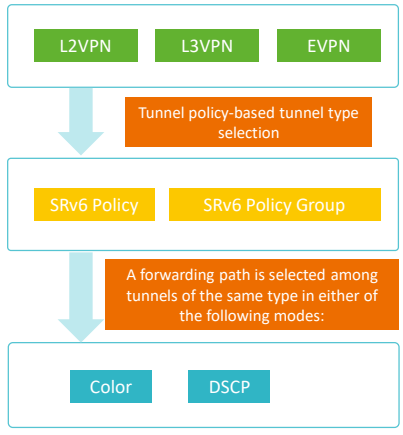
```
[P1]display bgp sr-policy ipv6 peer

BGP local router ID : 1.0.0.5
Local AS number : 65001
Total number of peers : 7              Peers in established state : 5

 Peer        V     AS  MsgRcvd  MsgSent  OutQ  Up/Down      State  PrefRcv
 2000::102   4   65001    11286    11195     0  0160h44m Established      2
```


HUAWEI

- BGP IPv6 SR Policy peer relationships can be established using IPv4 or IPv6 addresses. This course uses IPv6 addresses to establish such relationships.

VPN Service Forwarding over SRv6 Policies

- The following types of VPNs are available in enterprise network scenarios:
  - L2VPN: Customer IP addresses are on the same network segment.
  - L3VPN: Customer IP addresses are on different network segments.
  - EVPN: Customer IP addresses are either on the same network segment (L2VPN scenario) or on different network segments (L3VPN scenario).
- A tunnel policy is used by an application module to select tunnels for services. There are two types of tunnel policies:
  - (Preferred mode) Tunnel type prioritizing policy: recurses services to a tunnel based on the tunnel type priority and the number of tunnels participating in load balancing.
  - Tunnel binding policy: binds a destination address to a tunnel, so that the traffic of VPN services referencing the policy and destined for this address will be transmitted over the tunnel.
- VPN services first select tunnels in the up state based on the tunnel policy, and then select a forwarding path from qualified tunnels.



Quiz

1. (Short-answer question) An SRv6 SID has 128 bits. What are the three fields of an SRv6 SID?
2. (Short-answer question) In SIDs corresponding to SRv6 endpoint behaviors, which types of SIDs are similar to the node segments and adjacency segments in SR-MPLS?

HUAWEI

1. An SRv6 SID has 128 bits and consists of the Locator, Function, and Arguments fields.
2. End SIDs and End.X SIDs.

Summary

- This course describes the concept of SRv6 network programming, SRv6 instruction sets (endpoint node behaviors, source node behaviors, and flavors), SRv6 Policy, and basic SRv6 SID configurations on Huawei NetEngine series routers.
- Leveraging the programmability of 128-bit IPv6 addresses, SRv6 enriches the network functions expressed by SRv6 instructions. For example, in addition to identifying an instruction that can indicate a forwarding path, a network function can identify a VAS (e.g. firewall, application acceleration gateway, user gateway). To deploy a new network function, you only need to define a new instruction, without the need to change the protocol mechanism or deployment.
- SRv6 Policy information is carried by extending new NLRIs based on MP-BGP. The controller establishes BGP IPv6 SR Policy peer relationships with forwarders to deliver SRv6 Policies to them.

HUAWEI

More Information

- SRv6 Network Programming: Ushering in a New Era of IP Networks
- https://datatracker.ietf.org/doc/rfc8754/
- https://datatracker.ietf.org/doc/rfc8986/

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Enterprise Bearer WAN Design

HUAWEI

Foreword

- The enterprise bearer WAN is also called the enterprise backbone WAN, which is used for interconnection between enterprise branches, between enterprise branches and clouds, and between clouds.
- In the past, the enterprise bearer WAN only needs to forward traffic. However, as growing enterprise services depend increasingly on the network, service SLA assurance requirements are imposed on the network.
- To meet service requirements, the enterprise bearer WAN has employed various new technologies, which complicates the O&M and design of the bearer WAN.
- This course describes the technologies and network design roadmap of the bearer WAN based on Huawei's CloudWAN solution.

HUAWEI

- This course is based on Huawei's CloudWAN solution.

Objectives

- On completion of this course, you will be able to:
  - Describe the challenges faced by the enterprise bearer WAN.
  - Describe the basic functions of Huawei's CloudWAN solution.
  - Describe the basic design roadmap of the enterprise bearer WAN.
  - Describe the tunnel and VPN design roadmap of the enterprise bearer WAN.
  - Describe the SLA and reliability design roadmap of the enterprise bearer WAN.
  - Describe the optimization and O&M design roadmap of the enterprise bearer WAN.

HUAWEI

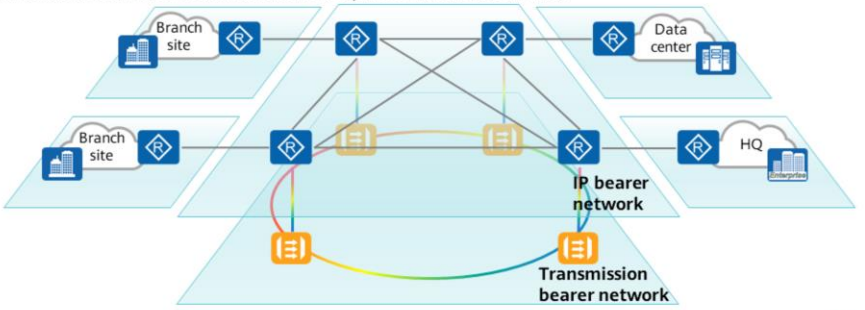
Contents

1. **Current Situation and Challenges of the Enterprise Bearer WAN**
2. Huawei CloudWAN Solution Overview
3. Basic Design for the Enterprise Bearer WAN
4. Tunnel and VPN Design for the Enterprise Bearer WAN
5. SLA and Reliability Design for the Enterprise Bearer WAN
6. Optimization and O&M Design for the Enterprise Bearer WAN

HUAWEI

Enterprise Bearer WAN Overview

- Bearer WANs can be classified into IP bearer networks and transmission bearer networks by network layer. For most enterprises, transmission bearer networks cannot be built by themselves (optical fiber layout requires qualification certificates). Therefore, they usually lease transmission lines from carriers. An enterprise can build its own IP bearer network over the transmission bearer network.
- The IP bearer network built by an enterprise is called the enterprise bearer WAN, which is also called the enterprise backbone network or enterprise core network.
- The enterprise bearer WAN carries the enterprise's internal cross-region interconnection services, such as synchronization services between data centers and voice services between enterprise branches and the HQ.

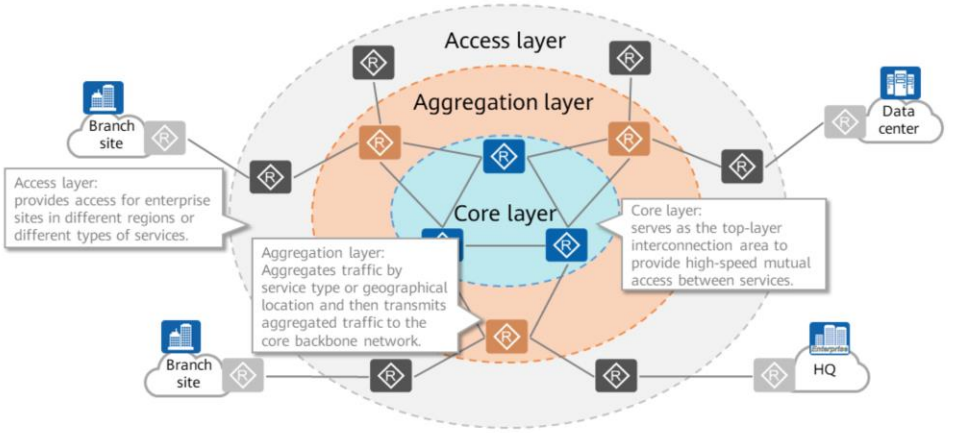


HUAWEI

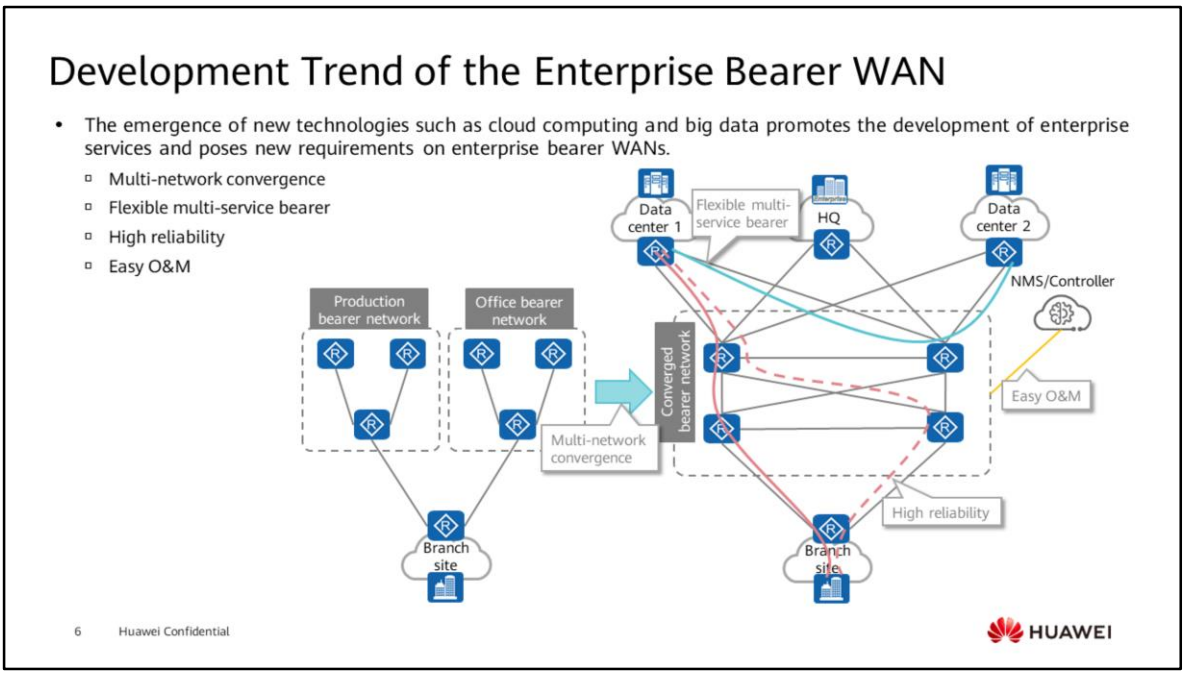
- The bearer WAN mentioned in this course mainly refers to the IP bearer WAN.

Enterprise Bearer WAN Architecture

- While the bearer WANs of different enterprises vary greatly in scale and actual topology, their architecture can be generally divided into the access, aggregation, and core layers.



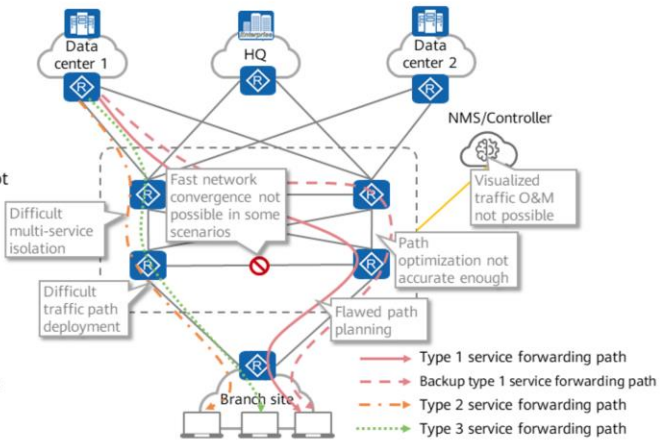
HUAWEI



- Multi-network convergence:
  - The bearer network uses one physical network to carry production, office, and other services. Powerful BGP routing policies are used to control traffic on the bearer network. Traffic diversion policies are deployed based on service attributes so that different types of services, such as production and office services, can run on different paths based on the customized policies.
- Flexible multi-service bearer:
  - As multiple services are migrated to the cloud and multiple networks are converged, issues such as low network resource utilization and unbalanced traffic distribution become more prominent. The segment routing technology can be used to flexibly plan paths to carry traffic, improving network utilization.
- High reliability:
  - Reliability is classified into network reliability and service reliability.
  - Network reliability is ensured from aspects such as network architecture and network stability.
  - Service reliability is mainly ensured by optimizing network bandwidth, delay, and packet loss rate based on service requirements.

Challenges Faced by Enterprise Bearer WANs

- The development trend of the enterprise bearer WAN brings many challenges.
  - Multi-network convergence
    - Challenge: It is difficult to isolate multiple services.
  - Flexible multi-service bearer
    - Challenge: Network resource planning is not precise enough, and traffic path planning and deployment are difficult.
  - High reliability
    - Challenge: Fast convergence cannot be implemented in all scenarios, and service SLA cannot be effectively guaranteed.
  - Easy O&M
    - Challenge: Precise traffic path optimization and visualized O&M cannot be provided.

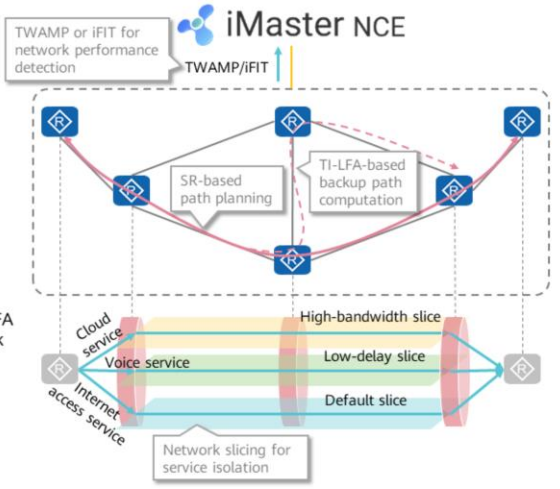


HUAWEI

- Multi-service isolation:
  - Due to the convergence of multiple networks, the originally physically isolated production and office networks are now carried on the same network. The service isolation requirements of some enterprises can be met using the VPN technology. However, for other enterprises, services need to be isolated using hard pipes.
- Network resource and traffic path planning:
  - To make full use of network resources (mainly bandwidth), network traffic needs to be planned. Traditional traffic engineering (mainly MPLS-TE) has many disadvantages, such as complex configuration and flawed path computation mechanism. Therefore, it cannot effectively solve the problem in multi-service bearing.
- Fast network convergence and service SLA assurance:
  - Traditional networks usually use FRR to shorten the convergence time. However, FRR based on the LFA or RLFA algorithm has restrictions in application scenarios and may fail to meet the high reliability of some networks.

Solutions to Challenges Faced by the Enterprise Bearer WAN

- To address challenges faced by the enterprise Bearer WAN, Huawei provides the CloudWAN solution.
  - Difficult multi-service isolation
    - Network slicing technology is deployed to carry different services over different slices.
  - Difficult traffic path planning and deployment
    - Based on SR, the controller plans paths globally and automatically deploys paths.
  - Lack of fast network convergence and service SLA assurance
    - TI-LFA is used for backup path computation, and TI-LFA works with FRR and HSB technologies for fast network convergence.
  - Difficult traffic path optimization and lack of visualized O&M
    - TWAMP and iFIT are used to precisely detect network quality, facilitating accurate traffic optimization.

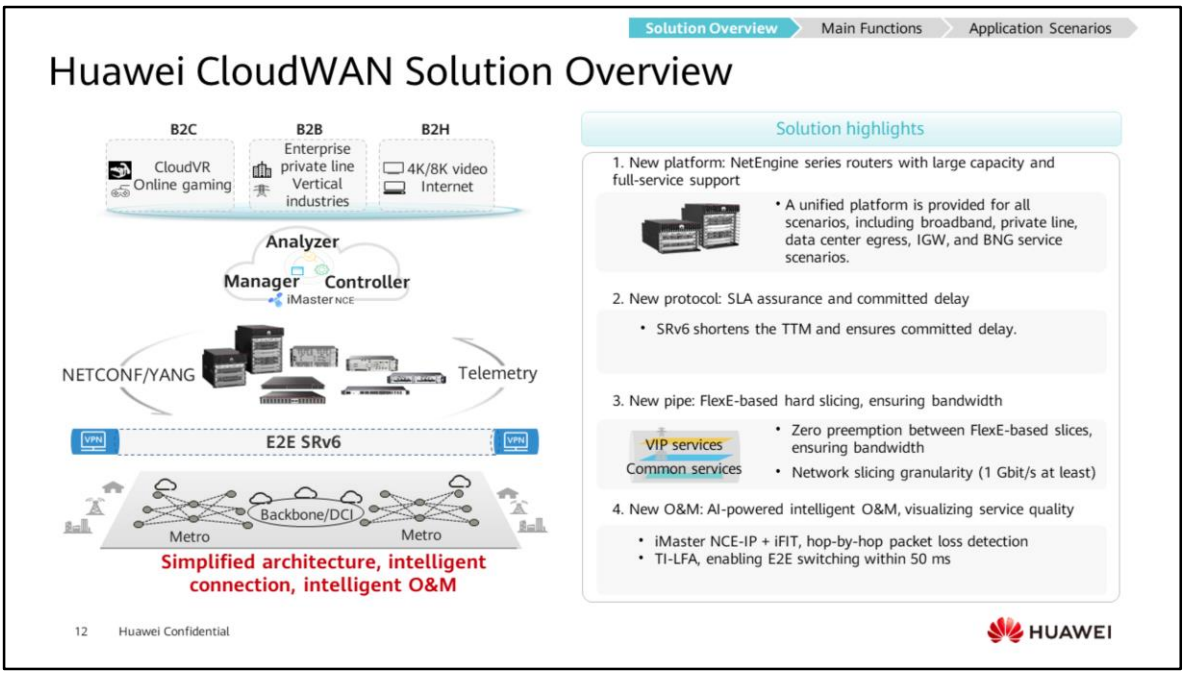


HUAWEI

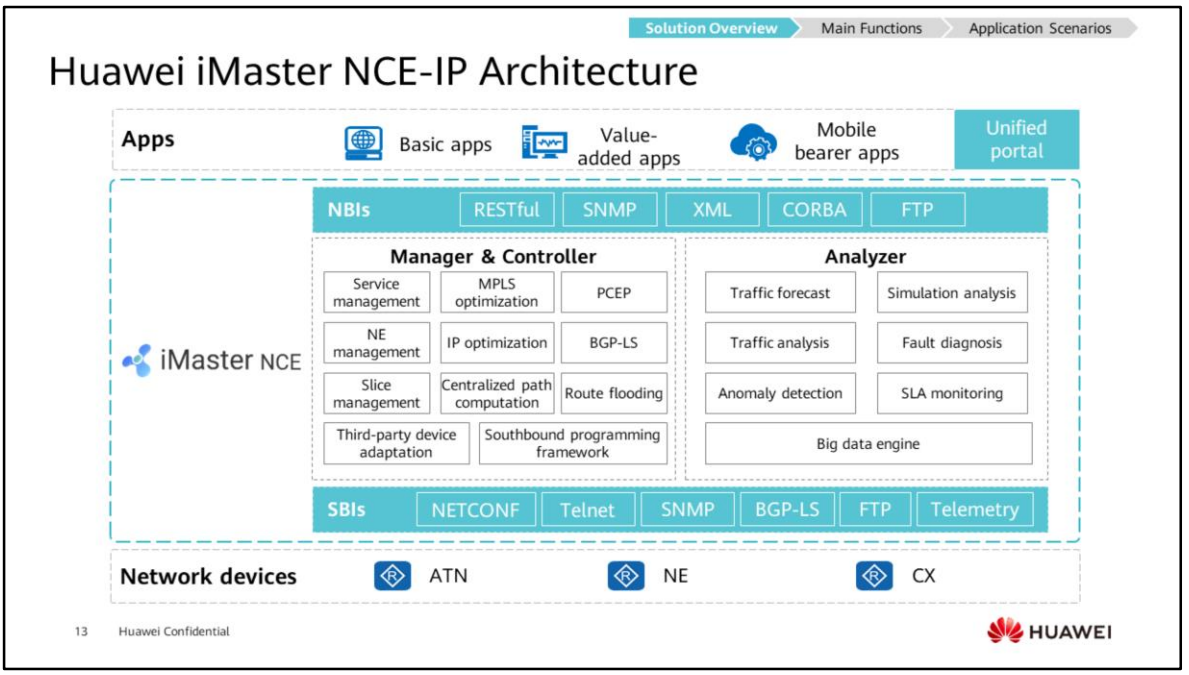
Contents

1. Current Situation and Challenges of the Enterprise Bearer WAN
2. **Huawei CloudWAN Solution Overview**
3. Basic Design for the Enterprise Bearer WAN
4. Tunnel and VPN Design for the Enterprise Bearer WAN
5. SLA and Reliability Design for the Enterprise Bearer WAN
6. Optimization and O&M Design for the Enterprise Bearer WAN

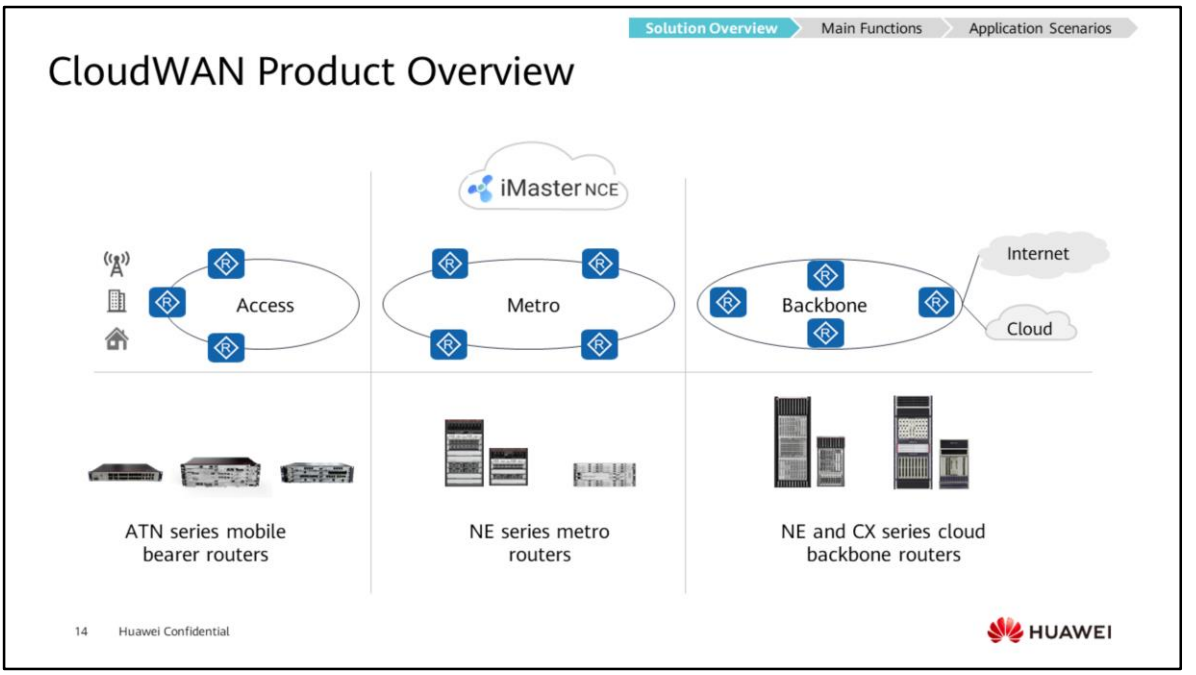
HUAWEI



- TI-LFA: Topology-Independent Loop-free Alternate
- BNG: broadband network gateway
- TTM: time to market



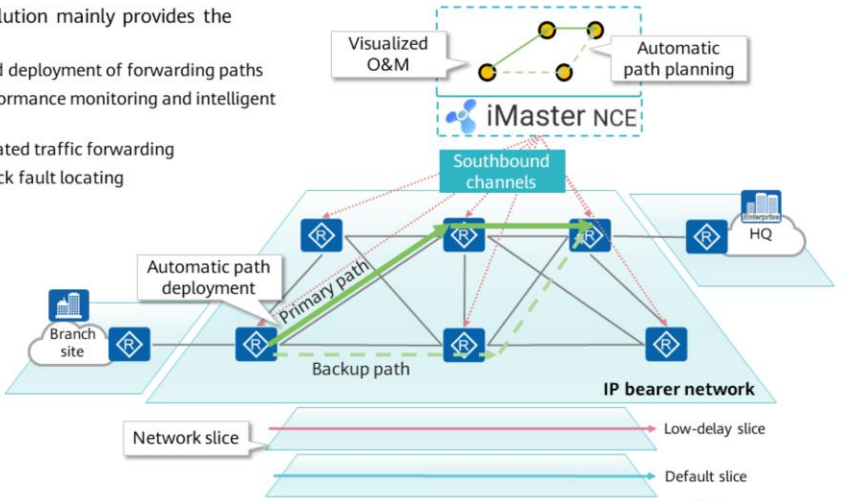
- Based on the cloud platform, iMaster NCE-IP provides three logical modules (Manager, Controller, and Analyzer) and various scenario-specific applications to achieve flexible modular deployment based on customer requirements.



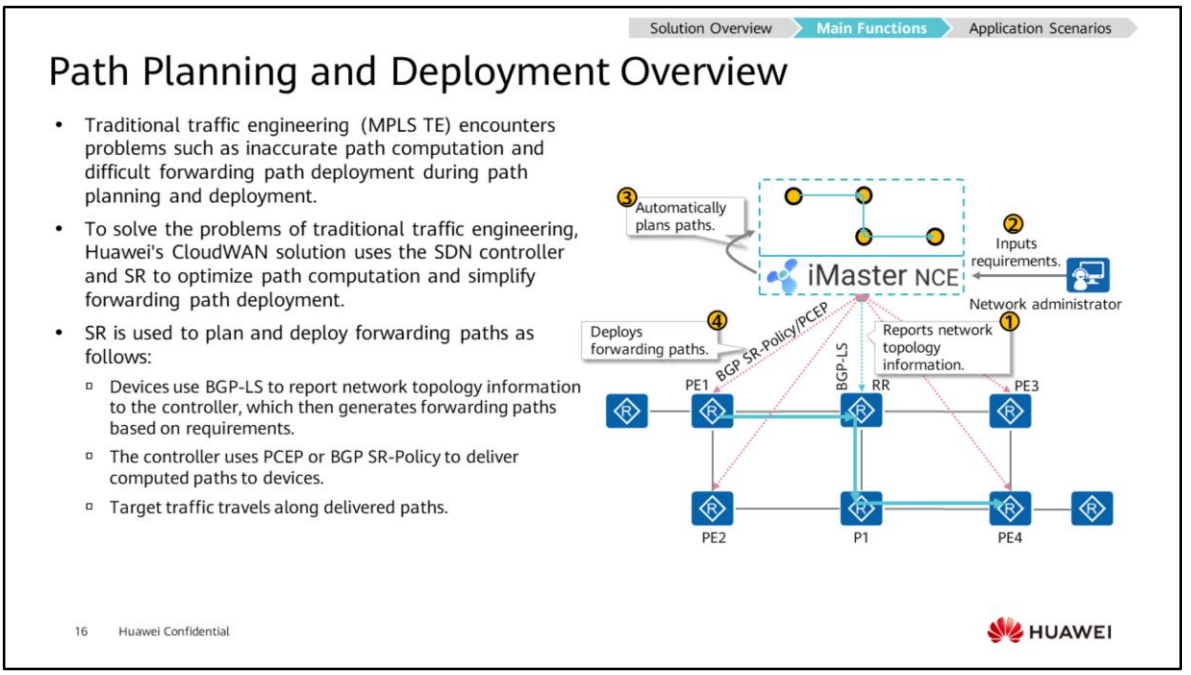
- Huawei has all-scenario SRv6 product capabilities and can provide access, metro, and backbone network routers for carriers and enterprises.

Main Functions of Huawei's CloudWAN Solution

- Huawei's CloudWAN solution mainly provides the following functions:
  - Automatic planning and deployment of forwarding paths
  - Real-time network performance monitoring and intelligent traffic optimization
  - Network slicing for isolated traffic forwarding
  - Visualized O&M for quick fault locating



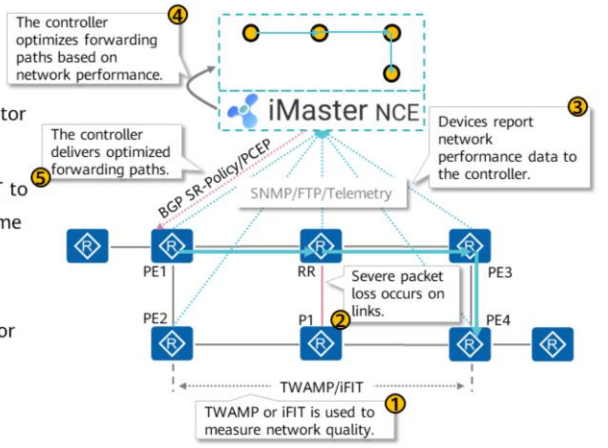
HUAWEI



- BGP-LS is a new method of collecting network topology information, which makes topology information collection simpler and more efficient. Using BGP-LS to report topology information has the following advantages:
  - Requirements on the computing capability of the upper-layer controller are lowered, and there is no requirement on the IGP capability of the controller.
  - BGP summarizes the topology information of each process or AS and directly sends the complete topology information to the controller, facilitating path selection and calculation.
  - All topology information on the network is sent to the controller through BGP, unifying the protocols for sending topology information.
- BGP SR-Policy delivers data forwarding path information to the headend through the BGP route. The headend then directs traffic to a specific SR Policy. Segment lists in SR Policies are used to guide traffic forwarding. A segment list is calculated based on a series of optimization objectives and constraints, such as delay, affinity, and SRLG.
- SR Policy is the mainstream SR implementation mode.

Network Performance Monitoring and Optimization

- As networks carry increasingly more enterprise services, they need to meet increasingly higher requirements. As such, a fast, flexible IP network performance measurement tool is required to monitor network performance in a timely manner.
- Huawei's CloudWAN solution uses TWAMP and iFIT to precisely and quickly measure historical and real-time network performance.
- Based on real-time network performance, the controller can optimize the network through PCEP or BGP SR-Policy.



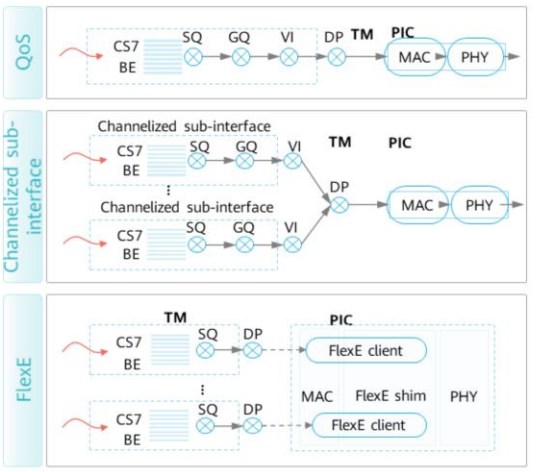
HUAWEI

- TWAMP is a standard protocol and can be deployed on IP, MPLS, and L3VPN networks. TWAMP is easy to obtain and deploy and does not require clock synchronization.
- iFIT measures the packet loss rate and delay of service packets transmitted on an IP network to determine network performance. It is easy to deploy and provides an accurate assessment of network performance.

Solution Overview | Main Functions | Application Scenarios

Network Slicing Overview

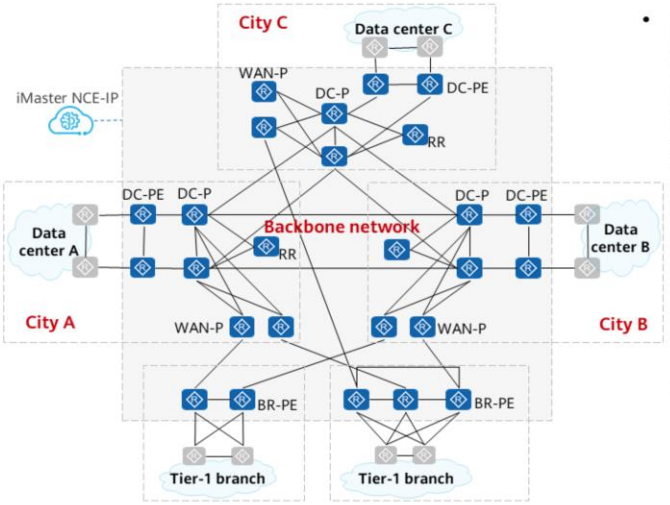
- Slice resource reservation technologies, such as FlexE, channelized sub-interface, and QoS queuing, can be used to direct services to respective service slices. These slices are isolated from each other and do not affect each other, providing different SLA levels.
- Slicing is an SLA assurance method in essence. It effectively integrates the network capabilities required by users to form a logical network called "slice".



HUAWEI

- Channelized sub-interfaces provide a mechanism to isolate different types of services. Different types of service traffic can be forwarded to different VLAN channelized sub-interfaces that use different dot1q encapsulation modes. Each channelized sub-interface can implement independent HQoS scheduling to isolate different types of services.
- FlexE is an Ethernet-based bearer technology for multi-rate sub-interfaces on multiple PHY links. FlexE supports bundling, channelization, and sub-rating.
- SQ: subscriber queue
- GQ: group queue
- VI: virtual interface
- DP: data plane
- TM: traffic manager
- PIC: physical interface card

Financial CloudWAN Scenario



- With the deployment of distributed cloud data centers, the cloud backbone network is elastically expanded by adding core and aggregation nodes to implement interconnection between these data centers. In addition, technologies such as SDN and SRv6 are introduced to implement functions such as intelligent network management and traffic optimization.
  - Fast service provisioning, centralized path computation, intelligent path selection
  - Intelligent network visualization, enabling real-time awareness of service status
  - Evolution to the distributed multi-city multi-center architecture
  - Capability openness (the controller provides open southbound and northbound interfaces)

HUAWEI

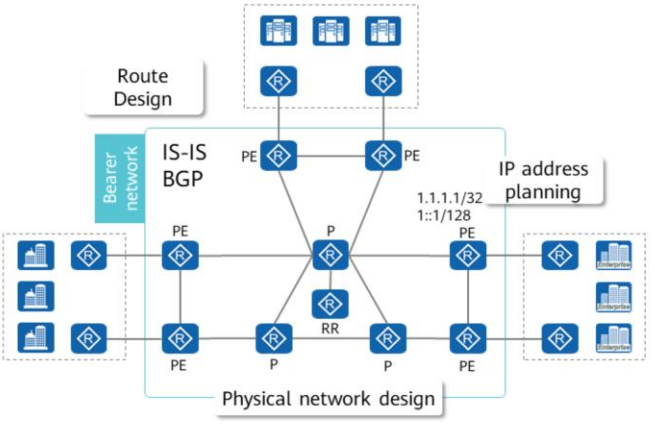
Contents

1. Current Situation and Challenges of the Enterprise Bearer WAN
2. Huawei CloudWAN Solution Overview
3. **Basic Design for the Enterprise Bearer WAN**
4. Tunnel and VPN Design for the Enterprise Bearer WAN
5. SLA and Reliability Design for the Enterprise Bearer WAN
6. Optimization and O&M Design for the Enterprise Bearer WAN

HUAWEI

Basic Design Overview for the Enterprise Bearer WAN

- Basic design for the enterprise bearer WAN includes physical network design, IP address planning, and routing design.



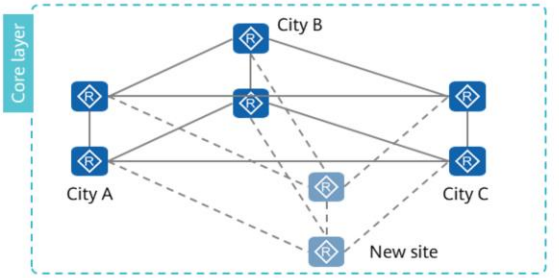
HUAWEI

Physical Network Design for the Enterprise Bearer WAN

HUAWEI

Core Layer Design

- The core layer, as the backbone of the entire network, aggregates and forwards various service traffic. When selecting core nodes, consider the following factors:
  - Service volume: Consider the current service volume and expected service growth.
  - Physical location: Ensure that core nodes are secure, easy to obtain, and easy to maintain, as they are the key infrastructure of the bearer network.
  - Number of nodes: The core layer usually adopts the full-mesh + dual-plane architecture to ensure stability.

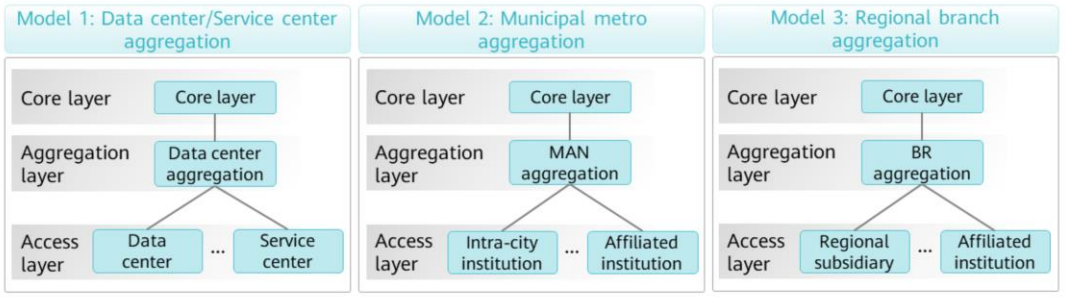


HUAWEI

- The service volume involves two aspects. The first is the service flow direction, that is, where the services concentrate. The second is the service volume size. The two aspects are complementary to each other. Generally, a greater concentration indicates a larger service volume. The core nodes must be nodes where services concentrate and the service volume is large.
- Core nodes in the same city are interconnected through WDM, and core nodes in different cities are interconnected through inter-provincial or inter-metro carrier private lines. The number of core nodes must be comprehensively considered and cannot be too large.

Aggregation Layer Design

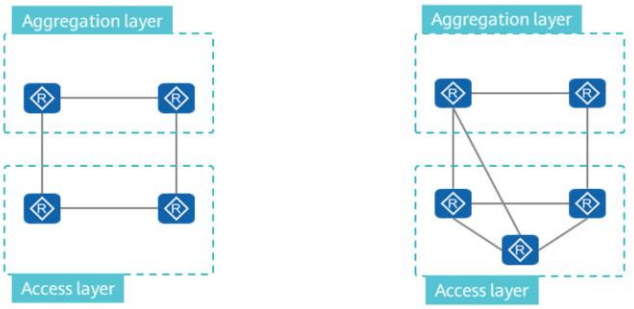
- Aggregation layer design must take into account the service types and scale on each aggregation node. In the initial phase of bearer network planning and construction, the aggregation layer can be planned based on existing enterprise services. The following three aggregation modes are available:
  - Data center aggregation (DC-PE): aggregates traffic from service units that provide services for the HQ in an enterprise. Such service units include data centers and service centers.
  - Metro aggregation (MAN-PE): aggregates the metro services of intra-city institutions and affiliated institutions.
  - Branch aggregation (BR-PE): aggregates the services of branches, provincial metro institutions, and affiliated institutions.



HUAWEI

Access Layer Design Factors

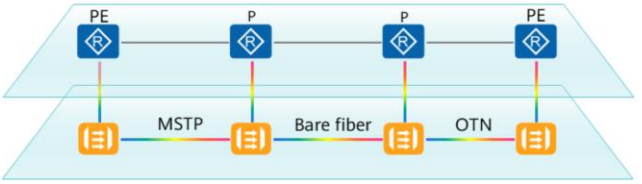
- When designing the access layer, you need to consider factors such as the bandwidth required by access services, required private lines, and private line prices. Moreover, you need to consider access reliability and traffic load balancing. For example:
  - You can use single-homed networking at the access layer to form dual planes working in active/standby mode. This helps improve network reliability.
  - On the active and standby planes, you can load-balance WAN traffic through service planning or routing policies.



HUAWEI

Bearer Network Private Line Selection

- Only a few enterprises have the capability to build their own transmission private lines. Most enterprises need to purchase transmission private lines from carriers and build their own bearer networks over these transmission private lines.
- Currently, the transmission private lines provided by carriers are mainly MSTP and OTN private lines:
  - MSTP private lines or MPLS VPNs can be used to transmit services from branches to the access layer of the bearer network.
  - MSTP private lines can be used between the access and aggregation layers of the bearer network.
  - If the aggregation layer and core layer are in different equipment rooms, MSTP/OTN private lines can be used between them.
  - MSTP/OTN private lines can be used between core-layer devices. DWDM private lines can also be used between core-layer devices if bare fibers are available.



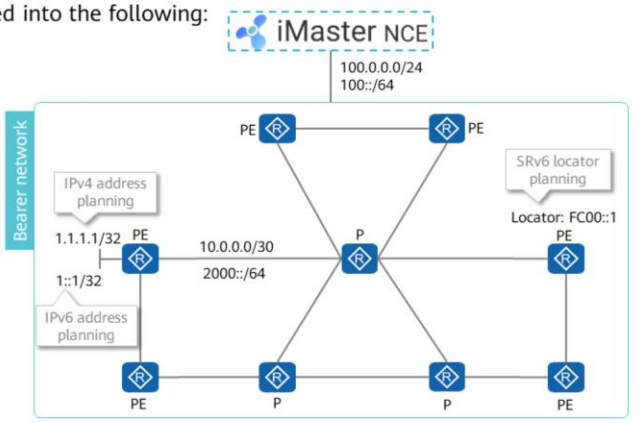
HUAWEI

IP Address Planning Roadmap for the Enterprise Bearer WAN

HUAWEI

IP Address Planning Overview

- With the development of services, increasingly more networks are deployed as IPv4/IPv6 dual-stack networks. Dual-stack address planning is essential to dual-stack networks.
- IP address planning is generally classified into the following:
  - IPv4 address planning
  - IPv6 address planning
  - SRv6 locator planning



HUAWEI

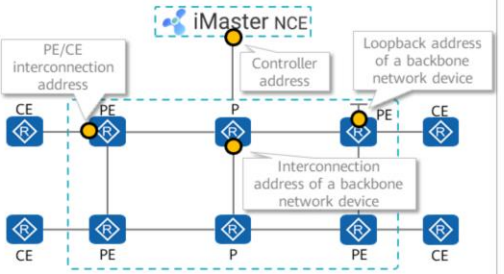
IPv4 Address Design Rules

- For the convenience of service inheritance and network management, the backbone network retains interconnection IPv4 addresses. IPv4 address allocation must comply with the existing IPv4 address allocation specifications of the customer.

IPv4 address design rules

- **Uniqueness**
  - Hosts on the backbone network must use unique IP addresses. Try to allocate a different address to each host even if they support VPN address overlapping.
- **Contiguity**
  - Routes with contiguous addresses can be easily summarized on a hierarchical network, reducing the routing table size and accelerating route calculation.
- **Scalability**
  - Addresses need to be reserved at each layer to ensure contiguity of addresses when the network is expanded.
- **Meaningfulness**
  - A well-planned IP address denotes the device to which the IP address belongs.

IPv4 address allocation range on the bearer WAN

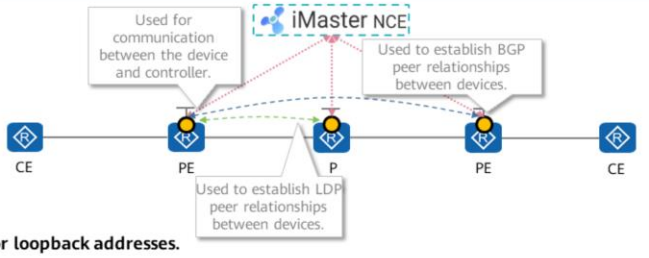


HUAWEI

Loopback Address Design Rules

- The loopback address of each router plays an important role in the normal running of the entire network. Therefore, a unified, dedicated address space must be used for the allocation and management of loopback addresses on each router.

Loopback address design rules and application scenarios



- **Use 32-bit masks for loopback addresses.**
  - Allocate loopback addresses based on physical locations and reserve sufficient address space.
  - Allocate loopback addresses for the same geographical location by plane. If there are two planes, allocate IP addresses to planes 1 and 2 in sequence.

HUAWEI

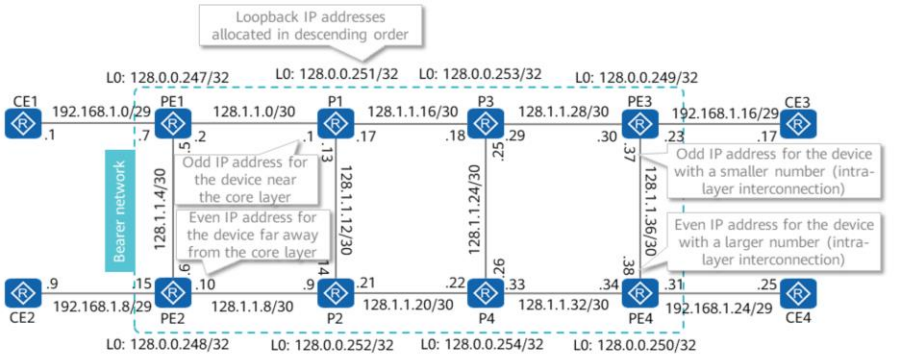
IPv4 Address Planning Suggestions

- It is recommended that the bearer WAN use a dedicated subnet. The allocated addresses include the management addresses and internal interconnection addresses of devices and access addresses of the bearer network. The following table describes the address allocation plan:
  - Internal interconnection addresses: Use a 30-bit mask for IP addresses used for internal interconnection between devices on the bearer network.
  - Access addresses: The access-layer device interfaces on the bearer network need to connect to the original network. Therefore, it is recommended that the IP addresses of these interfaces be allocated based on the original planning. For example, use a 29-bit mask for these IP addresses.
  - Address allocation sequence: Allocate interconnection addresses in ascending order and loopback addresses in descending order.
  - Interconnection between devices of the same layer: Assign an odd address to the device with a smaller number and an even address to the device with a larger number.
  - Interconnection between devices at different layers: Assign an odd address to the device close to the network core and an even address to the device far away from the network core.

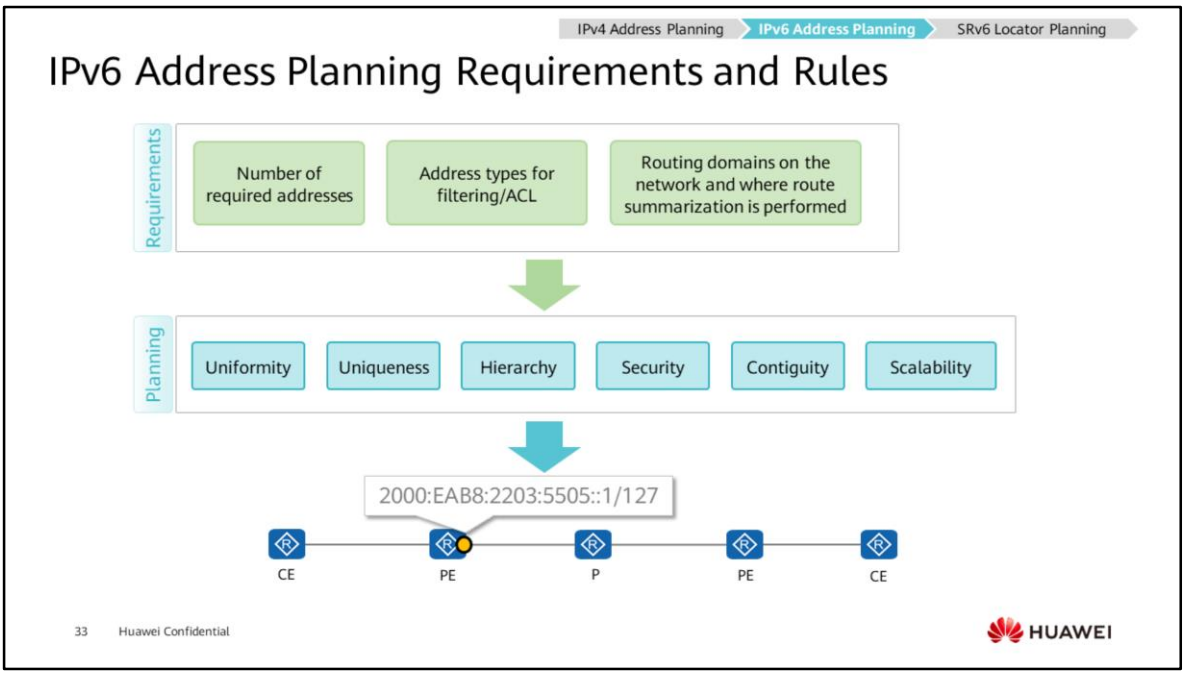
HUAWEI

Typical IPv4 Address Planning

- IPv4 addresses can be planned based on planning suggestions or the customer's planning rules.



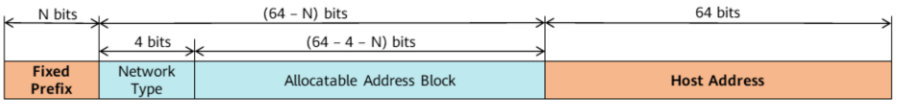
HUAWEI



- Uniformity: All IP addresses on the entire network are planned in a unified manner, including service addresses, platform addresses, and network addresses.
- Uniqueness: Each address is unique throughout the entire network.
- Hierarchy: The massive IPv6 address space poses higher requirements on the route summarization capability. The primary task of IPv6 address planning is to reduce network address fragments, enhance the route summarization capability, and improve the network routing efficiency.
- Security: Services with shared attributes have the same security requirements. Mutual access between services needs to be controlled. Services with shared attributes are allocated with addresses in the same address space, which facilitates security design and policy management.
- Contiguity: IPv6 addresses in an IPv6 address segment must be contiguous to prevent address wastes.
- Scalability: IP addresses must be planned and allocated based on network development requirements to reserve space for future capacity expansion. The addition of a small number of subnets does not require large-scale architecture or policy adjustment.

Structured IPv6 Address Planning

- Structured planning: The massive IPv6 address space poses higher requirements on the route summarization capability. Hierarchical IPv6 address allocation is recommended, which helps reduce the sizes of routing tables through route summarization. In addition, the subnet field of an IPv6 address can be divided into several independent fields to make the address more readable. (If possible, it is recommended that the subnet field be divided by 4 bits.) The following is a planning example:

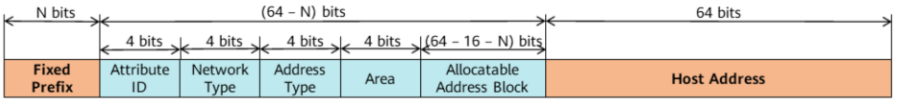


  - Fixed Prefix: indicates the fixed prefix applied for.
  - Network Type: indicates the type of a network. For example, 0 indicates a backbone network, 1 indicates a data center, and 2 to F are reserved.

HUAWEI

IPv6 Address Planning Suggestions

- The user-defined part of an IPv6 address needs to be planned based on its summarization characteristics. Fields that are easy to summarize, such as the address space and area, should be placed in the left-most part. Excessive layering results in strong coupling between services and address planning, which hinders subsequent service development and reduces address space utilization. Therefore, you need to determine the number of fields based on site requirements.

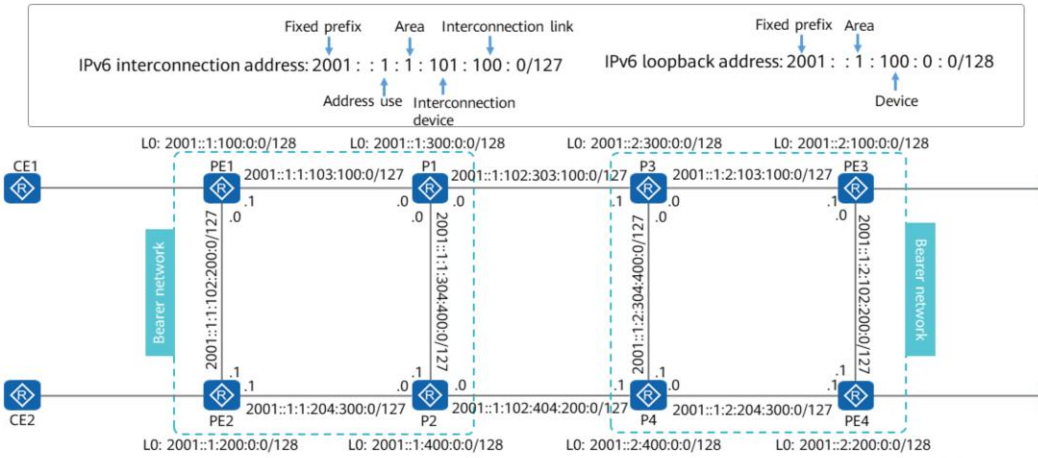


  - Fixed Prefix: indicates a fixed-length prefix applied for by an enterprise from an address allocation organization.
  - Subnet:
    - Attribute ID: is used to distinguish address types. It is used for level-1 address classification.
    - Network Type: identifies the type of a network.
    - Address Type: identifies the type of an address on the network.
    - Area ID: identifies an area on the network.
    - Allocatable Address Block: is reserved for future address allocation.
  - Interface Address: indicates the last 64 bits of an IP address. It is equivalent to the host ID in an IPv4 address.

HUAWEI

Typical IPv6 Address Planning

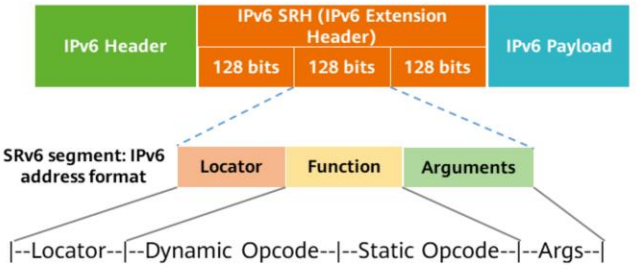
- IPv6 addresses can be planned based on planning suggestions or the customer's planning rules.



HUAWEI

SRv6 SID Overview

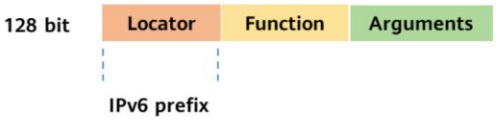
- An SRv6 SID usually consists of three fields: Locator, Function, and Arguments. The SRv6 SID is expressed in the *Locator:Function:Arguments* format. Note that the total length (Locator + Function + Arguments) is less than or equal to 128 bits. If the total length is less than 128 bits, the reserved bits are padded with 0s.
- If the Arguments field does not exist, the format is *Locator:Function*. The Locator field occupies the most significant bits of an IPv6 address, and the Function field occupies the remaining part of the IPv6 address.



- The locator is an IPv6 network segment. All IPv6 addresses in this network segment can be allocated as SRv6 SIDs. After a locator is configured for a node, the system generates a locator route. The node can be located based on the locator route. In addition, all SIDs advertised by the node can reach the node through the locator route.
- The Function field is also called opcode, which can be dynamically allocated using an IGP or statically configured using the **opcode** command. When configuring a locator, you can use the **static** *static-length* parameter to specify the length of the static segment, which determines the number of static opcodes that can be configured in the locator. When an IGP dynamically allocates opcodes, it applies for opcodes outside of the static segment range to ensure that SRv6 SIDs do not conflict.
- The Args field is determined by the **args** *args-length* parameter. The **Args** field is optional in SRv6 SIDs and is determined by the command configuration.

SRv6 Locator Overview

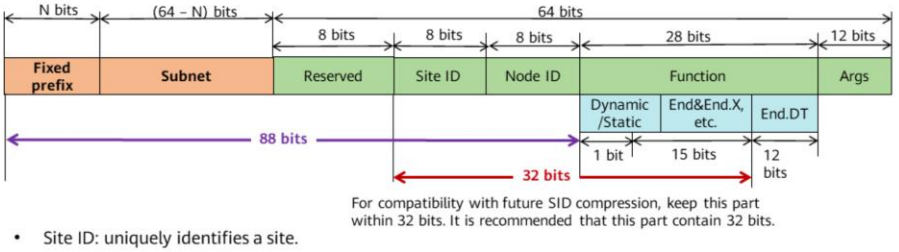
- The Locator field identifies the location of a network node, and is used for other nodes to route and forward packets to this identified node so as to implement network instruction addressing.
- A locator has two important characteristics: routable and aggregatable. After a locator value is configured for a node, the system generates a locator route and propagates the route throughout the SR domain using an IGP. A device advertising a locator route can be located by other devices on the same network based on the received locator route information, and all the SRv6 SIDs advertised by the device can be reached over the route.
- It is recommended that the same mask be used for the locators of all NEs on the entire network. After the mask is set, locators can be allocated in ascending order for the core, aggregation, and access layers in sequence.



HUAWEI

SRv6 Locator Planning Suggestions

- SRv6 locator addresses are allocated from the IPv6 host address field. Using hierarchical address allocation and reserving some address spaces are recommended for future expansion. An example is provided as follows:



For compatibility with future SID compression, keep this part within 32 bits. It is recommended that this part contain 32 bits.

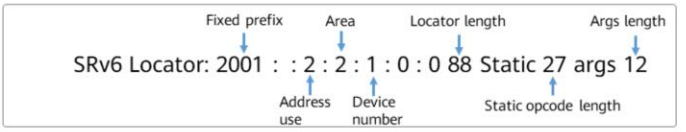
- Site ID: uniquely identifies a site.
- Node ID: uniquely identifies a device at a site.
- Function: indicates the Function field in an SRv6 SID.
  - Dynamic/Static: indicates whether a SID is a dynamic or static SID. 0 indicates a static SID (it is recommended that the End, End.X, and OAM-related SIDs be statically allocated). 1 indicates a dynamic SID (if many VPNs exist or VPNs change frequently, it is recommended that service SIDs, such as End.DT SIDs, be dynamically allocated).
  - End&End.X, etc: indicates the type of a SID. 0x000[1–F] indicates an End SID, and 0X1[peer site ID]X indicates an End.X SID.
- Args: indicates the parameter field in an SRv6 SID.

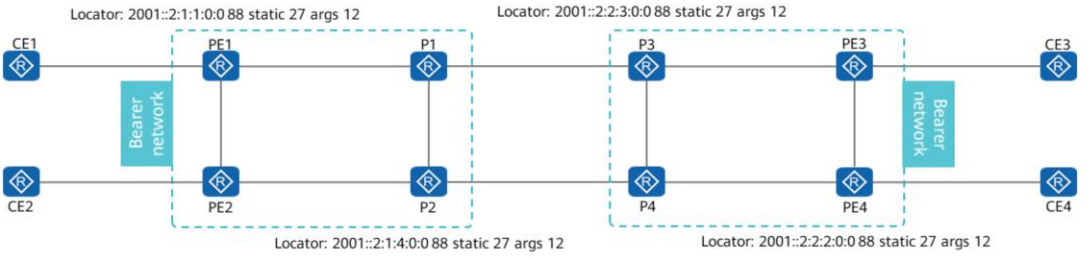
HUAWEI

- End.DT SIDs can be classified into End.DT4 SIDs and End.DT6 SIDs.
  - An End.DT4 SID (PE endpoint SID) identifies an IPv4 VPN instance on a network.
  - An End.DT6 SID (PE endpoint SID) identifies an IPv6 VPN instance on a network.
- In SRv6, End.OP SIDs are used to implement operation, administration and maintenance (OAM).
  - End.OP SIDs are mainly used in ping/tracert scenarios.

Typical SRv6 Locator Planning

- In SRv6 Locator Planning, it is recommended that the locator length not exceed 88 bits. Moreover, try to statically configure SRv6 SIDs, so that SRv6 SIDs remain unchanged after devices go online again. This facilitates management.





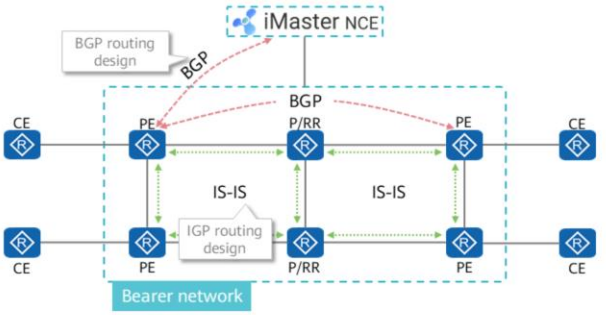
HUAWEI

Routing Design Roadmap for the Enterprise Bearer WAN

HUAWEI

Routing Design Overview

- Routing protocols are generally classified into IGP and BGP. These two types of protocols have different functions and therefore are designed for different purposes.
- Routing design can be classified into the following types:
  - IGP routing design
  - BGP routing design



HUAWEI

IGP Overview

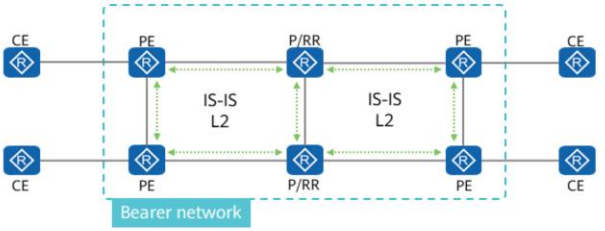
- On a bearer WAN, an IGP functions as a basic support protocol to collect and flood Layer 3 topology information on the entire network, and works with protocols such as TWAMP and iFIT to collect network status information, such as link delay.
- Generally, OSPF or IS-IS can be used on the backbone network for route reachability. However, the application scenarios of the two protocols are different to some extent.

	IS-IS	OSPF	Remarks
Network scale	Large	Small	Some devices support a maximum of 4K IS-IS routes and a maximum of 2K OSPF routes in a single area (with a single level).
IPv6 support	TLV-based packets are used, and no extra protocol needs to be independently deployed.	OSPFv3 needs to be independently deployed.	When OSPF supports both IPv4 and IPv6, OSPFv2 and OSPFv3 neighbor relationships need to be configured separately. The exchange of large numbers of protocol packets consumes a lot of resources.
SRv6 support	Related standards or drafts are available.	Related standards or drafts are available.	The standardization process of OSPF lags behind that of IS-IS in terms of SRv6 support. Not all vendors support SRv6-oriented OSPF extensions.
SR-MPLS support	Supported. The related functions are complete.	Supported. The related functions are basically complete.	

HUAWEI

IGP Route Planning

- Compared with OSPF, IS-IS supports SR-MPLS/SRv6 in a more comprehensive and scalable manner. Considering network expansion brought by enterprise service growth, large enterprise bearer networks generally start to use IS-IS. Therefore, it is recommended that IS-IS be preferentially used as the IGP for bearer networks.



- IS-IS: One IS-IS process is configured on the entire network, and an IS-IS level-2 area is configured in E2E mode.

HUAWEI

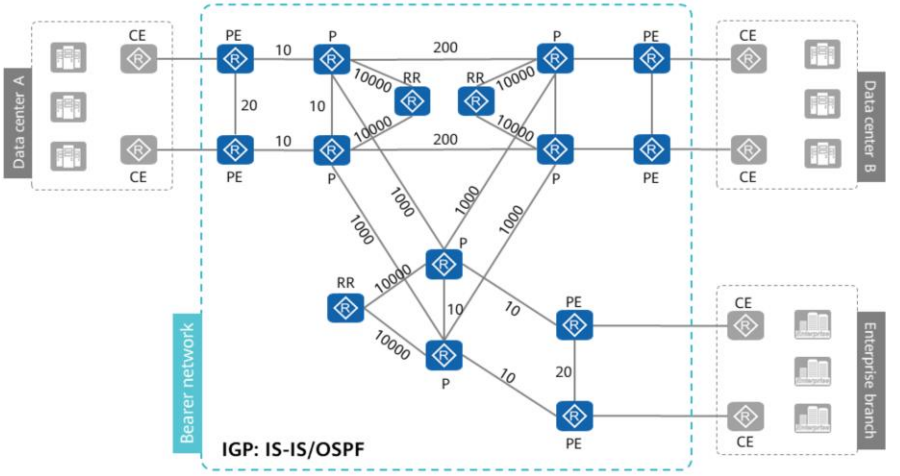
- It is recommended that some IGP parameters be set as follows:
  - IGP process ID: It is recommended that the IS-IS or OSPF process ID of a device on the backbone network be the same as the BGP AS number.
  - IS-IS NET: The recommended format is aa.bbbb.cccc.dddd.00. The loopback0 address of the device is used for NET derivation. For example, if the loopback0 address is 21.231.232.1, then the derived NET is 21.0231.0232.0001.00.
  - OSPF router ID: The global router ID is used. Generally, the router ID is the same as the loopback0 address.
  - Interface type: To speed up convergence, all interfaces are of the P2P type.
  - Route advertisement: IGP is mainly used to ensure the reachability of internal addresses on the WAN. Therefore, IGP advertises only interconnection interface addresses and device management addresses.

IGP Metric Planning

- The bearer network transmits different types of service traffic. When deploying an IGP on the bearer network, properly plan route metric to maximize bandwidth utilization, improve service quality, and ensure service reliability.
- IGP metric design rules:
  - Ensure that the metric of access-layer links is lower than that of aggregation-layer links.
  - Ensure that the metric of aggregation-layer links is lower than that of core-layer links.
  - Ensure that the metric of links between data centers is lower than that of WAN links between branches and data centers.
  - Ensure that inter-plane traffic between data centers preferentially traverses across planes through core nodes.
  - If a standalone RR is deployed, ensure that the metric of the link between the RR and core P is set to the maximum value (the RR only reflects routing information and does not forward data).

HUAWEI

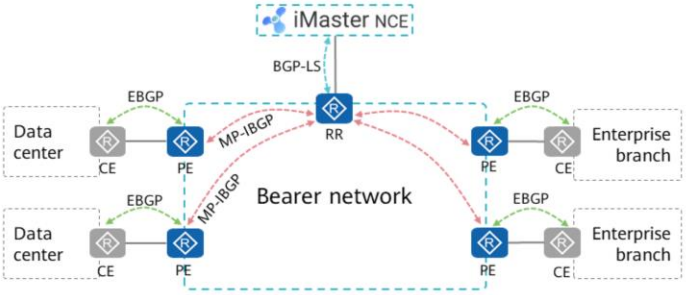
IGP Metric Planning Suggestions



HUAWEI

Overall BGP Planning

- After an enterprise bearer WAN is used to carry services that were used to be carried by multiple networks, it needs to use BGP/MPLS IP VPN to isolate these services.
- Generally, IBGP runs on the bearer network, and EBGP runs between the bearer network and other ASs (such as data centers and enterprise branches). PEs use BGP policies to control the transmission of VPN routes between ASs, achieving complex access control.
- If a controller is deployed on the bearer WAN, BGP-LS must be deployed between the controller and RR.

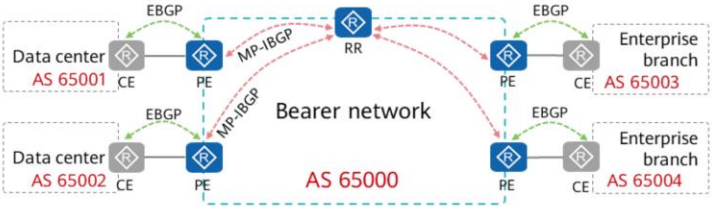


HUAWEI

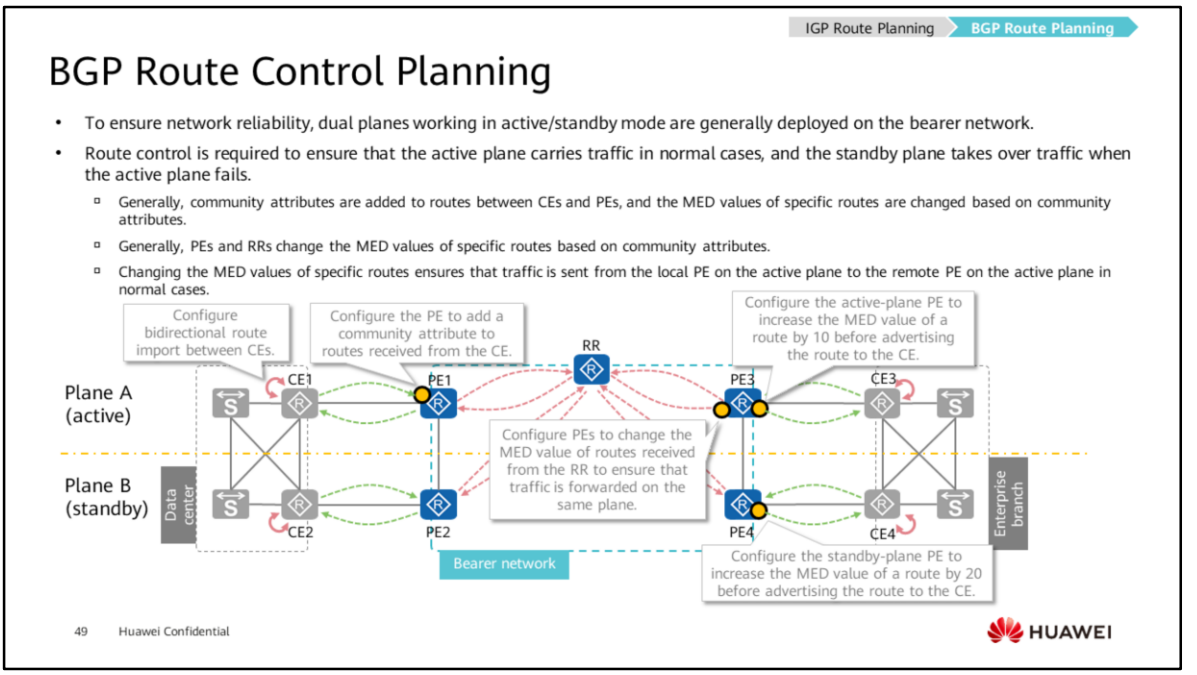
- In BGP peer relationship establishment, IBGP peer relationships are established using Loopback0 addresses, and EBGP peer relationships are established using interface addresses.
- AS: The bearer WAN can be classified as an independent AS.
- IBGP: PEs use loopback addresses to establish IBGP peer relationships with all RRs and use MP-IBGP to exchange VPN routes.
- EBGP: PEs use interface IP addresses to establish EBGP peer relationships with CEs. In inter-AS VPN route exchange scenarios, Option A is generally used.
- BGP-LS: The controller establishes BGP-LS peer relationships with all RRs to collect logical topology information on the backbone network.
- Deploy independent RRs and establish IBGP peer relationships for RRs on the backbone network.
- In addition to EBGP, IGPs such as OSPF, IS-IS, and RIP can also be used between PEs and CEs on the bearer network. Static routes can also be used to meet the requirements of flexible access in various scenarios.

BGP AS Planning

- An enterprise network usually uses a private AS number ranging from 64512 to 65534 during BGP deployment.
- It is recommended that one AS be deployed as the high-speed forwarding core of the entire bearer network, independent ASs be deployed for data centers and enterprise branches in different regions, and EBGP peer relationships be established between these ASs and the bearer network AS.



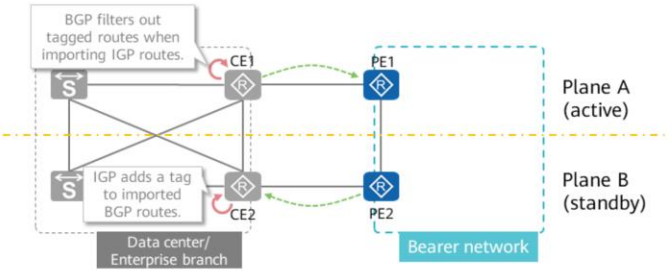
HUAWEI



- PE3 changes the MED value to 100 for the route whose next hop is PE1 (a PE on the same plane) and changes the MED value to 200 for the route whose next hop is PE2 (a PE on a different plane).
- PE4 changes the MED value to 100 for the route whose next hop is PE2 (a PE on the same plane) and changes the MED value to 200 for the route whose next hop is PE1 (a PE on a different plane).
- If PE1 and PE2 on the left learn the same VPN route and advertise the route to PE3 and PE4 on the right through the RR, PE3 and PE4 preferentially select the VPN route on the same plane as them. After the route is advertised to the CE, traffic from the CE preferentially travels along the route advertised by PE3 (because the MED value of the route advertised by PE3 is only increased by 10).

Design for BGP Routing Loop Prevention and Sub-optimal Route Prevention

- Generally, data center networks and branch networks do not run BGP. Their IGPs need to import routes learned by EBGP.
- When an IGP imports EBGP routes, it needs to add a tag to the routes, so that the IBGP peer can filter out imported BGP routes based on tags when importing local IGP routes to BGP.



HUAWEI

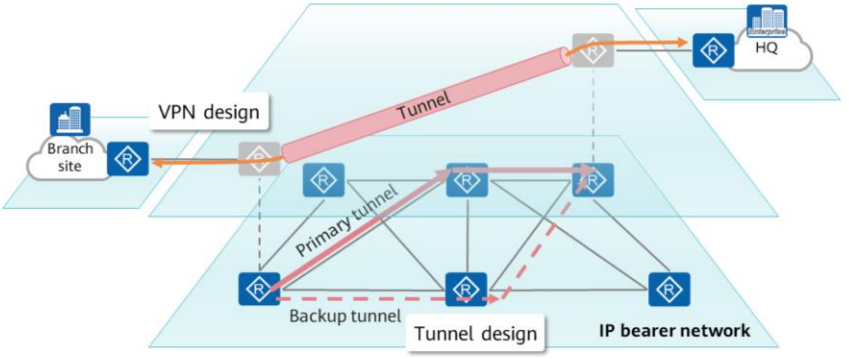
Contents

1. Current Situation and Challenges of the Enterprise Bearer WAN
2. Huawei CloudWAN Solution Overview
3. Basic Design for the Enterprise Bearer WAN
4. **Tunnel and VPN Design for the Enterprise Bearer WAN**
5. SLA and Reliability Design for the Enterprise Bearer WAN
6. Optimization and O&M Design for the Enterprise Bearer WAN

HUAWEI

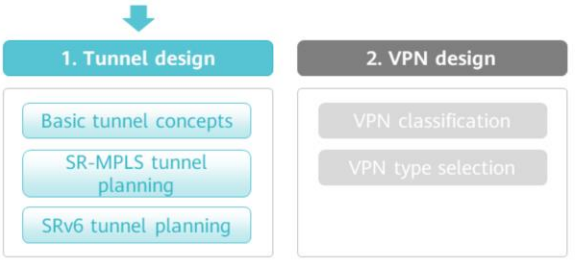
Tunnel and VPN Design Overview for the Enterprise Bearer WAN

- Enterprises usually use VPN to isolate services and SR to establish tunnels for traffic optimization and path planning.
- VPN traffic is carried over tunnels to isolate enterprise services while ensuring service quality.



HUAWEI

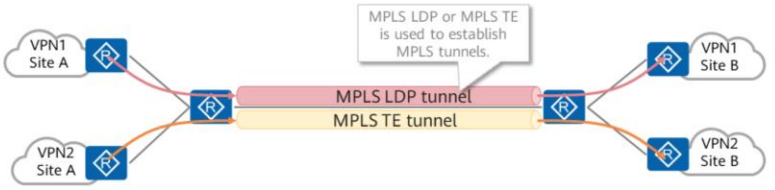
Tunnel Design Roadmap for the Enterprise Bearer WAN



HUAWEI

Tunneling Technology Overview

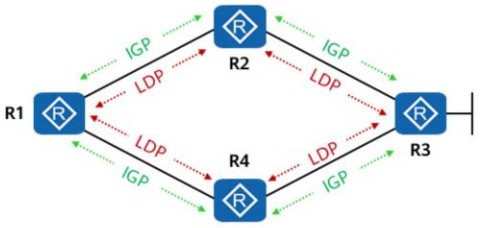
- The tunneling technology uses one protocol to encapsulate the packets of another protocol, and the carrier protocol itself can be encapsulated or carried by other protocols.
- For example, BGP/MPLS IP VPN commonly used on enterprise networks uses MPLS LDP or MPLS TE tunnels to carry VPN services.



HUAWEI

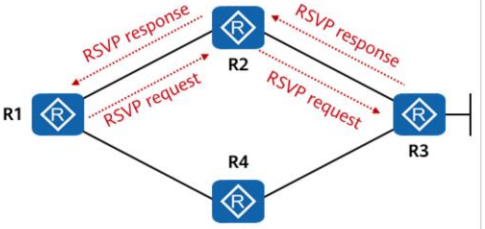
Issues with MPLS LDP and RSVP-TE

MPLS LDP



- LDP itself does not have the path computation capability and requires an IGP for path computation.
- Both the IGP and LDP need to be deployed for the control plane, and devices need to exchange a large number of packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- If LDP-IGP synchronization is not achieved, data forwarding may fail.

RSVP-TE

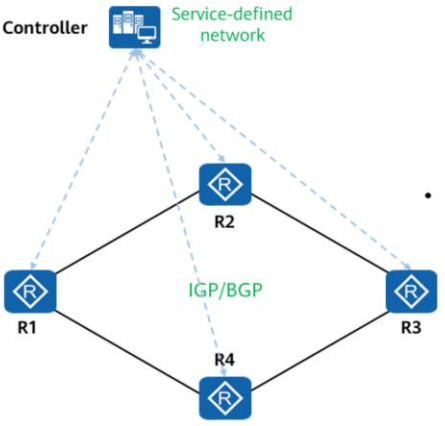


- RSVP-TE is complex to configure and does not support load balancing.
- To implement TE, devices need to exchange a large number of RSVP packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- RSVP-TE uses a distributed architecture, so that each device only knows its own state and needs to exchange signaling packets with other devices.

HUAWEI

- MPLS is a tunneling technology that guides data forwarding in essence and has complete tunnel creation, management, and maintenance mechanisms. The preceding mechanisms are driven by network operation and management requirements, not by applications.

Solution Provided by SR



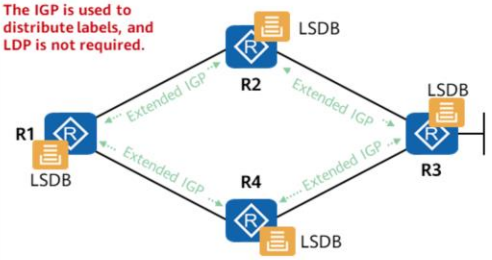
- Simplifying protocols and extending existing protocols
  - The extended IGP/BGP supports label distribution. Therefore, LDP is not required on the network, achieving protocol simplification. In addition, devices require only software upgrades instead of hardware replacement, protecting investments on the live network.
  - The source routing mechanism is introduced. The forwarding policy is instantiated into a segment list on the ingress to control the forwarding path of service traffic.
- Enabling networks to be defined by services
  - After an application raises requirements (e.g. delay, bandwidth, and packet loss rate), the controller collects information (e.g. network topology, bandwidth usage, and delay) and computes an explicit path according to the requirements.

HUAWEI

SR-MPLS Classification

SR-MPLS BE

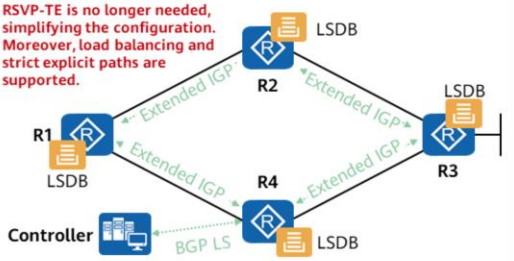
The IGP is used to distribute labels, and LDP is not required.



- An extended IGP has the label distribution capability. An LSDB can be formed based on the SRGB, prefix SID, link state, and other information distributed through the IGP, and the SPF algorithm can be used to compute the shortest forwarding paths based on labels. Therefore, LDP is no longer required on the network.
- Because only the IGP/BGP is used, the network does not have traffic blackholes and does not need to maintain LDP peer relationships.

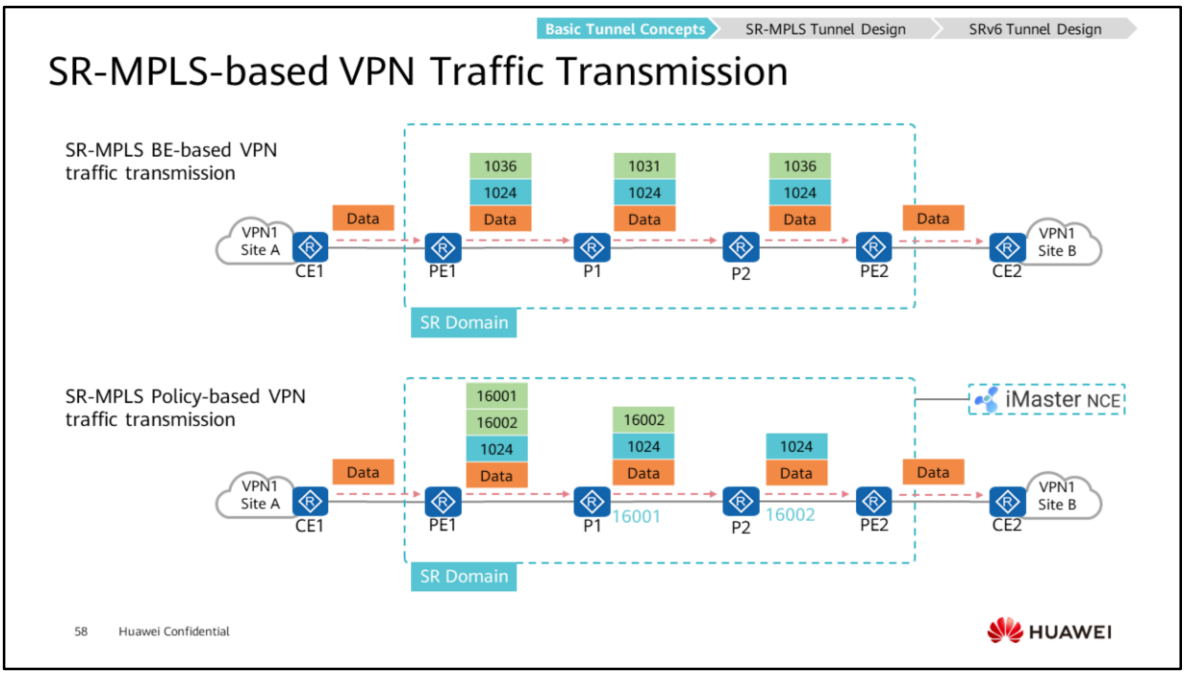
SR-MPLS Policy

RSVP-TE is no longer needed, simplifying the configuration. Moreover, load balancing and strict explicit paths are supported.



- Extended BGP/IGP sends the LSDB that carries label information to the controller, which then globally computes paths. There is no need to use RSVP-TE packets to apply for paths or maintain path status.
- Node SIDs and adjacency SIDs are used to implement load balancing and strict explicit paths.
- Path computation and service delivery are performed by the controller, simplifying configuration.

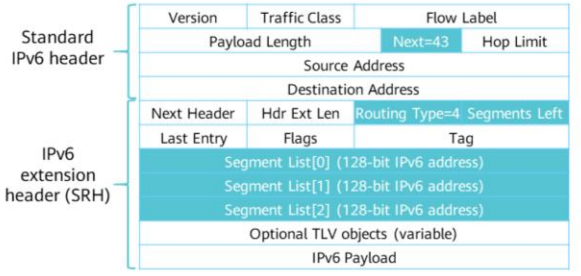
HUAWEI



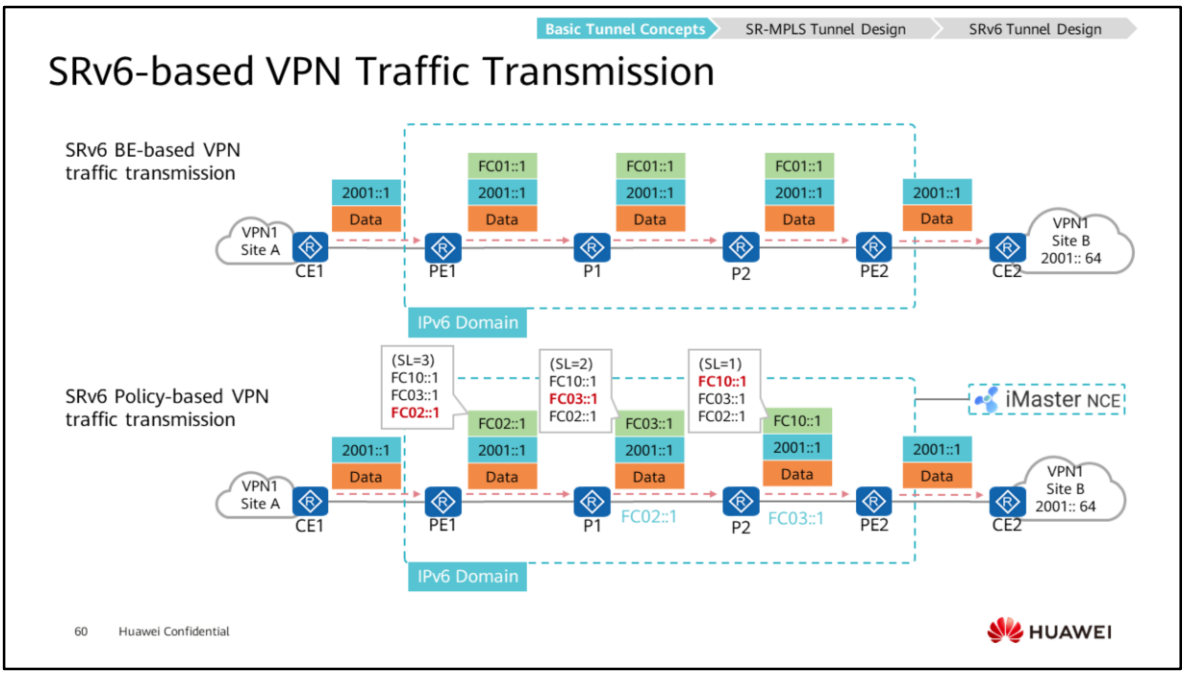
- The process of forwarding VPN traffic based on SR-MPLS BE is similar to the process of forwarding BGP/MPLS IP VPN traffic based on LDP.
  - After PE1 receives a VPN packet from CE1, PE1 searches the routing table and pushes two layers of labels into the packet. The outer label is a public network label, and the inner label is a private network label.
  - PE1 then sends the packet to P1, which swaps the outer label of the packet based on the SR-MPLS BE tunnel entry and sends the packet to P2. The process on P2 is similar to that on P1.
  - Upon receipt of the packet, PE2 sends the packet to a specific VPN site based on the inner label (PHP is not considered in this case).
- When an SR-MPLS Policy is used to carry VPN traffic, the forwarding path must be pre-computed and delivered to the ingress (PE1) as a segment list.
  - After receiving a VPN packet from CE1, PE1 searches the corresponding table and pushes the related segment list into the packet.
  - PE1 then sends the packet to P1, which determines the forwarding path based on the outer label, pops out the outer label, and sends the packet to P2.
  - After receiving the packet, P2 determines the forwarding path based on the outer label, pops out the outer label, and sends the packet to PE2.
  - PE2 sends the packet to the specified VPN site according to the inner label.

SRv6 Overview

- SRv6, or Segment Routing IPv6, is designed to forward IPv6 data packets on a network based on the source routing paradigm.
- SRv6 forwarding is no longer based on MPLS, simplifying the forwarding plane. SRv6 implements hop-by-hop forwarding by adding a Segment Routing header (SRH) into IPv6 packets, encapsulating an explicit IPv6 address stack into the SRH, and continuously updating destination addresses on transit nodes.



HUAWEI



- When SRv6 BE is used to carry VPN traffic, data packets carry two layers of IPv6 headers. The outer IPv6 header address is used to identify the VPN to which the data belongs, and the inner IPv6 header identifies the actual destination address of the data.
    - The outer IPv6 address is generated by the locator of PE2 and advertised to PE1 through BGP. PE2 advertises the locator to other devices in the form of a route.
    - After PE1 receives a packet destined for the destination network segment (2001:: 64), PE1 encapsulates the packet with an outer IPv6 header and forwards the packet based on the routing table.
    - Ps (P1 and P2) forward the packet based on the outer IPv6 header.
    - After receiving the packet, PE2 matches the packet with the corresponding VPN instance based on the outer IPv6 header and forwards the packet based on the routing table.
- When SRv6 Policy is used to carry VPN traffic, data packets carry two layers of IPv6 headers. The outer IPv6 header address is replaced by each hop based on the SRH information, and the inner IPv6 header identifies the actual destination address of the data.
    - Upon receipt of a packet destined for 2001:: 64, PE1 adds an outer IPv6 header (including the SRH) to the packet and sends the packet to the next hop based on the header.

SR-MPLS Tunnel Design Principles

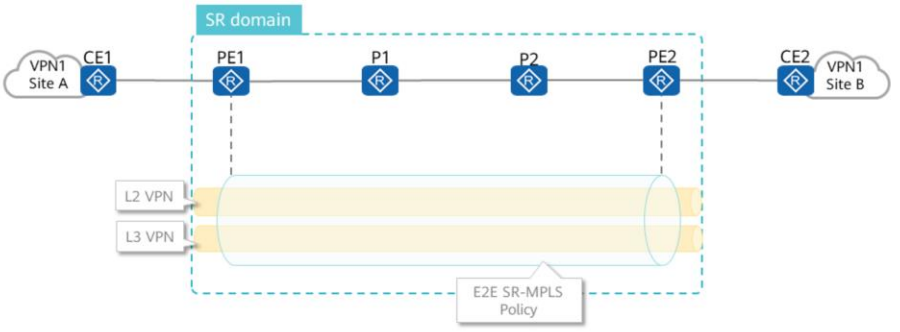
- Tunnels are intended to better serve services and focus on optimization requirements under the prerequisite that VPN bearer requirements are met. Comply with the following principles to achieve a balance among aspects such as service path visualization, service protection, network maintainability, and network scalability:
  - Service path visualization: Associate service traffic with tunnels to achieve some degree of path visualization.
  - Maintainability: Keep the total number of tunnels at an appropriate level to reduce live network maintenance pressure and shorten the optimization time.
  - Ease of optimization: Ensure that the traffic on each tunnel is not too heavy. Otherwise, bandwidth optimization will be difficult.
  - Reliability: Ensure that main services are under protection, and key services can be quickly converged.
  - Scalability: Consider possible network expansion in the future.

HUAWEI

- SR-MPLS BE tunnels are similar to LDP tunnels. Tunnel establishment depends on IGP design. Therefore, after IGP design is complete, SR-MPLS BE design is complete.
- This section mainly focuses on SR-MPLS Policy design.

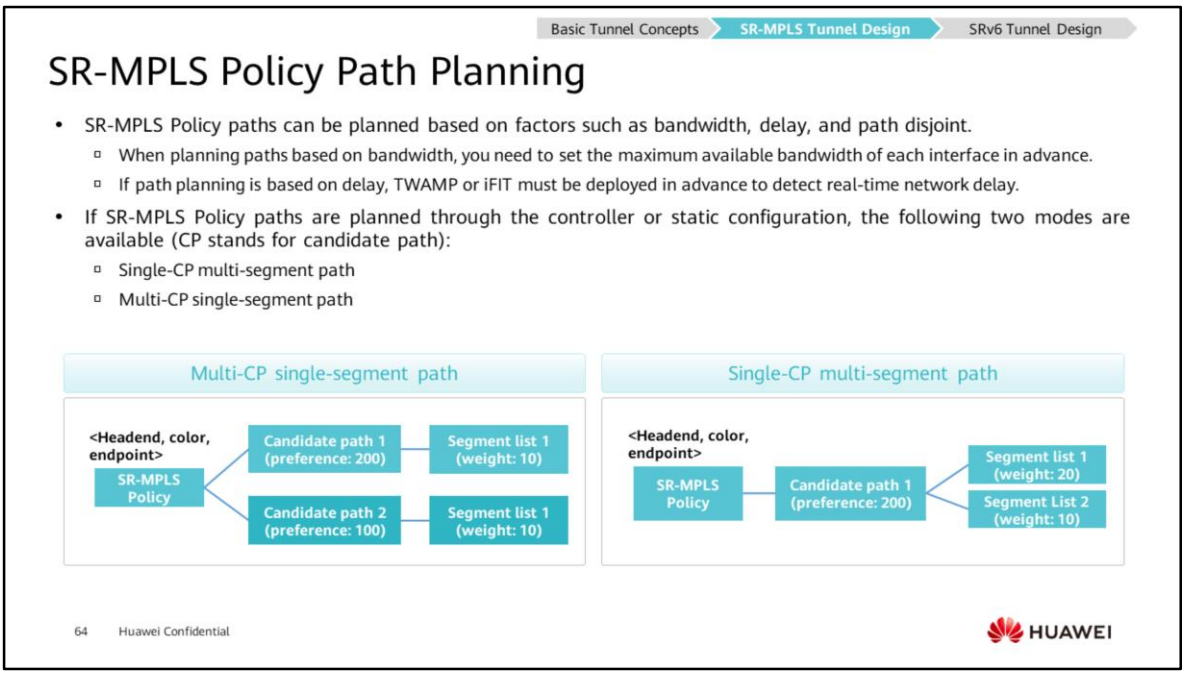
SR-MPLS Policy Deployment Design

- Generally, an enterprise bearer network belongs to an independent AS. Therefore, E2E SR-MPLS Policies can be deployed.
- An E2E tunnel can be deployed between the ingress (PE) where service traffic enters the bearer network and the egress (PE) where service traffic leaves the backbone network.

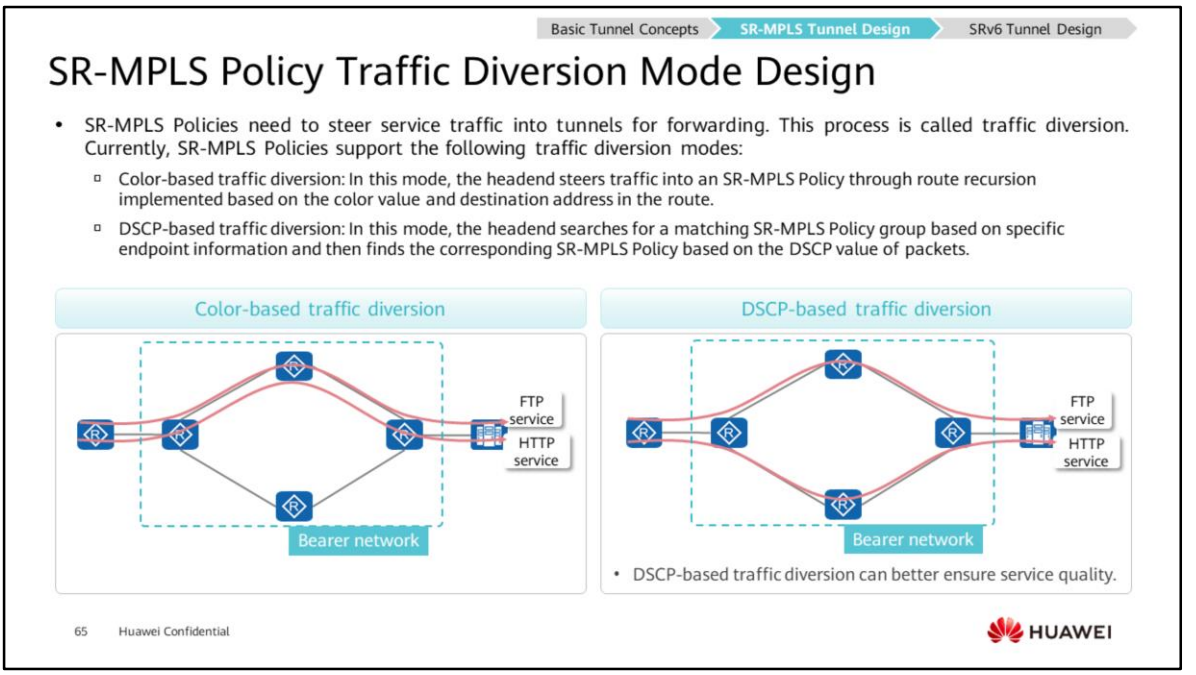


HUAWEI

- End-to-end deployment has the following characteristics:
  - Strong path control can be implemented through the planning of end-to-end paths (especially explicit paths).
  - Path visualization is relatively good.
  - If the network is large (for example, a network with multiple data centers and dozens of branches) and has 5,000 to 10,000 TE tunnels, the maintenance workload is heavy.
  - The scalability is fair. If a new data center or branch is added, a large number of end-to-end tunnels need to be added.



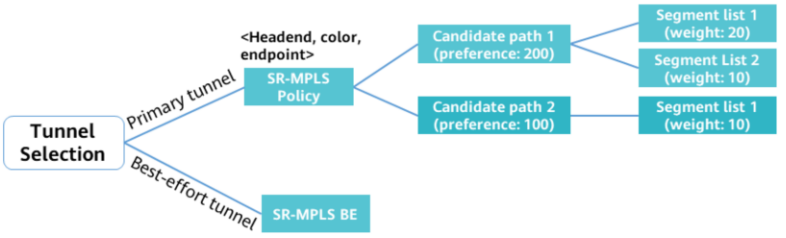
- If iFIT is used to measure the network delay, 1588v2 must be enabled on the entire network. Therefore, there are restrictions on application scenarios.
- TWAMP requires only NTP in network delay measurement.
- For a tunnel planned based on bandwidth, the actual traffic volume of the tunnel cannot be limited on devices after the tunnel is delivered. The traffic volume of a tunnel needs to be limited on the ingress, and the QoS or network slicing technology needs to be used.



- In color-based traffic diversion, different tunnels (including primary and backup tunnels) can only be selected based on endpoints. If different service traffic (such as HTTP and FTP traffic) is destined for the same address, color-based traffic diversion diverts the traffic to the same tunnel. As a result, the quality of some services deteriorates.
- In DSCP-based traffic diversion, different tunnels can be selected based on endpoint + DSCP information. If different service traffic (such as HTTP and FTP traffic) is destined for the same address, DSCP-based traffic diversion will steer different services into different tunnels based on the configuration, thereby ensuring the service quality.

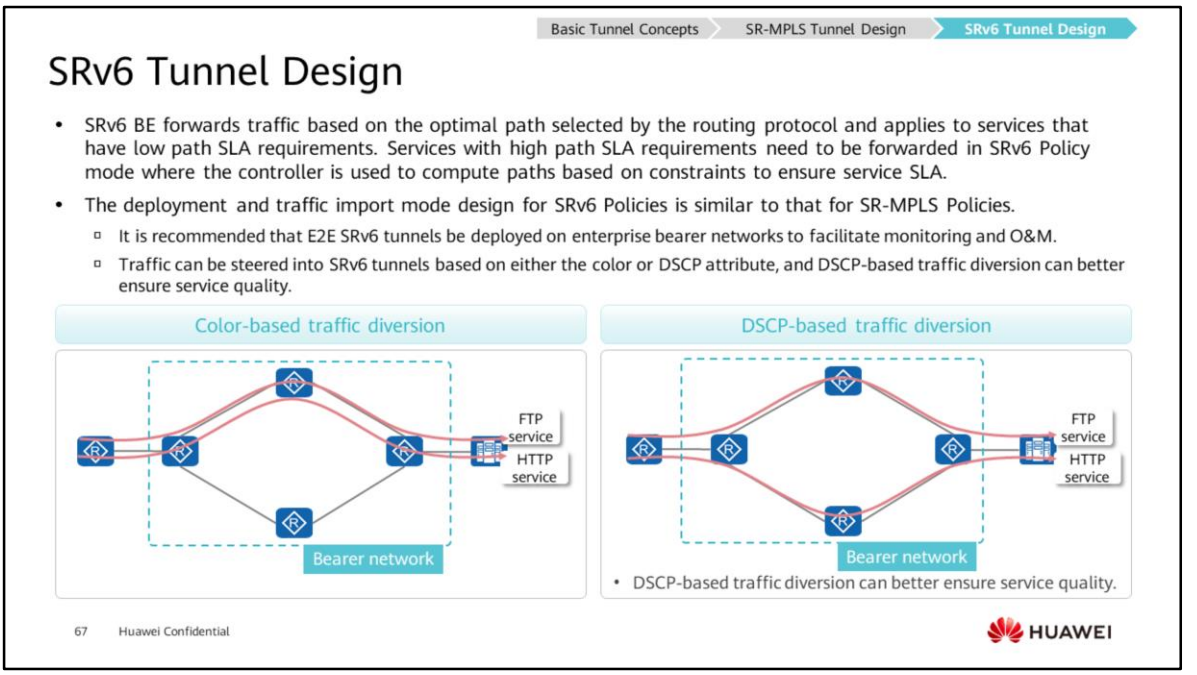
Best-Effort Forwarding Design for SR-MPLS Tunnels

- VPN services can be carried over either SR-MPLS Policies or SR-MPLS BE tunnels.
- An SR-MPLS Policy can contain multiple candidate paths with the preference attribute. The valid candidate path with the highest preference functions as the primary path of the SR-MPLS Policy, and the valid candidate path with the second highest preference functions as the HSB path.
- If all SR-MPLS Policies fail, VPN services can be carried over SR-MPLS BE tunnels.
- It is recommended that both SR-MPLS Policies and SR-MPLS BE tunnels be deployed. SR-MPLS Policies are preferentially used, and SR-MPLS BE tunnels serve as their backup.



HUAWEI

- An SR-MPLS Policy can have multiple candidate paths, such as CP1 and CP2. Each path is uniquely identified by a 3-tuple <protocol, origin, discriminator>.
- CP1 is the activated path because it is valid and has a higher priority. The two SID lists (also called segment lists) of CP1 are delivered to the forwarder, and traffic is balanced between the two tunnel paths based on weight. For example, traffic along the SID list <SID11, SID12> is balanced based on W1/(W1+W2). In the current mainstream implementation, a candidate path has only one segment list.
- If a controller is used to generate SR-MPLS Policies, only primary and backup tunnels can be established, and load balancing cannot be implemented for the primary tunnel.



- Similar to LDP tunnels, SRv6 BE tunnels are calculated based on IGP/BGP optimal paths. Unlike label-featured MPLS, SRv6 BE uses the shortest path first (SPF) algorithm to calculate forwarding paths based on SRv6 SIDs in an IGP domain. SRv6 BE requires only one segment to identify a forwarding path and the carried services. Traffic forwarding along paths depends on cost planning. Traffic is forwarded based on the least-cost route.
- This section describes SRv6 TE Policy design.

SRv6 Policy Path Planning

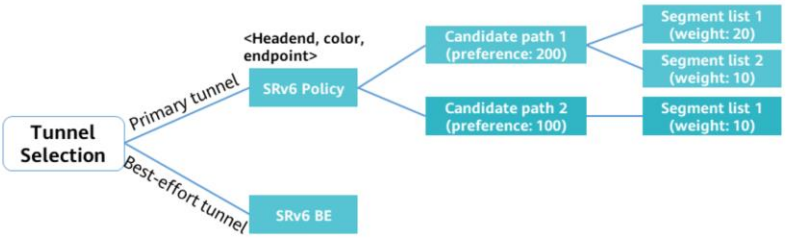
- SRv6 Policy paths can be planned based on factors such as bandwidth, delay, and path disjoint.
  - When planning paths based on bandwidth, you need to set the maximum available bandwidth of each interface in advance.
  - If path planning is based on delay, TWAMP or iFIT must be deployed in advance to detect real-time network delay.
- If SRv6 Policy paths are planned through the controller or static configuration, the following two modes are available:
  - Single-CP multi-segment path
  - Multi-CP single-segment path

HUAWEI

- If iFIT is used to measure the network delay, 1588v2 must be enabled on the entire network. Therefore, there are restrictions on application scenarios.
- TWAMP requires only NTP in network delay measurement.
- For a tunnel planned based on bandwidth, the actual traffic volume of the tunnel cannot be limited on devices after the tunnel is delivered. The traffic volume of a tunnel needs to be limited on the ingress, and the QoS or network slicing technology needs to be used.

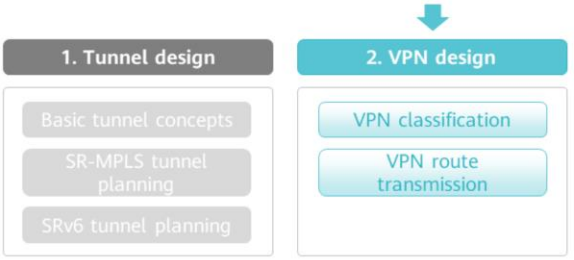
Best-Effort Forwarding Design for SRv6 Tunnels

- Similar to SR-MPLS, SRv6 can use either SRv6 BE tunnels or SRv6 Policies to carry VPN services.
- SRv6 BE tunnels are generally used as best-effort tunnels. If all SRv6 Policies fail, SRv6 BE tunnels are used to carry services.
- It is recommended that SRv6 BE tunnels and SRv6 Policies both be deployed. SRv6 Policies are preferentially used, and SRv6 BE tunnels serve as their backup.



HUAWEI

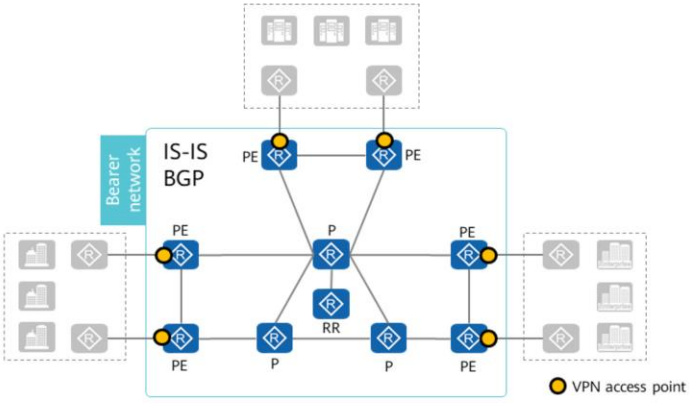
VPN Design Roadmap for the Enterprise Bearer WAN



HUAWEI

VPN Classification

- Different enterprises have different requirements for VPN division. Most enterprises use one VPN for major services (including production and office services), one VPN for non-major services, one VPN for external services, and one VPN for Internet services.



HUAWEI

IPv4 VPN Type Selection

- When SR-MPLS is used to carry IPv4 L3VPN traffic, the BGP VPNv4 or BGP EVPN address family can be used to transmit VPN routes.
- When SRv6 is used to carry IPv4 L3VPN traffic, it is recommended that the BGP EVPN address family be used to transmit VPN routes, so that IPv6 and Layer 2 services can be carried in the same manner.
- The L3VPN capabilities and implementation processes of VPNv4 and EVPN address families are basically the same.

	VPNv4	EVPNv4	Remarks
Forwarding plane	MPLS/SRv6	MPLS/SRv6	
VPN route learning	BGP	BGP	The BGP route formats are different.
Whether RR-based route reflection is supported	Supported	Supported	
Failover	Supported	Supported	VPN FRR, IP FRR, TE-HSB (MPLS), CBTS, TI-LFA (SRv6), Mirror-SID

HUAWEI

IPv6 VPN Type Selection

- When SR-MPLS is used to carry IPv6 L3VPN traffic, the MP-VPNv6 or BGP EVPN address family can be used to transmit VPN routes.
- When SRv6 is used to transmit IPv6 services, it is recommended that the BGP EVPN address family be used to transmit VPN routes.

	VPNv6	EVPNv6	Remarks
Forwarding plane	MPLS/SRv6	MPLS/SRv6	
VPN route learning	BGP	BGP	The BGP route formats are different.
Whether RR-based route reflection is supported	Supported	Supported	
Failover	Supported	Supported	VPN FRR, IP FRR, TE-HSB, CBTS, TI-FLA, Mirror-SID

HUAWEI

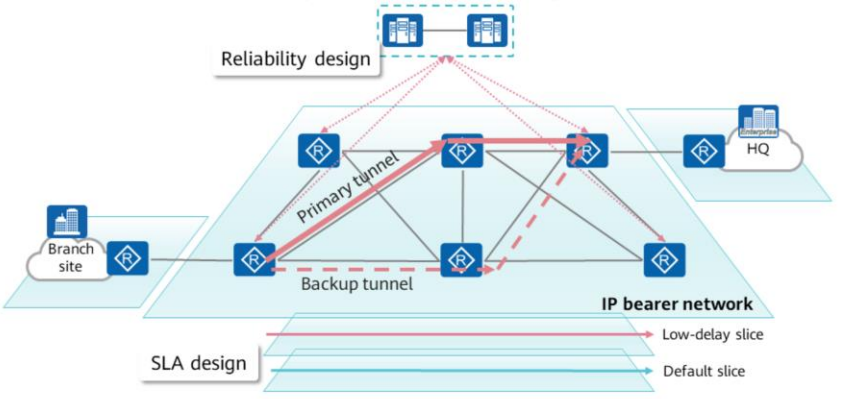
Contents

1. Current Situation and Challenges of the Enterprise Bearer WAN
2. Huawei CloudWAN Solution Overview
3. Basic Design for the Enterprise Bearer WAN
4. Tunnel and VPN Design for the Enterprise Bearer WAN
5. **SLA and Reliability Design for the Enterprise Bearer WAN**
6. Optimization and O&M Design for the Enterprise Bearer WAN

HUAWEI

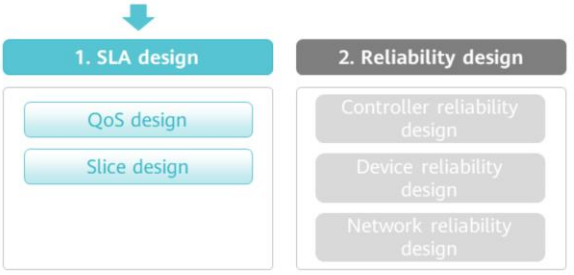
SLA and Reliability Design Overview for the Enterprise Bearer WAN

- The enterprise bearer WAN uses reliability and SLA technologies to effectively ensure the quality of carried services. Therefore, reliability and SLA design are very important.



HUAWEI

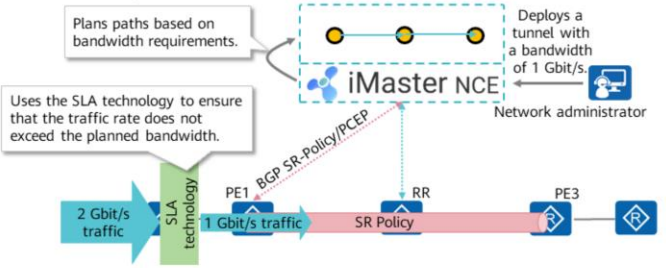
SLA Design Roadmap for the Enterprise Bearer WAN



HUAWEI

SLA Technology Overview

- In Huawei's CloudWAN solution, iMaster NCE-IP can compute paths based on bandwidth requirements and deliver forwarding paths (SR Policies) to network devices.
- Although the controller can compute paths based on bandwidth requirements, the delivered path information does not contain any traffic rate limiting policy. As a result, the traffic rate on the forwarding path (SR Policy) may exceed the planned bandwidth.
- To ensure that the traffic rate does not exceed the planned bandwidth, SLA technologies need to be deployed on the network to limit traffic bandwidth.
- SLA technologies mainly include QoS and network slicing.



HUAWEI

QoS Planning Principles

- The enterprise bearer network needs to provide differentiated service quality assurance for different services. QoS planning ensures that various services are properly forwarded on the bearer network. QoS planning mainly complies with the following four principles:
  - Reasonableness: Resources must be allocated appropriately based on the importance of services.
  - Consistency: QoS planning involves various behaviors (such as service classification, marking, scheduling, and rate limiting), which must be consistent on the entire network.
  - Scalability: Current QoS policies must take into account future service expansion.
  - Maintainability: Because services change rapidly in real-world situations, QoS policies may be frequently adjusted during routine maintenance. Ensure that QoS policies can be easily adjusted and maintained.

HUAWEI

QoS Design Suggestions

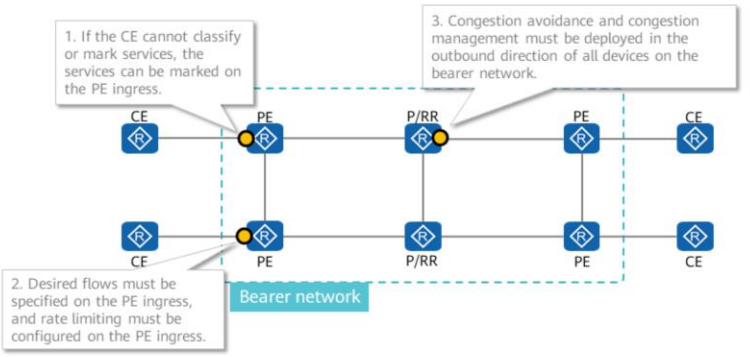
- When the network is normal, a different link can be planned for each service of an enterprise to ensure that these services do not affect each other. However, when a link is faulty, the affected services are switched to other links, and bandwidth competition may occur. QoS policies need to be deployed for key services in a unified manner.
- If there are more than eight types of enterprise services, they need to be properly classified and combined. The following is an example of common enterprise service planning.

Priority	Service	Scheduling Mode	Weight	Traffic Shaping	Drop Method
CS6/CS7	Protocol packet	PQ	NA	NA	Tail drop
EF	Production service	PQ	NA	NA	Tail drop
AF4	VoIP, video conference	PQ	NA	Configure rate limiting to prevent bandwidth starvation of low-priority services.	Tail drop
AF3	Video surveillance	WFQ	Determined based on live network conditions.	NA	WRED
AF2	Office service	WFQ	Determined based on live network conditions.	NA	WRED
AF1	OA service	LPQ	NA	NA	Tail drop
BE	Other workloads	LPQ	NA	NA	Tail drop

HUAWEI

QoS Deployment Design

- QoS planning requires that QoS features be properly deployed at different locations on the network. The following figure shows the deployment of different features on the network:

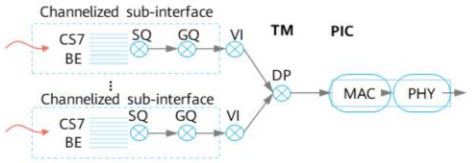


HUAWEI

Network Slicing Overview

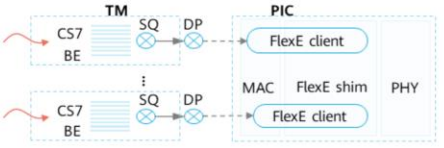
- Network slicing can be used to allocate dedicated network resources on a network to carry high-value service traffic.
- Network slicing and SR tunnels apply to different network layers. Network slicing reserves resources on the Layer 1.5 or Layer 2 network and can be used together with SR tunnels.
- Network slicing is generally implemented based on channelized sub-interfaces or FlexE.

Channelized sub-interface



- Channelized sub-interface: Queue resources are isolated. Hierarchical scheduling is used to implement flexible and refined management of interface resources, provide bandwidth guarantee, and work with the controller to provide E2E resource reservation.

FlexE

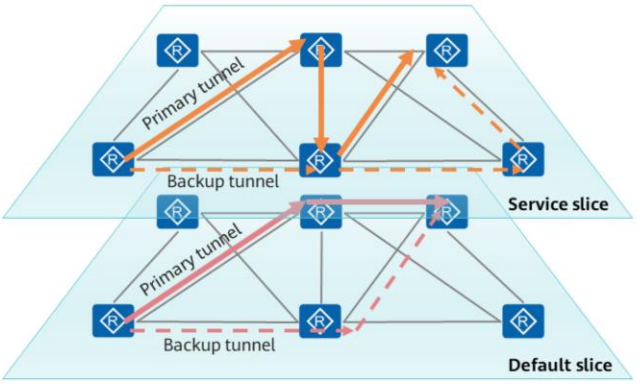


- FlexE: Queue and interface resources are isolated. Every resource is divided by TDM timeslot. This meets the requirements for exclusive resource use and resource isolation and provides flexible and refined management of interface resources.

HUAWEI

Network Slicing Design

- Currently, existing devices mainly implement network slicing based on bandwidth.
- A slice with the corresponding bandwidth is created based on actual service requirements. An SR tunnel is then bound to the slice for bearing.

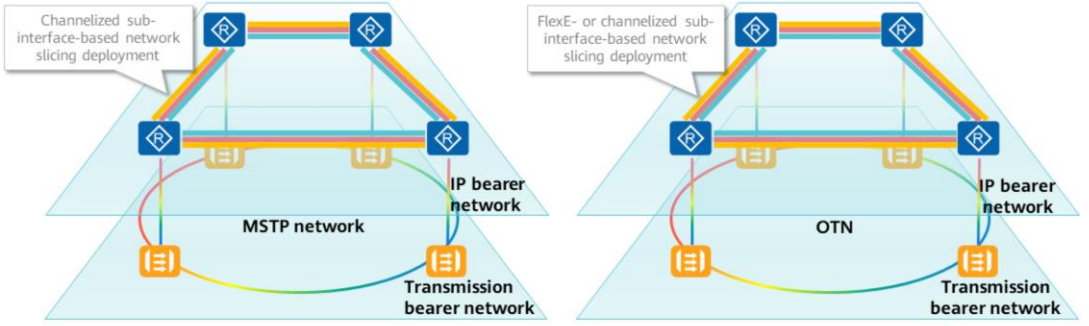


- SR BE tunnels or SR Policies can be deployed in the default slice.
- SR Policies are deployed in the service slice. BFD is deployed to detect tunnel status, implementing HSB protection and VPN FRR within slices.
- iMaster NCE-IP computes tunnels over slices based on affinity attributes.
- iMaster NCE-IP optimizes traffic within slices.

HUAWEI

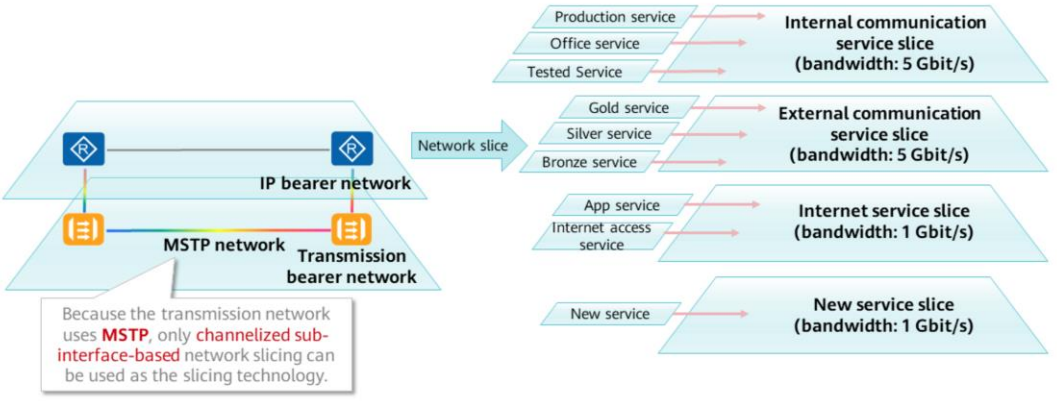
Network Slicing Application Scenarios

- FlexE-based network slicing and channelized sub-interface-based network slicing have different application scenarios.
  - It is recommended that FlexE be used to reserve resources for 50GE and higher-speed interfaces and channelized sub-interfaces be used to reserve resources for lower-speed interfaces.
  - Only channelized sub-interface-based network slicing can be deployed across MSTP devices.
  - FlexE- or channelized sub-interface-based network slicing can be deployed across OTN devices as required.



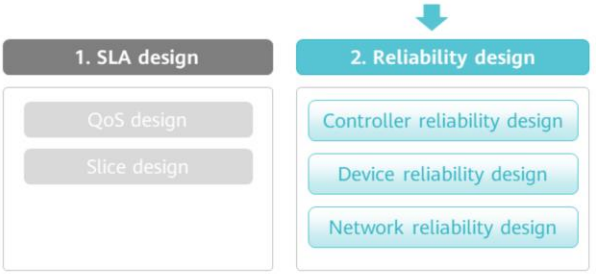
Typical Network Slicing Planning

- Services can be placed into different slices based on application types. Moreover, different slicing technologies can be used based on transmission lines (MSTP or OTN).



HUAWEI

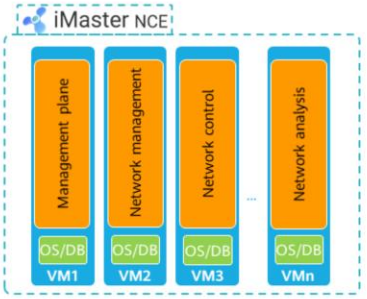
Enterprise Bearer WAN Reliability Design



HUAWEI

Controller Local Reliability Design

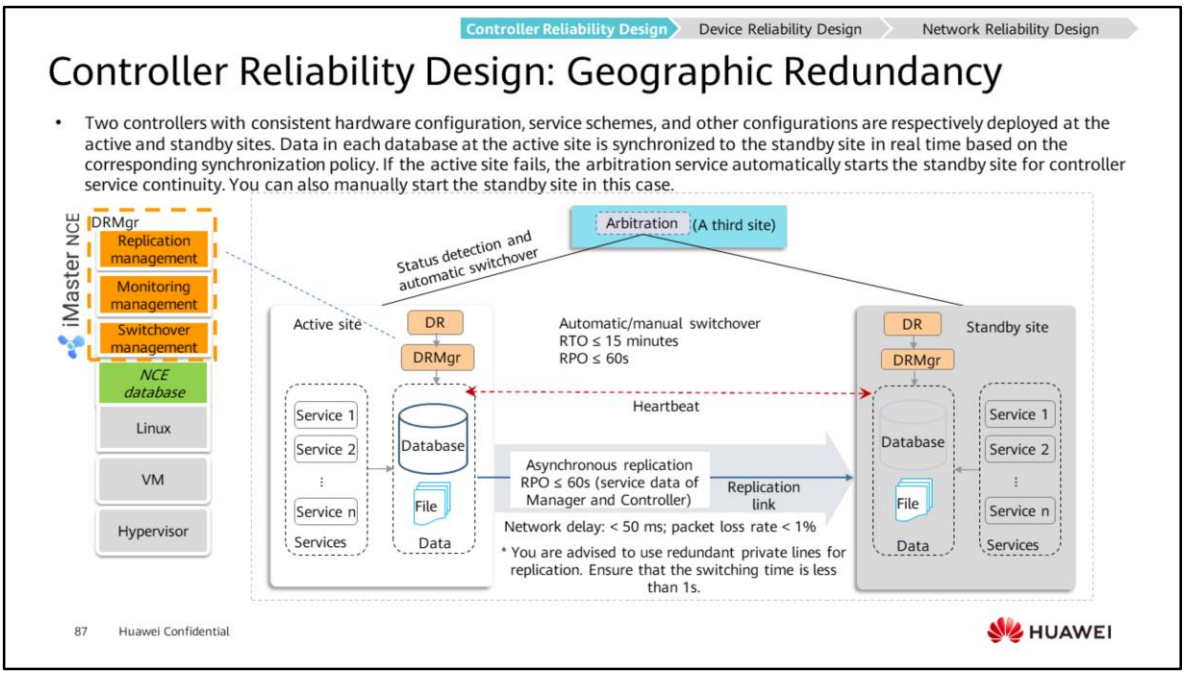
- iMaster NCE-IP has two high reliability modes:
  - Active/standby protection mode: Only the services on the active node are in the running status. If a service process on the active node encounters a fault, the controller automatically starts the service process on the standby node to provide services.
  - Cluster protection mode: When running properly, all cluster nodes are in the all-active state. If one node is faulty, other nodes share the load of the faulty node and continue to provide services evenly.



Protection Layer	Component	Remarks
Application layer	Manager	Protection mechanism: application active/standby protection Performance indicators: RPO: 0s; RTO: ≤ 5 minutes
	Controller	Protection mechanism: application active/standby protection Performance indicators: RPO: 0s; RTO: ≤ 60s
	Analyzer	Protection mechanism: application active/standby protection Performance indicators: RPO: 0s; RTO: ≤ 5 minutes
Database	Database	Protection mechanism: database active/standby protection Performance indicators: RPO: 60s; RTO: ≤ 60s
Virtualization layer	FusionCompute	N/A
Server	TaiShan/RH/E9000 server	Management module: 1+1 backup; power supply: 1+1 backup Network port: 1+1 backup; hard disk: RAID1/RAID10

HUAWEI

- RPO: recovery point objective
- RTO: recovery time objective

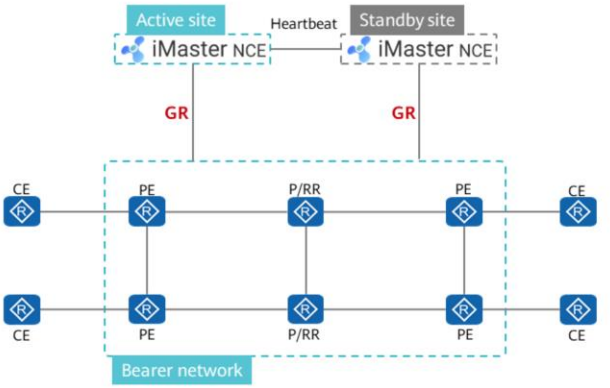


- During the operation of the DR system, NCE monitors the association status of the primary and secondary sites over the heartbeat link and synchronizes data over the replication link. If the heartbeat or replication link between the primary and secondary sites is abnormal, the controller reports an alarm. The fault can be either manually rectified or automatically processed by the arbitration service.
- The DRMgr service is developed by Huawei. Customers do not need to purchase third-party HA management software.

Controller Reliability Design | Device Reliability Design | Network Reliability Design

Control Network Reliability Design for the Controller

- GR needs to be deployed on both iMaster NCE-IP and devices, so that policy entries on forwarders can be retained for a longer time if the controller is faulty, maintained, or upgraded, or an active/standby controller switchover occurs.



HUAWEI

- Currently, the active/standby switchover of iMaster NCE-IP can be completed within 10 minutes. Therefore, the GR time cannot be less than 20 minutes.
- GR: graceful restart

Device Reliability Design

- High device reliability is essential to the effective running of the network and needs to be guaranteed through the hardware, software, and protection mechanisms.
- Device reliability deployment:
  - 1+1 active/standby protection for main control boards (NSR is recommended for smooth active/standby switchovers)
  - Load balancing among multiple microengines for interface boards
  - Maximum backup for power supplies
  - Redundancy protection for other key components (such as fan modules)

HUAWEI

- NSR: non-stop routing

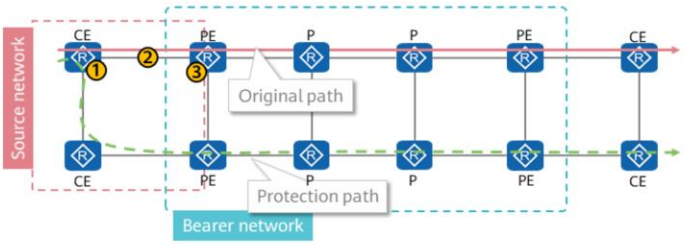
Network Reliability Solution Overview

- Multiple reliability technologies can be deployed on the bearer network to implement E2E reliability protection.
- Network reliability technologies include:
  - BFD/SBFD: is mainly used to check network connectivity.
  - IP FRR: is mainly used to provide link and node protection for transit networks and can work with the LFA/RLFA/TI-LFA algorithm.
  - Anycast FRR: is mainly used to provide protection for specific nodes, including transit and egress nodes.
  - HSB: is mainly used to provide E2E tunnel protection.
  - Mirror SID: is mainly used to provide protection for egress nodes.
  - Microloop avoidance: is mainly used to prevent temporary loops caused by inconsistent route convergence time on the entire network.

HUAWEI

Source Network Reliability Design

- The source network mainly includes source CEs, source PEs, and links between source CEs and source PEs. The source network may encounter the following faults:
  1. Source CE fault
  2. Link fault between a source CE and a source PE
  3. Source PE fault
- These three types of faults can be quickly detected through BFD, and IP FRR (mainly based on the LFA/RLFA algorithm) can be used to compute backup links, so that services can be quickly switched to the protection paths.

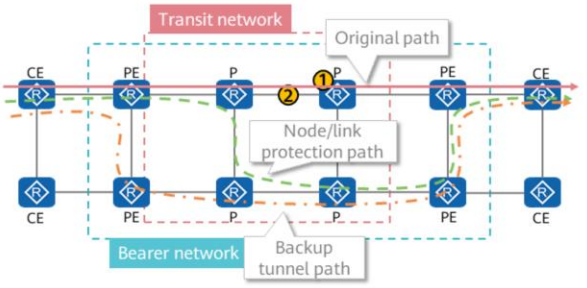


HUAWEI

Controller Reliability Design > Device Reliability Design > Network Reliability Design

Transit Network Reliability Design

- The transit network mainly includes Ps, links between Ps and source PEs, and links between Ps.
- The transit network may encounter the following faults:
  1. P fault
  2. Link fault between a P and a PE or between Ps
- The following method can be used to protect the network against these two types of faults:
  - Deploy BFD to quickly detect network link faults.
  - Use IP FRR (mainly based on the TI-LFA algorithm) to compute backup paths.
  - If the transit network cannot meet service requirements due to a fault, use the tunnel HSB technology to switch traffic to the backup path.

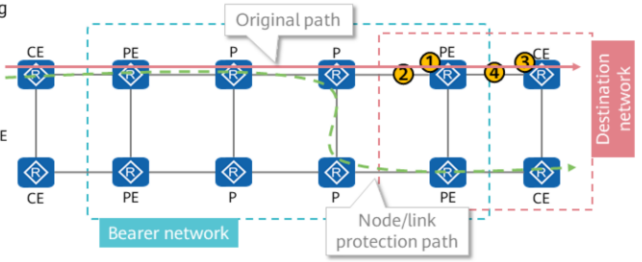


HUAWEI

- The transit network may fail to meet service requirements due to insufficient bandwidth or long delay. To detect the network bandwidth or delay, network quality detection technologies such as TWAMP or iFIT need to be deployed.

Destination Network Reliability Design

- The destination network mainly includes destination PEs, destination CEs, links between destination PEs and Ps, and links between destination PEs and destination CEs.
- The destination network may encounter the following faults:
  1. Destination PE fault
  2. Link fault between a destination PE and a P
  3. Destination CE fault
  4. Link fault between a destination PE and a destination CE
- The following method can be used to protect the network against the type 1 and type 2 faults:
  - Deploy BFD to quickly detect network link faults.
  - Use mirror SID or anycast FRR to switch traffic to the standby destination PE.
- The following method can be used to protect the network against the type 3 and type 4 faults:
  - Deploy BFD to quickly detect network link faults.
  - Compute backup paths through IP FRR (mainly based on the LFA/RLFA algorithm).



HUAWEI

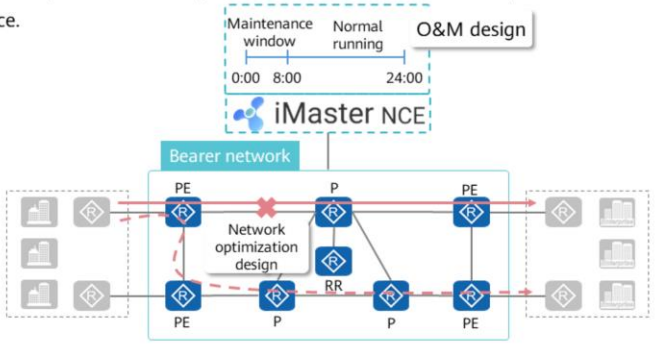
Contents

1. Current Situation and Challenges of the Enterprise Bearer WAN
2. Huawei CloudWAN Solution Overview
3. Basic Design for the Enterprise Bearer WAN
4. Tunnel and VPN Design for the Enterprise Bearer WAN
5. SLA and Reliability Design for the Enterprise Bearer WAN
6. **Optimization and O&M Design for the Enterprise Bearer WAN**

HUAWEI

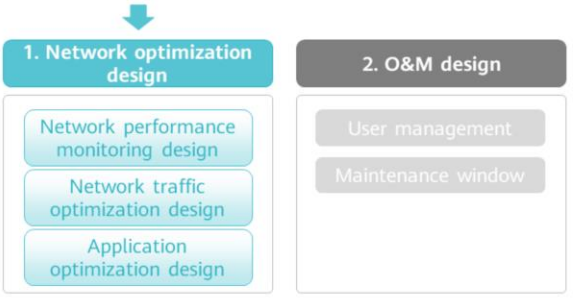
Optimization and O&M Design Overview for the Enterprise Bearer WAN

- Network operation becomes the new focus after the initial stage of network construction is complete. Network optimization and O&M are essential to smooth network operation.
- To better support subsequent network optimization and O&M, network optimization and O&M design must be performed in advance.



HUAWEI

Optimization Design Roadmap for the Enterprise Bearer WAN



HUAWEI

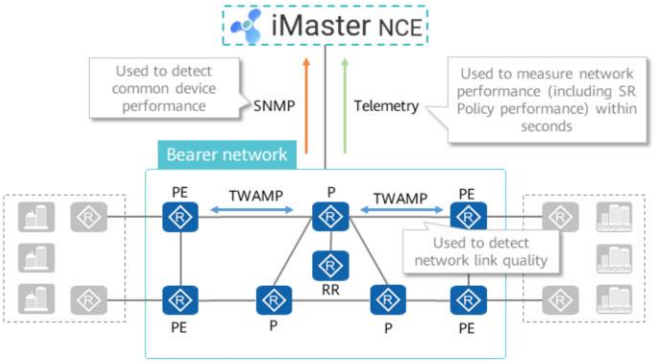
Network Performance Monitoring Overview

- Network performance monitoring can be implemented in multiple ways, one of the most common being SNMP. However, as services impose increasingly more requirements on networks, new technologies are required to quickly obtain network performance indicators.
- Network performance monitoring technologies commonly used on live networks include:
  - SNMP: During performance monitoring, SNMP obtains network performance information in a query-reply manner. When frequently obtaining information, SNMP imposes heavy pressure on the device.
  - Telemetry: Telemetry mainly obtains network performance information in a subscription-reporting manner. When frequently obtaining information, telemetry does not impose much pressure on the device.
  - NQA: NQA is mainly used to detect network quality. NQA uses simulated traffic to test the network environment. Therefore, the test result is not so accurate.
  - TWAMP: TWAMP is mainly used to detect network quality. Compared with NQA, TWAMP has a unified detection model and packet format, and is easy to deploy.

HUAWEI

Application Scenarios of Network Performance Monitoring

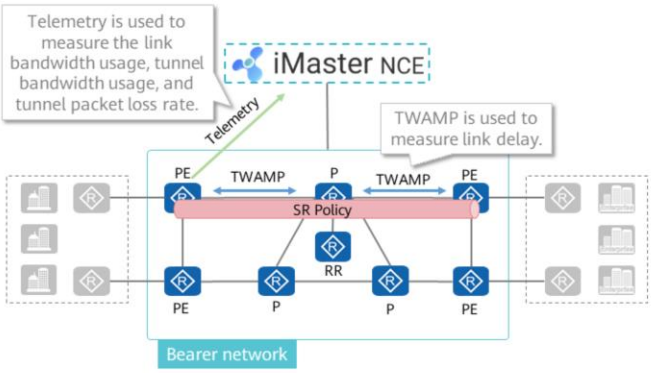
- In Huawei's CloudWAN solution, SNMP is used to collect common device performance information, telemetry is used to collect information that needs to be measured within seconds, and TWAMP is used to collect link quality-related performance data (including packet loss, delay, and jitter).



HUAWEI

Network Performance Monitoring Design

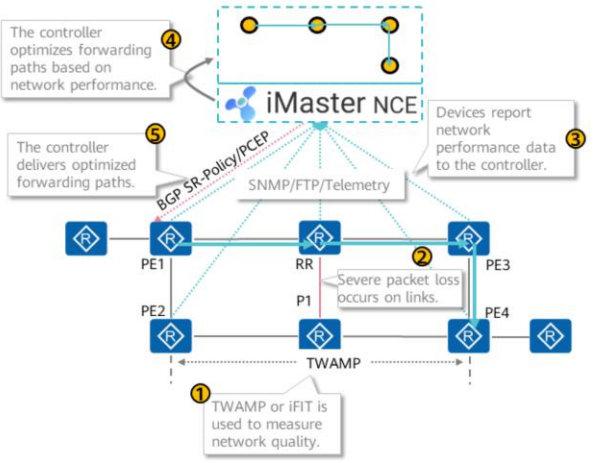
- The major performance indicators for WAN services include:
  - Link bandwidth usage (measured through telemetry)
  - Link delay (measured through TWAMP)
  - Tunnel bandwidth usage (measured through telemetry)
  - Tunnel packet loss rate (measured through telemetry)



HUAWEI

Network Traffic Optimization Overview

- MPLS TE uses the Constrained Shortest Path First (CSPF) algorithm to compute paths. This distributed computation and management mode, however, cannot manage network bandwidth resources in a coordinated manner.
- Huawei's CloudWAN solution uses BGP-LS (SRv6 Policy) or BGP-LS + SR-MPLS Policy (PCEP) to obtain tunnel status and adjusts tunnel paths based on the network status.

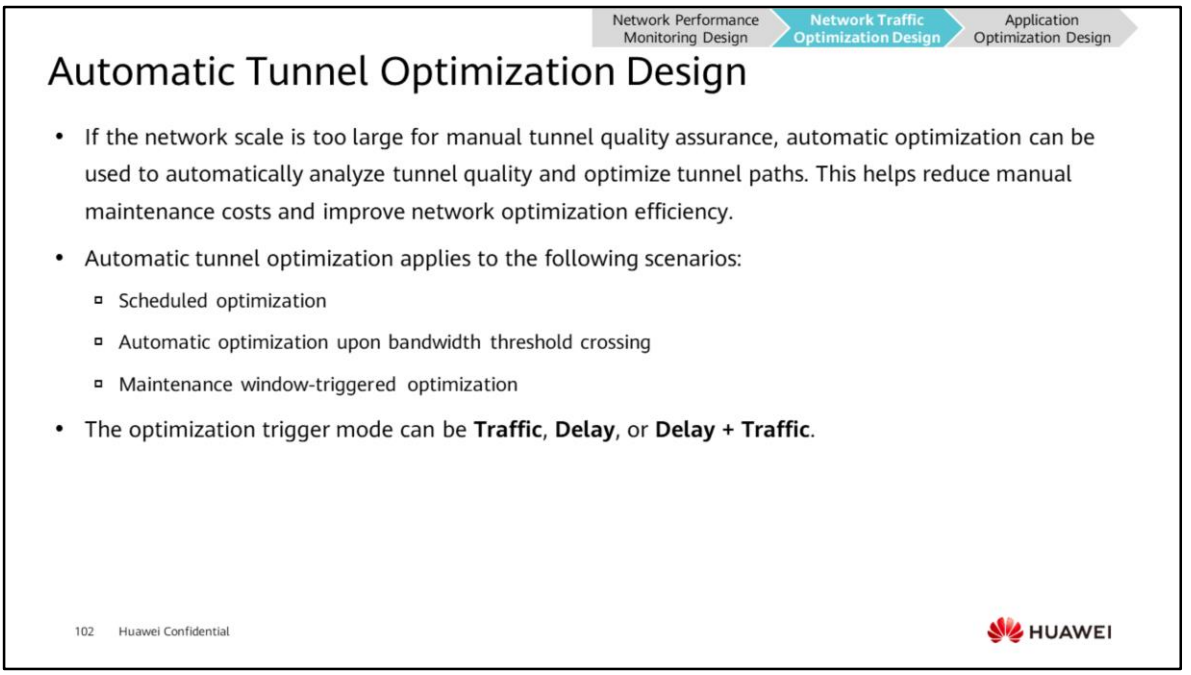


HUAWEI

Network Traffic Optimization Design

- Network optimization design focuses on two aspects: what to optimize and how to optimize.
- What to optimize
  - Simply put, optimization is to optimize service paths (LSPs). An LSP is a logical path equivalent to a tunnel, and a link is a physical path. One link can carry multiple tunnels, and one tunnel (LSP) corresponds to one service. Therefore, service paths can be changed based on either links or tunnels.
    - Link optimization: When one or more links are selected for optimization, all LSPs carried by the selected links are involved in path computation.
    - Tunnel optimization: When one or more tunnels are selected for optimization, the LSPs corresponding to the selected tunnels are involved in path computation.
- How to optimize
  - Optimization can be performed either automatically or manually:
    - In an automatic optimization scenario, traffic is automatically analyzed, and the optimization is automatically performed at scheduled times.
    - In a manual optimization scenario, traffic is manually optimized on demand based on network conditions.

HUAWEI



- Automatic optimization:
  - Scheduled optimization: You can set the interval for automatically optimizing network paths to 5 minutes or longer to ensure that the current service paths are optimal.
  - Automatic optimization upon bandwidth threshold crossing: You can set the link threshold. Then, when the bandwidth usage of a link exceeds the threshold, the system automatically adds tunnels over the link to the path computation queue and performs optimization when the optimization period arrives.
  - Maintenance window-triggered optimization: You can maintain a node or link. During the maintenance, the node or link is unavailable. After the maintenance starts or ends, the controller automatically recomputes the paths of tunnels that pass through the node or link.
  - The automatic optimization interval is an integer multiple of 5 minutes, for example, 10 minutes or 15 minutes.
- Optimization policy:
  - Traffic-based optimization: In this mode, the link threshold must be set. If the bandwidth usage of links exceeds the threshold, the controller determines whether to perform local or global optimization based on the number of threshold-crossing links.

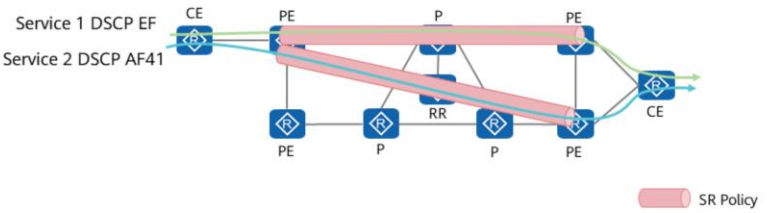
Manual Tunnel Optimization Design

- In manual optimization, the operator analyzes the network status and then adjusts the network manually. Manual optimization can be performed on demand at any time.
- Manual optimization is typically performed in the following scenarios:
  - If the current traffic paths do not meet requirements, you can modify link constraints and then manually trigger optimization to change traffic paths.
  - After configuring a service, you can manually perform network re-optimization to ensure that all tunnel paths are optimal.
  - When the link quality deteriorates (or the bandwidth usage is high but does not reach the automatic optimization threshold) or the bandwidth usage of links is uneven (some links have high bandwidth usage while others are idle), you can manually perform local/global optimization.

HUAWEI

Application Optimization Overview

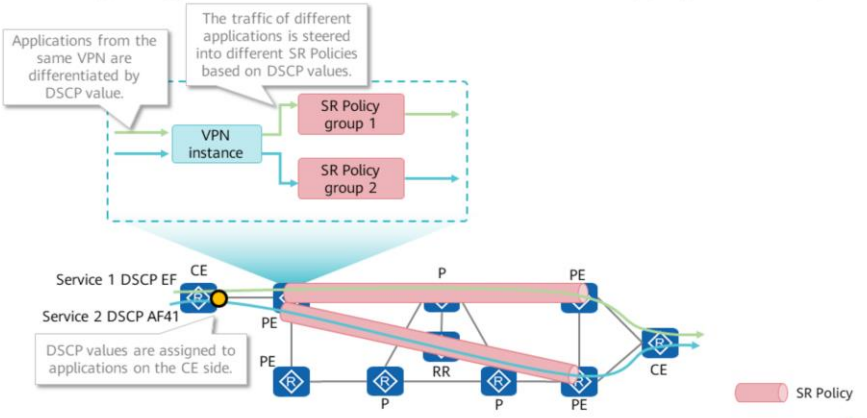
- On a network where VPN is deployed and application-based differentiated SLA path assurance is required, applications can be classified into different types and identified based on differentiated services code point (DSCP) values. The DSCP value corresponds to the color of a tunnel. Application traffic can be steered into the corresponding tunnel based on the DSCP value.



HUAWEI

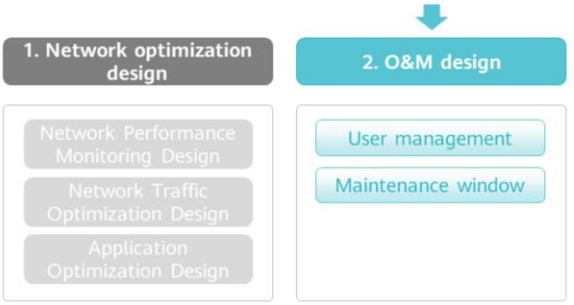
Application Optimization Design

- A different DSCP value must be assigned to each application prior to application optimization. Generally, DSCP values are assigned to applications on the CE side.
- To enable different types of applications to be carried over different tunnels, the SR Policy group+DSCP mode is generally used.



HUAWEI

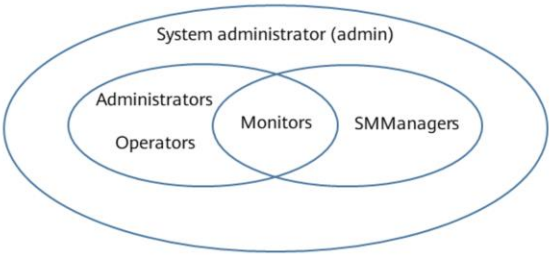
Optimization Design Roadmap for the Enterprise Bearer WAN



HUAWEI

User Type Overview

- Operations that can be performed by a user on the cloud WAN vary according to the user type. iMaster NCE-IP provides the following default user roles:
  - Administrators: Users attached to this role have permission to perform operations except managing users, querying security logs, querying personal security logs, and viewing online users.
  - SMManagers: Users attached to this role have permission to manage users, query security logs, and view online users.
  - Operators: Users attached to this role have permission to perform non-security-related operations.
  - Monitors: Users attached to this role have permission to view non-security-related functions, but do not have permission to perform the corresponding operations.



HUAWEI

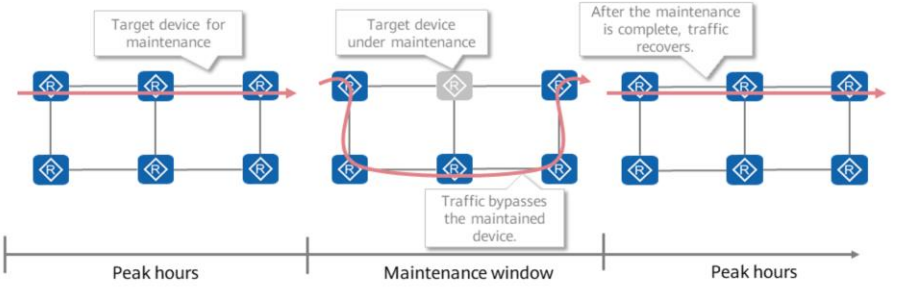
User Monitoring Overview

- By monitoring user sessions, the security administrator can learn information such as online users in the user system, IP addresses used by these users to access the system, access time, and roles of these users. When detecting that a user is attempting to perform unauthorized operations, the security administrator can send an instant message to the user or forcibly deregister the user.
  - A user session refers to the connection between a user and the system. A session starts when a user logs in and ends when the user deregisters or logs out. A user can generate multiple sessions.
  - The maximum number of online sessions allowed for a user is specified by the **Max. online sessions** parameter.
  - The function of monitoring user sessions does not involve users' personal information.

HUAWEI

Maintenance Window Overview

- When cutting over or maintaining devices and links, you can configure maintenance policies for them. The Network Path Navigation app allows you to configure maintenance policies for NEs and links.
- After the maintenance starts, the system generates new service tunnel paths that bypass the maintained devices and links. After the maintenance is complete, the system triggers path computation for traffic optimization.
- During the maintenance window, you can set the maintenance time and maintain the target devices. During the maintenance, the traffic does not pass through the maintained devices.



HUAWEI

Quiz

1. (Multiple-answer question) Which of the following tunnels can be used to carry IPv6 services? ( )
   A. SR-MPLS BE tunnel
   B. SRv6 BE tunnel
   C. SR-MPLS Policy
   D. SRv6 Policy

HUAWEI

1. ABCD

Summary

- When designing the enterprise bearer WAN, you need to design the infrastructure network first, including the physical network, IPv4/IPv6 addresses, and IGP/BGP.
- Tunnels, including SR-MPLS BE tunnels, SR-MPLS Policies, SRv6 BE tunnels, and SRv6 Policies, can be established on the infrastructure network. During tunnel design, pay attention to the tunnel path planning, traffic diversion, and best-effort forwarding mode.
- After a tunnel is established, it can carry VPN services. VPN planning must be based on enterprise service types and requirements.
- Reliability and SLA assurance are also very important for services. QoS and network slicing design helps ensure service SLA, and high reliability design for controllers, devices, and networks helps ensure service reliability.
- A network needs optimization and maintenance after it is constructed. Optimization design needs to cover both performance monitoring and traffic optimization. The maintenance scope needs to be divided based on user roles, and the network needs to be maintained at a proper time.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

IPE Key Technologies and Evolution Trends

HUAWEI

Foreword

- Given that IPv6 addresses are abundant, secure, and scalable, they have become an inevitable trend for network evolution. IP Enhanced Innovation (IPE) is a further development and application of IPv6 technologies and an upgrade of the IPv6-based next-generation Internet, and can create greater value for users.
- This course introduces the basic concepts, development trends, and key technical applications of IPE.
- This course also introduces IPv6 network evolution solutions for scenarios such as DCs, WANs, and campus networks, covering IPv6 network evolution phases and technologies and network architectures adopted in each phase.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the technical concepts related to IPE.
  - Describe the key technical applications of IPE.
  - Describe IPv6 network evolution trends in scenarios such as DCs, WANs, and campus networks.

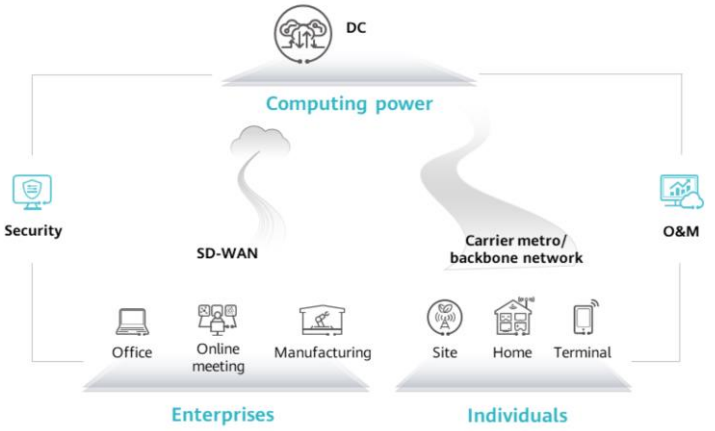
HUAWEI

Contents

1. **Technical Background of IPE**
2. Key Technical Applications of IPE
3. IPv6 Network Evolution Solutions

HUAWEI

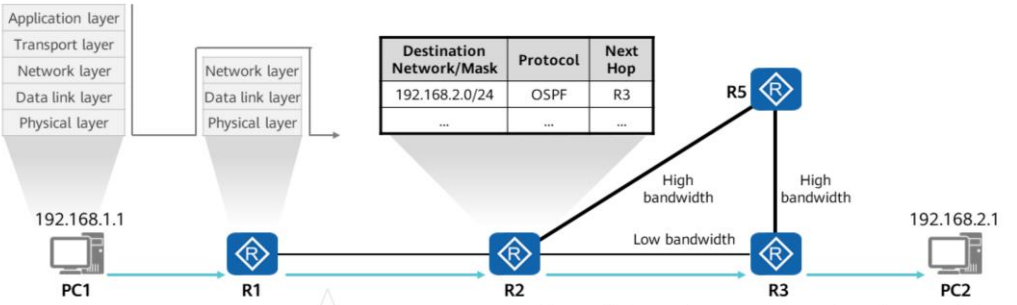
Datacom Network: Cornerstone for Digitalization



- The datacom network comprises a variety of datacom devices.
- The digital world connects individuals, enterprises, and computing power.
- The datacom network is the cornerstone for the digital world.

HUAWEI

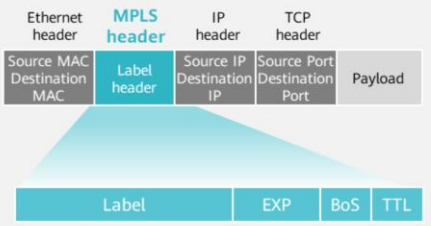
Understanding IP Forwarding on a Datacom Network Through a Network Model



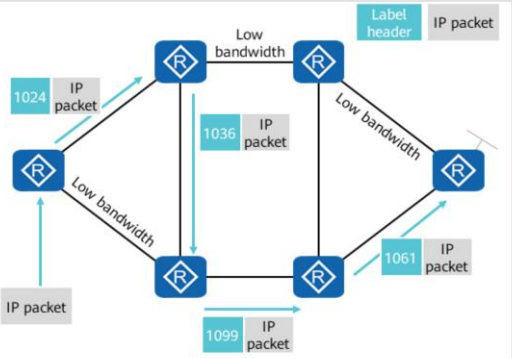
- IP is one of the most important protocols at the network layer.
- A device performs IP encapsulation on the payload and forwards the payload on the network.
- Network devices search their routing tables for a matching forwarding entry based on the destination IP address of the packet and forward the packet along the shortest paths.

HUAWEI

Emergence and Application of MPLS

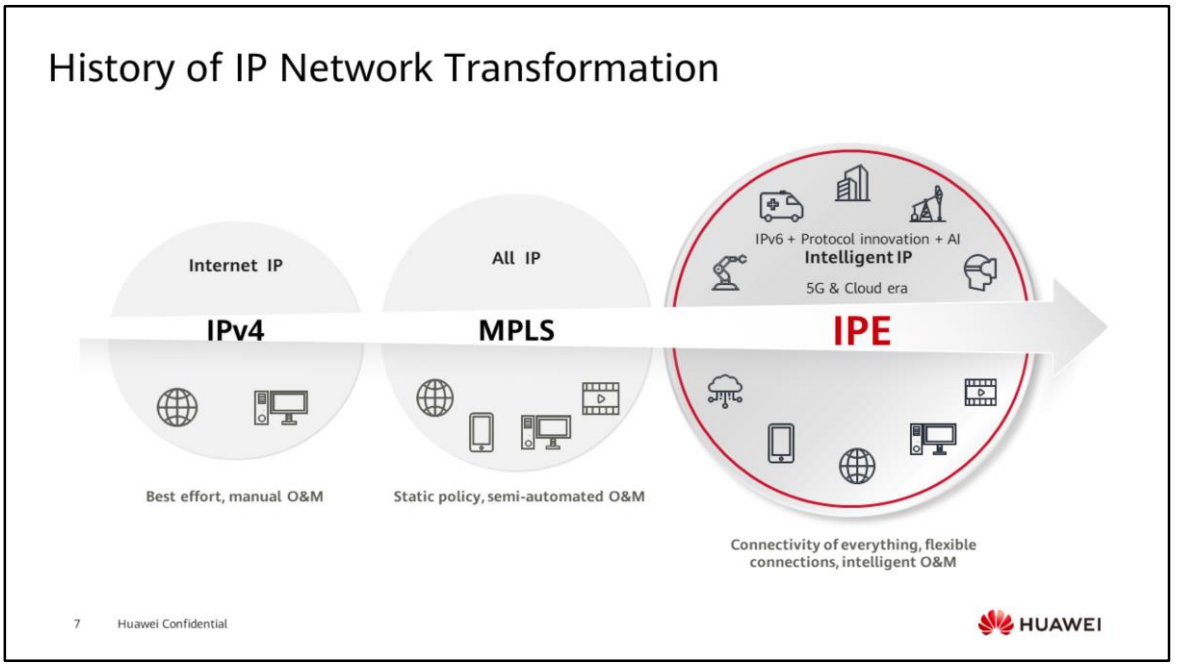


- Multiprotocol Label Switching (MPLS) is used on IP backbone networks.
- MPLS combines Layer 3 routing with Layer 2 switching by applying connection-oriented label switching on connectionless IP networks. Therefore, it has flexibility of IP routing and simplicity of Layer 2 switching.
- MPLS is essentially a tunneling technology and supports multiple upper-layer protocols and services.



- MPLS headers are added to IP packets for data forwarding based on MPLS labels.
- MPLS has higher forwarding efficiency, and MPLS labels can carry instruction information to guide data forwarding.
- It supports applications such as traffic engineering (TE), virtual private network (VPN), and QoS.

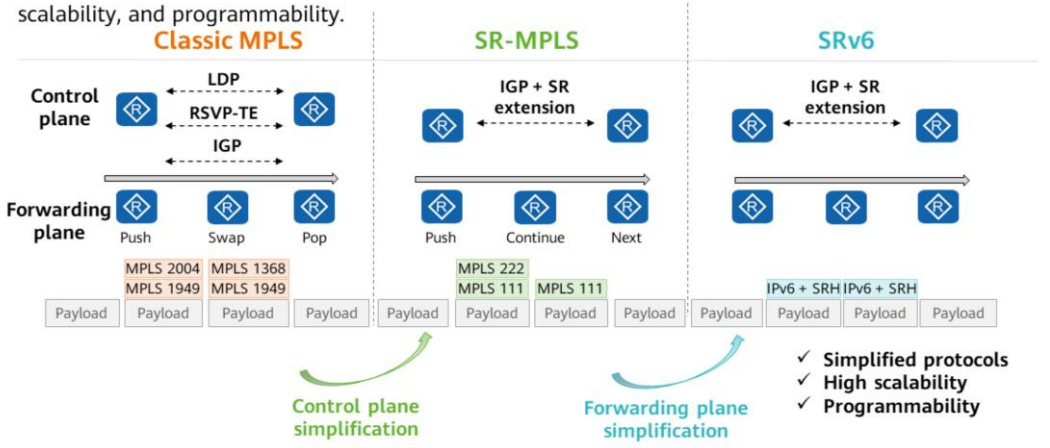
HUAWEI



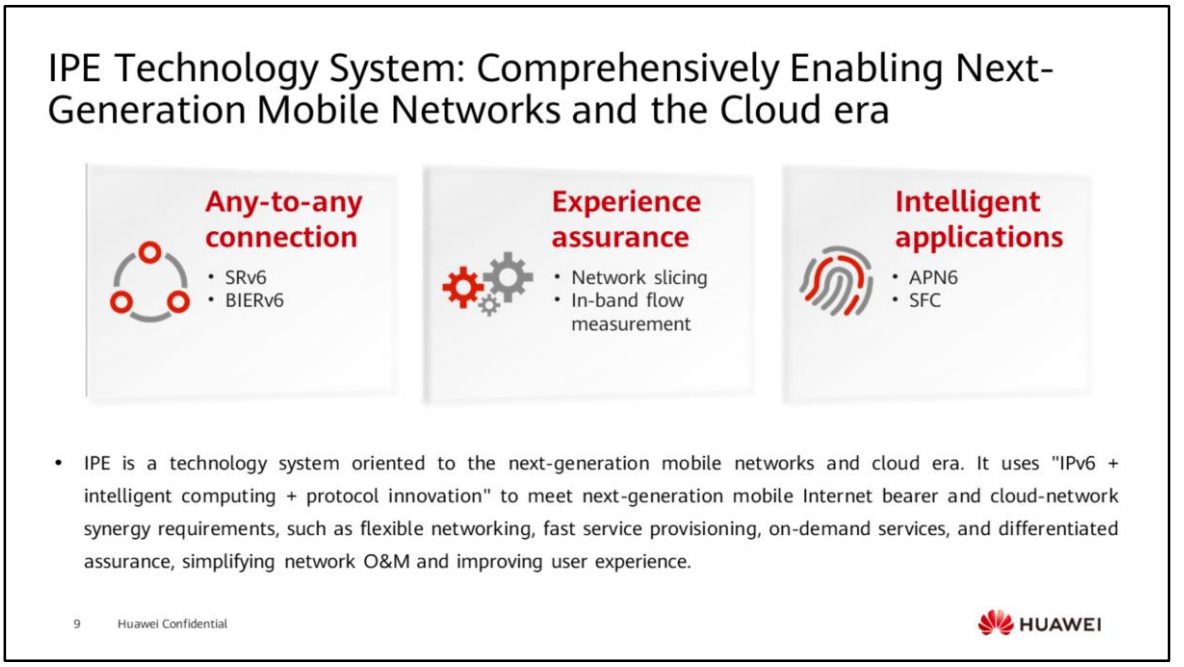
- After more than 40 years of development, IP networks have actually gone through three eras.
  - The first era is the Internet era. The iconic technology of this era is best-effort forwarding represented by IPv4.
  - The second era is the all-IP era. The core technology of this era is MPLS, which supports applications such as TE and E2E VPN.
  - Currently, we are experiencing the intelligence era of Internet of Everything, and the core technology of this era is IPE. This era is featured by the shift from human-to-human and human-to-machine communication to the Internet of Everything. There are huge bandwidth demands, large numbers of connections, and flexible connection models. The requirements for Service Level Agreement (SLA) assurance extend from providing only connectivity to providing comprehensive capabilities (such as strict delay, jitter, and packet loss guarantee capabilities), and O&M complexity increases accordingly. As a result, a new technology is required to empower this era, and the heavy responsibility of leading networks into the IPE era falls onto the SRv6 technology family.

From MPLS to SRv6

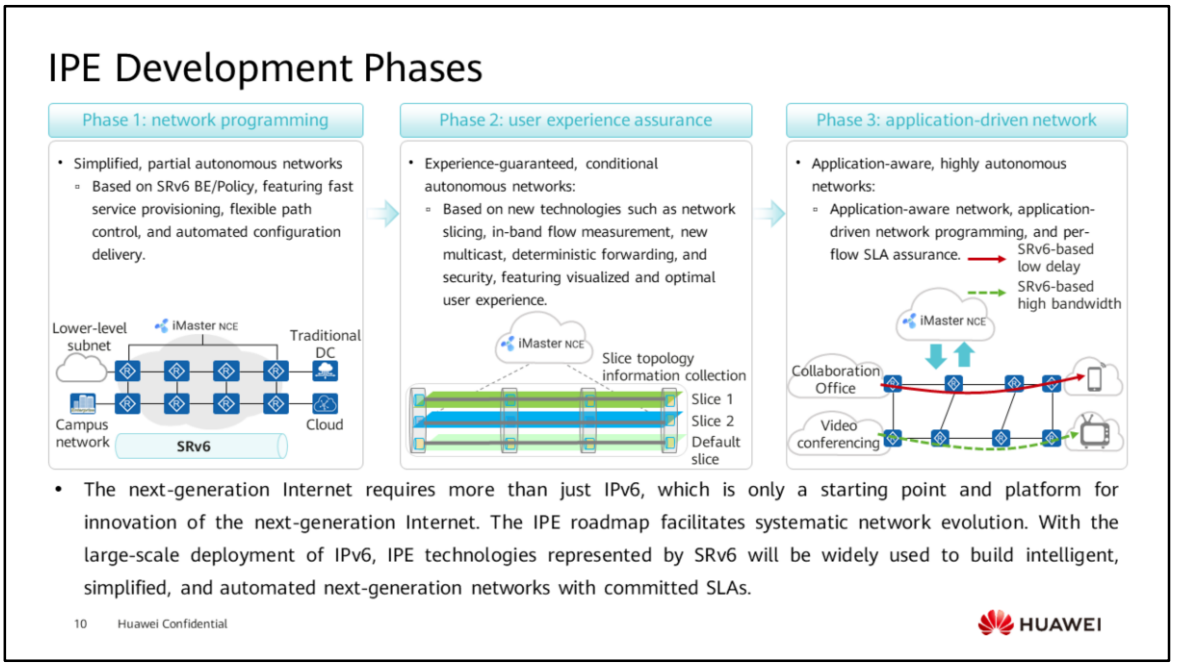
- SRv6 provides a unified forwarding plane and has advantages such as simplified protocols, high scalability, and programmability.



HUAWEI



- IPE technologies are innovations to the network technology system, represented by Segment Routing over IPv6 (SRv6), network slicing (FlexE), deterministic network (DetNet), in-band flow measurement (iFIT), new multicast (BIERv6), and application-aware IPv6 networking (APN6), and more.
- The IPE technology system is classified into the following three categories:
  - The first category provides basic connectivity and is represented by SRv6 (used to carry unicast services) and BIERv6 (used to carry multicast services). Both protocols carry and forward traffic based on the native IPv6 architecture, which eliminates the need of MPLS on the control and forwarding planes and takes advantage of IPv6's inherent scalability and other benefits.
  - The second category provides experience assurance and is represented by technologies used to reserve network resources, deliver slice-based deterministic network capabilities, and provide in-band flow measurement.
  - The third category comprises APN6 and SFC. APN6 provides application-level network service capabilities from the application perspective, and SFC provides flexible programming of network capabilities.
- FlexE: Flexible Ethernet
- DetNet: Deterministic Networking
- iFIT: in-situ Flow Information Telemetry
- BIERv6: Bit Index Explicit Replication IPv6 Encapsulation
- APN6: Application-aware IPv6 Networking

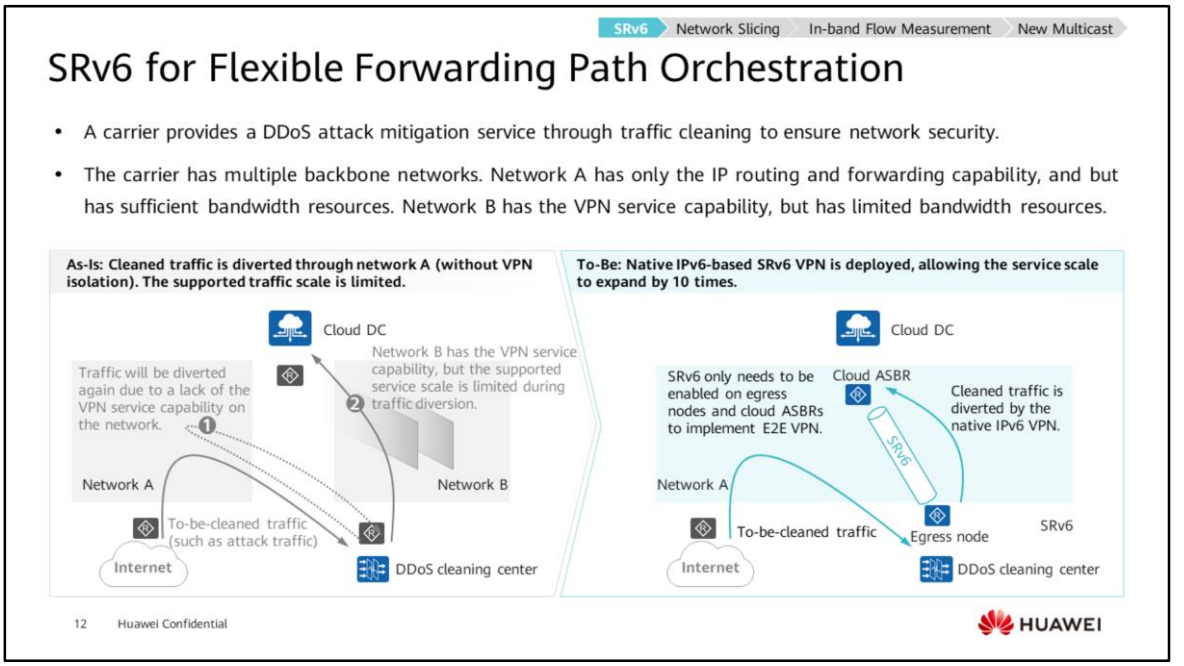


- Phase 1:
  - SRv6 is used to simplify protocols, resolve the low efficiency problem caused by segment-based configuration and manual cross-domain configuration on traditional MPLS networks, and achieve fast E2E service provisioning across domains.
  - In addition, SRv6 Policies can quickly adjust and optimize paths, improving network-wide flexibility and laying a foundation for autonomous networks.
- Phase 2:
  - Network slicing and FlexE are introduced to provide deterministic forwarding and differentiated dedicated network experience. Layer 2 FlexE and Layer 3 network slicing provide dedicated network resources for different services, E2E service isolation, and manageable and controllable paths. In this way, different services can be transmitted on the same network. Abundant slices ensure that each service can enjoy dedicated network experience and deterministic forwarding with controllable delay and jitter.
  - Moreover, the in-band flow measurement mechanism reports service traffic status in real time, enabling network experience visualization. This allows us to determine whether service quality deteriorates in a timely manner. If service quality deteriorates, we can adjust service configurations in a timely manner to ensure optimal user experience.
- Phase 3:
  - In this phase, we not only need to ensure service experience, but also need to deliver experience assurance at a finer granularity (application level).
  - IPv6 apps can directly carry application information in IPv6 extension headers, enabling the network to identify user and application information and provide fine-grained differentiated network channels.

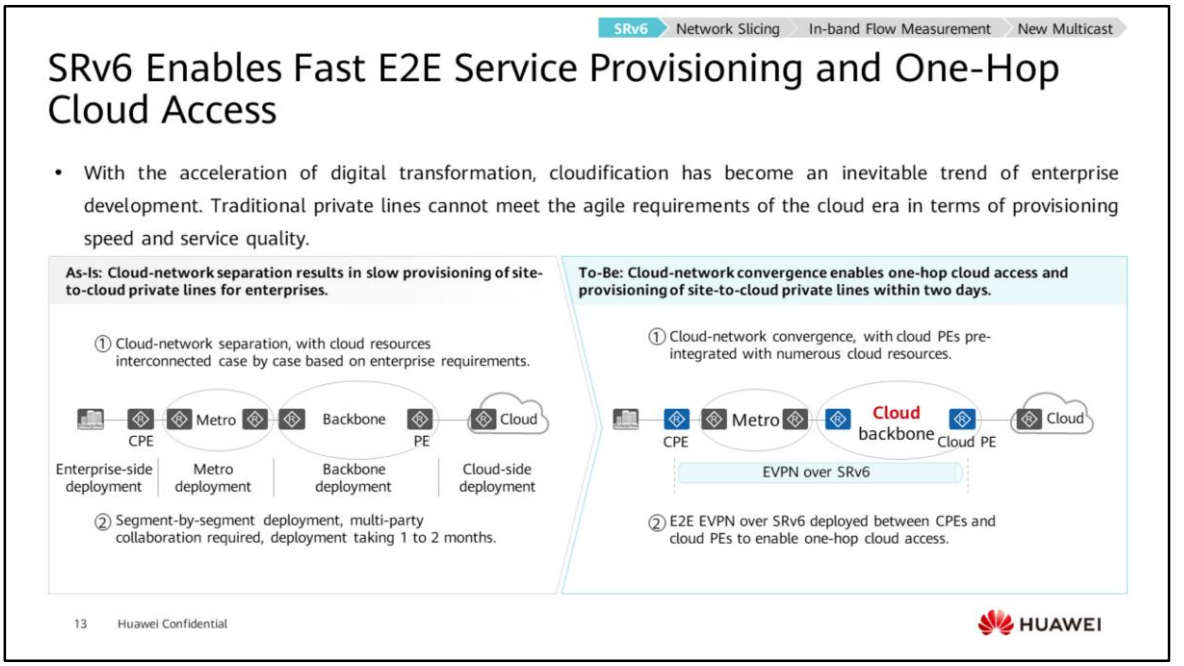
Contents

1. Technical Background of IPE
2. **Key Technical Applications of IPE**
3. IPv6 Network Evolution Solutions

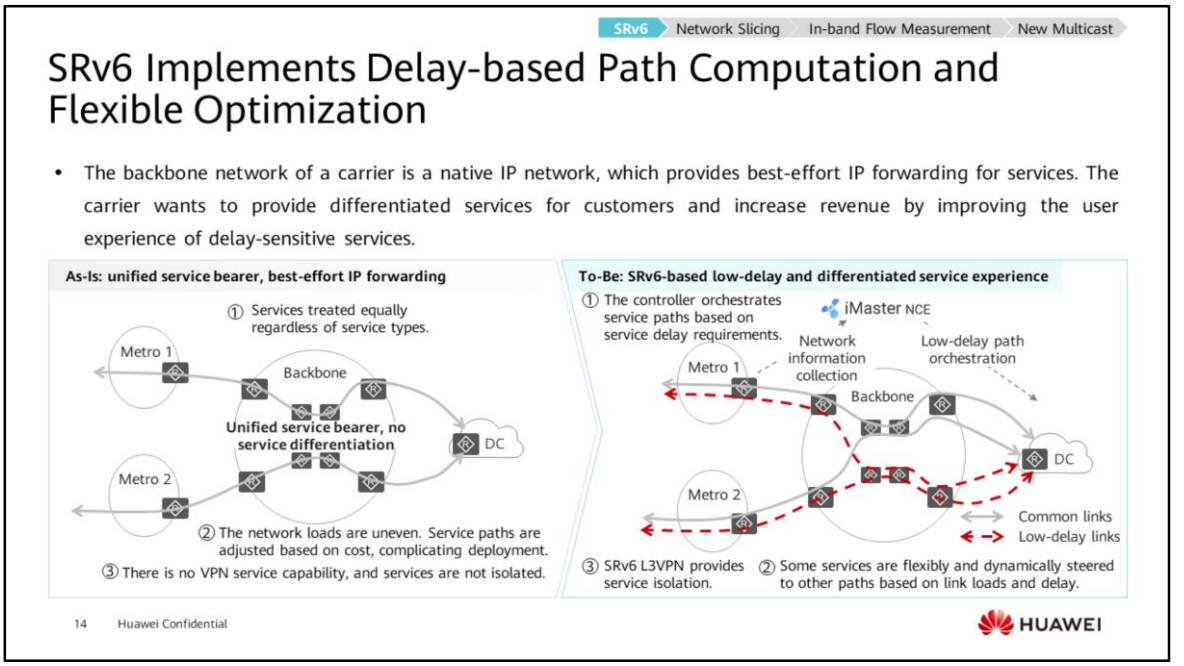
HUAWEI



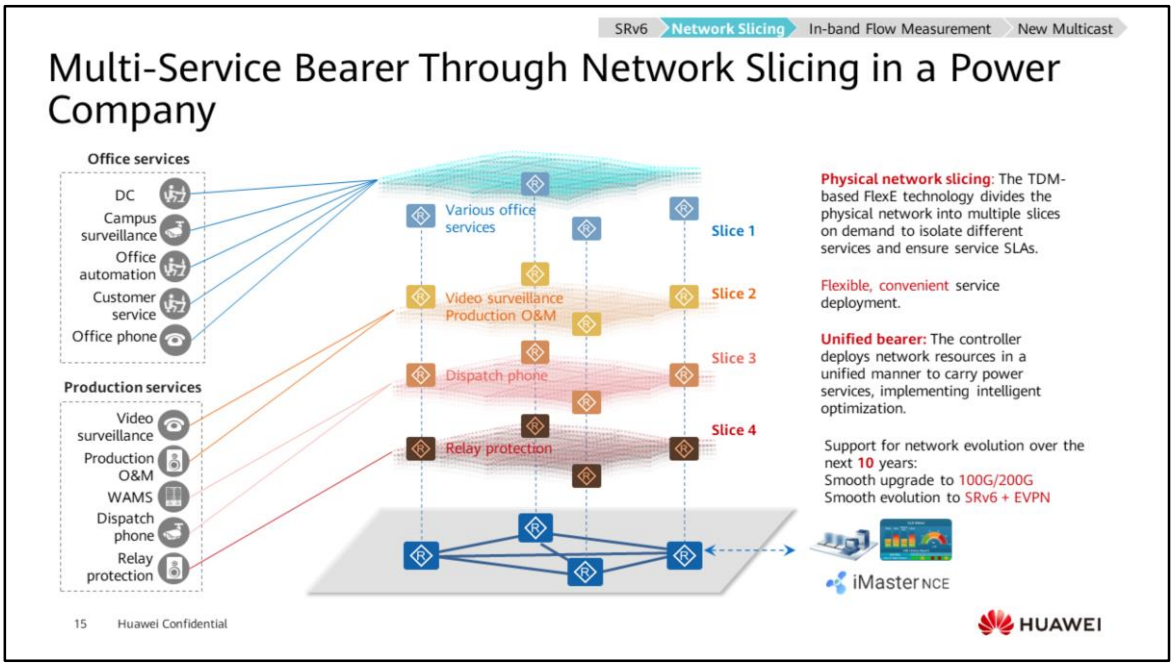
- This figure shows a typical SRv6 deployment case. A carrier provides network security services for customers. To prevent DDoS attacks, the traffic to the DC needs to be cleaned by the DDoS cleaning center first. Such traffic traverses the backbone network.
    - The carrier has two backbone networks. Network A has high bandwidth but does not have the VPN service capability. As a result, cleaned traffic may be sent back to the cleaning center if network A is used.
    - Network B has the VPN service capability but does not have high bandwidth. It is unable to support the cleaning of all service traffic.
    - The volume of to-be-cleaned traffic is large. If such traffic is diverted through network A (without VPN isolation), cleaned traffic may be diverted back to the cleaning center, causing a loop. If such traffic is diverted through network B, the bandwidth resources of network B become a bottleneck as the service scale expands.
- In this situation, SRv6 can be deployed on network A to direct traffic to this high-bandwidth backbone network. This implementation can provide protection for hundreds of DCs. SRv6 is easy to deploy and supports fast service provisioning.



- Currently, large numbers of enterprises are deploying their services on the cloud. Carriers must consider how to quickly provision site-to-cloud private lines for enterprises.
    - Traditional private lines, such as MPLS private lines, may involve multiple ASs. Different ASs are managed and maintained by different management teams. The provisioning of a private line for one-hop cloud access involves collaboration and coordination among multiple departments.
    - With SRv6, the deployment is easy. An E2E SRv6 logical private line can be established between the enterprise CPE and cloud PE to carry cloud-based services, achieving one-hop cloud access. Moreover, private line provisioning is very fast in this case.
- As shown in the figure, carriers want to build a cloud backbone network to enhance their market competitiveness in the cloud era by delivering one-hop cloud access and cloud-network convergence capabilities to enterprises.



- As shown in the figure, the backbone network of a carrier is a native IP network, which provides best-effort IP forwarding for services. The carrier's DC provides cloud services for users, and some of these services are delay-sensitive. The current network does not differentiate delay-sensitive services from common services during service transmission, resulting in poor user experience of delay-sensitive services. Moreover, because services are forwarded along least-cost paths, network loads are uneven, and network management and maintenance are complex. Against this backdrop, the carrier wants to provide differentiated services for customers and increase revenue by improving the user experience of delay-sensitive services.
- The iMaster NCE + SRv6 solution measures the delay of each link on the network and computes paths based on the shortest delay. This solution allows public cloud-based services to travel along the shortest-delay paths, improving the competitiveness of the public cloud.

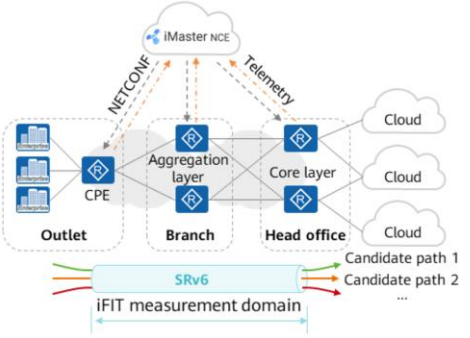


- WAMS: Wide Area Measurement System

SRv6 > Network Slicing > In-band Flow Measurement > New Multicast

Application of iFIT to One-Financial-WAN-for-All-Services Scenarios

- In a one-financial-WAN-for-all-services scenario, SRv6 can quickly and conveniently establish basic network connections between clouds and various access points, ensuring efficient service provisioning.



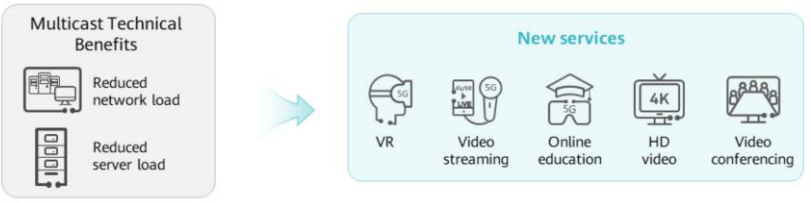
- Solution highlights:
  - In SRv6 scenarios, tunnel-level iFIT can be enabled to measure the quality of each SRv6 segment list and select the optimal link. The link currently in use is periodically compared with the optimal link for path selection and optimization, implementing intelligent traffic steering.
  - One controller is deployed to perform centralized O&M on the entire financial network and implement E2E management and scheduling.

HUAWEI

- One WAN for all services is a technology that provides cross-domain network services through coordination among different networks. In the financial industry, tier-2 banks, outlets, subsidiaries, and external organizations access the head office DC through tier-1 banks, which aggregate service traffic and forward aggregated traffic to the bank core network.
- The financial industry has high requirements on SLA performance. With the development of banking services, diversified service types have emerged in outlets. In addition to traditional production and office services, there are also security protection, IoT, public cloud, and other services. This poses higher O&M requirements on the one-financial-WAN-for-all-services scenario. Against this backdrop, Huawei proposes the iFIT tunnel-level measurement solution.

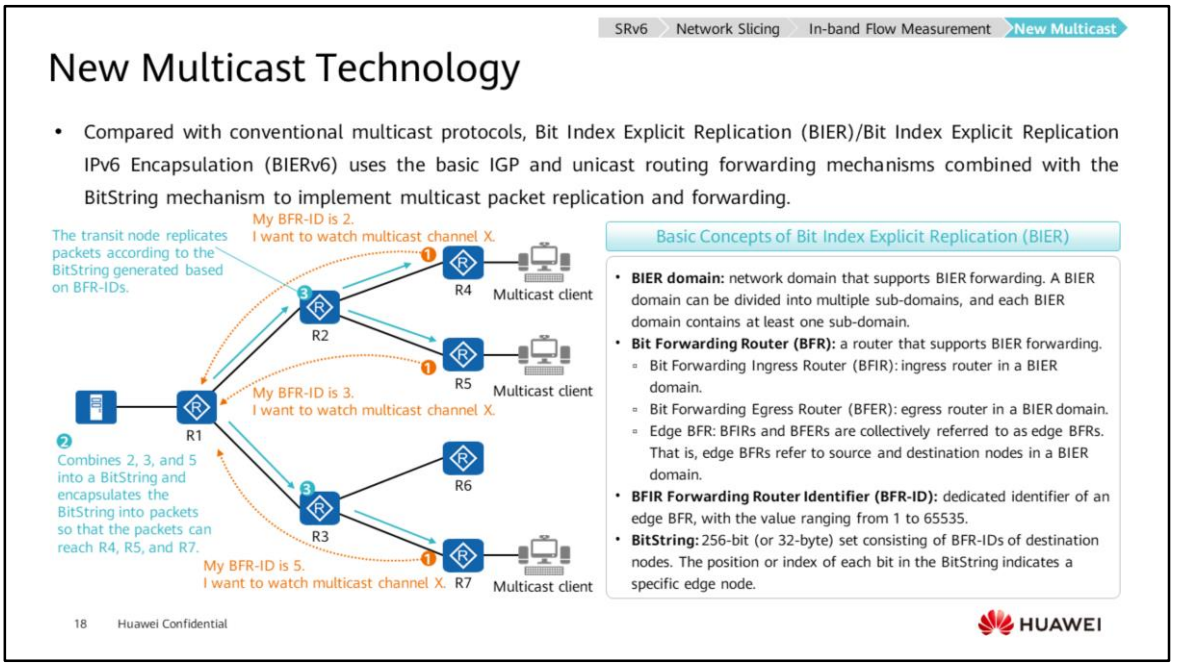
Application Trends of Multicast Technologies

- Currently, video traffic accounts for a large proportion of Internet traffic, including video calls, video sharing, and video conferences. High Definition (HD) visual and new interactive videos may become the main social means in the future, and media gradually evolve to Virtual Reality (VR)/Augmented Reality (AR). These new services pose new requirements on network bandwidth and user experience.



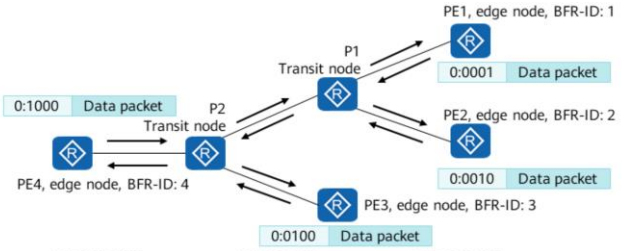
- While the potential applications of multicast are booming, the trend of adopting IPv6-based networks becomes more and more prominent. As IPv6 becomes more widely adopted and new service scenarios require higher bandwidth and better user experience, multicast technologies on IPv6 networks need to continuously evolve in order to keep pace with new service scenarios and technology development trends.

HUAWEI



- BIER overview:
  - This multicast technology encapsulates a set of destination nodes of multicast packets in a BitString in the packet header before sending the packets. With this multicast technology, transit nodes do not need to establish a multicast distribution tree (MDT) for each multicast flow, or maintain the states of multicast flows. Instead, the transit nodes replicate and forward packets according to the BitString in the packet header.
  - In BIER, each destination node is a network edge node. For example, on a network with no more than 256 edge nodes, each node needs to be configured with a unique value ranging from 1 to 256. In this case, the set of destinations is represented by a 256-bit (32-byte) BitString, and the position or index of each bit in the BitString indicates an edge node. This explains the meaning of Bit Index Explicit Replication.
- Advantages of BIER:
  - Supports large-scale multicast service scenarios and reduces resource consumption as BIER does not need to establish an MDT for each multicast flow or maintain the states of multicast flows.
  - Improves multicast group joining efficiency of multicast users in SDN network scenarios because requests of the multicast users do not need to be forwarded along the MDT hop by hop, and instead their requests are directly sent by leaf nodes to the ingress node. This is more suitable for the controller on an SDN network to directly deliver the set of destinations to which multicast packets are to be sent after collecting the set.
- BIERv6 inherits the advantages of BIER and uses IPv6 to program paths, functions, and objects, facilitating multicast forwarding on SRv6-based networks.

BIER Multicast Traffic Forwarding Mechanism



BFR-PE4 BIFT

ID	F-BM	NBR
1	0111	P2
2	0111	P2
3	0111	P2
4	1000	PE4

BFR-P2 BIFT

ID	F-BM	NBR
1	0011	P1
2	0011	P1
3	0100	PE3
4	1000	PE4

BFR-P1 BIFT

ID	F-BM	NBR
1	0001	PE1
2	0010	PE2
3	1100	P2
4	1100	P2

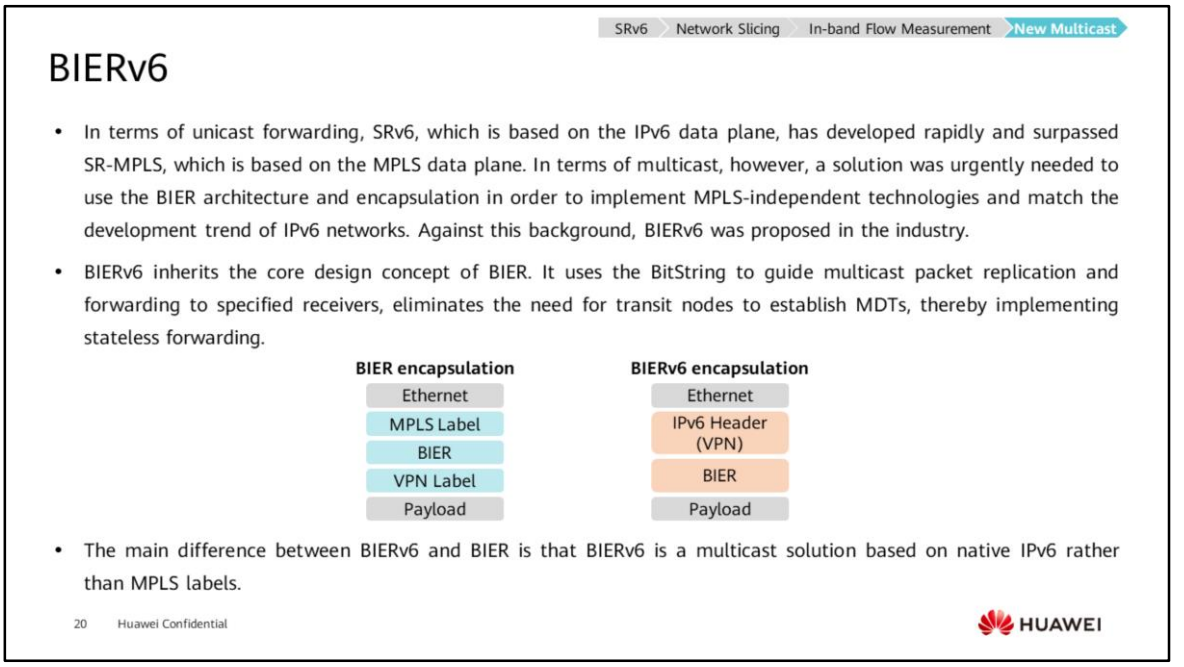
- The ID refers to a BFR-ID. To forward packets to the node of a specified BFR-ID, the device needs to query the entry corresponding to the BFR-ID.
- F-BM is short for Forwarding BitMask. It indicates the set of BIER domain edge nodes that are reachable through the next hop after packets are replicated and sent to the next hop.
- NBR is short for neighbor. It indicates the next hop neighbor through which packets can reach a node of a specified BFR-ID.

- To send multicast traffic of (S1, G1) to PE1, PE2, and PE3, PE4 encapsulates a multicast packet with the BitString (0111). The multicast traffic is sent as follows:

1. PE4 -> P2: In the packet sent from PE4 to P2, the BitString is 0111 (set of the BFR-IDs of PE1, PE2, and PE3).
2. P2 -> PE3: In the packet sent from P2 to PE3, the BitString is 0100 (containing only the BFR-ID of PE3).
3. P2 -> P1: In the packet sent from P2 to P1, the BitString is 0011, which contains the set of BFR-IDs of PE1 and PE2, with the BFR-ID of PE3 removed.
4. P1 -> PE1: In the packet sent from P1 to PE1, the BitString is 0001 (containing only the BFR-ID of PE1).
5. P1 -> PE2: In the packet sent from P1 to PE2, the BitString is 0010 (containing only the BFR-ID of PE2).

HUAWEI

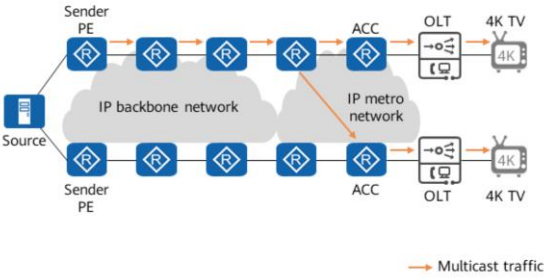
- Bit allocation fundamentals: BIER floods the mapping between bit positions (BFR-IDs) of nodes and prefixes through IS-IS LSPs (IS-IS for BIER is used as an example). Devices learn the complete BIFT (BIER neighbor table) through flooding. The BIFT has the following characteristics:
  - In the neighbor table, each directly connected neighbor has one entry.
  - Each entry contains information about the edge nodes that are reachable to a neighbor.
- BIFT: Bit Index Forwarding Table



- Advantages of BIERv6:
  - Simplified network protocols:
    - BIERv6 uses IPv6 addresses to carry Multicast VPN (MVPN) and GTM services, further simplifying protocols and eliminating the need to allocate, manage, and maintain MPLS labels.
    - BIERv6 has high extensibility as it complies with the design concept of SDN and network programming.
  - Simplified deployment and O&M:
    - BIERv6 uses an IPv6 extension header to carry BIER forwarding instructions, eliminating the need for MPLS label-based forwarding. IPv6 extensibility facilitates the evolution and addition of new features.
    - Because BIERv6 is deployed only on the ingress and egresses, transit nodes are unaware of multicast service changes. When the network topology changes, there is no need to withdraw or re-establish numerous MDTs, thereby greatly simplifying O&M.

BIERv6 Application in an IPTV Scenario

- IP video services include live TV (such as news, sports events, movies and TV series, and live webcasting), video on demand (VOD), and video surveillance. Video transmission service applications such as 4K IPTV, 8K VR, smart city, smart home, autonomous driving, telemedicine, and safe city are expected to grow dramatically in the near future, especially in countries and regions with fast economic development.



- In the figure, MVPN over BIERv6 is deployed on the public network to carry IPTV traffic. BIERv6 multicast technology greatly reduces the network load and improves user experience (for example, it ensures fast VOD, clear images, and smooth playback).
- In addition, this solution features simple deployment, O&M, and capacity expansion and is suitable for large-scale deployment.

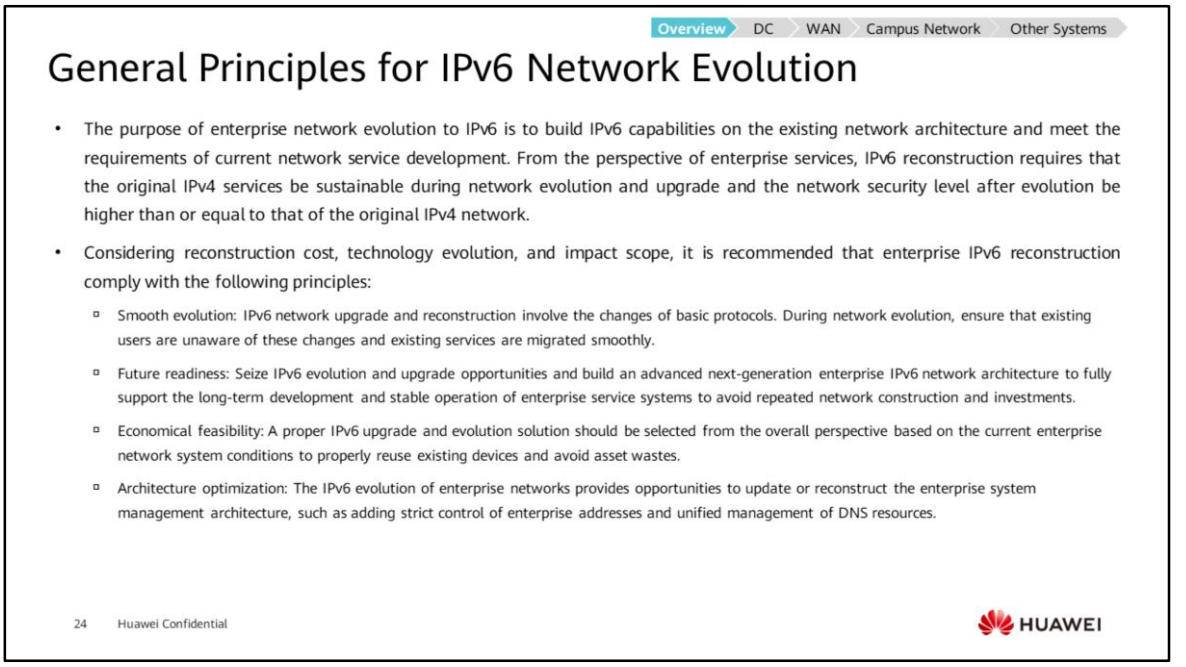
HUAWEI

- IPTV: Internet Protocol Television
- OLT: Optical Line Termination

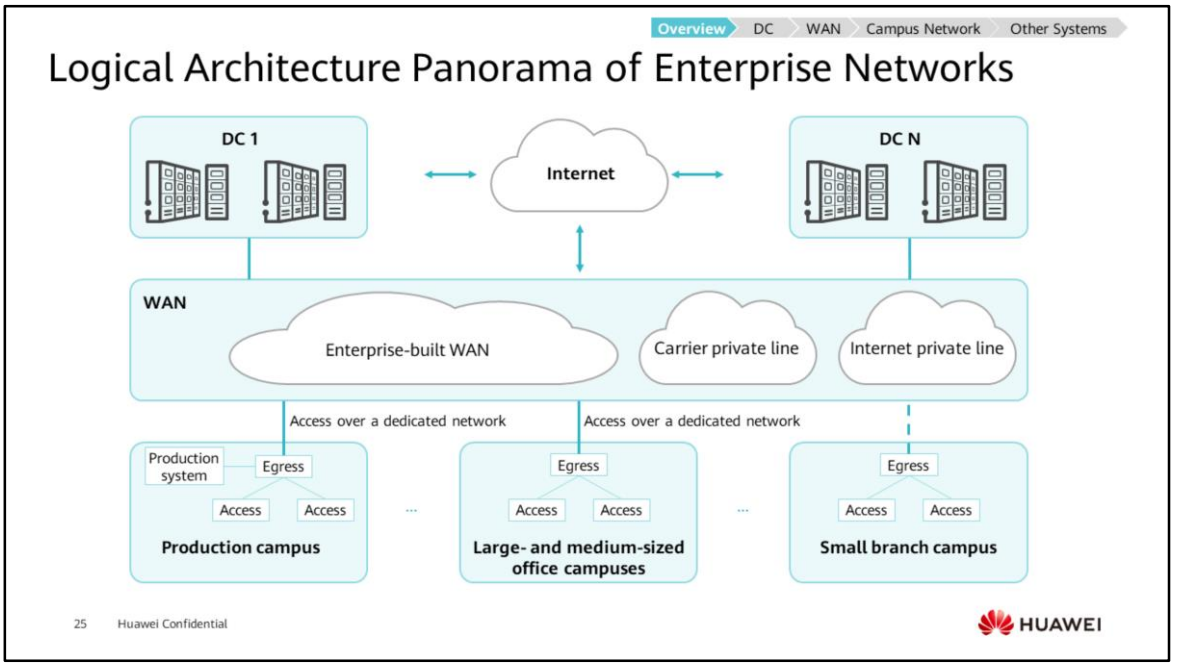
Contents

1. Technical Background of IPE
2. Key Technical Applications of IPE
3. **IPv6 Network Evolution Solutions**

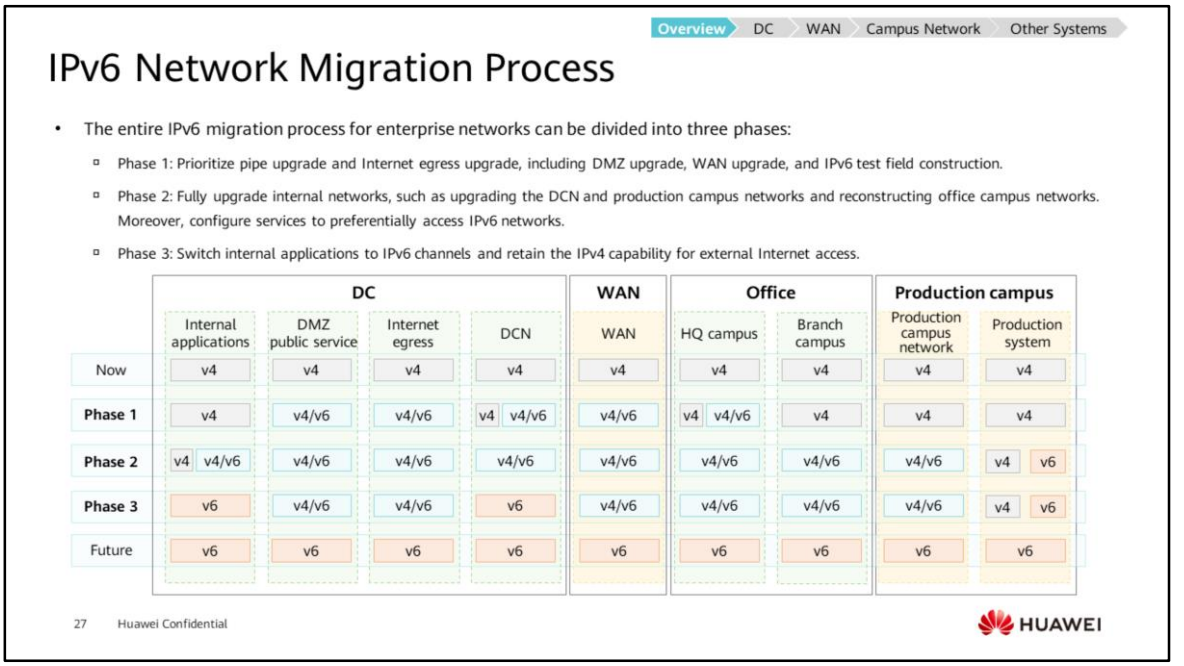
HUAWEI



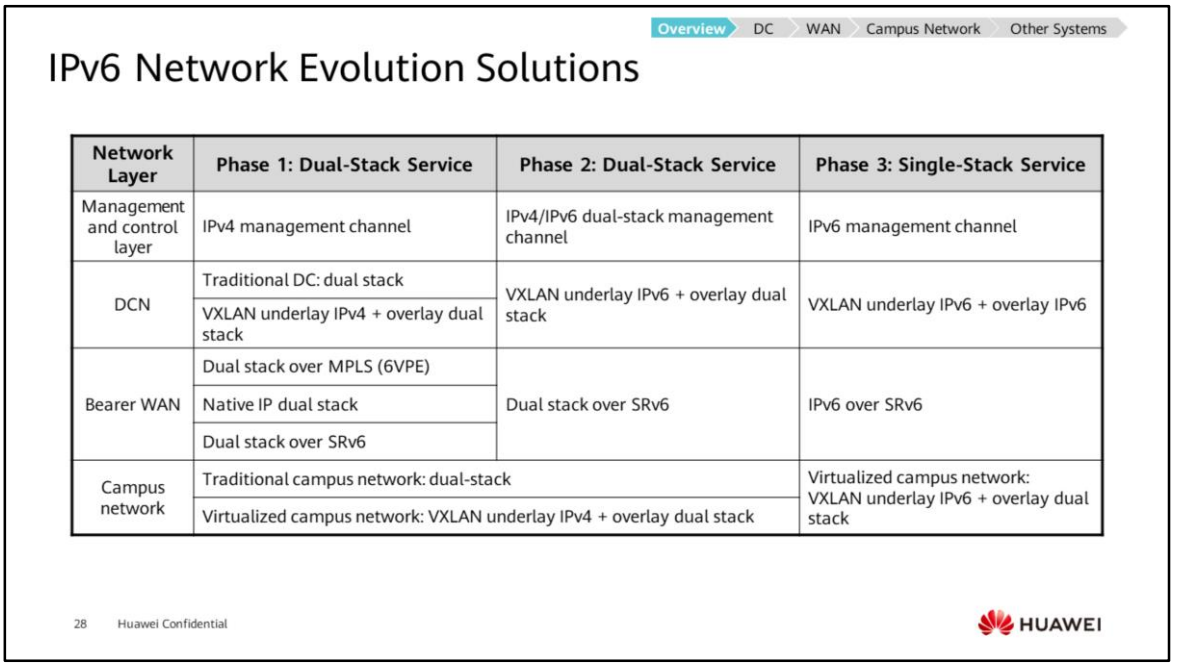
- Enterprise services can be classified into Internet services, DMZ services, and enterprise-built service systems. Internet services and DMZ services depend on external network applications and user environments and require long-term coexistence of IPv4 and IPv6 user access. This factor must be considered during IPv6 reconstruction of enterprise networks.



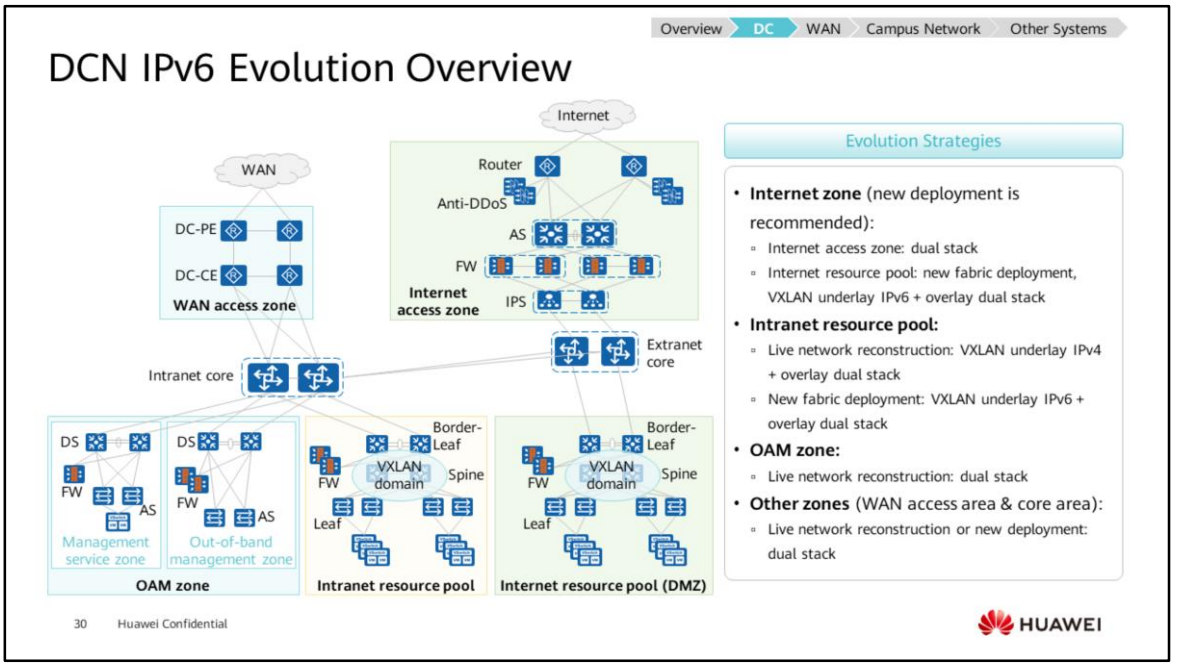
- DCN:
  - Based on the service scope, enterprise DC services can be classified into external services and internal applications.
    - The front-end servers for external services are generally deployed in the DMZ of the DC and connect to external user devices through the Internet egress zone.
    - Internal applications mainly involve the DC's internal service networks (such as networks in the WAN egress zone and service zone).
  - From the perspective of service reconstruction, the DCN can be classified into the internal application network, DMZ (for external services), Internet egress, and internal service network during IPv6 evolution and upgrade planning.
- WAN:
  - Multiple backhaul solutions are available for interconnection between enterprise branches, such as enterprise-built WAN, carrier MPLS L2VPN/L3VPN, and Internet private line.
  - The IPv6 evolution of enterprise WANs mainly considers the WAN upgrade policy.



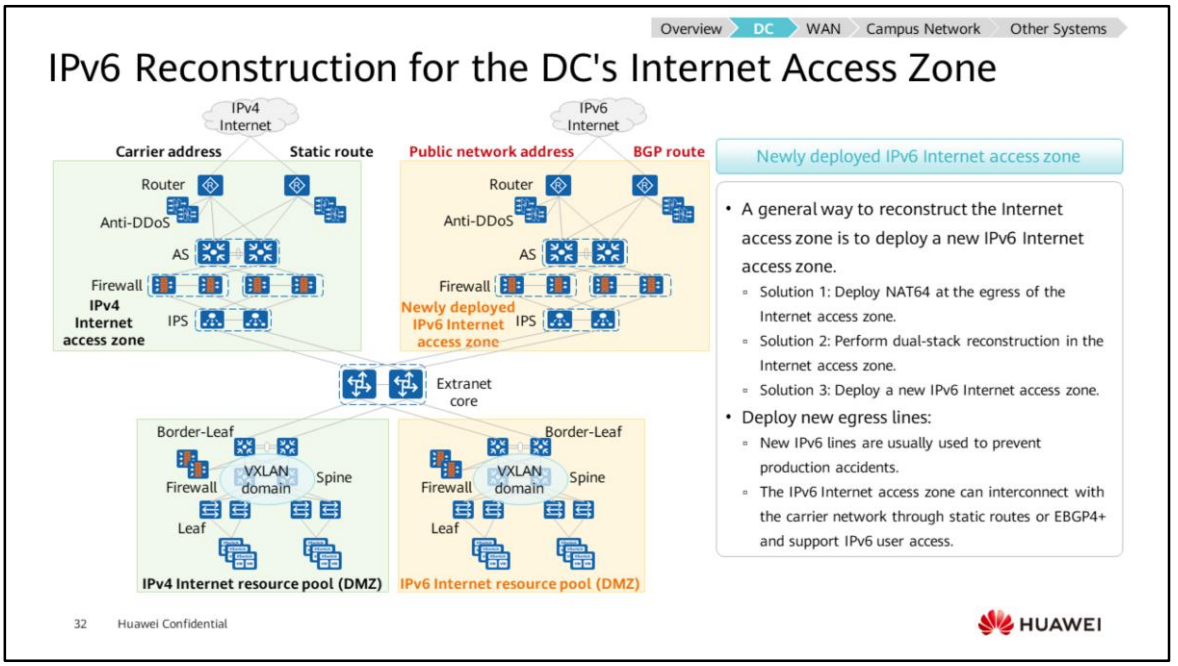
- The overall IPv4-to-IPv6 network migration principle is "DCN first, WAN second, and campus network reconstruction on-demand".
  - Phase 1: Deploy dual-stack services in the DC's public service and test zones and IPv4 single-stack services on the WAN's underlay network and dual-stack services on the WAN's overlay network, and pilot dual-stack services on the campus network .
  - Phase 2: Gradually apply dual-stack to the DC's internal applications and the campus network's office campus part and apply dual-stack to the campus network's production campus part.
  - Phase 3: Comprehensively apply IPv6 single stack to the DC's internal applications and ensure that the WAN gradually evolves to IPv6-only networks.



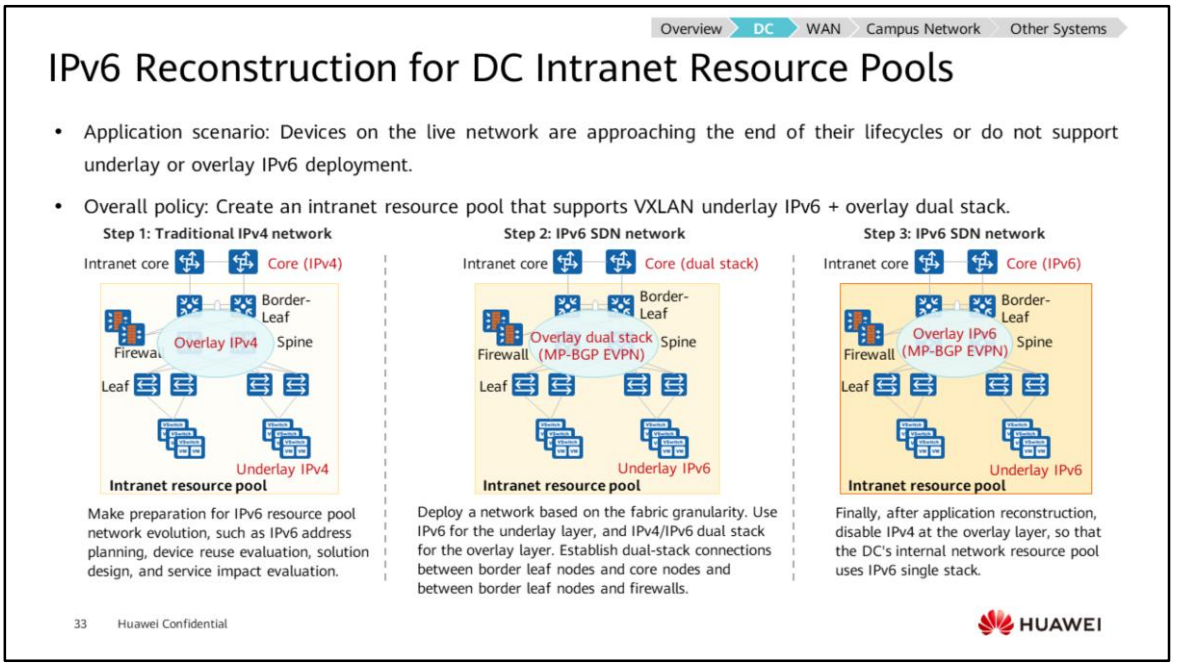
- DCN:
  - The reconstruction solutions for the Internet access zone include NAT64, IVI, and dual-stack reconstruction. It is recommended to use the dual-stack solution to provide IPv6 addresses and service capabilities.
    - NAT64 is limited by session table specifications and consumes a lot of resources. With the increase of IPv6 terminals, NAT64 will become a performance bottleneck for IPv6 service development. Therefore, NAT64 is applicable only to the early deployment of IPv6 services and is not recommended to be a target solution.
    - The IPv6 address structure of the IVI is limited and does not meet the IPv6 address planning principles. Therefore, the IVI is not recommended for large-scale deployment.
  - Internal network resource pool reconstruction mainly uses dual-stack solutions, including VXLAN underlay IPv4 + overlay dual stack and VXLAN underlay IPv6 + overlay dual stack.
    - VXLAN underlay IPv4 + overlay dual stack can be used for initial dual-stack reconstruction to quickly provide IPv6 service bearer capabilities.
    - VXLAN underlay IPv6 + overlay dual stack can be used for new DCN deployment and existing DCN reconstruction. This facilitates gradual evolution to IPv6-only networks.



- DCN architecture:
  - The underlay network refers to the physical network or infrastructure network. It is required that any two nodes on the physical network be routable to each other. The spine-leaf architecture is recommended for underlay networking.
    - A spine node is a core node on a VXLAN fabric network, which provides high-speed IP forwarding and connects to leaf nodes using high-speed interfaces.
    - A leaf node is an access node on a VXLAN fabric network, which connects various network devices to the VXLAN network. Different access objects are classified into border leaf nodes, server leaf nodes, and service leaf nodes.
  - The overlay network is a virtualized network built on the underlay network to carry applications over the physical network while remaining isolated from other service networks.



- Internet zone evolution strategies:
  - Solution 1: Deploy NAT64 at the egress of the Internet access zone. If existing DC services remain in IPv4 single-stack mode but IPv6 services need to be quickly provided due to other factors, NAT64 can be deployed for IPv4 servers in the DC's DMZ to temporarily provide IPv4/IPv6 dual-stack services through the NAT64 gateway.
  - Solution 2: Perform dual-stack reconstruction in the Internet access zone and DMZ. If the Internet zone (including the DMZ) supports IPv6 well and devices are far from reaching the end of their lifecycle, it is recommended that these devices be reused. You can deploy the dual-stack solution in the existing Internet access zone and DMZ to complete the reconstruction at a low cost.
  - Solution 3: Deploy a new Internet access zone and DMZ. Reusing existing devices during IPv6 network reconstruction may involve software and hardware upgrades or partial hardware replacement, which affects IPv4 services. To ensure enterprise DMZ service continuity and zero impact on live network IPv4 services, deploy an IPv6-only Internet access zone and DMZ. IPv4 users access the IPv4 DMZ through the IPv4 Internet access zone, and IPv6 users access the IPv6 DMZ through the IPv6 Internet access zone.
- Egress route selection:
  - Single-egress Internet access: Static routes are preferred.
  - Multi-egress Internet access: BGP routes are preferred. To ensure load balancing and reliability, BGP route attributes can be used to control route selection.



- Resource pool/server deployment:
  - A general way to prevent production accidents and service interruption caused by the upgrade of original IPv4 services is to deploy new IPv6 resource pools/servers.
  - If the DC is a traditional DC, it is recommended that new technologies, such as SDN, be deployed in the new resource pools.

WAN IPv6 Evolution Overview

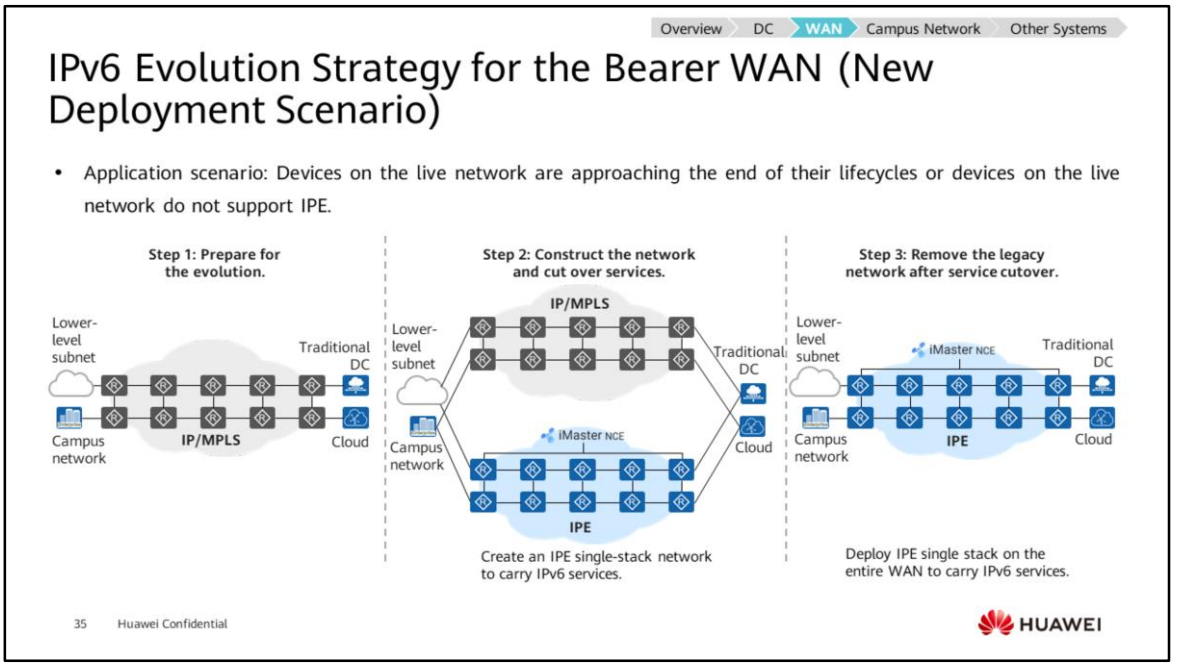
- The WAN is mainly used to carry services. Currently, the industry has reached a consensus that IPv4 and IPv6 will coexist for a long time.
  - For an existing WAN where MPLS is deployed, using 6VPE for reconstruction is recommended. Only edge nodes need to be upgraded or have their hardware replaced. P nodes do not need to be reconstructed. Deploying IPv6 is like introducing new services and has little impact on original maintenance. Therefore, this solution is suitable for evolution from existing networks to IPv6.
  - SRv6 can be considered during new WAN planning and design. SRv6 is the preferred technology for SDN-based networks in the IPv6 era. SRv6 features high scalability, good application compatibility, and strong programming capabilities. SRv6 can meet the requirements of future service changes and is suitable for evolving new networks to IPv6.

- IPv6-only networks are deployed once and for all, with all network traffic carried using IPv6.
- IPE supports service isolation.

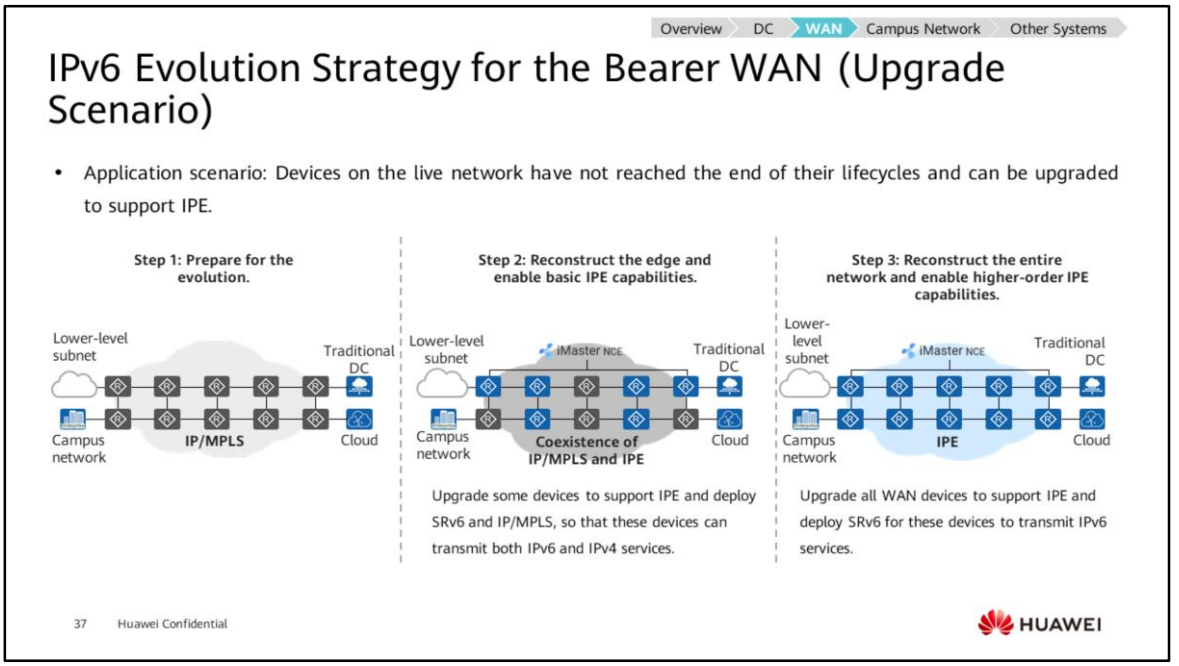
- IPv4/IPv6 dual stack must be enabled on all Layer 3 devices.
- Layer 3 devices must be reachable to each other through dual-stack routes.
- The network needs to maintain two protocol stacks, increasing complexity.

- IPv4/IPv6 dual stack must be enabled on edge nodes. The edge nodes must be enabled to exchange VPNv6 routing information with their original MP-IBGP peers.
- The intermediate nodes only require IPv4 single stack (MPLS) deployment and do not need to be reconstructed.

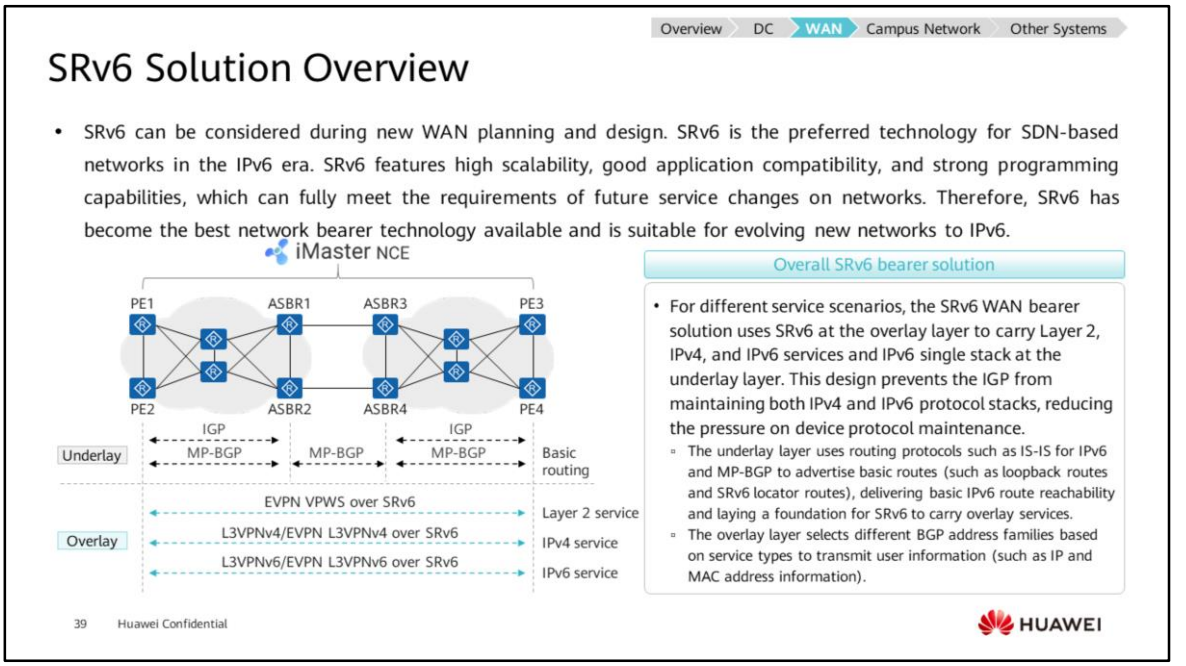
HUAWEI



- Overall strategy: Deploy IPE single stack to carry both IPv4 and IPv6 services, and gradually evolve office, production, management, and other services from IPv4 to IPv6.
- Step 1: Prepare for the evolution.
  - Address application: Apply for IPv6 addresses based on enterprise requirements from address management organizations or carriers.
  - Live network evaluation: Evaluate live network conditions, covering the network infrastructure, security infrastructure, service support systems, and service applications.
  - Integration verification: Design and verify the IPE evolution solution and prepare a feasibility report.
  - Network planning: Formulate the IPE network construction specifications and evolution solution to support smooth evolution to IPv6-only networks.
- Step 2: Construct the network and cut over services.
  - Network construction: Build an IPE network based on the network plan and deploy capabilities such as SRv6, network slicing, in-band flow measurement, and SDN.
  - Security construction: Security devices must support IPv6 security protection. Network devices interwork with security devices to ensure that IPv6 has the same or even stronger security protection capabilities compared with IPv4.



- Overall strategy: Upgrade and replace nodes one by one from the edge to the core. Gradually deploy IPE features (simple features first, then complex features) and cut over services (common services first, then critical services).
- Step 1: Prepare for the evolution.
  - Address application: Apply for IPv6 addresses based on enterprise requirements from address management organizations or carriers.
  - Live network evaluation: Evaluate live network conditions, covering the network infrastructure, security infrastructure, service support systems, and service applications.
  - Integration verification: Design and verify the IPE evolution solution and prepare a feasibility report.
  - Network planning: Formulate the IPE network construction specifications and evolution solution to support smooth evolution to IPv6-only networks.
- Step 2: Reconstruct the edge and enable basic IPE capabilities.
  - Edge device upgrade: Upgrade edge devices (cloud, Internet egress, and campus egress PEs preferred) to support IPE.
  - IPE basic capability deployment: Deploy both SRv6 and traditional IP/MPLS on upgraded devices. Configure new and old devices to interwork through IP/MPLS, and use SRv6 between new devices. Deploy a controller for network-wide management and control.

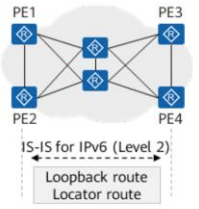


- Overlay service deployment suggestions:
  - Layer 2 services: Use SRv6 EVPN as the bearer protocol to provide P2P and P2MP connection models. Compared with traditional Virtual Switching Instances (VSIs) and Pseudo Wires (PWs), EVPN features simple deployment, high bandwidth utilization, and fast convergence. EVPN is the best choice for L2VPN service bearer.
  - Layer 3 IPv4 services: Use either SRv6 L3VPN or SRv6 EVPN L3VPN. Using SRv6 EVPN L3VPN is recommended.
  - Layer 3 IPv6 services: Use either SRv6 L3VPNv6 or SRv6 EVPN L3VPNv6. Using SRv6 EVPN L3VPNv6 is recommended.

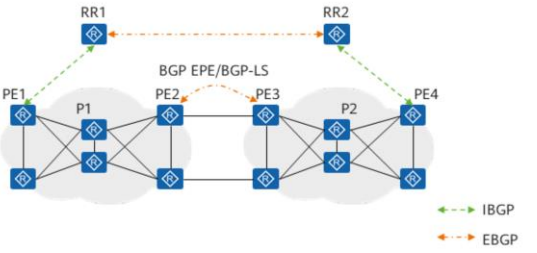
SRv6 Solution: IGP & BGP Design

- IGP design: Use the existing network design for IGP planning. The existing IPv4 network does not need to be redesigned. It is advised to use IS-IS and enable IPv6 (IS-IS for IPv6).
- BGP design: BGP focuses on the control and advertisement of routes. After an IGP is deployed, PEs need to establish BGP peer relationships to advertise service routes through different address families.

IS-IS for IPv6 (single area)



BGP-based advertisement of service routes & SRv6 locators/SRv6 SIDs

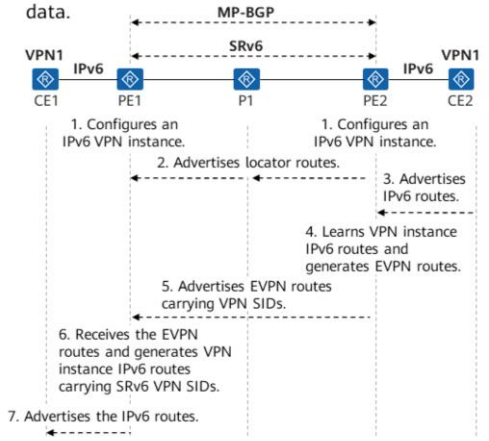


HUAWEI

- Loopback routes and SRv6 locator routes need to be advertised in an IGP domain. Loopback routes are used for network management or BGP peer relationship establishment. Locator routes are used to guide the forwarding of data traffic over SRv6 tunnels in an IGP domain.
- BGP Egress Peer Engineering (EPE) is used to allocate SRv6 SIDs to BGP peers between ASs in inter-AS scenarios.

SRv6 Solution: Layer 3 IPv6 Services (1)

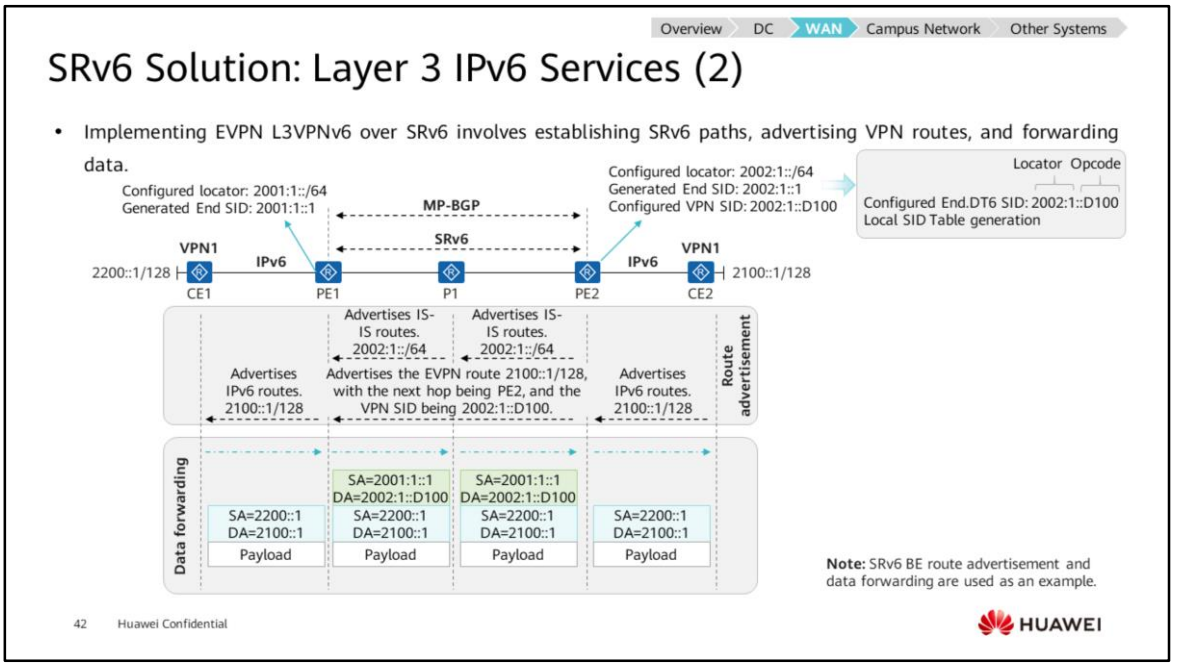
- Implementing EVPN L3VPNv6 over SRv6 involves establishing SRv6 paths, advertising VPN routes, and forwarding data.



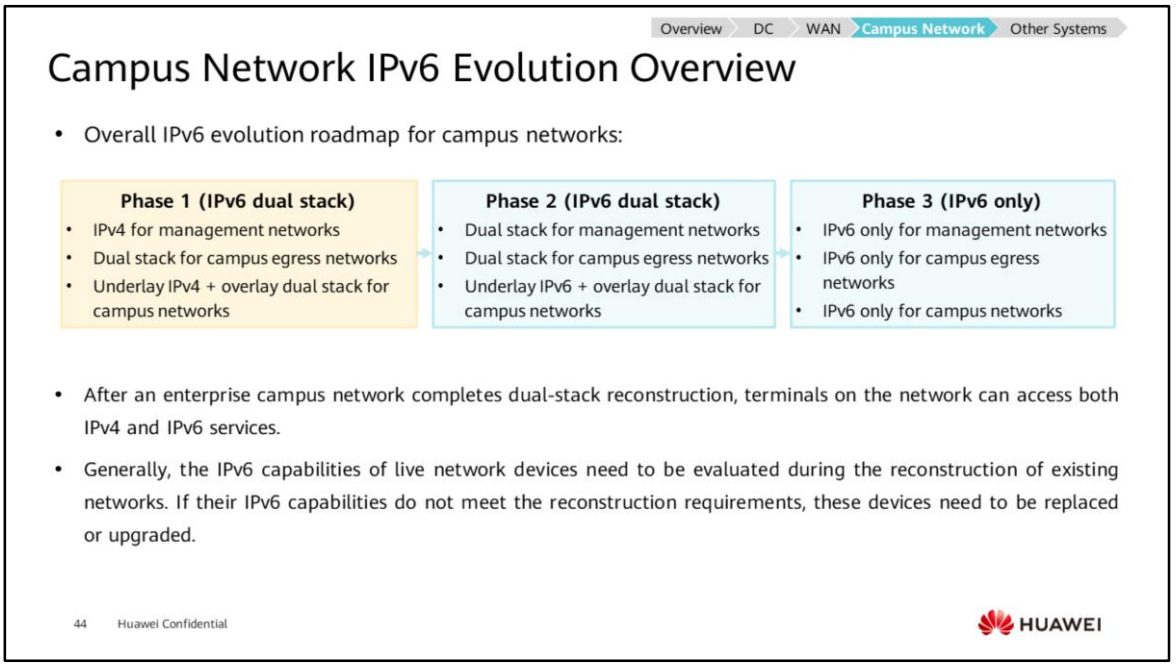
Service establishment procedure (SRv6 BE)

1. A user configures SRv6 and IPv6 VPN instances on PEs, and configures transit devices to support IPv6.
2. PE2 advertises an SRv6 locator route to PE1.
3. CE-to-PE route advertisement: CE2 advertises its IPv6 routes to PE2. Static routes, OSPFv3, IS-IS for IPv6, or BGP4+ can be used between CEs and PEs.
4. After learning the VPN route advertised by CE2, PE2 installs it in the IPv6 routing table of the corresponding VPN instance and converts it into an EVPN IP prefix route.
5. Inter-PE route advertisement: PE2 advertises an EVPN route to PE1 (BGP EVPN peer) through an update message carrying RT and SRv6 VPN SID attributes.
6. PE1 receives the EVPN route. If the next hop in the EVPN route is reachable and the route matches the BGP route import policy, PE1 performs a series of actions to determine whether to install the route in the IPv6 routing table of the corresponding EVPN instance. These actions include route leaking, route recursion to an SRv6 path, and route selection. The VPN route is associated with an SRv6 VPN SID when being delivered.
7. PE-to-CE route advertisement: CE1 can learn VPN routes from PE1 in multiple modes, including static routing, OSPFv3, IS-IS for IPv6, and BGP4+. The route exchange between CE1 and PE1 is similar to that between CE2 and PE2.

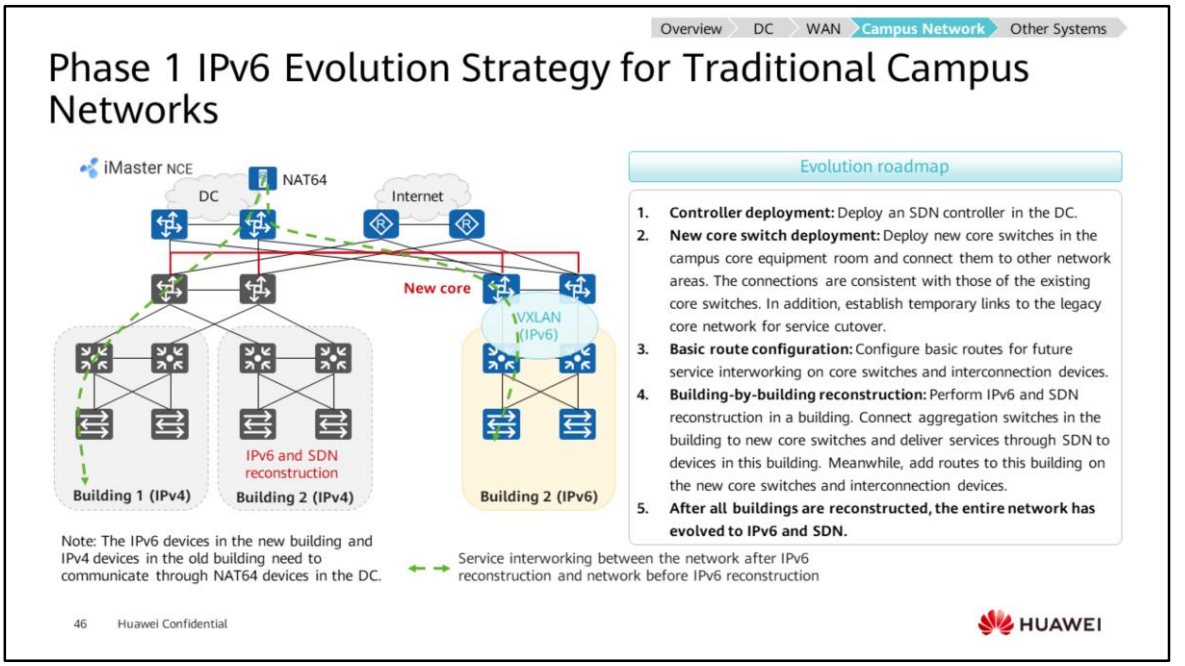
HUAWEI



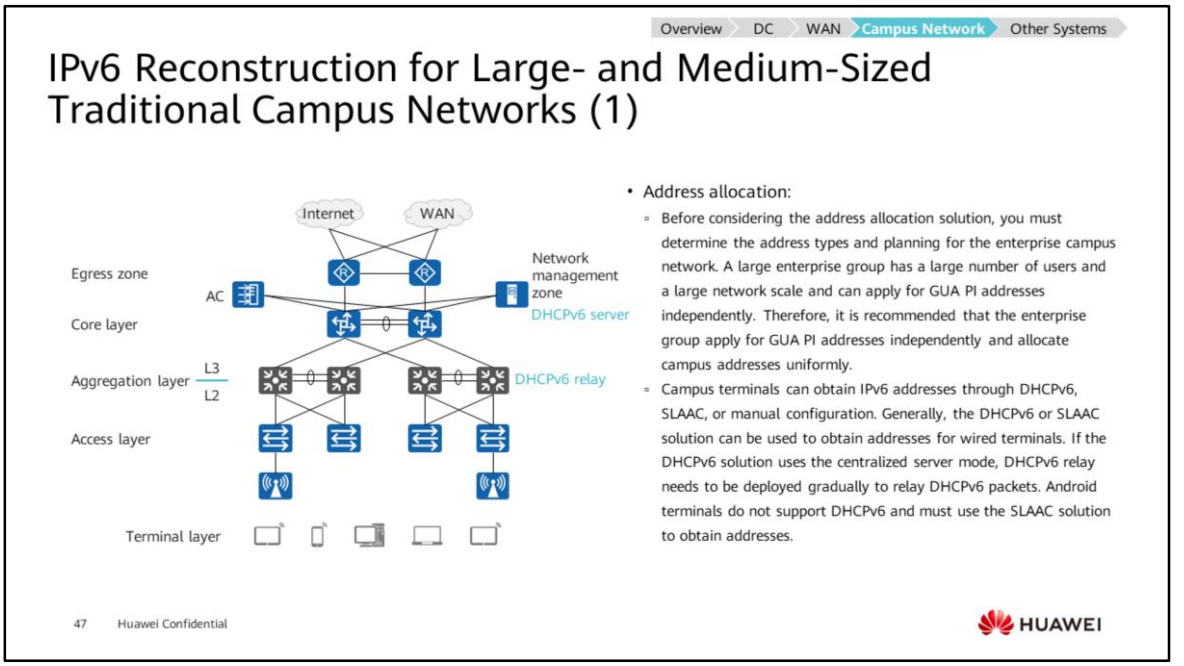
- Route advertisement:
    1. PE2 generates an End SID based on the configured SRv6 locator.
    2. PE2 advertises the locator route 2002:1::/64 corresponding to the specified End SID to PE1 through an IGP. PE1 installs the route to its IPv6 routing table.
    3. A VPN End.DT6 SID 2002:1::D100 within the locator range is configured on PE2, which then generates Local SID Table.
    4. After receiving the VPN IPv6 route advertised by CE2, PE2 converts it into an EVPN IP prefix route and advertises it to PE1 through the BGP EVPN peer relationship. The route carries an SRv6 VPN SID — VPN End.DT6 SID 2002:1::D100.
    5. After receiving the EVPN route, PE1 leaks it to the IPv6 routing table of the corresponding VPN instance, converts it into a common IPv6 route, and advertises it to CE1.



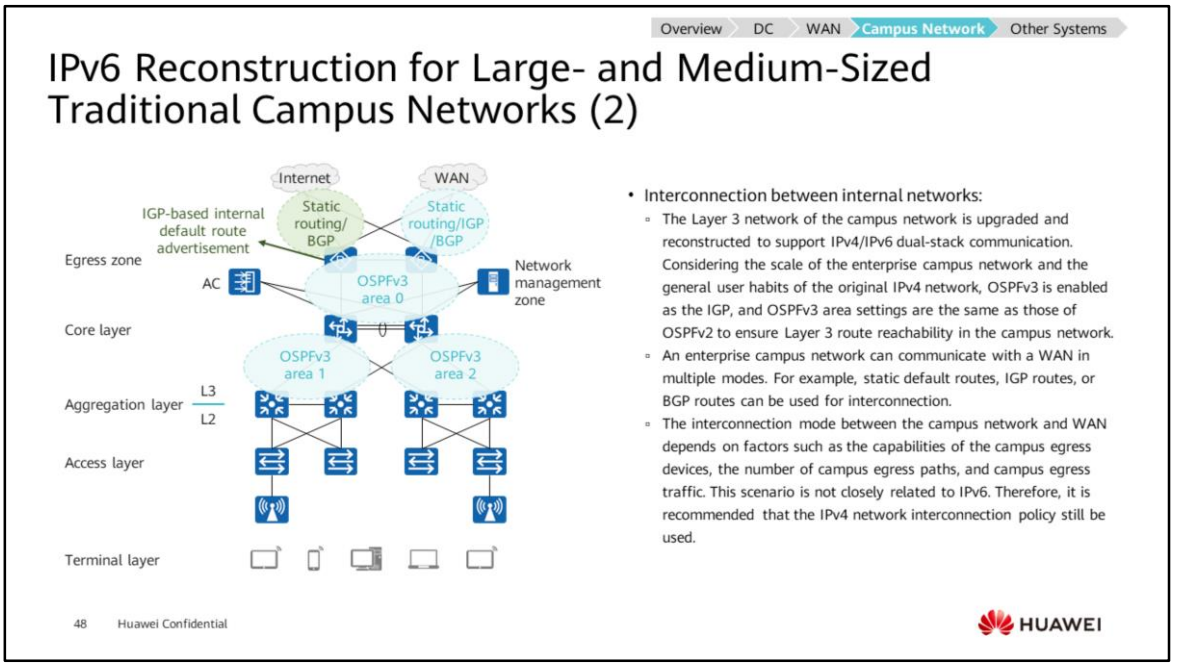
- Campus networks mainly involve large- and medium-sized office campuses, small-sized branch office campuses, and industrial production campuses. There are three scenarios for intra-campus mutual access: mutual access between internal office systems, Internet access, and mutual access between production systems.
    - During dual-stack reconstruction for large- and medium-sized office campus networks, the original IPv4 network architecture and configurations are retained, and IPv6 configurations and IPv6-related support systems are added.
    - In large- and medium-sized virtualized campus scenarios, underlay IPv4 + overlay VXLAN dual stack can be used to evolve VXLAN to IPv6 based on live network device capabilities.
    - When a small office campus network is upgraded to IPv4/IPv6 dual-stack, services from terminals inside the campus to the enterprise HQ need to be transmitted through WAN dual-stack channels. Considering the campus network scale and WAN line leasing solution, we can determine which option to choose based on live network conditions:
        - In the scenario where only a single Internet private line is leased for backhaul, the campus egress generally connects to the intranet over an IPsec tunnel. One of the following solutions can be used: dual-stack traffic over IPsec6, dual-stack traffic over IPsec4, and dual-stack traffic over GRE over IPsec6/4.



- Phase 1 evolution strategy: Perform IPv6 and SDN evolution concurrently to achieve two objectives at one time.
    - Overall solution roadmap: Create a core network that supports SDN and IPv6 network virtualization, and connect the core network to the campus egress network. Reconstruct the existing network building by building (aggregation + access) and connect the network to the new core network. Use the SDN controller for automated deployment in reconstructed buildings.
    - Solution advantages: Operations are simple, the deployment pace is controllable, and the impact on live network services is minimized.



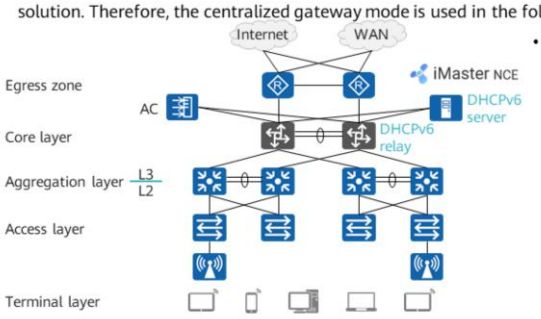
- Large- and medium-sized traditional campus networks generally use Layer 3 IP forwarding and use Layer 2 VLANs to isolate broadcast domains. If Layer 3 isolation is required, ACLs can be deployed. For external services such as Internet access and enterprise WAN interconnection, intra-campus support systems include AAA authentication servers and DHCP servers. During dual-stack reconstruction for large- and medium-sized traditional campus networks, the original IPv4 network architecture and configurations are retained, and IPv6 configurations and IPv6-related support systems are added.
- GUA: global unicast address
- The PI address space is a block of IPv6 addresses assigned by a regional Internet registry to enterprises.



- Interconnection with external networks:
  - For example, if the campus network IPv6 addresses are PI addresses independently applied for, static routes can be used for interconnection with external networks and IGPs can be used to advertise default routes on the internal network. Carrier networks advertise specific routes to the campus network, and default egress routes are configured on the campus network.
  - If multiple egresses are involved, BGP can be used for interconnection with external networks and IGPs can be used to advertise default routes on the internal network. The campus network internally advertises default egress routes to ensure that internal service packets can reach egress routers. The egress routers connect to the Internet using BGP to implement optimal path selection and load balancing.

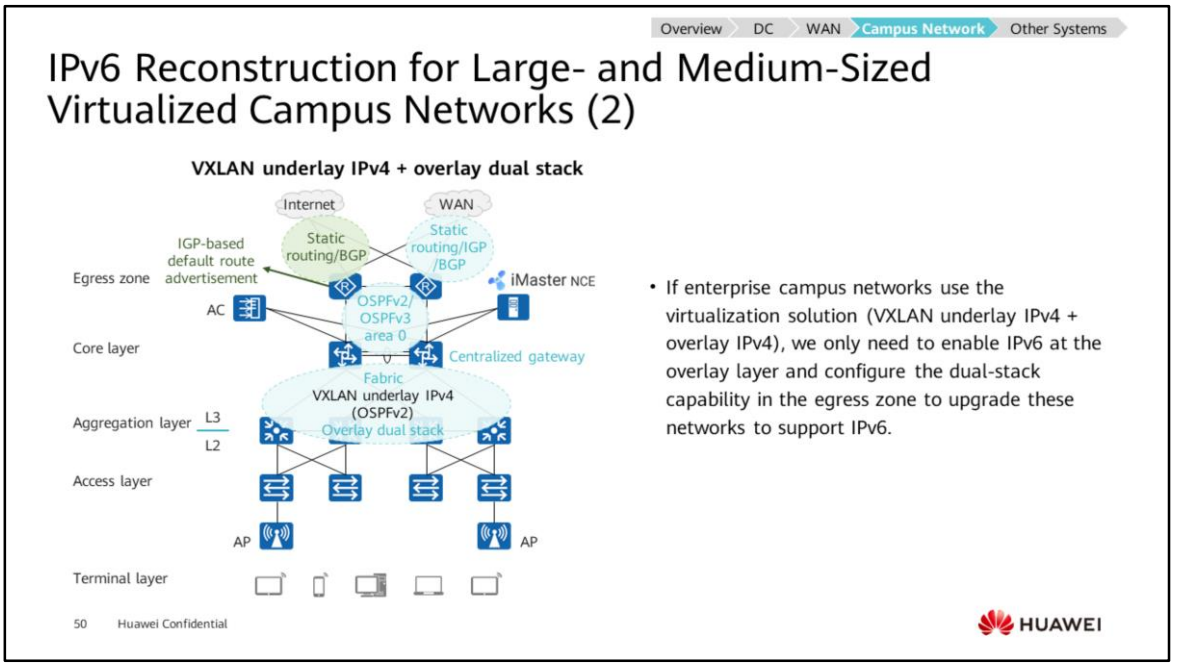
IPv6 Reconstruction for Large- and Medium-Sized Virtualized Campus Networks (1)

- Currently, the virtualized network solution (VXLAN) is deployed on some large- and medium-sized campus networks to meet the requirements of these networks, such as one network for multiple purposes and fast, automated service provisioning.
- VXLAN underlay IPv4 + overlay dual stack can be used for the IPv6 evolution of virtualized campus networks. The horizontal traffic between networks in an enterprise campus is light. The centralized gateway mode is generally used in the virtualized network VXLAN solution. Therefore, the centralized gateway mode is used in the following examples.

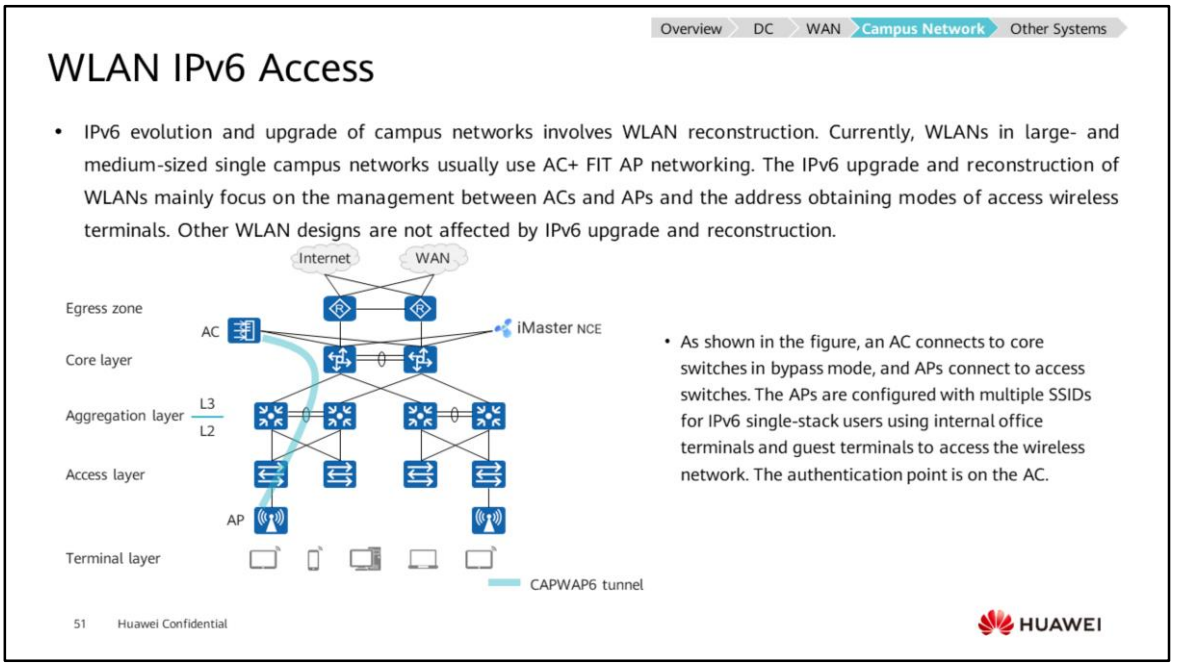


- Address allocation:
  - There are three network address allocation modes: DHCPv6, SLAAC, and manual configuration. If the DHCPv6 or SLAAC mode is used for automatic address allocation, the virtualized network design is different from that of the traditional network.
  - In DHCPv6 address allocation mode, the DHCPv6 server needs to be deployed in centralized mode, and DHCPv6 relay needs to be enabled on the VBDIF interface serving as the centralized VXLAN IPv6 gateway to relay DHCPv6 packets. Android terminals do not support DHCPv6 and must use the SLAAC solution to obtain addresses.
  - In SLAAC address allocation mode, RA parameters need to be configured on the centralized gateway, so that the centralized gateway pushes RA messages carrying address prefixes and DNS information to terminals.

HUAWEI



- Service forwarding:
  - The overlay IPv6 design is similar to the original overlay IPv4 design. The underlay configuration from the access layer to the core layer remains unchanged. The VXLAN control plane uses BGP EVPN, and core switches are configured as RRs. Enable IPv6 on the centralized gateway and configure an IPv6 address for the VBDIF interface to ensure Layer 2 IPv6 communication.
  - Edge nodes on the access side can associate forwarded packets with overlay BDs based on interface VLANs, enabling terminals to be assigned to different gateway areas. In VXLAN Layer 3 forwarding, terminal packets are first advertised to the centralized gateway, and horizontal and vertical traffic are forwarded by the gateway in a unified manner.
  - For design details about internal network interconnection and external network interconnection involved in the network egress zone, see the traditional campus network solution.
- Access authentication:
  - Edge nodes on the campus network provide access for dual-stack users. Single authentication and dual-stack service policy association needs to be implemented for dual-stack users. This helps prevent dual-stack terminals from undergoing two separate authentications when accessing IPv4 and IPv6 services.
  - A campus network with VXLAN underlay IPv4 + overlay dual stack must support various authentication modes (such as 802.1x, Portal, and MAC address authentication) for IPv6 users, so that authentication schemes can be implemented flexibly based on user terminals (such as Portal authentication for guest terminals and 802.1x authentication for internal office terminals). The deployment of network authentication points, policy enforcement points, and access points in the IPv6 solution is consistent with that in the original IPv4 solution. The authentication server uses a unified controller to provide authentication policy services.

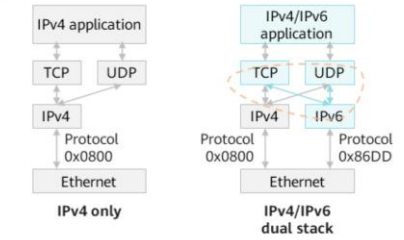


- Wireless terminal access solution:
  - The core switches or AC functions as the gateway for wireless access of office terminals. The SLAAC hybrid solution can be used to configure IPv6 addresses for terminals, and 802.1x authentication can be used.
  - Different VLANs are configured for guest SSIDs. The core switches or AC functions as the gateway for wireless access of guest terminals. The SLAAC hybrid solution is used to allocate IPv6 addresses. Portal authentication is recommended.
- AP management solution:
  - The CAPWAP tunnel supports both IPv4 and IPv6. However, only IPv4 or IPv6 can be selected at one time. That is, the AC can manage APs only in either IPv4 or IPv6 mode. The default mode is IPv4.
  - APs can go online in either IPv4 or IPv6 mode. That is, an AP can obtain only one IP address. The AC's IP address is manually configured. The AC can use DHCPv6 or SLAAC to assign IP addresses to APs.
- Note: You can run the **capwap ipv6 enable** command to enable the IPv6 function for the CAPWAP tunnel.

IPv6 Reconstruction of Other Related Systems (1)

- The key to internal application system reconstruction is to adjust address-related modules.

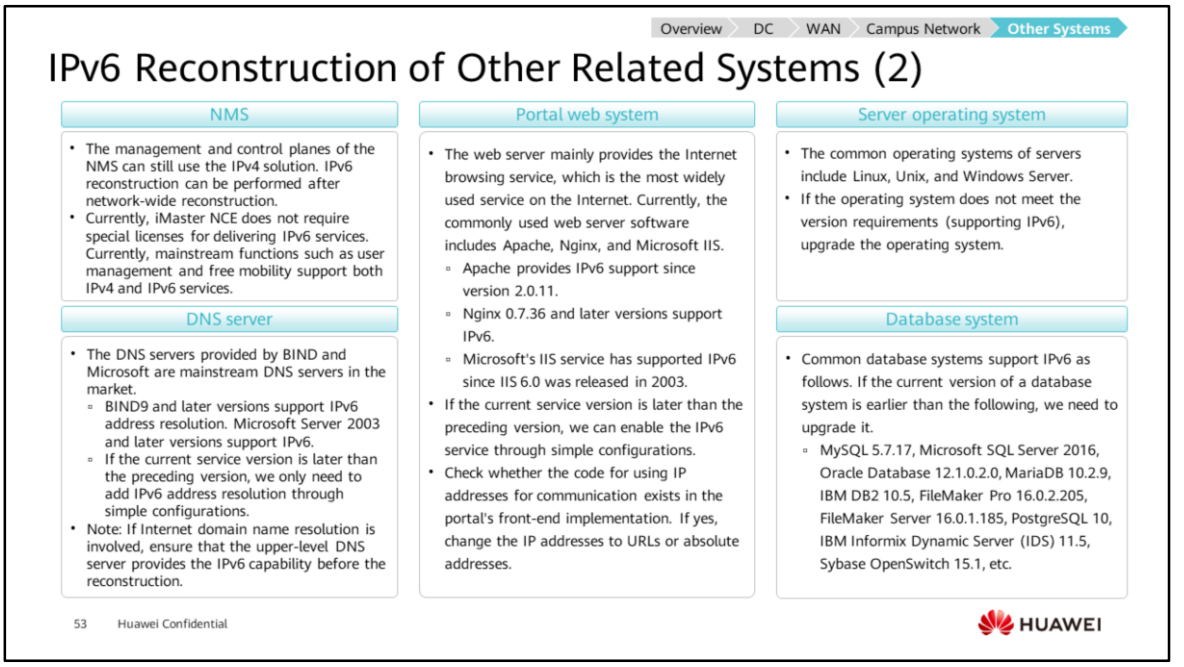
IPv6 reconstruction of application systems mainly involves the network layer.



- The network layer is used to transparently transmit data between two end systems and is a channel for communication between application systems.
- Application programs connect to application processes and network protocol stacks through sockets to implement binding between application programs and network protocol interfaces.

- General suggestions on IPv6 reconstruction of application systems:
  - Deploy an IPv6-capable DNS system and add AAAA records to the system. All service systems identify remote hosts by domain names instead of using IP addresses as host identifiers. Check whether applications can correctly parse the AAAA records of URLs.
  - Check whether the application system directly uses IP addresses. Ensure that IP addresses are used only as addresses, not as user IDs or key service attributes, and ensure that services are independent of IP address types.
  - Check the socket communication interface. Ensure that this interface has shifted from an IPv4-oriented programming interface to one that supports IPv4&IPv6 compatibility. Use IPv6-oriented functions, macros, and libraries to check whether the code for verifying IP address validity needs to be modified.

HUAWEI



- NMS IPv6 reconstruction:
  - The NMS is not user-oriented and requires only a small number of IP addresses. Therefore, IPv4 addresses can still be reserved for internal communication, and IPv6 addresses are only optional.
  - Through reconstruction, the NMS can also provide the following functions:
    - Identifies and manages various IPv6 address types.
    - Supports DNS resolution monitoring in IPv6.
    - Collects performance, resource, and fault data of IPv6 and dual-stack devices, and provides functions such as performance management, resource management, and fault management and analysis for these devices.
    - Associates IPv4 with IPv6 for dual-stack devices, so that the resource, performance, and fault data of these devices can be smoothly associated with historical data.
    - Accesses the IPv6 MIBs of devices through IPv4-based SNMP and obtains IPv6 configuration, traffic, and other information from the IPv6 MIBs.
- IPv6 reconstruction for server operating systems:
  - Common Linux versions that support IPv6 are as follows (IPv6 installed by default):
    - Fedora 13, Red Hat Enterprise Linux 6, Ubuntu 12.04, Oracle Solaris 10, SUSE Linux Enterprise Server 11, etc.
  - Windows Server supports IPv6 as follows:
    - Windows Server 2003 supports IPv6, but IPv6 is not installed by default and needs to be manually installed.
    - Windows Server 2008 and later versions have IPv6 installed by default.

Quiz

1. (Multiple-answer question) Which of the following deployment solutions may be used during WAN IPv6 evolution? ( )
   A. Dual stack
   B. 6VPE
   C. SRv6
   D. VXLAN underlay IPv4 + overlay dual stack
2. (True or false) IPE technologies include network programming, network slicing, deterministic networking, in-band flow measurement, new multicast, and application awareness. ( )
   A. True
   B. False

HUAWEI

1. ABC
2. A

Summary

- Enterprise networks generally involve DCNs, WANs, and campus networks. IPv6 evolution needs to take into consideration the overall capability upgrade and coordination of applications, terminals, and networks to ensure that the user experience of existing services is not affected during IPv6 evolution. IPv6 solution design must ensure smooth and continuous network evolution. During the solution design, consider using the optimal IPv6 technologies to build next-generation IPv6 informatization infrastructure and facilitate sustainable service development.
- This course describes the evolution trends of IPv6 networks in scenarios such as DCs, WANs, and campus networks and the technical application of IPE technologies.

HUAWEI

Thank you.

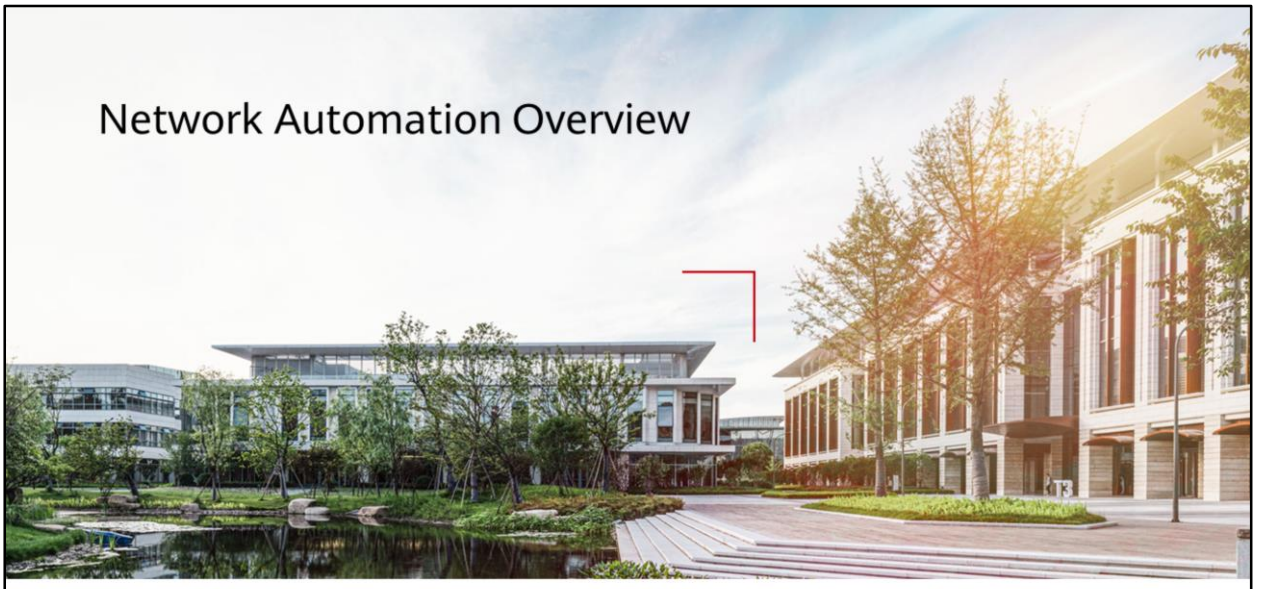
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Network Automation Overview



HUAWEI

Foreword

- Networks are evolving to be more open, more flexible, and simpler. Huawei's vision is to help customers build intent-driven networks (IDNs).
- Network openness and programmability are prerequisites of intent-driven networks.
- AI is enabling a wide range of industries. The combination of AI and networks is an inevitable trend.
- This course briefly introduces the current situation of the AI industry and Huawei's network AI engine iMaster NAIE.

HUAWEI

Objectives

- On completion of this course, you will be able to describe:
  - Background of network programming and automation.
  - Network openness levels.
  - Open capabilities of Huawei iMaster NCE.
  - Positioning and capabilities of network automation engineers.
  - Describe the requirements of network AI.
  - Describe the iMaster NAIE services of Huawei network AI engine.

HUAWEI

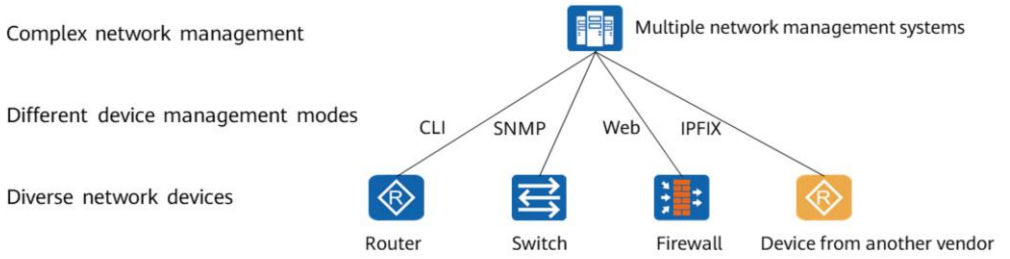
Contents

1. **Network Programming and Automation**
2. Network Automation Engineers
3. Network Automation Classification
4. Network AI

HUAWEI

Background: Complex Networks

- From the first day when the computer network is generated, the network is complex and difficult to manage, which is reflected in the following aspects:
  - There are various network devices such as routers, switches, firewalls, and intrusion detection system (IDS).
  - Devices from different or even the same vendors are managed in different ways.
  - Complex devices result in complex network management.

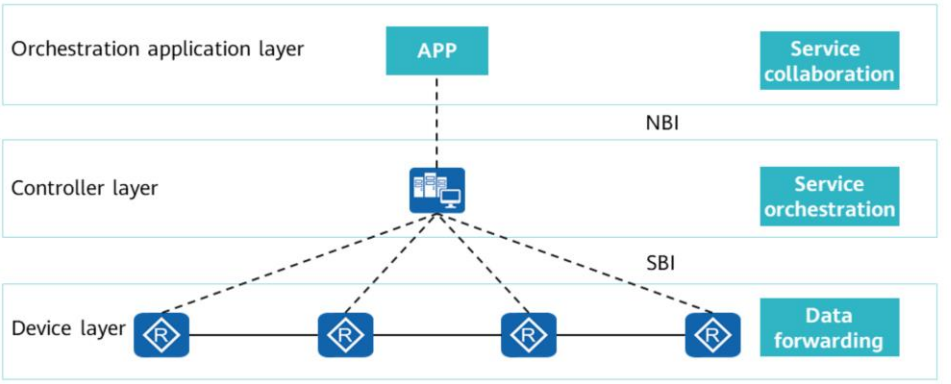


HUAWEI

- There are different device management modes, such as SNMP, CLI, IPFIX, and Web UI.

Background: Network Architecture Transformation

- Software defined networking (SDN) brings about network architecture transformation. It introduces a network controller to implement centralized control from a global perspective, achieving objectives such as fast service deployment, traffic optimization, and network service openness.

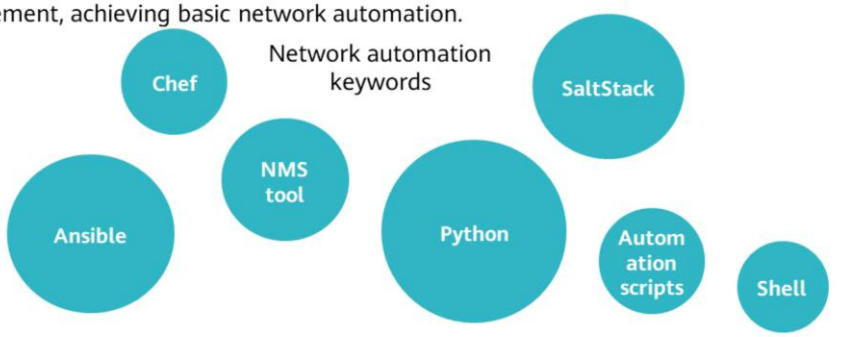


HUAWEI

- Orchestration application layer: implements various upper-layer applications of user intents. Typical orchestration applications include OSS and OpenStack. The OSS is responsible for service collaboration on the entire network. The OpenStack is used for network, computing, and storage service collaboration in a data center. There are other orchestration-layer applications. For example, a user wants to deploy a security application. The security application does not care about the device deployment location but invokes a controller NBI, for example, Block (Source IP, DestIP). Then the controller delivers an instruction to network devices. This instruction varies according to the southbound protocol.
- Controller layer: The entity at the controller layer is the SDN controller, which is the core of the SDN network architecture. The control layer is the brain of the SDN system. Its core function is to implement network service orchestration.
- Device layer: The network devices receive instructions from the controller and forward the instructions.
- NBIs: NBIs are used by the controller to interconnect with the orchestration application layer. The main NBIs are RESTful interfaces.
- SBIs: SBIs are protocols used for interaction between the controller and devices, including NETCONF, SNMP, OpenFlow, and OVSDB.

Basic Network Automation

- Network automation means that tools are used to implement automatic network deployment, operation, and maintenance, gradually reducing the dependency on human resources.
- There are many open-source tools that implement network automation in the industry, such as Ansible, SaltStack, Puppet, and Chef. These tools connect to devices through SSH to implement batch operation and management, achieving basic network automation.



HUAWEI

- Network automation tools implement basic network automation. That is, tools connect to devices through SSH to implement batch operation and management.

Development of Network Automation

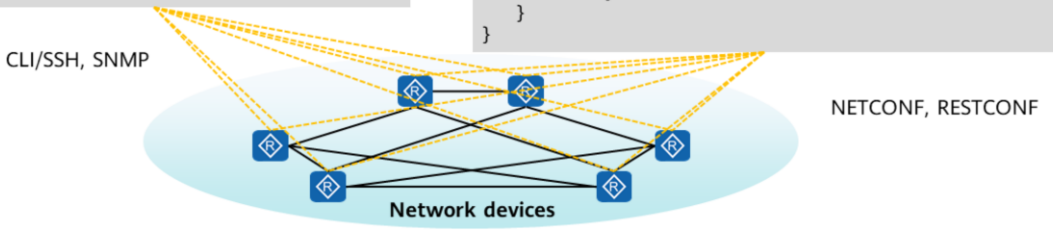
- In basic network automation, networks are managed on the CLI. The pain point is that network devices return unstructured data (text display), which is inconvenient for computers to process. The basic requirement of network automation development is that devices provide structured data, which can greatly promote the development of network automation. Devices provide NETCONF/RESTCONF interfaces to provide data in XML or JSON format.

Unstructured data: easy to understand for humans

Interface	InUti	OutUti	inErrors	outErrors
GigabitEthernet0/0/0	10%	20%	0	0
GigabitEthernet0/0/1	20%	30%	0	0
...				

Structured data: easy to understand for machines

```
{  "Interfaces":
     { "GigabitEthernet0/0/0":
        { "InUti": "10", "OutUti": "20", "inErrors": "0", "outErrors":
"0" },
       "GigabitEthernet0/0/1":
        {"InUti": "20", "OutUti": "30", "inErrors": "0",
"outErrors":"0"}
     }
}
```



HUAWEI

- Unstructured data can be easily understood by humans, but it is difficult for machines to understand and difficult for automatic data collection.

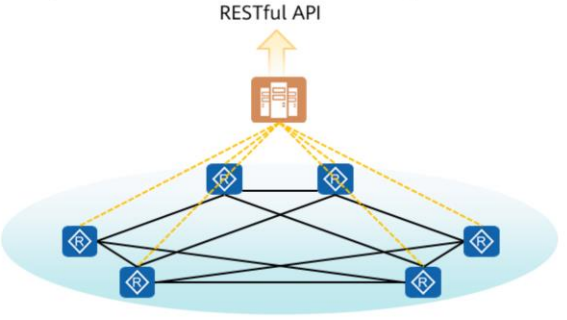
Network Openness and Programmability

- Network openness and programmability is to use programming methods to implement automated networks on the premise of open networks.
- In the early stage of SDN commercial use and in the future, traditional networks and SDN networks will coexist on a large scale. Network openness is implemented at two levels: device openness and SDN platform openness.

CLI/SSH, SNMP, NETCONF, RESTCONF, Telemetry, OPS, ...

Device-based openness and programmability

RESTful API

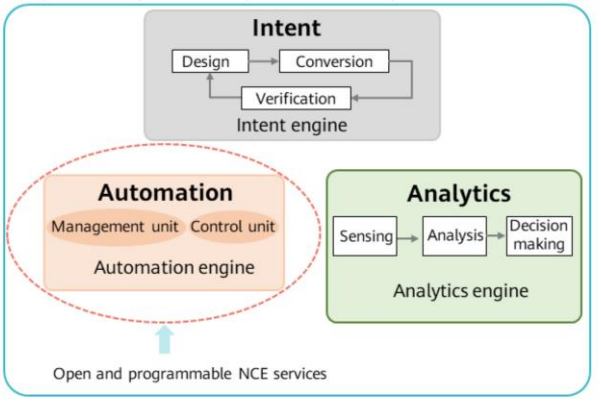


SDN-based openness and programmability

HUAWEI

Huawei iMaster NCE, Open and Programmable

- iMaster NCE (NCE for short) is an innovative network cloud engine of Huawei. The overall open programming capabilities of NCE include automation, analytics, and intent engines. The goal is to build a full-lifecycle open programmable architecture to meet customer requirements for openness and programmability. NCE service openness and programmability are designed to help customers implement service automation.



HUAWEI

- iMaster NCE is not only a controller, but also provides analysis and network management functions.

Contents

1. Network Programming and Automation
2. **Network Automation Engineers**
3. Network Automation Classification
4. Network AI

HUAWEI

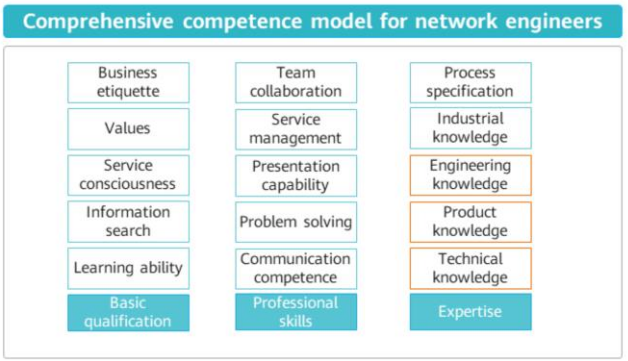
Network Automation Engineers

- Network engineers:
  - Network engineers are professionals who master network technologies, have professional skills, competence, and project management experience in the network engineering field. They are able to fully communicate with customers or other project stakeholders onsite. In addition, they can develop implementation solutions and project plans (recognized by project stakeholders) based on customer requirements and environment factors, fully mobilize resources of all parties to ensure timely and high-quality project implementation, and provide training for stakeholders and deliver engineering documents after the project is implemented.
- Network automation engineers:
  - They are professionals who have network skills, competence, and project management experience in the network automation field. They aim to meet the requirements of automatic deployment, development, and O&M of enterprise networks.

HUAWEI

Comprehensive Competence Model

- Compared with the comprehensive competence model of network engineers, network automation engineers have the same basic competence and professional skills, but have different skill requirements in terms of expertise.



In addition to basic network technologies, network automation engineers need to master development-related technology, product, and engineering knowledge.

HUAWEI

Expertise

- The expertise of network automation engineers must be all-rounders who master skills of network engineers, system engineers, and development engineers to some extent, including but not limited to the following capabilities:

Source code management	Source code control to help developers manage and store code
Programming language	A good command of at least one programming language
Operating system (OS)	Understanding the basic principles and mechanism of the OS
Network technology	A good command of basic network protocol principles and network engineering technologies
Open network architecture	A good command of the data structure and resource structure of open networks
Other professional knowledge	Other related professional knowledge, such as database knowledge

HUAWEI

- Network automation developers may need to have more professional knowledge, such as database, algorithm, cryptography, software development lifecycle management, development framework, big data, cloud computing, and artificial intelligence (AI), depending on the specific work content and scenario.

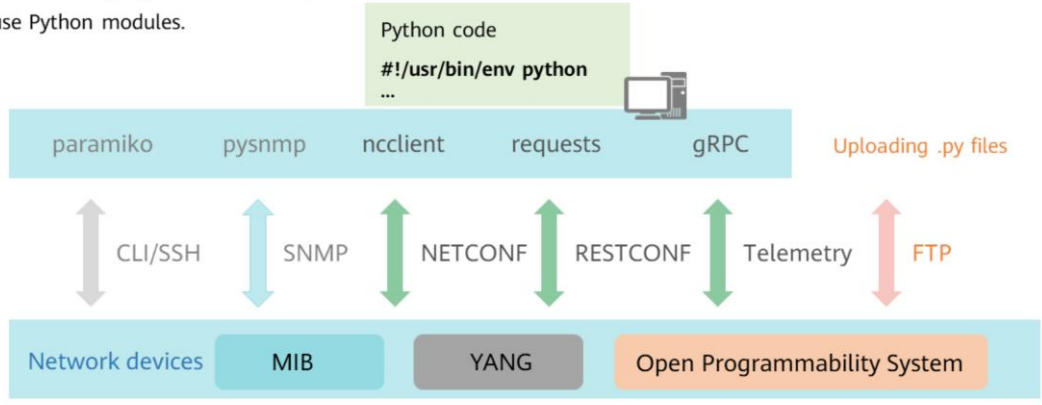
Contents

1. Network Programming and Automation
2. Network Automation Engineers
3. **Network Automation Classification**
4. Network Automation Engineers

HUAWEI

Device Openness and Programmability

- Device openness and programmability aim to provide engineers with guidance on how to implement network automation through programming based on device openness capabilities.
- The following figure shows the open capabilities of Huawei network devices. This course module focuses on how to use Python modules.

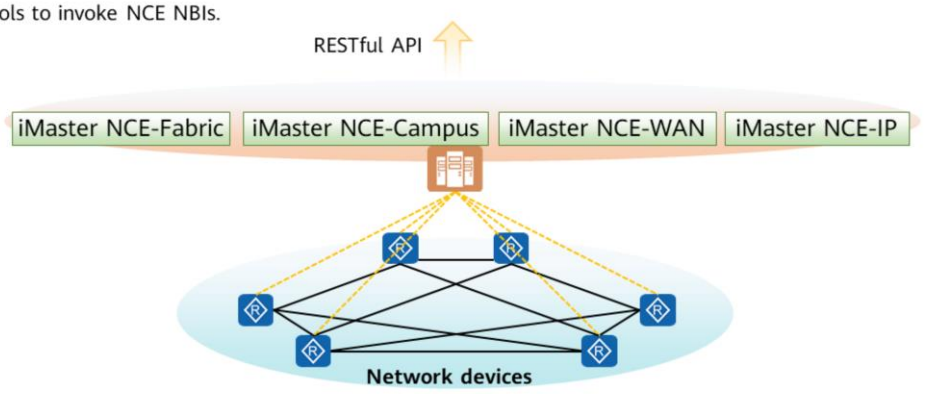


HUAWEI

- Part 1 of this course module describes how to use Python modules, including paramiko, pysnmp, ncclient, requests, and grpc, to communicate with devices.
- Part 2 focuses on the OPS. The OPS refers to open programmability provided by Huawei devices. You can upload Python code to a device, and the device runs the code to implement specified functions.

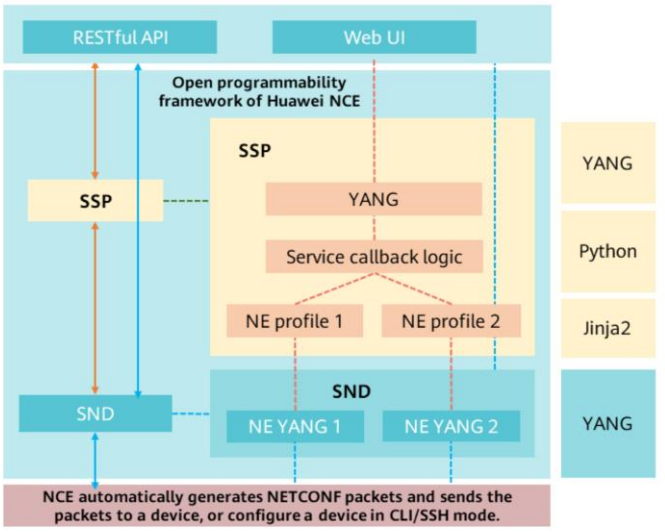
NCE Northbound Openness

- The openness and programmability of the controller provide engineers with guidance on how to implement network automation through programming based on the open capabilities of the SDN controller.
- Huawei iMaster NCE includes controllers and provides northbound RESTful APIs. This course module focuses on how to use tools to invoke NCE NBIs.



HUAWEI

Huawei NCE Service Openness and Programmability



- NCE service openness and programmability is a subsystem of NCE. It provides E2E programing capabilities including the openness of NE-layer functions and the openness of network service functions.
- Engineers compile an NE driver package (specific NE driver, SND) and a service driver package (specific service plugin, SSP) to implement the following functions:
  - Multi-vendor device management
  - User-defined network services
  - Web-based service management
  - Automatic exposure of northbound RESTful APIs

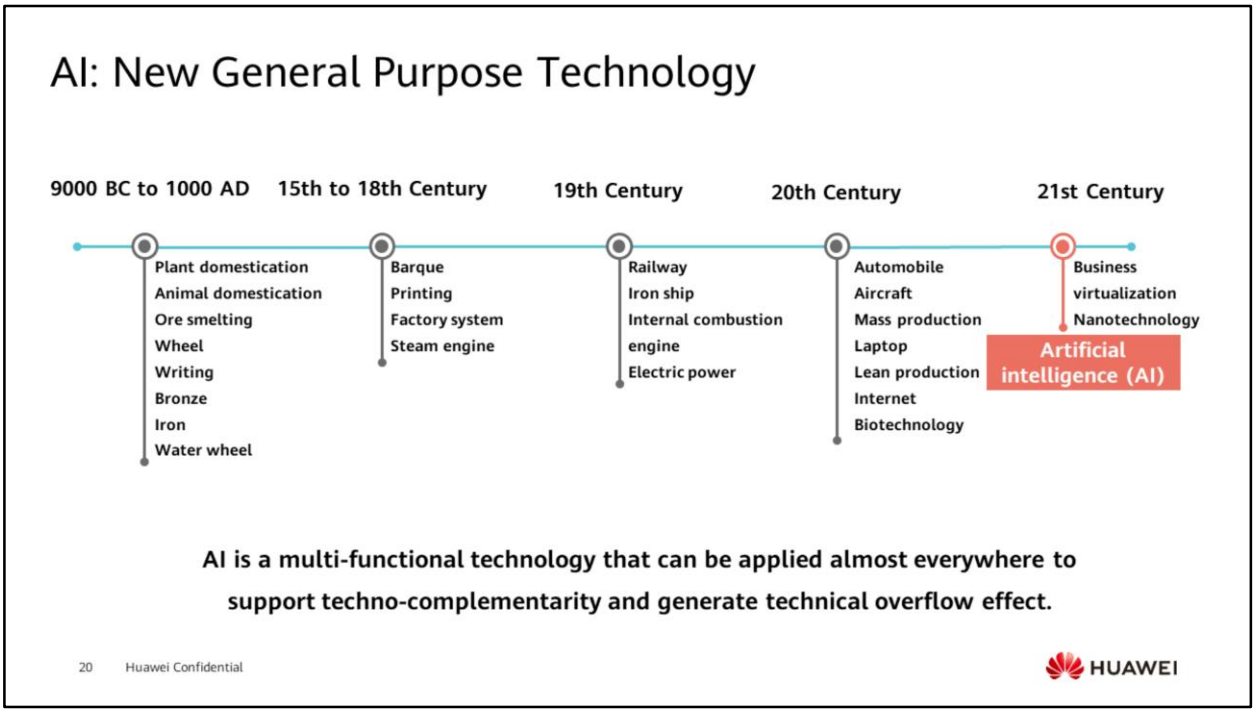
HUAWEI

- An SND abstracts device capabilities based on a device YANG model. A user can generate an SND based on device YANG files and a few Python code. After the SND is uploaded to NCE, device management and service provisioning can be implemented. SND types include NETCONF SNDs, CLI SNDs, and customized SNDs.
  - NETCONF SND: provides the capability of converting YANG files into NETCONF files.
  - CLI SND: provides the conversion capability from YANG to CLI.
  - Customized SND: provides the capability of converting YANG to other protocols such as RESTCONF.
- An SSP allows user to customize network services (apps), for example, quickly provision L3VPN services. These types of services or application involve multiple devices and protocols and are presented as an SSD. To compile an SSD, an engineer needs to compile service YANG files, Python scripts (service callback logic) for service mapping, and Jinja2 template. The basic principles are as follows (from north to south):
  - A service model automatically generates northbound interfaces or UIs, which are invoked by an external system to initiate a service request.

Contents

1. Network Programming and Automation
2. Network Automation Engineers
3. Network Automation Classification
4. **Network AI**

HUAWEI



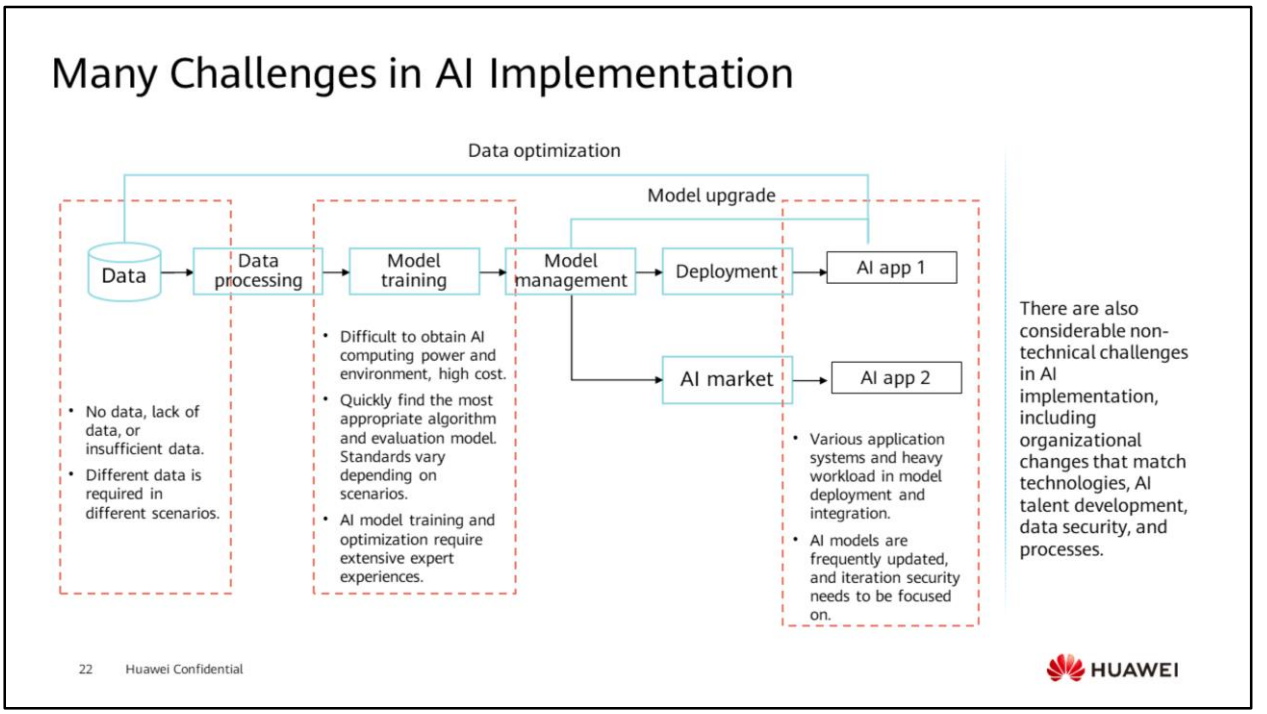
- General Purpose Technology (GPT) is the main driving force for economic and social transformation. From the agricultural society to the industrial society and then to the information society, the production mode, life mode, and management mode of the human society have undergone tremendous changes and experienced unprecedented economic and social transformation. For a long time, people have been thinking and exploring the drivers of economic and social development and transformation. From the first technological revolution represented by steam engine to the second technological revolution represented by electricity technology, looking at the industrial and technological revolution of the past 300 years, we can see that science and technology are important sources for promoting sustained economic growth. AI has become a new general purpose technology. Currently, popular AI technologies are being implemented, enabling a wide range of industries.
- https://en.wikipedia.org/wiki/General_purpose_technology
- Richard G. Lipsey, etc., Economic Transformations: General Purpose Technologies and Long-Term Economic Growth.

AI-based Network Becomes a Consensus in the Industry

Manual Processing -> Tool Assistance -> Intelligent Processing

Definition/Level	L0 Manual operation	L1 Tool-assisted automation	L2 Partially autonomous networks	L3 Restricted autonomous networks	L4 Highly autonomous networks	L5 Fully autonomous networks
Repeated execution	Man	Man-Machine	Machine	Machine	Machine	Machine
Context awareness	Man	Man	Man-Machine	Machine	Machine	Machine
Analysis and decision-making	Man	Man	Man	Man-Machine	Man-Machine	Machine
Service experience	Man	Man	Man	Man	Man-Machine	Machine
Scope	N/A	Subtask level	Unit level	Domain level	Service level/Cross-domain	Public ICT infrastructure

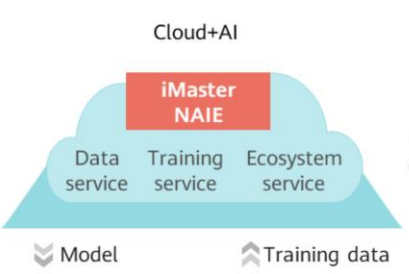
HUAWEI



- Technically, AI case training requires joint development across domains (such as data, algorithm, and expert experience). Model optimization requires continuous iterative training, which has the following difficulties:
  - The success of AI projects requires the cooperation of service experts and AI experts.
  - It is difficult for service experts to transform into AI experts.
  - There are many data issues, such as a few data sources and the requirement for data governance (labor-intensive).
  - There are many algorithm engineering issues, such as conversion from paper to code and the efficiency of open-source algorithms.
  - The computing power is difficult to obtain. The computing power is used during peak hours. (Nvida does not allow the use of G series GPUs in data centers.)
- Essentially, AI will bring about organizational transformation, from human resources to human-machine coexistence in the AI Ops phase.

Network AI Engine iMaster NAIE

- iMaster NAIE is a one-stop AI application development cloud platform that Huawei introduces AI technologies to telecom networks. Based on the public cloud mode, iMaster NAIE provides the data service, model training service, model generation service, and communication model service, covering the most complex parts of network AI app development, such as data preparation, data feature exploration, and model optimization. This helps developers quickly obtain NAIE capabilities.



2

Data processing

Efficient governance, reducing data preparation time by 90%.

1

Cloud data collection

3

Model development and training

Embedded Huawei experiences and improved training efficiency.

Low-threshold model generation and zero code development.

HUAWEI

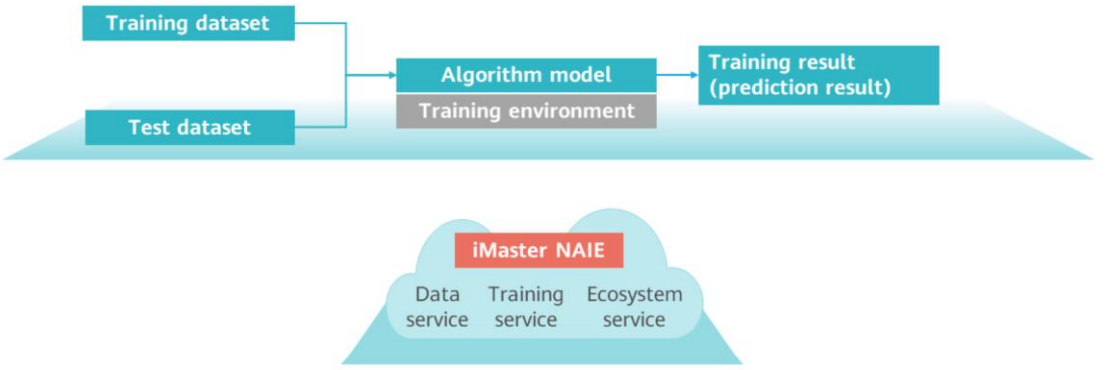
Example: Intelligent Traffic Sorting

- Requirement description:
  - Data center networks carry various services, including big data, distributed storage, high-performance computing, and GPU cluster services. Traffic sorting is a process in which data center switches identify and classify traffic packets and determine the service to which each data flow or data packet belongs.
  - It is required that an AI algorithm for traffic classification be implemented based on the embedded AI platform APIs (compatible with TensorFlow and Caffe) of Huawei's AI Fabric switches.

HUAWEI

Training of the Intelligent Traffic Sorting Model

- The dataset, algorithm model, and training environment required for intelligent traffic sorting can be provided by the iMaster NAIE.



HUAWEI

- The final objective of the experiment is to deploy the trained model in a real environment. Therefore, it is expected that the trained model can obtain a good prediction effect on real data. That is, it is expected that a smaller error between a prediction result of the model and the real result on real data is better. The best method is to divide real data into a training dataset and a test dataset. We can use the training dataset to train the model, and then use the error of the test dataset as the error of the final model in actual scenarios. With the test dataset, to verify the final effect of the model, we can calculate the error of the trained model only based on the test dataset. A smaller error indicates a better algorithm model.
- For detailed operations, see the following website: https://devstar.developer.huaweicloud.com/devstar/code-templates/e9078ee2d7024ffabbac3f8fd1bad806
- For more information about AI, refer to Huawei AI certification documents.

Quiz

1. (Multiple-answer question) Which of the following protocols are used by the controller to interact with devices? ( )

   A. NETCONF

   B. SNMP

   C. OpenFlow

   D. RESTful

HUAWEI

1. ABC

Summary

- Network programing and automation is to use programming methods to implement automated networks on the premise of open networks. Network openness is implemented at two levels: device openness and SDN platform openness.
- Network automation engineers must be all-rounders who master skills of network engineers, system engineers, and development engineers to some extent to support enterprise network automation.
- This course consists of four modules: programming basics, device openness and programmability, NCE northbound openness, and NCE service openness and programmability. Next, let's learn network programmability.
- The combination of network and AI is the development trend of the network industry.
- Huawei iMaster NAIE provides network-based AI services, making network AI openness easier.

HUAWEI

Thank you.

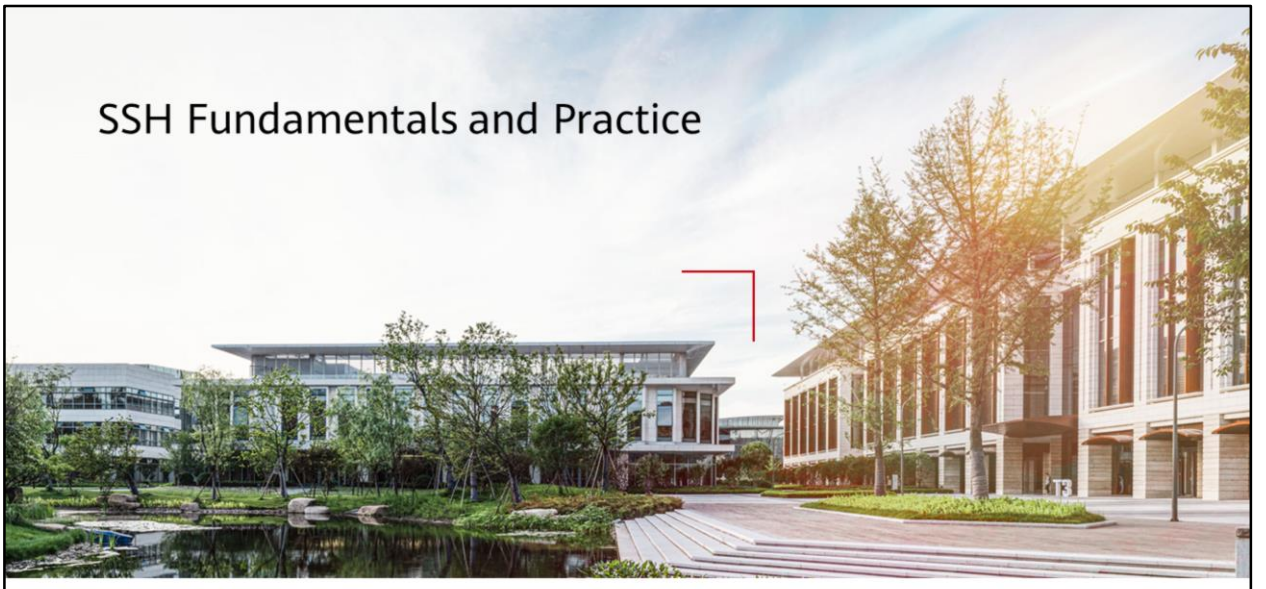
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

SSH Fundamentals and Practice



HUAWEI

Foreword

- To achieve efficient O&M and enhance agility amid increasingly complex service requirements and network architecture, network automation is gaining momentum and ever evolving. Currently, Secure Shell (SSH) is the most common method used by engineers to log in to devices for remote management. As such, engineers are expected to learn about and use an automation tool to implement SSH remote login, simulate man-machine interaction with O&M personnel, and automatically transfer files.
- In this course, we will use the Python Paramiko module to write automation scripts to implement SSH-based preliminary network automation.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the basic concepts and working principles of SSH.
  - Understand the concept of Paramiko.
  - Master the composition and common methods of Paramiko.
  - Grasp the common methods for implementing Paramiko.

HUAWEI

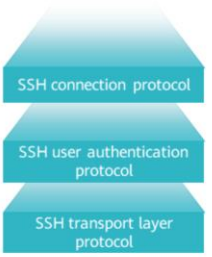
Contents

1. **Introduction to SSH**
   - Overview of SSH
     - Working Principles of SSH
     - Overview of Paramiko
2. Paramiko Component Architecture
3. SSH Practices

HUAWEI

Overview of SSH

- Secure Shell (SSH) is a protocol for secure remote login and other secure network services over an insecure network.
- SSH consists of the following sub-protocols: SSH transport layer protocol, SSH user authentication protocol, and SSH connection protocol.



Establishes a session connection.

Authenticates users (password and key).

Negotiates the version and algorithm and exchanges keys.

HUAWEI

SSH Transport Layer Protocol

- SSH transport layer protocol is a secure transport protocol. The SSH transport layer is usually established over TCP/IP connections. It can also be established over any other reliable data flow.
- The SSH transport layer protocol negotiates all key exchange algorithms, public key algorithms, symmetric encryption algorithms, and message authentication algorithms.

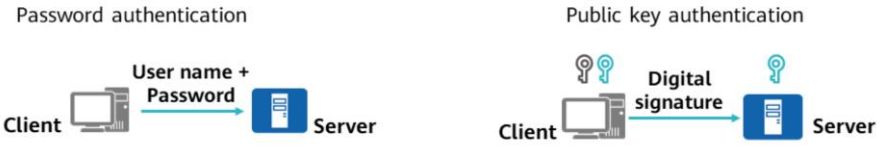
Algorithm Type	Function	Name
Key exchange algorithm	Generates session keys.	diffie-hellman-group14-sha1, diffie-hellman-group1-sha1, etc.
Public key algorithm	Performs digital signature and user authentication.	ssh-rsa, ssh-dss, etc.
Symmetric encryption algorithm	Encrypts sessions.	aes128-ctr, 3des-cbc, etc.
Message authentication algorithm	Verifies data integrity.	hmac-sha1, hmac-md5, etc.

HUAWEI

- Perfect Forward Secrecy (PFS) is a property of secure communication protocols. PFS was proposed by Christoph G. Gunther in 1990. PFS is essentially defined as the cryptographic property of a key-establishment protocol in which the compromise of a session key or long-term private key after a given session does not cause the compromise of any earlier session. PFS ensures that any future disclosure of passwords or keys cannot be used to decrypt any communications sessions recorded in the past, even if the attacker proactively intervenes.
- The SSH transport layer protocol uses the Diffie-Hellman key exchange algorithm to implement PFS.
- For details, see section 9.3.7 "Forward Secrecy" in RFC 4251 (https://www.ietf.org/rfc/rfc4251.txt).

SSH User Authentication Protocol

- The SSH user authentication protocol authenticates the client-side user to the server. It runs over the transport layer protocol.
- The SSH user authentication protocol provides two authentication methods: password authentication and public key authentication.
  - Password authentication: The client uses the user name and password for authentication before successfully logging in to the server.
  - Public key authentication: The server decrypts the digital signature of the client by using a public key.

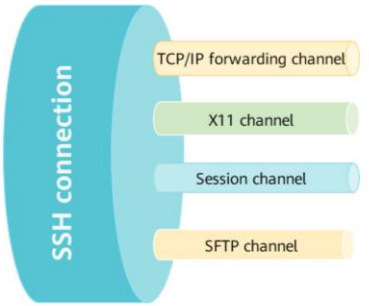


HUAWEI

- A public key is used to decrypt information to ensure message authenticity and integrity. Therefore, the receiver knows that the information comes from someone who has a private key. The encrypted information is called a digital signature. The public key is in the form of a digital certificate.
- When the SSH user authentication protocol is started, it receives a session ID from the SSH transport layer protocol. The session ID uniquely identifies a session and is a part of the digital signature to indicate the ownership of the private key.

SSH Connection Protocol

- The SSH connection protocol multiplexes several logical channels into a single encrypted tunnel. It provides interactive login sessions, remote execution of commands, forwarded TCP/IP connections, and forwarded X11 connections.
- The SSH connection protocol runs on top of the SSH transport layer protocol and user authentication protocol.



HUAWEI

- A TCP/IP connection can forward network data of other TCP ports through SSH channels, ensuring security.
- Data of Telnet, SMTP, IMAP, and other TCP/IP-based insecure protocols can be forwarded through SSH, which prevents the transmission of user names, passwords, and privacy information in plaintext and therefore enhances security. In addition, if the firewall restricts the use of some network ports but allows the SSH connection, communication can be implemented through the SSH TCP/IP connection.
- In X11, X refers to the X protocol, and 11 is the eleventh version of the X protocol. The Linux graphical user interface (GUI) is based on the X protocol at the bottom layer. When remote interaction with graphical applications on the Linux server is required, a method for enhancing communication security is to use SSH to display the GUI on the local client through the X11 tunnel.
- A session is a remote execution of a program. A program can be a shell, an application, a system command, or some built-in subsystems. Multiple session channels can be active at the same time. An interactive login session can be implemented using the invoke_shell() method, and the remote command can be implemented using the exec_command() method, which will be described in detail later.

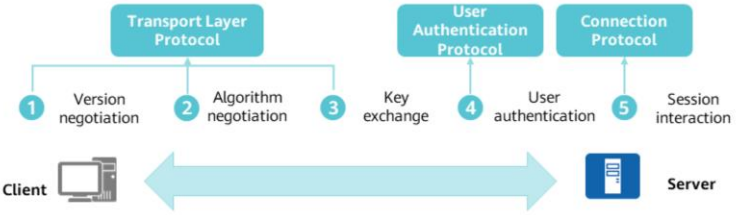
Contents

1. **Introduction to SSH**
   - Overview of SSH
   - Working Principles of SSH
   - Overview of Paramiko
2. Paramiko Component Architecture
3. SSH Practices

HUAWEI

Working Principles of SSH

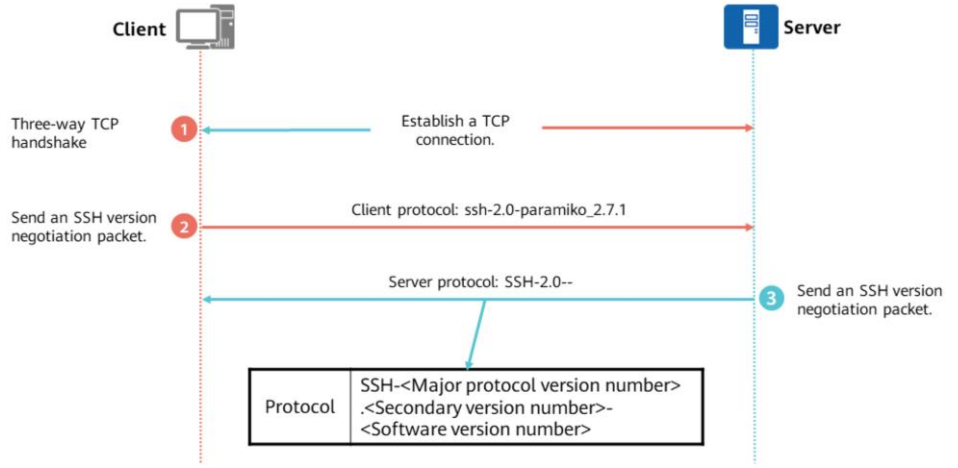
- In the entire communication process, to implement a secure SSH connection, the server and client go through the following five phases:
  - Version negotiation phase: Two versions of SSH are available: SSH Version 1 (SSHv1) and SSH Version 2 (SSHv2). The server and client determine the version to be used through negotiation.
  - Algorithm negotiation phase: SSH supports multiple encryption algorithms. The server and client negotiate the encryption algorithm to be used based on the algorithms that they support.
  - Key exchange phase: A session key is generated by using a key exchange algorithm. The subsequent sessions between the server and client are encrypted by using the session key.
  - User authentication phase: The SSH client sends an authentication request to the server, and the server authenticates the SSH client.
  - Session interaction phase: After the authentication succeeds, the server and client exchange information.



HUAWEI

Version Negotiation Phase

- The client and server exchange SSH version negotiation packets to determine whether to use SSHv1 or SSHv2.

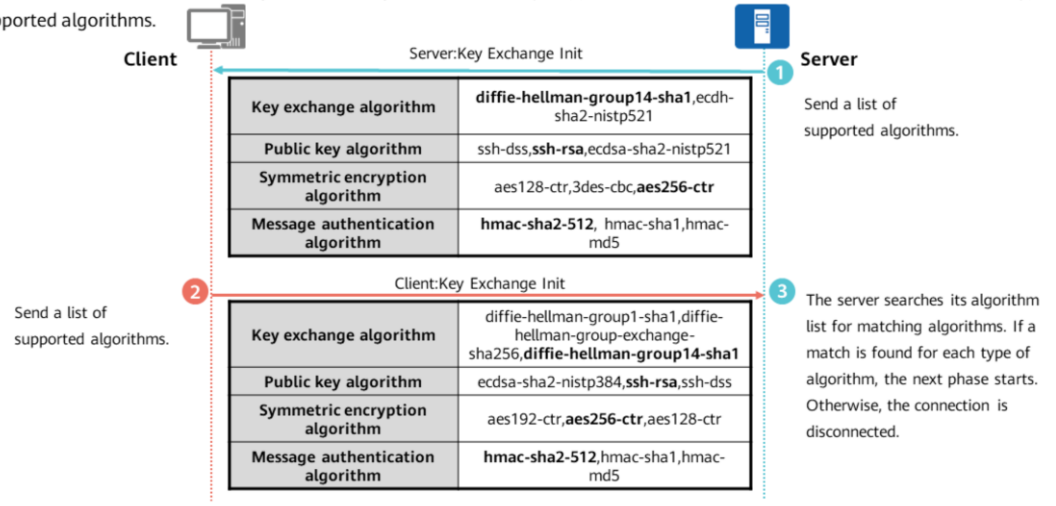


HUAWEI

- Port 22 is enabled on the server, waiting for the client to connect. The client initiates a TCP connection to the server. The two parties complete the handshake and establish a connection. The client sends a packet to the server. The packet contains the version field, in the format of Major version number.Secondary version number-Software version number. After receiving the packet, the server parses it to obtain the protocol version number. If the protocol version number of the client is earlier than that of the server and the server supports the earlier version of the client, the server uses the protocol version number of the client. Otherwise, the server uses its own protocol version number.

Algorithm Negotiation Phase

- The client and the server exchange a list of algorithms that they support. The list includes specific names of the four types of supported algorithms.

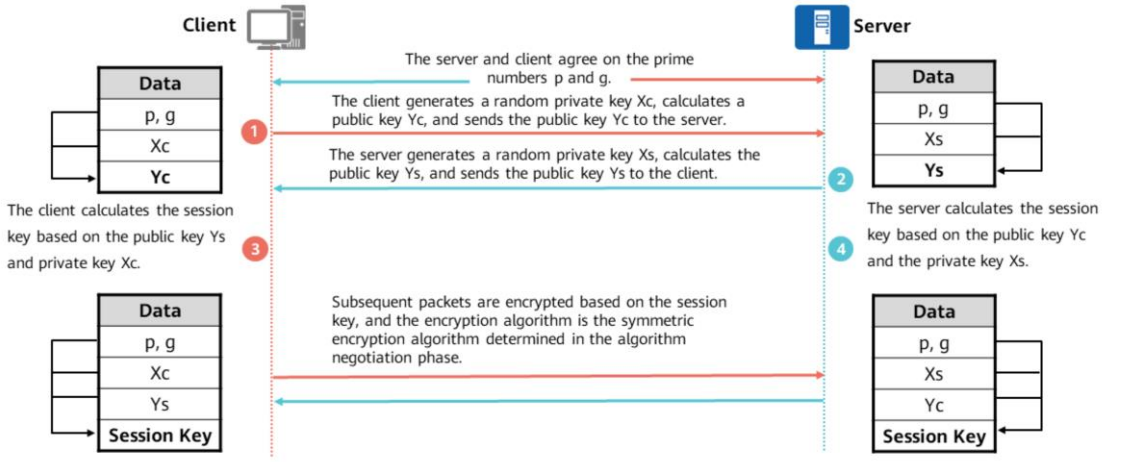


HUAWEI

- The algorithm negotiation process is as follows: The server obtains the first algorithm from the algorithm list of the client and searches its own algorithm list for the same algorithm. If the same algorithm is found, the negotiation succeeds, and the server continues to negotiate the algorithm of the next type. Otherwise, the server searches its own algorithm list for the next algorithm in the client's algorithm list until a match is found.

Key Exchange Phase

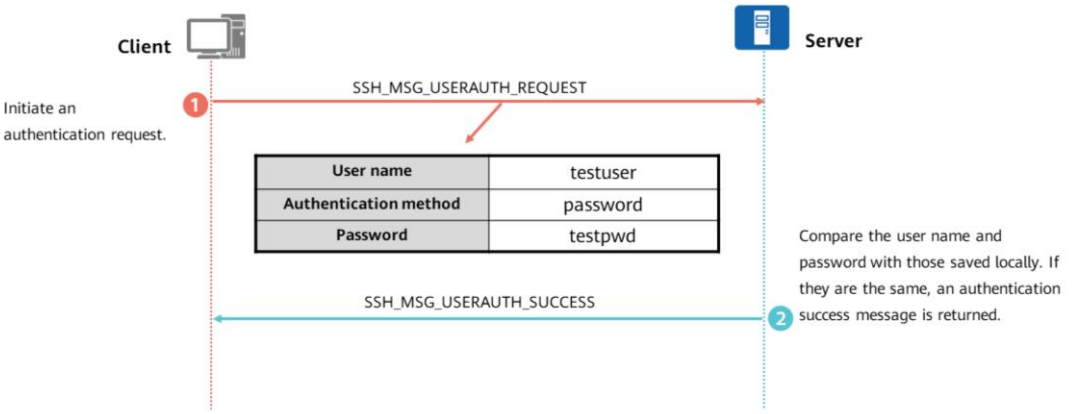
- Based on the key exchange algorithm, the server and client dynamically generate a session key for subsequent session encryption. The session key cannot be intercepted by a third party, enhancing security and reliability.



- The client and the server first agree on two public prime numbers p and g.
- The client and server each randomly generate a private key Xc and Xs, respectively.
- The client and server each calculate their own public key Yc and Ys, respectively.
- The client and server exchange their own public key.
- The client and server calculate the session key for encryption based on the public and private keys.
- The Diffie-Hellman key exchange algorithm is used for key exchange, which is based on the mathematical discrete logarithm and is not described in this course. During key exchange, the private keys Xc and Xs are not transferred and, due to the difficulty in computing discrete logarithms, they cannot be decrypted by other users even if p, g, Yc, and Ys are obtained. This ensures the confidentiality of the session keys.
- Note that the public and private keys generated in this phase are used only to generate session keys and are irrelevant to subsequent user authentication. After the key exchange phase is complete, all subsequent packets are encrypted based on the session keys.

User Authentication Phase: Password Authentication

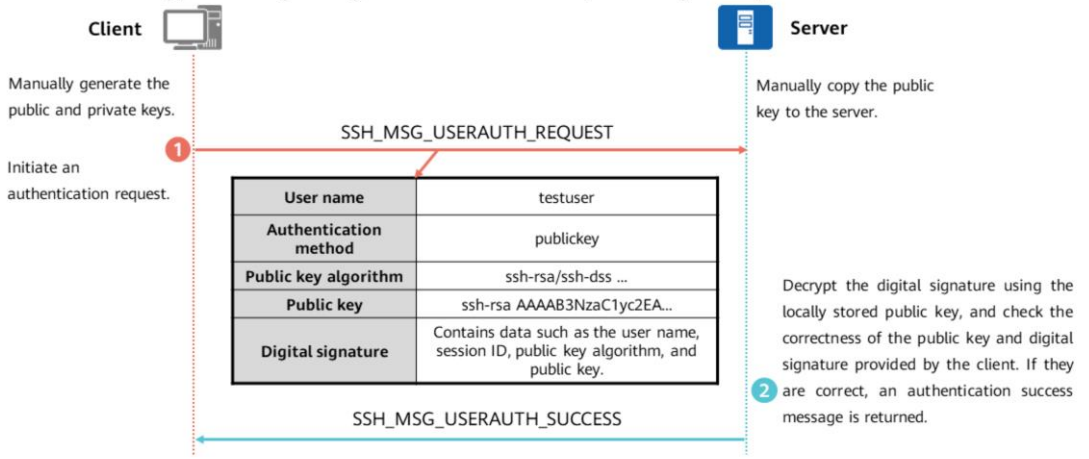
- There are two user authentication modes: password authentication and public key authentication.
- During password authentication, the client sends an authentication request carrying the user name and password, and the server authenticates the received user information against the local user information.



HUAWEI

User Authentication Phase: Public Key Authentication

- During public key authentication, a client sends an authentication request carrying a digital signature, and the server decrypts the digital signature based on the public key to implement authentication.

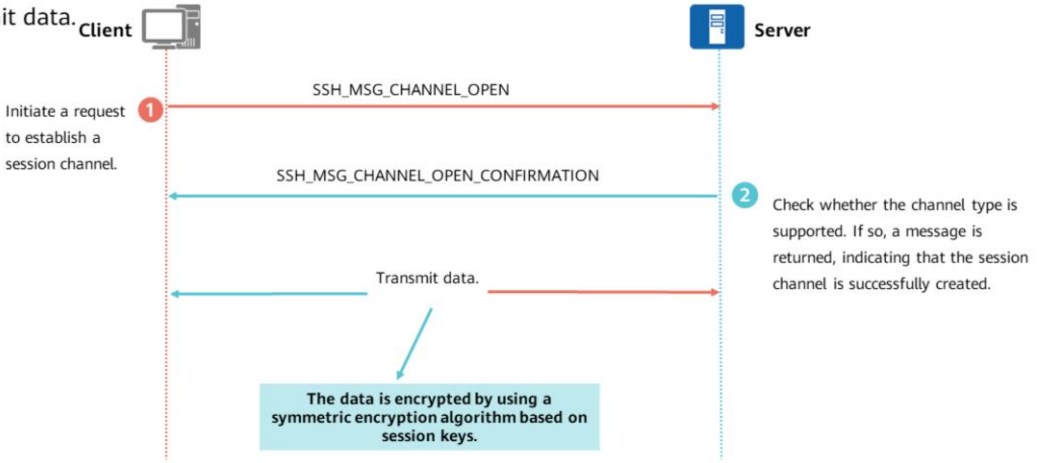


HUAWEI

- The digital signature is encrypted by client's private key. To see the content, we need public key to decrypt it.

Session Interaction Phase

- After the user is authenticated, the client sends a request to the server for establishing a channel to transmit data.



HUAWEI

- The channel types include session, x11, forwarded-tcpip, and direct-tcpip.
- For details, see section 4.9.1 "Connection Protocol Channel Types" in RFC4250 at https://www.ietf.org/rfc/rfc4250.txt.
- Different ssh logical channels can multiplex one ssh session.

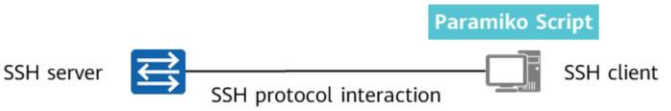
Contents

1. **Introduction to SSH**
   - Overview of SSH
   - Working Principles of SSH
   - Overview of Paramiko
2. Paramiko Component Architecture
3. SSH Practices

HUAWEI

Overview of Paramiko

- Paramiko is a Python module that implements the SSHv2 protocol. It supports password authentication and public key authentication and implements functions such as secure remote command execution and file transfer.
- Engineers can compile Python code based on the Paramiko module to implement SSH functions.



HUAWEI

- In HCIA courses, we learned how to use the telnetlib module for Telnet remote connections. In the production environment, the more secure Paramiko module is recommended for SSH remote connections.

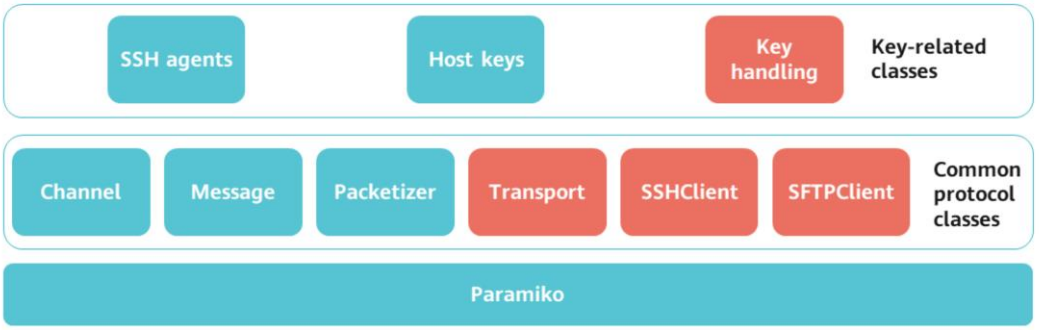
Contents

1. Introduction to SSH
2. **Paramiko Component Architecture**
   - Paramiko Component Architecture
   - Transport Class and Its Methods
   - Key Handling Class and Its Methods
   - SFTPClient Class and Its Methods
   - SSHClient Class and Its Methods
3. SSH Practices

HUAWEI

Paramiko Component Architecture

- The following figure shows the components of the Paramiko module. SSHClient and SFTPClient are its most commonly used classes, which provide the SSH and SFTP functions, respectively.
- This course describes the methods of the Transport, key handling, SSHClient, and SFTPClient classes.



HUAWEI

Common Paramiko Classes

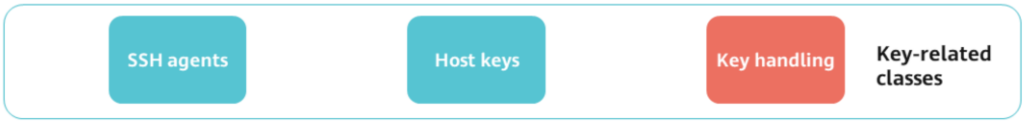
- Channel: This class is used to create a secure channel over the SSH transport layer.
- Message: An SSH message is a stream of bytes that encodes some combinations of strings, integers, bools, and infinite-precision integers (known in Python as longs).
- Packetizer: This class is used for packet handling.
- Transport: This class is used to create a transport session object over an existing socket or socket-like object.
- SFTPClient: This class creates an SFTP session connection through an open SSH transport session and performs remote file operations.
- SSHClient: This class is an advanced representation of a session with the SSH server. This class integrates the Transport, Channel, and SFTPClient classes.

HUAWEI

- The Channel class provides methods for executing commands, requesting X11 sessions, sending data, and opening interactive sessions. Generally, these common methods from the Channel class have been packaged in the SSHClient class.
- The Message class provides methods for writing bytes to a stream and extracting bytes.
- The Packetizer class provides methods for checking handshakes and obtaining channel IDs.
- The Transport class provides methods such as public key authentication, private key authentication, and channel opening.
- The SSHClient class provides methods for establishing connections and opening interactive sessions.
- The SFTPClient class provides methods such as file upload and download.

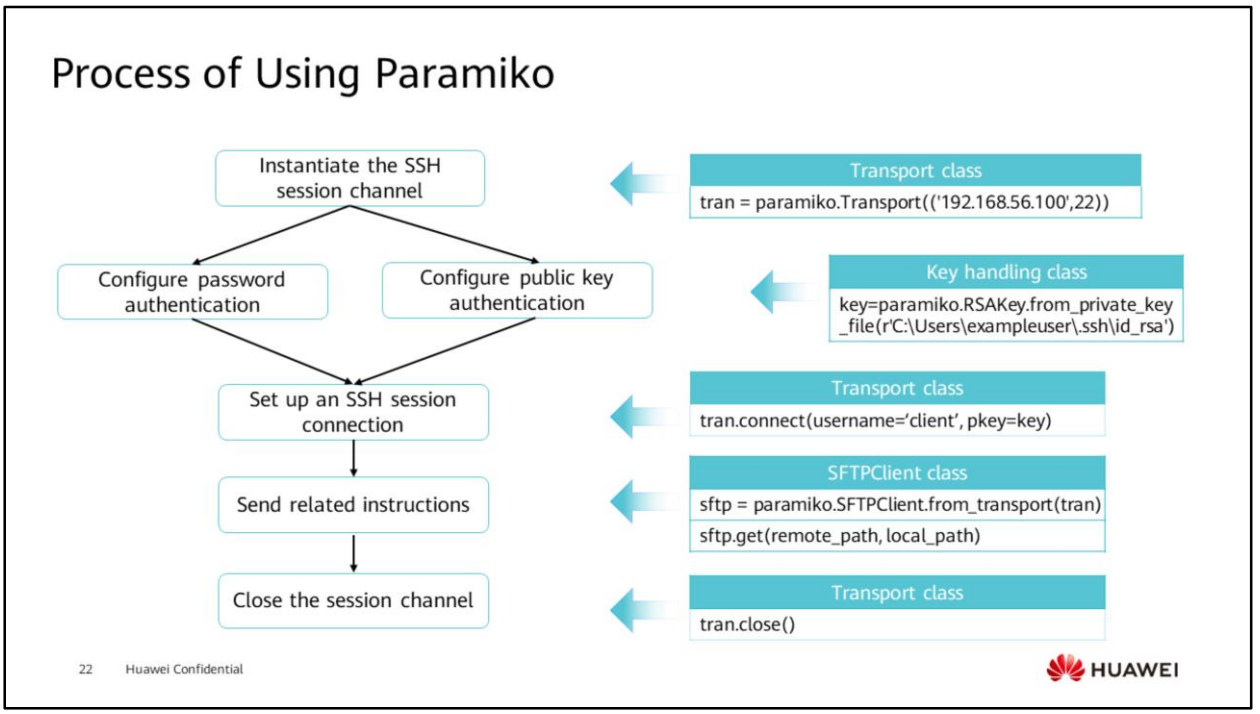
Key-Related Classes of the Paramiko Module

- SSH Agent: This class is used for the SSH agent.
- Host keys: This class is related to the OpenSSH known_hosts file and is used to create a host keys object.
- Key handling: This class is used to create instances of the corresponding key type, for example, RSA keys and DSS (DSA) keys.



HUAWEI

- OpenSSH is a free open-source implementation of the SSH protocol. It provides server programs and client tools. OpenSSH is integrated in all Linux operating systems. OpenSSH records the public key of each computer that a user has accessed in ~/.ssh/known_hosts. When the same computer is accessed next time, OpenSSH checks the public key. If the public keys are different, OpenSSH generates a warning to prevent man-in-the-middle attacks.



- This course describes methods of four classes: Transport, key handling, SSHClient, and SFTPClient.
- This process uses the Paramiko SFTP session as an example. Because the SSHClient class integrates the Transport, Channel, and SFTPClient classes, the preceding methods can be implemented by the SSHClient class. This is especially true for SSH sessions.

Contents

1. Introduction to SSH
2. **Paramiko Component Architecture**
   - Paramiko Component Architecture
   - Transport Class and Its Methods
   - Key Handling Class and Its Methods
   - SFTPClient Class and Its Methods
   - SSHClient Class and Its Methods
3. SSH Practices

HUAWEI

Transport Class and Its Methods

- Transport class: An SSH transport connects to a stream (usually a socket) to negotiate and encrypt sessions and perform authentication. Channels can then be created based on the encrypted sessions. Multiple channels can be multiplexed in a single session connection (in fact, this is often the case, such as port forwarding).
- The following is an example of the method:

```
tran = paramiko.Transport(('192.168.56.100', 22))
tran.connect(username='client',  password='test')
```

Common Method	Function
Transport(sock)	Creates a Transport object and instantiates the SSH session channel.
connect(username='',password=None,pkey=None)	Establishes an SSH session connection and uses a password or private key for identity authentication.
close()	Closes the session.

HUAWEI

- For ease of use, you can use an address (as a tuple) or a host string as the sock parameter. The host string is the host name with an optional port, separated by a colon (:). If a port is transferred, it is converted to a tuple in the format (host name, port).

Contents

1. Introduction to SSH
2. **Paramiko Component Architecture**
   - Paramiko Component Architecture
   - Transport Class and Its Methods
   - Key Handling Class and Its Methods
   - SFTPClient Class and Its Methods
   - SSHClient Class and Its Methods
3. SSH Practices

HUAWEI

Key Handling Class and Its Methods

- The key handling class is used to create instances of the corresponding key type, for example, RSA keys and DSS (DSA) keys. This class provides methods for reading and writing keys.
- The following is an example of the method:

```
key=paramiko.RSAKey.from_private_key_file(r'C:\Users\exampleuser\.ssh\id_rsa')
```

Common Method	Function
RSAKey.from_private_key_file(filename)	Reads the RSA private key from a file to create a key object.
DSSKey.from_private_key_file(filename)	Reads the DSS private key from a file to create a key object.

HUAWEI

Contents

1. Introduction to SSH
2. **Paramiko Component Architecture**
   - Paramiko Component Architecture
   - Transport Class and Its Methods
   - Key Handling Class and Its Methods
   - SFTPClient Class and Its Methods
   - SSHClient Class and Its Methods
3. SSH Practices

HUAWEI

SFTPClient Class and Its Methods

- The SFTPClient class creates an SFTP session connection through an open SSH transport session and performs remote file operations.
- The following is a typical example:

```
key=paramiko.RSAKey.from_private_key_file(r'C:\Users\exampleuser\.ssh\id_rsa')
tran = paramiko.Transport(('192.168.56.100', 22))
tran.connect(username='client', pkey=key)
sftp = paramiko.SFTPClient.from_transport(tran)
local_path=r'C:\Users\exampleuser\.ssh\vrptest.cfg'
remote_path= '/vrpcfg.cfg'
sftp.get(remote_path, local_path)
```

Common Method	Function
from_transport()	Creates an SFTP session connection through an open Transport session channel.
get()	Downloads a specified file.
put()	Uploads a specified file.

HUAWEI

Method from_transport

- from_transport(): This method creates an SFTP client channel from the enabled Transport session channel.
- The following is an example of the method:

```
t = paramiko.Transport(('192.168.56.100', 22))
sftp = paramiko.SFTPClient.from_transport(t)
```

Parameter	Description
T	An authenticated and enabled Transport session, in the format of (hostname,port).
windows_size	Size of the SFTP session window. This parameter is optional.
max_packet_size	Maximum size of the SFTP session window. This parameter is optional.

HUAWEI

Method get

- get(): This method copies a remote file (specified by **remotepath**) from the SFTP server to the destination path (specified by **localpath**) on the local host. Any exception raised by operations will be passed through.
- The following is an example of the method:

```
local_path=r'C:\Users\exampleuser\.ssh\vrptest.cfg'
remote_path= '/vrpcfg.cfg'
sftp.get(remote_path, local_path)
```

Parameter	Description
remotepath	Remote file.
localpath	Destination path on the local host. The path must contain the file name. If only a directory is specified, an error may occur.

HUAWEI

Method put

- put(): This method copies a local file (specified by **localpath**) from the local host to the destination path (specified by **remotepath**) on the SFTP server. Any exception raised by operations will be passed through.
- The following is an example of the method:

```
local_path=r'C:\Users\exampleuser\.ssh\vrptest.cfg'
remote_path= '/vrpcfg.cfg'
sftp.put(localpath, remotepath)
```

Parameter	Description
localpath	Local file.
remotepath	Destination path on the SFTP server. The path must contain the file name. If only a directory is specified, an error may occur.

HUAWEI

Contents

1. Introduction to SSH
2. **Paramiko Component Architecture**
   - Paramiko Component Architecture
   - Transport Class and Its Methods
   - Key Handling Class and Its Methods
   - SFTPClient Class and Its Methods
   - SSHClient Class and Its Methods
3. SSH Practices

HUAWEI

SSHClient Class and Its Methods

- The SSHClient class is an advanced representation of a session with an SSH server. This class contains the Transport, Channel, and SFTPClient classes for session channel establishment and authentication. The following is a typical example:

```
client=paramiko.client.SSHClient()
client.connect(hostname=192.168.56.100',port=22,username='client',password='123456')
stdin,stdout,stderr=client.exec_command('ls  -l')
```

Common Method	Function
connect()	Connects to the remote server and implements authentication.
set_missing_host_key_policy()	Specifies a policy to be used when the connected server does not have a known host key.
load_system_host_key()	Loads the host key from the system file.
exec_command()	Runs Linux commands on the remote server.
invoke_shell()	Starts an interactive shell session on the remote server.
open_sftp()	Creates an SFTP channel in a session connection.
close()	Closes a connection.

HUAWEI

Method connect

- connect(): This method is used to connect to a remote server and implement authentication.
- The following is an example of the method:

```
client.connect(hostname='192.168.56.100',port=22,username='client',key_filename='id_rsa')
client.connect(hostname='192.168.56.100',port=22,username='client',password='123456')
```

Parameter	Description
hostname	Target host to be connected. Only this parameter is mandatory.
port	Specified port. The default value is 22.
username	User name for authentication. This parameter is left empty by default.
password	Password of the user to be authenticated. This parameter is left empty by default.
key_filename	Private key file name or list. This parameter is left empty by default.
pkey	Private key used for identity authentication.
...	...

HUAWEI

Method set_missing_host_key_policy

- set_missing_host_key_policy(): This method specifies a policy to be used when the connected server does not have a known host key.
- The following is an example of the method:

```
client.set_missing_host_key_policy(paramiko.client.AutoAddPolicy())
```

Parameter	Description
AutoAddPolicy	Automatically adds the host name and host key to the local HostKeys object, without depending on the configurations of the load_system_host_keys method. That is, when a new SSH connection is set up, you do not need to enter **yes** or **no** for confirmation.
WarningPolicy	Logs a Python-style warning for an unknown host key and accepts it. This method provides functions similar to AutoAddPolicy. The difference lies in that this method will display a message, indicating that the connection is a new connection.
RejectPolicy	Automatically rejects the unknown host name and key. This method depends on the configuration of the load_system_host_keys method. This is the default option.

HUAWEI

- OpenSSH records the public key of each computer that a user has accessed in ~/.ssh/known_hosts. When the same computer is accessed next time, OpenSSH checks the public key. If the public keys are different, OpenSSH generates a warning to prevent man-in-the-middle attacks. Generally, when a client connects to the SSH server for the first time, you need to enter Yes or No for confirmation.

Method load_system_host_keys

- load_system_host_keys(): This method loads the host key from the system file. If no parameter is specified, the system attempts to read the key from the known hosts file on the local host.
- The following is an example of the method:

```
client.load_system_host_keys()
```

Parameter	Description
filename	File name. This parameter is left empty by default.

HUAWEI

Method exec_command

- exec_command(): This method is used to run Linux commands on a remote server.
- The following is an example of the method:

```
stdin,stdout,stderr=client.exec_command('ls  –l')
```


HUAWEI

Method invoke_shell

- invoke_shell(): This method starts an interactive shell session based on the SSH session connection.
- The following is an example of the method:

```
cli = client.invoke_shell()
```


HUAWEI

Method open_sftp

- open_sftp(): This method creates and opens an SFTP session on the SSH server.
- The following is an example of the method:

```
sftp=client.open_sftp()
```


HUAWEI

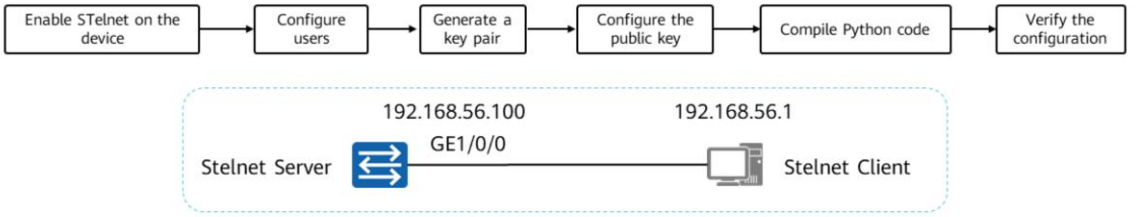
Contents

1. Introduction to SSH
2. Paramiko Component Architecture
3. **SSH Practices**
   - Practices in SSH Python Scripts
   - Practices in SFTP Python Scripts

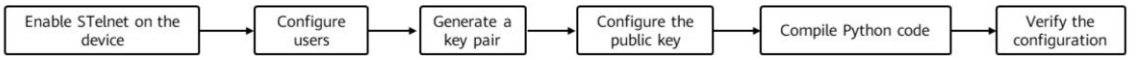
HUAWEI

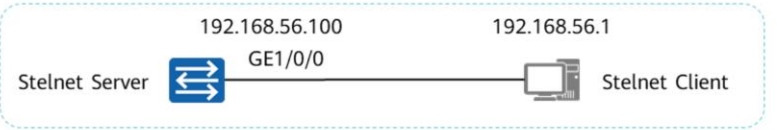
Case: Using SSH to Log In to a Device

- Description:
  - As shown in the figure below, after the STelnet server function is enabled on the switch that functions as the SSH server, the PC functioning as the SSH client can log in to the SSH server in password or RSA authentication mode.
  - This case uses RSA user authentication as an example to describe how to configure a client so that it logs in to a server through SSH using the Paramiko module of Python.



Configuration Roadmap

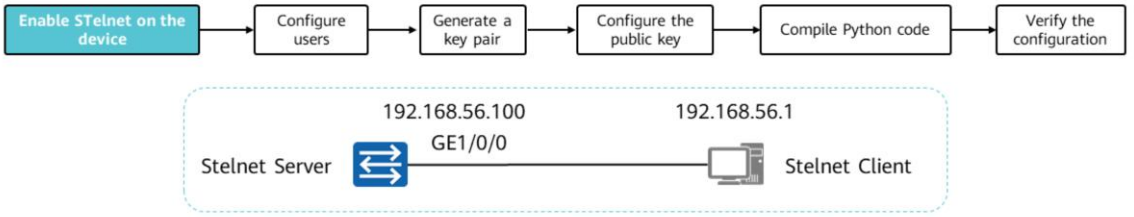




- Configuration on the server:
  - Configure STelnet. Specially, configure a management IP address, enable the STelnet function, and configure the user interface.
  - Configure users. Specially, create a local user and an SSH user, and configure the service type and authentication mode for the users.
  - Configure a public key. Specially, add the public key generated by the client and allocate it to the user.
- Configuration on the client:
  - Generate a key pair. Specifically, generate a public key and a private key locally.
  - Compile Python code.
  - Verify the configuration.

HUAWEI

Case: Configuring STelnet on the Server



1. Configure an IP address for the management network port on the server.

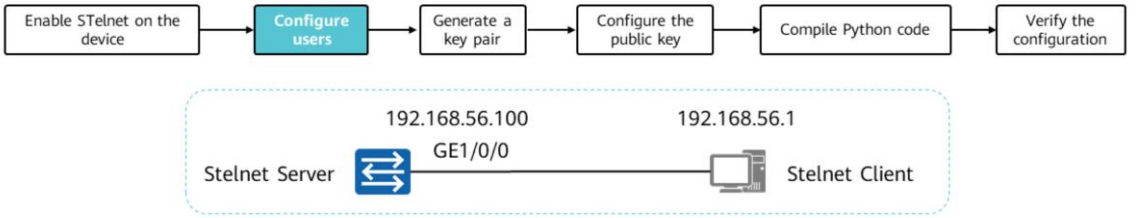
```
<HUAWEI>system-view immediately
[HUAWEI] sysname SSH Server
[SSH Server] interface GE 1/0/0
[SSH Server-GE1/0/0] ip add 192.168.56.100 24
[SSH Server-GE1/0/0] quit
```

2. Enable STelnet on the server and configure the VTY user interface.

```
[SSH Server] stelnet server enable
[SSH Server] user-interface vty 0 4
[SSH Server-ui-vty0-4] authentication-mode aaa
[SSH Server-ui-vty0-4] protocol inbound ssh
[SSH Server-ui-vty0-4] user privilege level 3
[SSH Server-ui-vty0-4] quit
```


HUAWEI

Case: Configuring Users on the Server



3. Create a local user on the server, add the user to the administrator group, and configure the service type for the user.

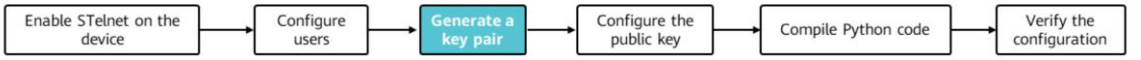
```
[SSH Server] aaa
[SSH Server-aaa] local-user client password irreversible-cipher Huawei@123
[SSH Server-aaa] local-user client user-group manage-ug
[SSH Server-aaa] local-user client service-type ssh
[SSH Server-aaa] quit
```

4. Create an SSH user on the server and configure the authentication mode and service type for the user.

```
[SSH Server] ssh user client
[SSH Server] ssh user client authentication-type rsa
[SSH Server] ssh user client service-type stelnet
```


HUAWEI

Case: Creating an RSA Key Pair on the Client



5. On the client, use Git Bash to create an RSA key pair (private key id_rsa and public key id_rsa.pub) and check the public key.

```
exampleuser@exampleuser  MINGW64 ~
$ ssh-keygen  -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/c/Users/exampleuser/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /c/Users/exampleuser/.ssh/id_rsa
Your public key has been saved in /c/Users/exampleuser/.ssh/id_rsa.pub
```

Generate an RSA public/private key pair --

Set the path for storing the key. (Press **Enter** to use the default path.) --

Enter the pass phrase. (Press **Enter** to use the default pass phrase.) --

Enter the pass phrase again. (Press **Enter** to use the default pass phrase.) --

Path for storing the private key file of the client. --

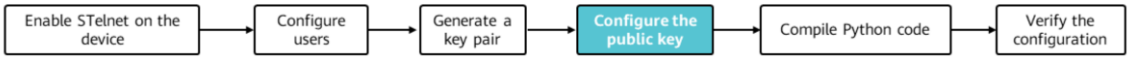
Path for storing the public key file of the client. --

Check the public key. --

```
$ cat /c/Users/exampleuser/.ssh/id_rsa.pub
```


HUAWEI

Case: Configuring a Public Key on the Server



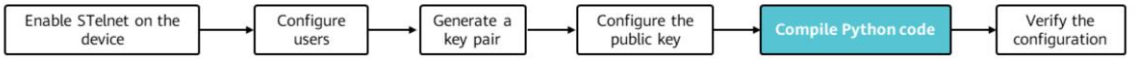
6. On the server, add the public key generated by the client and allocate it to the user.

```
[SSH Server] rsa peer-public-key rsa01 encoding-type openssh
[SSH Server-rsa-public-key] public-key-code begin
[SSH Server-rsa-public-key-rsa-key-code] ssh-rsa
AAAAB3NzaC1yc2EAAAADAQABAAABgQDwLRx8MmuNs500dRemhFHdDbBmxco8Bp+wyqwaGuHJZBCjyFQV6AB+ezu
5t0eWE3mw57IZfgmvR+MjBcliZv/x3l8oUMLcQKlKslYQDtvfUCZd+za1suXAPB/dyPKMhYPAzSDA7K+xqCWlmU3q06vxH
EPLMv4A5IX54rKtBnK92fWjl9ACU+ak0ZlHxbKwOFn1tr0GJBazcInEs9DKGwkTTqJdu9+5hI5NxXTSbM3an53805ZbCU18
xPy57g7MZC89vbdsag/uvQmFkLJ3arts/Om2R7fhR92EU/SNPmVy+qDEdwZEVdubdqJInW+8zzVkPGlnb2oH5hwH78Kskl
bxb0fEfmGR0mS1ZAi3ZHUGcEEjuFZona3+5Z0Un2OPxfXwvoljVDusbYcugJHo9Ssurz05GzVuamQZlcO2JYY6FhtLUAImt
XGQ80MpTjB0lcprkAZCib8agYOtVQNTZ7iB0g2EcBN9UTyMz7sh8RtrBDj445r+XPaDE8LmpDRKHMk=
[SSH Server-rsa-public-key-rsa-key-code] public-key-code end
[SSH Server-key-code] peer-public-key end
[SSH Server] ssh user client assign rsa-key rsa01
```


HUAWEI

- For details about the commands, refer to the product documentation at https://support.huawei.com/enterprise/en/doc/EDOC1000097293/466984de?idPath=24030814|21432787|21430822|22318704|9794900.

Case: Compiling Python Code on the Client

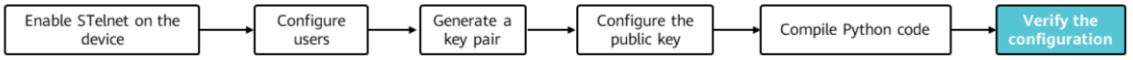


7. Compile and run Python code on the client to log in to the server through SSH.

```
Import module                                          --  import paramiko
                                                           import time
Instantiate SSH objects.                               --  ssh = paramiko.SSHClient()
Allow connections to unknown hosts.                    --  ssh.set_missing_host_key_policy(paramiko.AutoAddPolicy())
Set up an SSH session connection.                      --  ssh.connect(hostname='192.168.56.100',port=22,username='client',key_filename=r'C:
                                                           \Users\exampleuser\.ssh\id_rsa')
Open an interactive session.                           --  cli = ssh.invoke_shell()
Send the command for canceling screen splitting.       --  cli.send('screen-length  0 temporary\n')
Send the command for displaying the current configuration. --  cli.send('display cu\n')
Set the pause duration to 3 seconds.                   --  time.sleep(3)
Instantiate the received data.                         --  dis_cu = cli.recv(999999).decode()
Print the command output.                              --  print(dis_cu)
Close the SSH connection.                              --  ssh.close()
```


HUAWEI

Case: Verifying the Configuration on the Client



8. Run the code. The current configuration of the SSH server is displayed.

```
Info: The max number of VTY users is 5, the number of current VTY users online is 1,
and total number of terminal users online is 2.
<SSH Server>screen-length 0 temporary
Info: The configuration takes effect on the current user terminal interface only.
<SSH Server>display cu
!Software Version V200R005C10SPC607B607
#
......
```


HUAWEI

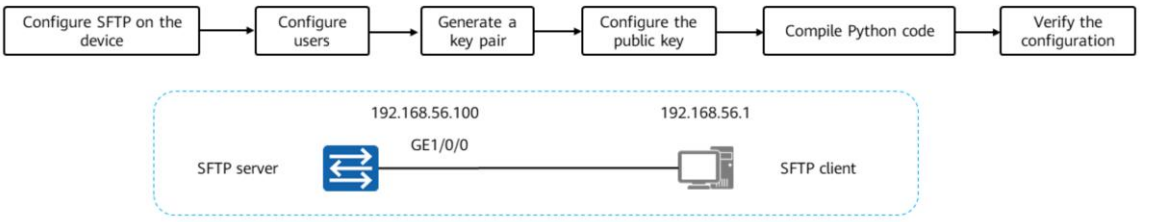
Contents

1. Introduction to SSH
2. Paramiko Component Architecture
3. **SSH Practices**
   - Practices in SSH Python Scripts
   - Practices in SFTP Python Scripts

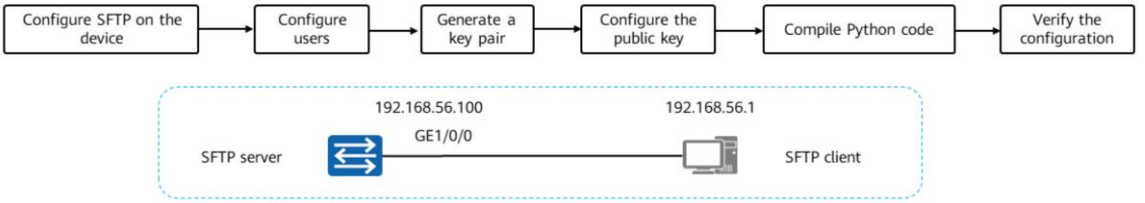
HUAWEI

Case: Using SFTP to Upload and Download Files

- Description:
  - SSH File Transfer Protocol (SFTP) is a secure file transfer protocol based on SSH. SFTP not only provides all functions of FTP, but also has higher security and reliability.
  - As shown in the figure below, after the SFTP server function is enabled on the switch that functions as the SFTP server, the PC functioning as a client can log in to the SFTP server in password or RSA authentication mode to upload or download files.
  - This case uses RSA user authentication as an example to describe how to upload and download files on the client through SFTP using the Paramiko module of Python.



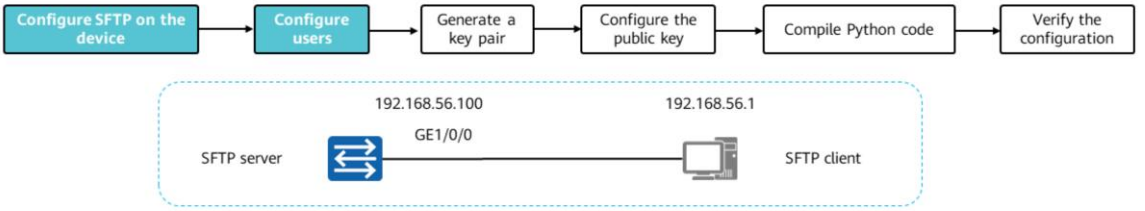
Configuration Roadmap



- Configuration on the server:
  - Configure SFTP. Specifically, configure the management IP address and enable SFTP on the device.
  - Create a user. Specifically, create an SSH user and configure the service type, authentication mode, and SFTP path.
  - Configure a public key. Specially, add the public key generated by the client and allocate it to the user.
  - Verify the configuration. Specifically, check the uploaded files.
- Configuration on the client:
  - Generate a key pair. Specifically, generate a public key and a private key locally.
  - Compile Python code.
  - Verify the configuration. Specifically, check the downloaded files.

HUAWEI

Case: Configuring SFTP and Users on the Server



1. Configure the management IP address for the SFTP server and enable the SFTP server function.

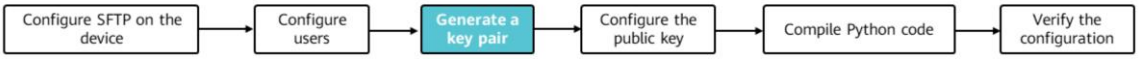
```
<HUAWEI>system-view immediately
[HUAWEI] sysname SFTP Server
[SFTP Server] interface GE 1/0/0
[SFTP Server-GE1/0/0] ip add 192.168.56.100 24
[SFTP Server-GE1/0/0] quit
[SFTP Server] sftp server enable
```

2. Create the SSH user **client** and configure the authentication type and service type for the user.

```
[SFTP Server] ssh user client
[SFTP Server] ssh user client authentication-type rsa
[SFTP Server] ssh user client service-type sftp
[SFTP Server] ssh user client sftp-directory cfcard:
[SFTP Server] ssh authorization-type default root
```


HUAWEI

Case: Creating an RSA Key Pair on the Client

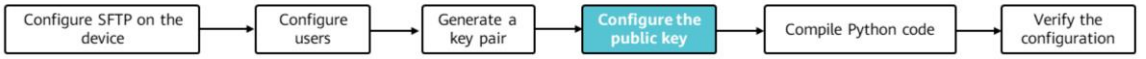


3. On the client, use Git Bash to create an RSA key pair (private key **id_rsa** and public key **id_rsa.pub**) and check the public key.

Generate an RSA public/private key pair.	--	exampleuser@exampleuser MINGW64 ~ $ ssh-keygen -t rsa Generating public/private rsa key pair.
Set the path for storing the key. (Press **Enter** to use the default path.)	--	Enter file in which to save the key (/c/Users/exampleuser/.ssh/id_rsa):
Enter the pass phrase. (Press **Enter** to use the default pass phrase.)	--	Enter passphrase (empty for no passphrase):
Enter the pass phrase again. (Press **Enter** to use the default pass phrase.)	--	Enter same passphrase again:
Path for storing the private key file of the client.	--	Your identification has been saved in /c/Users/exampleuser/.ssh/id_rsa
Path for storing the public key file of the client.	--	Your public key has been saved in /c/Users/exampleuser/.ssh/id_rsa.pub
Check the public key.	--	$ cat /c/Users/exampleuser/.ssh/id_rsa.pub

HUAWEI

Case: Configuring a Public Key on the Server



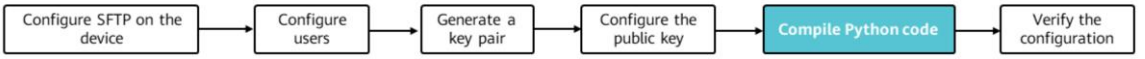
4. On the server, add the public key generated by the client and allocate it to the user.

```
[SFTP Server] rsa peer-public-key rsa01 encoding-type openssh
[SFTP Server-rsa-public-key] public-key-code begin
[SFTP Server-rsa-public-key-rsa-key-code] ssh-rsa
AAAAB3NzaC1yc2EAAAADAQABAAABgQDwLRx8MmuNs500dRemhFHdDbBmxco8Bp+wyqwaGuHJZBCjyFQV6AB+ezu
5t0eWE3mw57IZfgmvR+MjBcliZv/x3l8oUMLcQKlKslYQDtvfUCZd+za1suXAPB/dyPKMhYPAzSDA7K+xqCWlmU3q06vxH
EPLMv4A5IX54rKtBnK92fWjl9ACU+ak0ZlHxbKwOFn1tr0GJBazcInEs9DKGwkTTqJdu9+5hI5NxXTSbM3an53805ZbCU18
xPy57g7MZC89vbdsag/uvQmFkLJ3arts/Om2R7fhR92EU/SNPmVy+qDEdwZEVdubdqJInW+8zzVkPGlnb2oH5hwH78Kskl
bxb0fEfmGR0mS1ZAi3ZHUGcEEjuFZona3+5Z0Un2OPxfXwvoljVDusbYcugJHo9Ssurz05GzVuamQZlcO2JYY6FhtLUAlmt
XGQ80MpTjB0lcprkAZCib8agYOtVQNTZ7iB0g2EcBN9UTyMz7sh8RtrBDj445r+XPaDE8LmpDRKHMk=
[SFTP Server-rsa-public-key-rsa-key-code] public-key-code end
[SFTP Server-key-code] peer-public-key end
[SFTP Server] ssh user client assign rsa-key rsa01
```


HUAWEI

- For details about the commands, refer to the product documentation at https://support.huawei.com/enterprise/en/doc/EDOC1000097293/466984de?idPath=24030814|21432787|21430822|22318704|9794900.

Case: Compiling Python Code on the Client

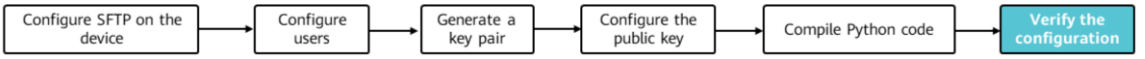


5. Compile and run Python code on the client to log in to the server through SFTP and upload and download files.

```
Import module. --  import paramiko
Create an RSA key object. --  key=paramiko.RSAKey.from_private_key_file(r'C:\Users\exampleuser\.ssh\id_rsa')
Instantiate the session channel. --  tran = paramiko.Transport(('192.168.56.100', 22))
Set up an SSH session connection. --  tran.connect(username='client', pkey=key)
Set up an SFTP channel. --  sftp = paramiko.SFTPClient.from_transport(tran)
Set the local path. --  local_path=r'C:\Users\exampleuser\.ssh\vrptest.cfg'
Set the remote path. --  remote_path='/vrpcfg.cfg'
Perform the download operation. --  sftp.get(remote_path, local_path)
Perform the upload operation. --  sftp.put(local_path,'/test.cfg')
Close the session. --  tran.close()
```


HUAWEI

Case: Verifying the Configuration on the Server and Client



6. Run the code. The client successfully downloads the specified file to the local host.

7. Run the dir command on the server. The specified file is successfully uploaded to the server.

```
<SFTP Server>dir
Directory of cfcard:/

 Idx  Attr     Size(Byte)  FileName
   5  -rw-         2,428  test.cfg
   6  -rw-         2,493  vrpcfg.cfg
```


HUAWEI

Quiz

1. (Multiple-answer question) Which of the following phases are involved in the SSH protocol? ( )
   A. Version negotiation phase
   B. Algorithm negotiation phase
   C. Key exchange phase
   D. User authentication phase
   E. Session interaction phase

HUAWEI

1. ABCDE

Summary

- This course describes the concepts of Paramiko and SSH, and illustrates the working principles of SSH.
- This course also describes the components and common methods of Paramiko. In the last part, this course uses example scripts of Python SSH and SFTP to show the use and practices of Paramiko methods, thereby implementing preliminary network automation based on SSH.
- For more information, visit Paramiko's official website, read SSH RFC documents, and learn upper-layer SSH libraries such as Fabric. Fabric is developed based on Paramiko and is further encapsulated to improve SSH-based application deployment and system management efficiency.

HUAWEI

More Information

- Paramiko official websites
  - docs.paramiko.org/en/latest/index.html
  - www.paramiko.com
- SSH RFC documents
  - https://tools.ietf.org/html/rfc4251.html
  - https://tools.ietf.org/html/rfc4252.html
  - https://tools.ietf.org/html/rfc4253.html
  - https://tools.ietf.org/html/rfc4254.html
  - Fabric official website
  - https://fabric-chs.readthedocs.io/zh_CN/chs/tutorial.html

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

NETCONF YANG Principles and Practices

HUAWEI

Foreword

- For device configuration management, engineers are more accustomed to using the CLI to interact with devices. This mode is simple, direct, and easy to understand. However, in the network automation field, the CLI mode has problems such as low efficiency and difficult command understanding.
- To simplify device configuration and management, Internet Engineering Task Force (IETF) has set up work groups and released NETCONF, YANG, and RESTCONF standards.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the problems and challenges of traditional network configuration management.
  - Describe the basic principles of NETCONF, YANG, and RESTCONF.
  - Describe the differences and relationships between NETCONF, YANG, and RESTCONF.
  - Understand the working principles of NETCONF and RESTCONF.

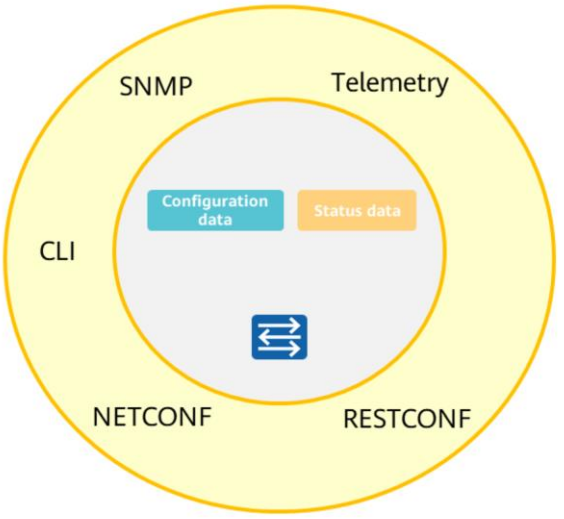
HUAWEI

Contents

1. **Network Management Technology Background**
2. NETCONF Protocol
3. Modeling Language of YANG
4. RESTCONF Protocol

HUAWEI

Introduction to Device Data



- Data is the carrier of information. Information that can be obtained from a running device is classified into configuration data and status data.
  - Configuration data is writable. It can change the system status, for example, from the initial state to the current state.
  - Status data is read-only non-configuration data in a system, for example, status information and statistics information.

There are multiple methods to modify device data, either based on command lines or based on specific protocols. The CLI and SNMP are the most typical methods for obtaining device data.

HUAWEI

Configuring a Network Device

- You can configure and manage a device in multiple ways. You can use the console cable to directly connect to the device or use SSH to remotely log in to the device, and then use the CLI to configure the device. You can also use the NMS server to set parameters in the MIB node of the device through SNMP set.
- With the increase of network scale and complexity, the preceding two methods cannot satisfy the configuration management requirements. To resolve this problem, NETCONF based on Extensible Markup Language (XML) is introduced.

HUAWEI

- SNMP is based on UDP and is stateless, unordered, and unreliable for configuration management.
- SNMP can be configured for only one object, not for one service. During the concurrent configuration of multiple objects, if some objects are successfully configured but some objects fail to be configured, unknown impacts will be caused on the network.
- The SNMP interface is difficult to understand.

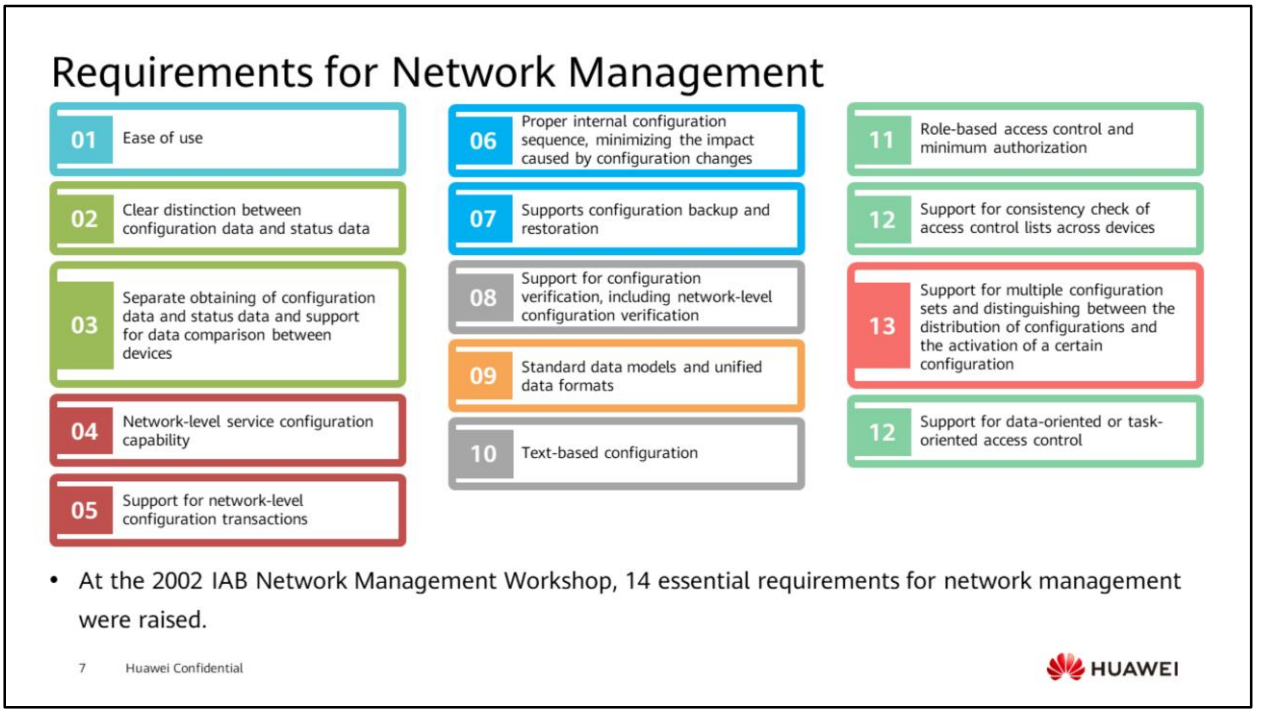
Typical Configuration Management Methods

	Advantage	Disadvantage
CLI	1. It uses a text interface, which is easy to understand. 2. Based on Telnet/SSH, it is easy to use.	1. Vendors have different definitions. Carriers need to learn and develop adaptation scripts for each vendor. 2. Configuration scripts are unstructured, unpredictable, and easy to change, which makes parsing complex. CLI scripts are difficult to maintain, and automatic parsing is difficult to implement.
SNMP	1. It uses an interface between machines. The data model file (MIB) is available. 2. It is mainly used for network monitoring.	1. Based on the UDP protocol, it is stateless, unordered, and unreliable. 2. Configuration can be performed on objects only one by one, but does not orient to a service. During the concurrent configuration of multiple objects, if some objects are successfully configured but some objects fail to be configured, unknown impacts may be caused on the network. 3. SNMP manages only a single device and does not support network-level configuration or multi-device configuration collaboration. 4. The binary interface is difficult to understand. 5. It is insecure.

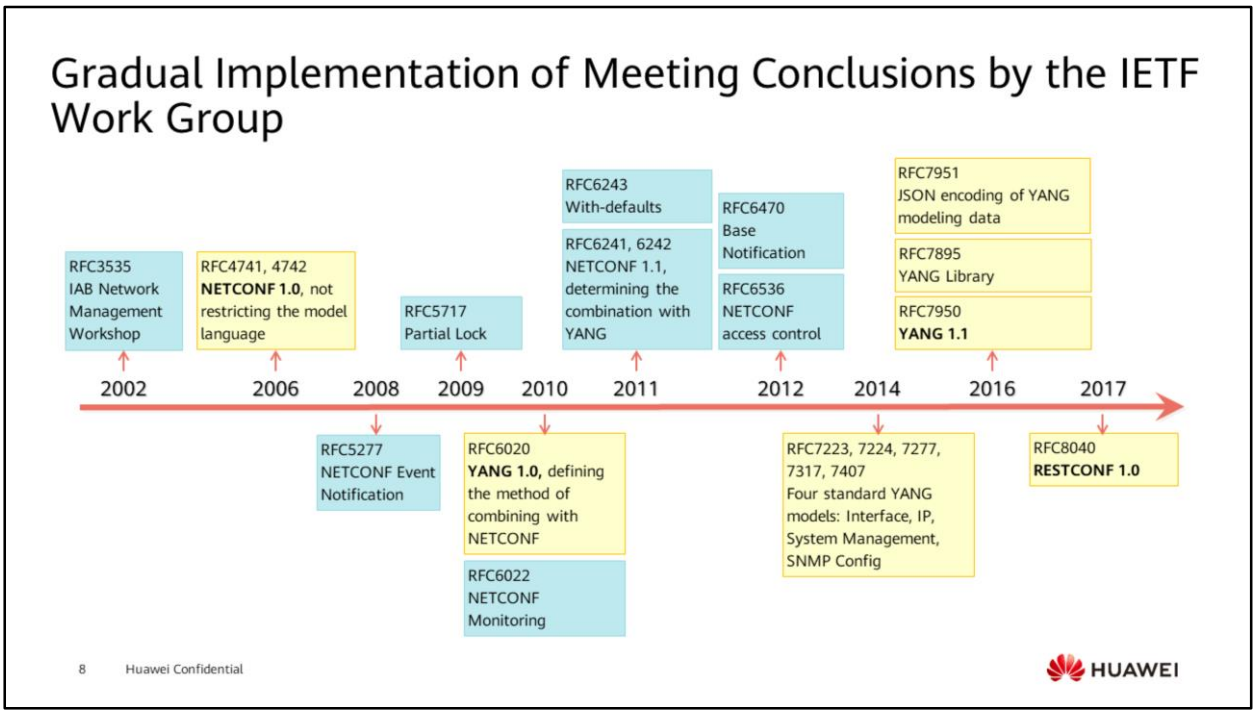
No matter how configuration automation technologies develop, the CLI is still used in essence.

SNMP is not suitable for configuration management.

HUAWEI



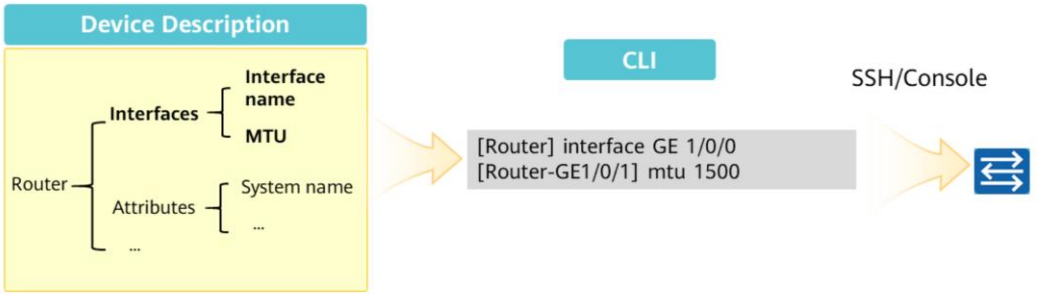
- For details, see RFC3535.
- Different IETF work groups and drafts gradually meet 14 requirements.



- The IETF gradually implements the conclusions of the IAB meeting. Different work groups gradually improve the 14 requirements.
- NETCONF 1.0 has no requirements on the model language. The combination between NETCONF 1.1 and YANG is determined.

Process for an Engineer to Configure Devices

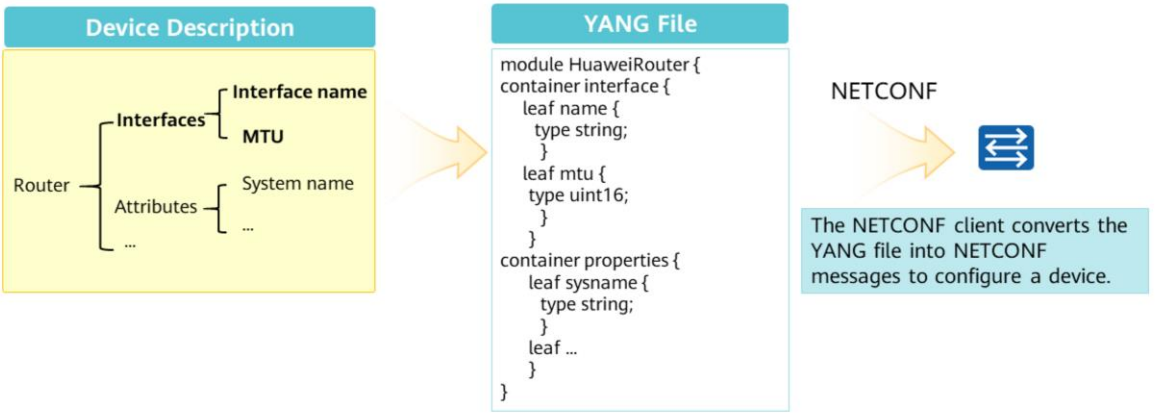
- The command line is an interactive language between humans and devices. Engineers query the product documentation and configure a device using the CLI.
- Command lines are nested to implement device configuration.



HUAWEI

Process for NETCONF to Configure Devices

- A YANG file describes device data in another way.
- The YANG model uses the module-container-leaf structure to describe devices. For example, the YANG model defines field types and specifications for router interfaces and attributes.



HUAWEI

- This example is not a real example. The YANG model does not take the entire device as one YANG file. Instead, the YANG model splits it into multiple YANG files by function.

Contents

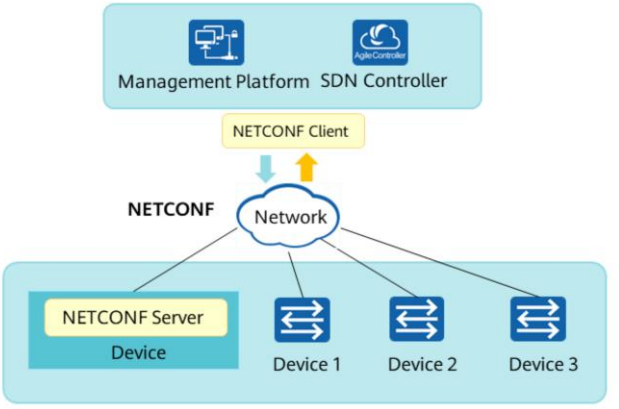
1. Network Management Technology Background
2. **NETCONF Protocol**
3. Modeling Language of YANG
4. RESTCONF Protocol

HUAWEI

NETCONF Overview

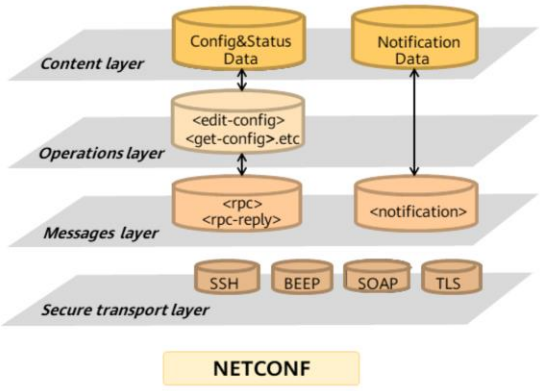
- NETCONF provides a network device management mechanism. You can use NETCONF to add, modify, or delete configurations of network devices, and obtain configurations and status of network devices.

- NETCONF has three objects:
  - NETCONF client
  - NETCONF server
  - NETCONF message



HUAWEI

NETCONF Protocol Framework



NETCONF is partitioned into four layers, as described in the following table.

- Secure transport layer: provides a communication path for interaction between the client and server. Currently, Huawei uses SSH as the transport protocol of NETCONF.
- Messages layer: provides a simple RPC request and response mechanism independent of the transport protocol layer. The NETCONF client uses an <rpc> element to encapsulate RPC request information and sends the information to the NETCONF server. Upon receipt, the NETCONF server uses an <rpc-reply> element to encapsulate RPC response information and returns the information to the NETCONF client.
- Operations layer: defines a group of basic operations as RPC invoking methods. These operations constitute basic NETCONF capabilities.
- Content layer: describes configuration data involved in network management. The configuration data depends on vendors' devices. Currently, mainstream data models include the Schema model and YANG model.

HUAWEI

XML Encoding

- XML is the encoding format of NETCONF. NETCONF uses text files to represent complex hierarchical data.

```
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to>Learners</to>
  <from>Huawei</from>
  <heading>Reminder</heading>
  <body>Don't forget Reading!</body>
</note>
```

The header of an XML encoding file is as follows:

<?xml version="1.0" encoding="UTF-8"?>

In the preceding format:

<?: indicates the start of an instruction.

xml: identifies an XML file.

version: indicates the XML version. "1.0" indicates that the XML1.0 standard version is used.

encoding: indicates the character set encoding format. Only UTF-8 encoding is supported.

?>: indicates the end of an instruction.

The XML document forms a tree structure, unfolding from the root.

The <note><to><from><heading><body> tag is private. The XML language does not have predefined tags, allowing users to customize tags and document structures.

Content is nested in the tag format. The slash (/) indicates the end of the current tag.

HUAWEI

Transport layer > Messages layer > Operations layer > Content layer

Transport Layer and Messages Layer

- NETCONF uses SSH to implement secure transmission and uses Remote Procedure Calls (RPCs) to implement communication between the client and server.

The RPC framework is independent of the transport layer, and is used to indicate NETCONF requests and responses.

- <rpc>: encapsulates NETCONF requests from the client to the server. The header defines the message-id identification sequence.
- <rpc-reply>: response message sent by the user server to the <rpc>. The values of message-id in the headers are the same.
  - <rpc-error> is sent in <rpc-reply>. One <rpc-reply> element can contain multiple <rpc-error> elements.
  - If <ok> is sent in <rpc-reply>, no error or data is returned.

HUAWEI

- For details, see RFC 6241.

Basic Operations of NETCONF

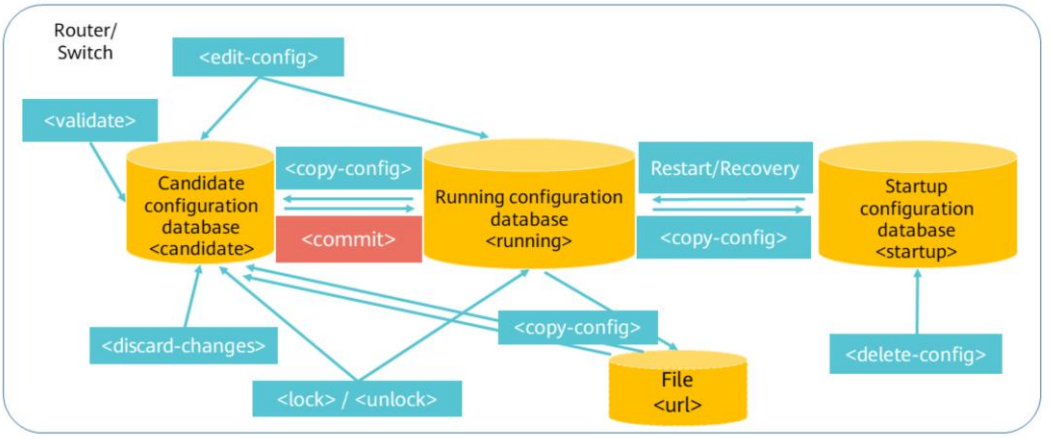
- NETCONF defines a series of operations:

Scenario Type	Operation	Function Description
Querying data	<get-config>	Queries configuration data.
	<get>	Queries the current configuration and status data of the device.
Editing data	<edit-config>	Creates, modifies, or deletes configuration data.
Backup/Recovery	<copy-config>	Exports configuration data, or replaces one set of configuration data with another set of configuration data.
	<delete-config>	Deletes the configuration data set and clears the startup configuration.
Locking/Unlocking	<lock>	Locks and exclusively occupies the permission to modify the configuration data set.
	<unlock>	Unlocks the configuration data set and cancels the exclusive permission to modify the configuration data set.
Transaction	<commit>	Commits configuration data in the <candidate> data set to the current running configuration data.
	<cancel-commit>	Cancels the trial run of configuration committing.
	<discard-changes>	Discards the uncommitted configuration data in <candidate>.
	<validate>	Checks whether the syntax and semantics of the specified configuration data are correct.
Session	<close-session>	Normally ends the NETCONF session.
	<kill-session>	Forcibly terminates other NETCONF sessions. You must have the administrator rights to perform this operation.

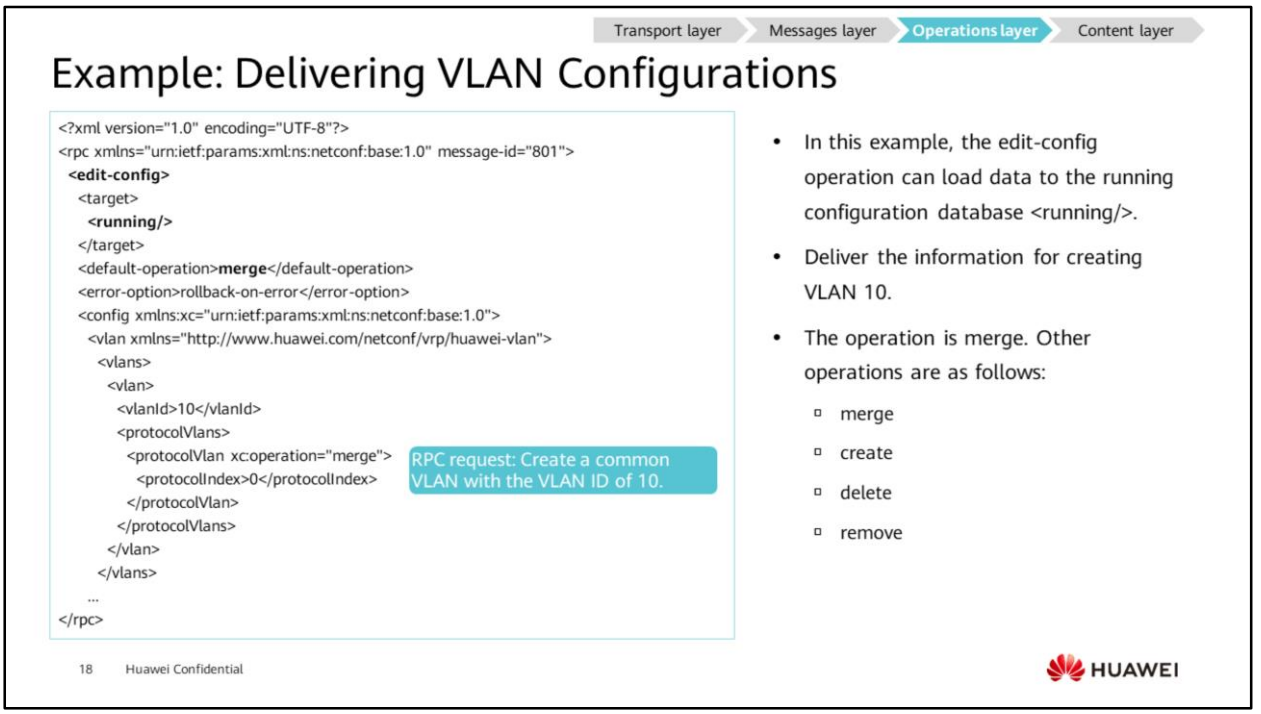
HUAWEI

Operation Objects of NETCONF

- Operation objects have three configuration databases. You can flexibly read and edit the candidate, running, and startup configuration databases to deliver, verify, and roll back the overall configuration.



HUAWEI



- <config> may contain the optional attribute <operation>, which is used to specify an operation type for a configuration item. If the <operation> attribute is not carried, the <merge> operation is performed by default. The <operation> attribute values are as follows:
  - merge: In the database, modify the existing data or create data that does not exist. This is the default operation.
  - create: Add configuration data to the configuration database only when the configuration data to be created does not exist in the configuration database. If the configuration data exists, <rpc-error> is returned, in which the <error-tag> value is data-exists.
  - delete: Delete a specified configuration data record from the configuration database. If the data record exists, the data record is deleted. If the data record does not exist, <rpc-error> is returned, in which the <error-tag> value is data-missing.
  - remove: Delete a specified configuration data record from the configuration database. If the data exists, the data is deleted. If the data does not exist, a success message is returned.

Transport layer > Messages layer > Operations layer > Content layer

Content Layer of NETCONF

- The NETCONF content layer configures data for devices. Configuration data requires a modeling language, which depends on the implementation of each vendor. NETCONF 1.0 has no requirements on the model language. NETCONF 1.1 clearly defines the combination with YANG and starts to standardize the NETCONF content format.

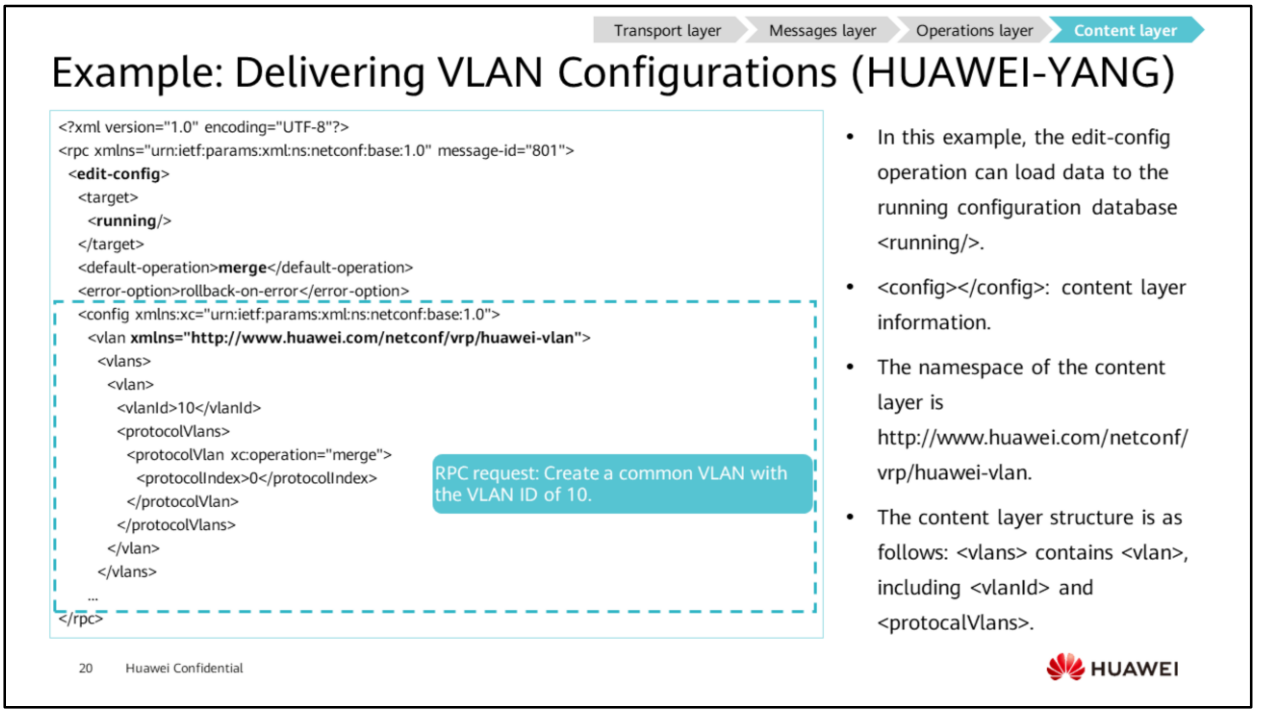
NETCONF has two modeling languages: Schema and YANG.

- Schema is a set of defined rules for describing XML documents. A schema file defines all the management objects on a managed device, the constraints and hierarchical relationships between these management objects, and the read and write permissions of these management objects.
- YANG is a data modeling language developed to design NETCONF-oriented configuration data, status data models, RPC models, and notification mechanisms.

- Currently, Huawei devices support the following content layers:
  - Huawei-YANG
  - NETCONF Schema
  - IETF-YANG
  - OpenConfig-YANG

HUAWEI

- Schema is a language that Huawei extends private syntax based on the W3C XML standard. Before the NETCONF standard is bound to the YANG model, VRPV8 has implemented Schema.
- Huawei-YANG has the most abundant content.



- For details, see the NETCONF YANG API Reference released at the official website.

Example: Delivering VLAN Configurations (Schema)

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="801">
 <edit-config>
  <target>
   <running/>
  </target>
  <default-operation>merge</default-operation>
  <error-option>rollback-on-error</error-option>
  <config>
   <vlan xmlns="http://www.huawei.com/netconf/vrp" content-version="1.0" format-version="1.0">
    <vlans>
     <vlan operation="merge">
      <vlanId>10</vlanId>
      <vlanType>common</vlanType>
     </vlan>
    </vlans>
   </vlan>
  </config>
 </edit-config>
</rpc>
```

RPC request: Create a common VLAN with the VLAN ID of 10.

- This example also describes how to create VLAN 10, but the content is in the Schema mode.
- HUAWEI-YANG is recommended, because it has more comprehensive support features.

HUAWEI

NETCONF Configuration

1. Enter the NETCONF view.

   [Huawei] **netconf**

   Run the **netconf** command in the system view to enter the NETCONF view.

2. Configure the port number for NETCONF services.

   [Huawei-netconf] **protocol inbound ssh port 830**

   Enable NETCONF services on port 830 of the SSH server.

3. Enable NETCONF.

   [Huawei] **snetconf server enable**

   Enable the NETCONF services of the SSH server on TCP port 22.

   - Either **snetconf server enable** or **protocol inbound ssh port 830** command can be used to enable NETCONF. If both commands are executed, the client can use port 22 or 830 to establish a NETCONF connection with the server.
   - NETCONF uses SSH as the transport protocol. Therefore, you must configure SSH before using NETCONF.

HUAWEI

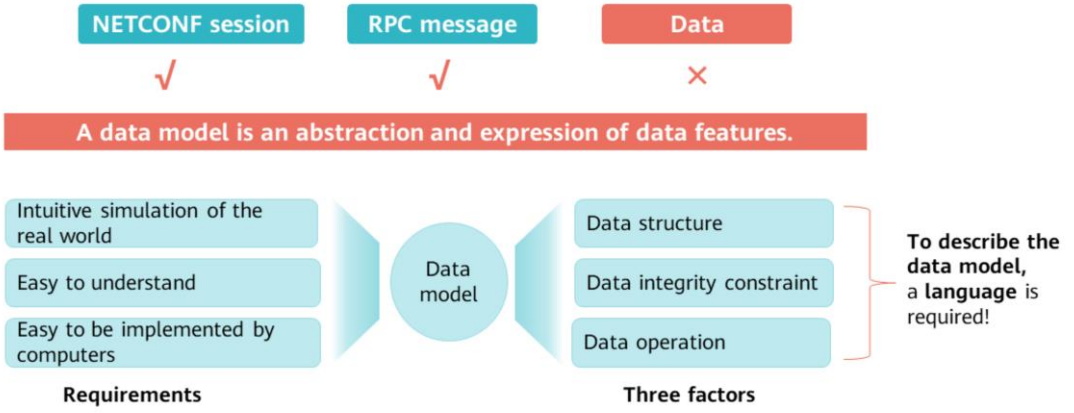
Contents

1. Network Management Technology Background
2. NETCONF Protocol
3. **Modeling Language of YANG**
4. RESTCONF Protocol

HUAWEI

Origin of YANG

- Although the NETCONF protocol is standardized, data content is not standardized. How can we describe a piece of data?



HUAWEI

YANG Language Overview

- **Yet Another Next Generation (YANG) is a data modeling language.**
- The YANG model defines the hierarchical structure of data and can be used for NETCONF-based operations. Modeling objects include configuration, status data, remote procedure calls, and notifications. This allows a complete description of all data exchanged between a NETCONF client and server. YANG has the following features:
  - Hierarchical tree-like structure modeling.
  - Data models are presented as modules and sub-modules.
  - It can be converted to the YANG Independent Notation (YIN) model based on the XML syntax without any loss.
  - Defines built-in data types and extensible types.

The YANG model is presented as a .yang file.

HUAWEI

- YANG originates from NETCONF but is not only used for NETCONF. Although the YANG modeling language is unified, YANG files are not unified.
- YANG files can be classified into three types: vendor-specific YANG files, IETF-defined YANG files, and OpenConfig YANG files.
- The Config&Status Data, Notification Data, and bottom-layer RPC messages in NETCONF can be modeled using the YANG model. YANG model files can be converted into XML/JSON files using a tool and then encapsulated into NETCONF/RESTCONF messages.
- For details, see RFC 7950.

Introduction to YANG Files – Module

```
module example-system {
    yang-version 1.1;
    namespace "urn:example:system";
    prefix "sys";

    organization "Example Inc.";
    contact "joe@example.com";
    description
      "The module for entities implementing the Example system.";

    revision 2007-06-09 {
      description "Initial revision.";
    }

    container system {
        leaf host-name {...}
        container login{...}
    }
    container X {...}
  }
```

- A YANG file can be defined as a module or submodule. Modules and submodules can reference other modules' model files to use the data types and structures defined by other modules.
- Each module contains multiple declarations, including:
  - YANG version, which is 1.0 or 1.1.
  - YANG namespace, which is IETF YANG or Openconfig YANG.
  - Organization information
  - Contact
  - File description
  - Change history

HUAWEI

- For details, see RFC 7950.

Introduction to YANG Files – Leaf Node

- A leaf node is used to define a variable of a simple specified type and is declared using the **leaf** keyword.
- A leaf node has sub-declarations and values but no sub-nodes. In this example, **host-name** contains two sub-declarations: type and description.
  - **type** indicates the value type. In this example, the value is **string**.
  - **description** indicates the description.

YANG model

```
leaf host-name {
      type string;
      description
        "Hostname for this system.";
    }
```

XML encoding

```
<host-name>my.example.com</host-name>
```


HUAWEI

- For more information, see RFC 7950.

Introduction to YANG Files – Leaf List

- Leaf List is used to define an array variable and is declared using the **leaf-list** keyword.
- In this example, **domain-search** has two sub-declarations: type and description.
  - **type** indicates the value type. In this example, the value is **string**.
  - **description** indicates the description.

YANG model

```
leaf-list domain-search {
      type string;
      description
        "List of domain names to search.";
    }
```

XML encoding

```
<domain-search>high.example.com</domain-search>
<domain-search>low.example.com</domain-search>
<domain-search>everywhere.example.com</domain-search>
```


HUAWEI

Introduction to YANG Files – List Node

- A list node is used to define a higher-level data node. Each list node is uniquely identified by a key and can contain multiple leaf nodes.

YANG model

```
list user {
      key "name";
      leaf name {
        type string;
      }
      leaf full-name {
        type string;
      }
      leaf class {
        type string;
      }
    }
```

XML encoding

```
<user>
  <name>glocks</name>
  <full-name>Goldie Locks</full-name>
  <class>intruder</class>
</user>
<user>
  <name>snowey</name>
  <full-name>Snow White</full-name>
  <class>free-loader</class>
</user>
<user>
  <name>rzell</name>
  <full-name>Rapun Zell</full-name>
  <class>tower</class>
</user>
```


HUAWEI

Introduction to YANG Files – Container Node

YANG model

```
container system {
    container login {
      leaf message {
        type string;
        description
          "Message given at start of login session.";
      }
      list user {
         key "name";
         leaf name {
           type string;
         }
         leaf full-name {
           type string;
         }
         leaf class {
           type string;
         }
       }}
}
```

XML encoding

```
<system>
  <login>
    <message>Good morning</message>
  </login>
</system>
```

- Container nodes are used to define data sets in a larger scope. Each container node has only different sub-nodes, but does not have any value. These sub-nodes can be container, leaf, leaf-list, or list nodes.

HUAWEI

Introduction to YANG Files – Grouping

```
grouping target {
  leaf address {
    type inet:ip-address;
    description "Target IP address.";
  }
  leaf port {
    type inet:port-number;
     description "Target port number.";
  }
 }
 container peer {
  container destination {
    uses target;
  }
 }
```

- Grouping is used to define nodes that can be reused. Generally, grouping is used together with **uses**. In this example, **target** defines **leaf address** and **port**. If **use target** is specified in the container peer, the leaf model is reused.

HUAWEI

Introduction to YANG Files – Configuration Data and Status Data

```
list interface {
    key "name";
    config true;
    leaf name {
      type string;
    }
    leaf speed {
      type enumeration {
        enum 10m;
        enum 100m;
        enum auto;
      }
    }
    leaf observed-speed {
      type uint32;
      config false;
    }
}
```

- YANG files can declare data types.
  - The config declaration is used to distinguish configuration data from status data.
    - **config true** indicates configuration data.
    - **config false** indicates status data.

HUAWEI

- For details, see RFC 7950.

Introduction to YANG Files – Data Type

Name	Description
binary	Any binary data
bits	A set of bits or flags
boolean	"true" or "false"
decimal64	64-bit signed decimal number
empty	A leaf that does not have any value
enumeration	Enumerated strings
identityref	A reference to an abstract identity
instance-identifier	References a data tree node
int8	8-bit signed integer
int16	16-bit signed integer
int32	32-bit signed integer
int64	64-bit signed integer
leafref	A reference to a leaf instance
string	Human-readable string
uint8	8-bit unsigned integer
uint16	16-bit unsigned integer
uint32	32-bit unsigned integer
uint64	64-bit unsigned integer
union	Choice of member types

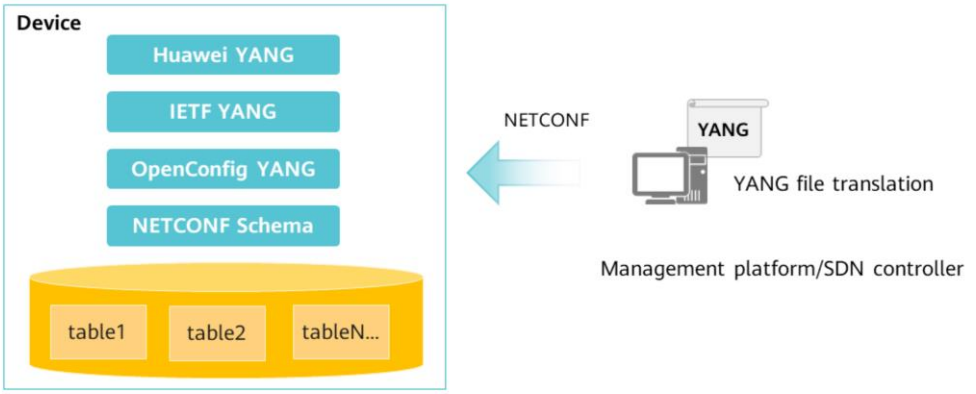
- The YANG model supports built-in default data types and extended data types. The **typedef** declaration is used to define extended data types.

```
typedef percent {
  type uint8 {
    range "0 .. 100";
  }
}

leaf completed {
  type percent;
}
```


HUAWEI

FAQ: How Do I Load a YANG Model?

- A YANG file is loaded on the NETCONF client (such as the NMS or SDN controller). A tool is used to convert YANG files into NETCONF messages and deliver the messages to devices.



HUAWEI

Contents

1. Network Management Technology Background
2. NETCONF Protocol
3. Modeling Language of YANG
4. **RESTCONF Protocol**

HUAWEI

Origin of RESTCONF

- Doctoral thesis of Roy Thomas Fielding in 2000
  - "Architectural Styles and the Design of Network-based Software Architectures“
- REST: Representational State Transfer
  - The state of a resource is transferred in a representation mode on the network.

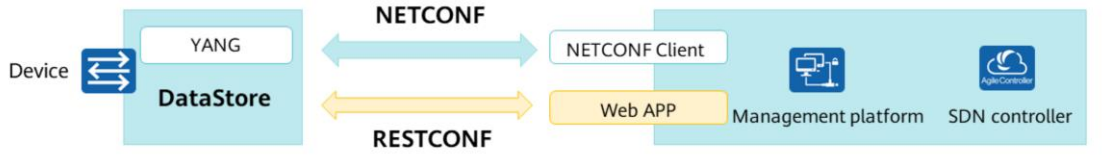
Resource: data. Uniue URI	Representation mode	State transfer: HTTP operation

REST: Client-server interaction framework

HUAWEI

RESTCONF Overview

- RESTCONF allows web applications to access configuration data, status data, and event notifications of network devices in a modular and scalable manner. It has the following features:
  - RESTCONF uses HTTP methods to perform operations (CRUD) on data defined by YANG.
  - YANG files can be shared by NETCONF and RESTCONF.
  - The data encoding format can be XML or JSON.

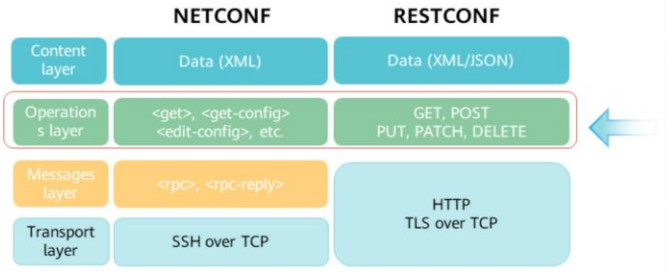


Compared with NETCONF, RESTCONF uses different operation methods and data encoding formats.

HUAWEI

- NETCONF and RESTCONF can coexist.
- CRUD: Create, Remove, Update, Delete.

Comparison Between RESTCONF and NETCONF



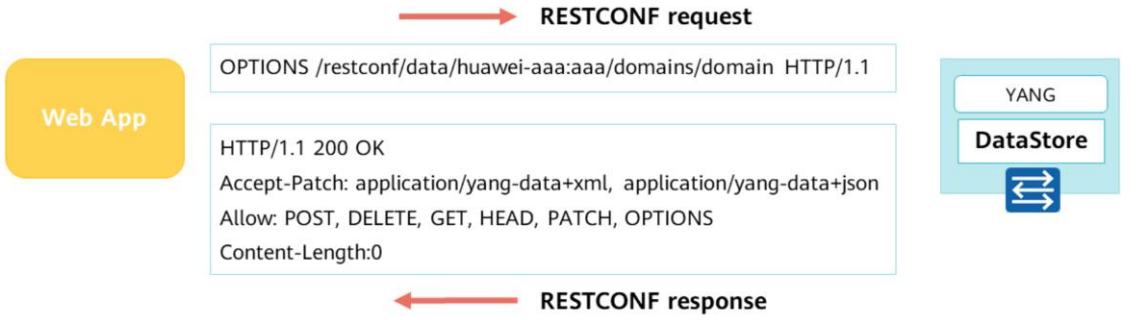
RESTCONF	NETCONF Operation
GET	<get-config>, <get>
POST	<edit-config> (nc:operation="create")
	invoke an RPC operation
PUT	<edit-config> (nc:operation="create/replace")
	<copy-config> (PUT on datastore)
PATCH	<edit-config> (nc:operation depends on PATCH content)
DELETE	<edit-config> (nc:operation="delete")

- NETCONF operates multiple configuration libraries of a device. The transaction mechanism and rollback mechanism are available.
- RESTCONF uses HTTP operations, which are stateless, have no transaction mechanism, and does not support rollback.

HUAWEI

Typical RESTCONF Interaction

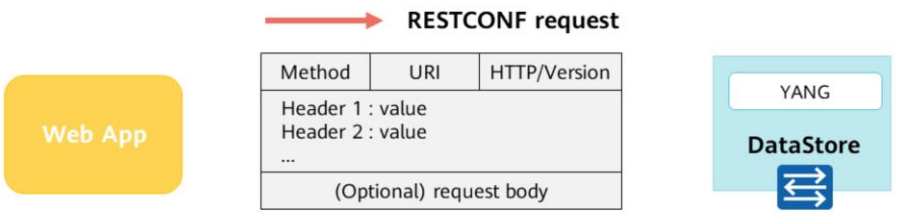
- A complete RESTCONF interaction includes a request and a response.
- In this example, the client uses the OPTIONS method to obtain the operations supported by the device.
- The device response supports the following operations: POST, DELETE, GET, HEAD, PATCH, and OPTIONS.



HUAWEI

- NETCONF and RESTCONF can coexist.

RESTCONF Request Packet



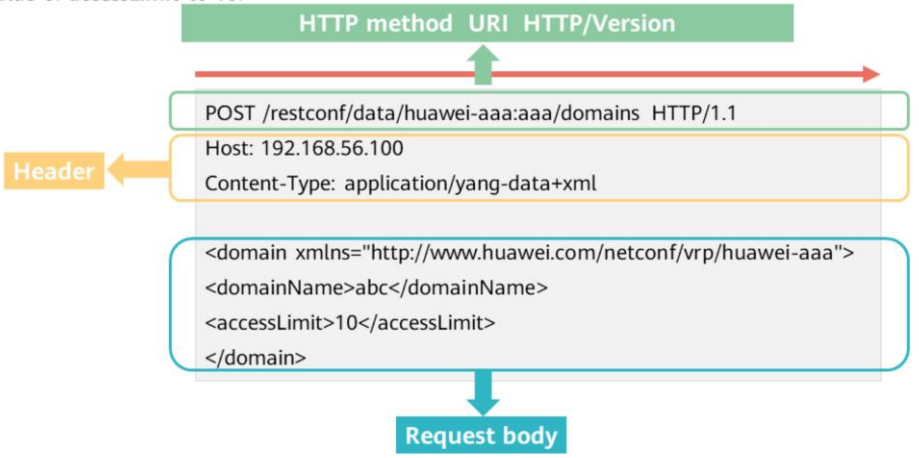
Description	
Method	HTTP method, performed on the resource identified by the request URI.
URI	Uniform Resource Identifier.
HTTP/Version	HTTP version.
Header : value	Header of a request packet, which has specific field requirements. The format is header field and value.
Request body	(Optional) Request body. Some methods do not carry body information.

HUAWEI

- A request header may contain multiple fields, such as Accept, Authorization, Host, and From. For details, see RFC 2916.

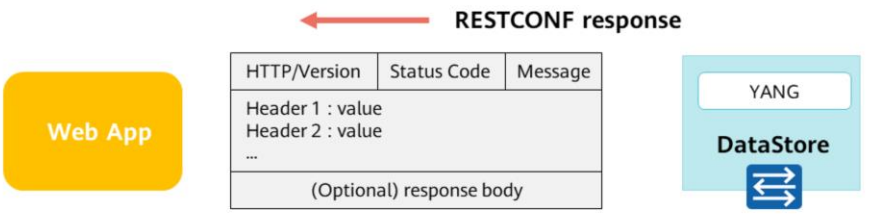
RESTCONF Request Example

- The IP address of the RESTCONF request object is 192.168.56.100. Change the value of **domainName** to **abc** and the value of **accessLimit** to **10**.



HUAWEI

RESTCONF Response Packet



Description	
HTTP/Version	HTTP method, performed on the resource identified by the request URI.
Status code	HTTP status code.
Message	HTTP status message.
Header : value	Header of a response packet, which has specific field requirements. The format is header field and value.
Response body	(Optional) Response body. Some methods do not carry body information.

HUAWEI

- Header information contains details about Response Header and Entity Header. For details, see sections 6.2 and 7.1 in RFC 2916.

HTTP Status Code

- An HTTP status code is a three-digit number indicating the status of a response from the server. It is used to return the operation result to the client.

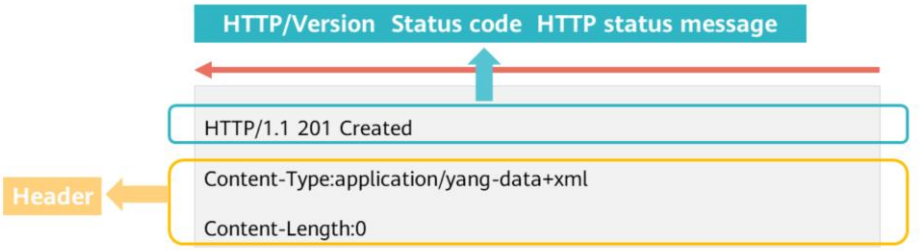
Status Code		Description	
1XX	(Informational) The request is received.	100 Continue	If the request is received, go to the next step.
2XX	(Successful) The request is successful.	200 OK	Success. The response body is available.
		201 Created	A resource is successfully created, and the URI of the newly created resource exists in the location field.
		204 No Content	Success. The response body is unavailable.
3XX	(Redirection) Further operations need to be performed.	301 Moved Permanently	A new URI is allocated to the target resource, and all future resources will be associated with the new URI.
4XX	(Client Error) Request error.	400 Bad Request	The request body is incorrect and carries error information.
		401 Unauthorized	Authorization failed. For example, the certificate does not match.
		403 Forbidden	Access denied. The possible cause is that the user attempts to perform operations beyond the permission or the login user name or password is incorrect.
		404 Not Found	The requested resource cannot be found.
5XX	(Server Error) Server error.	500 Internal Server Error	The request cannot be executed due to an internal server error. The user needs to resend the request later.
		501 Not Implemented	The function has not been implemented.

HUAWEI

- For details, see 6 Response Status Code in RFC 7231.

RESTCONF Response Example

- RESTCONF response packet
  - If status code 201 is returned, the resource is successfully created.
  - **Content-Type** and **Content-Length** in the header describe body information. The body data type is XML, and the content length is 0.



HUAWEI

Quiz

1. (Short-answer question)Which protocol is used at the transport layer of Huawei NETCONF?
2. (Short-answer question)What is the relationship between NETCONF, RESTCONF, and YANG?

HUAWEI

1. Huawei uses SSH as the transport layer protocol. Before enabling the NETCONF function on a device, you need to create an SSH user as the NETCONF user for login.
2. YANG is a modeling language used to describe the content layer of NETCONF and RESTCONF. The difference between NETCONF and RESTCONF is as follows: RESTCONF constructs the transport layer, messages layer, and operations layer based on HTTP, while NETCONF has defined the operations layer and uses SSH as the transport layer and RPC as the messages layer.

Summary

- NETCONF provides a network device management mechanism. You can use NETCONF to add, modify, or delete configurations of network devices, and obtain configurations and status of network devices.
- The YANG model defines the hierarchical structure of data and can be used for NETCONF-based operations.
- RESTCONF uses HTTP methods to perform operations (CRUD) on data defined by YANG.

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

Telemetry Fundamentals and Practices

HUAWEI

Foreword

- Telemetry is a network device monitoring technology that periodically samples statistics and status data on network devices. This course describes the technical background, concepts, framework, and technical fundamentals of telemetry, as well as the configuration procedure of static telemetry subscription.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the background of the telemetry technology.
  - Describe the telemetry concepts and framework.
  - Distinguish static and dynamic telemetry subscriptions.
  - Describe the concepts and functions of gRPC.
  - Configure telemetry data sampling.

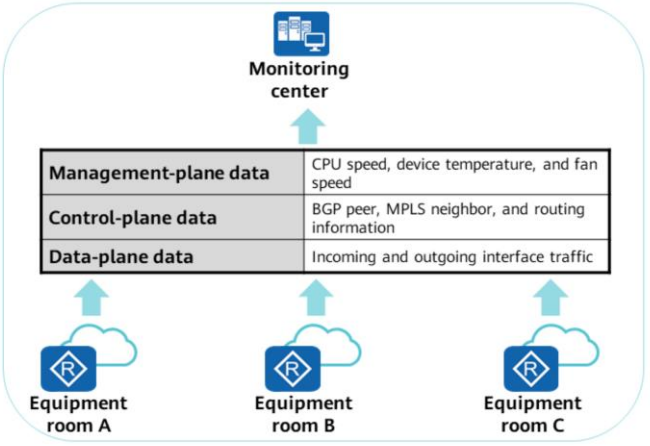
HUAWEI

Contents

1. **Introduction to Telemetry**
   - Technical Background
     - Overview
2. Telemetry Technical Fundamentals
3. Telemetry Configuration and Practice

HUAWEI

Technical Background: Introduction to Network Device Monitoring



- An O&M platform mainly provides unified monitoring and performance management for network devices. It monitors device data on the data, control, and management planes.

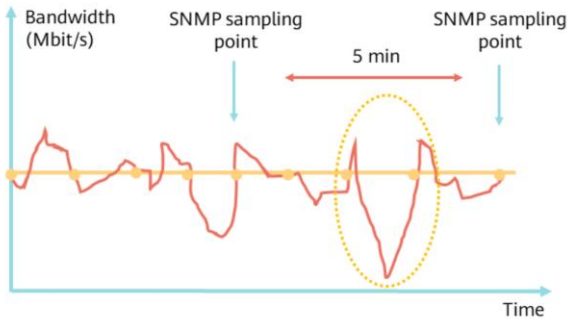
- Device monitoring data can be obtained through SNMP, CLI, Syslog, NetStream, and sFlow.
- NetSteam and sFlow deal with network traffic monitoring and are mainly applicable to data-plane data.
- SNMP is the most popular protocol for device monitoring.

HUAWEI

Technical Background: Bottlenecks of Traditional Data Sampling Technologies

- The Nyquist Sampling Theorem states that a signal must be sampled at more than twice the highest frequency component of the signal to ensure that no information will be lost.

Long sampling period



- The red wavy line represents a normal interface bandwidth change trend. Traffic exceptions cannot be detected through periodic SNMP-based data collection.

HUAWEI

Summary: Traditional Network O&M Faces Many Challenges

- On a traditional network, data is sampled in pull mode at an interval of 5 to 15 minutes on average. Data sampling in pull mode at a shorter interval will result in network breakdown.

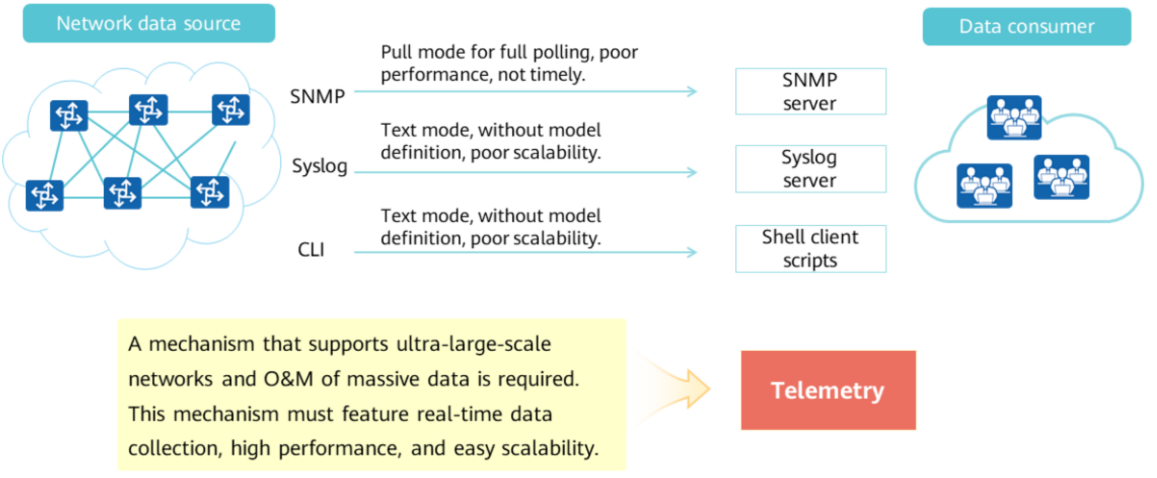
Problems in traditional network O&M

- The large sampling interval results in low network visibility.
- SNMP-based O&M systems are inefficient.
- Real-time end-to-end monitoring, like IT O&M, cannot be achieved, and complete historical data cannot be retained.
- Network problems caused by a large number of microbursts cannot be detected.

HUAWEI

- A microburst refers to a situation in which a large amount of burst data is received within a very short time (milliseconds), so that the burst data rate is tens or hundreds times higher than the average rate or even exceeds the port bandwidth. The NMS or network performance monitoring software calculates the real-time network bandwidth at an interval of seconds to minutes. At such an interval, the network traffic seems to be stable. However, packet loss may have occurred due to microbursts.

Summary: Traditional Data Collection Mechanisms Cannot Cope with Massive Data



A mechanism that supports ultra-large-scale networks and O&M of massive data is required. This mechanism must feature real-time data collection, high performance, and easy scalability.

Telemetry

HUAWEI

- SNMP queries are performed in a question-answer manner. If 1000 interactions are performed within 1 minute, SNMP parses 1000 query request packets. Telemetry avoids repeated queries. This is because subscription needs to be performed only once and then devices can continuously push data to the NMS.

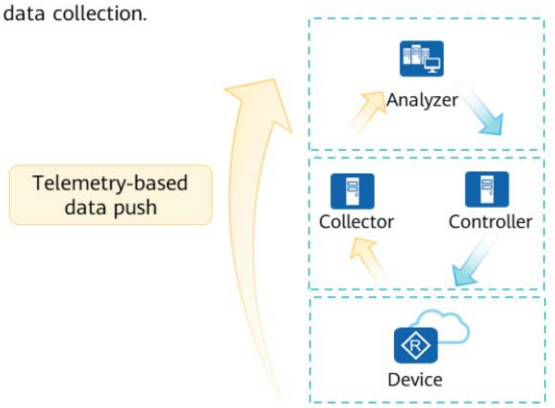
Contents

1. **Introduction to Telemetry**
   - Technical Background
   - Overview
2. Telemetry Technical Fundamentals
3. Telemetry Configuration and Practice

HUAWEI

Telemetry Overview

- Telemetry, also called network telemetry, is a technology that remotely collects data from physical or virtual devices at a high speed. Devices periodically send interface traffic statistics, CPU usage, and memory usage to collectors in push mode. Compared with the traditional pull mode (question-answer interaction), the push mode provides faster and real-time data collection.

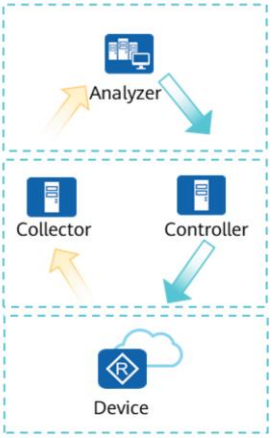


HUAWEI

- There is also a view in the industry that SNMP is considered as a traditional telemetry technology, and telemetry is currently referred to as streaming telemetry or model-driven telemetry.
- Telemetry packs the data to be sent, improving transmission efficiency.

Telemetry Technical Features

- Telemetry works in push mode. YANG models define the telemetry data structure, and data sampling can be performed at an interval of sub-seconds.

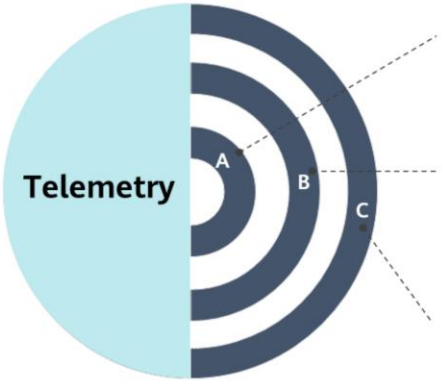


- Open APIs are available to third-party data analysis and processing companies, providing stronger data storage and processing capabilities.
- The traditional NMS is decoupled into a collector and a controller. In addition, communication protocols and management applications are decoupled.
- Improves devices' data collection capabilities.
- Supports the subscription mechanism.
- Constructs standard data models.

HUAWEI

Telemetry Advantages

- Telemetry obtains abundant monitoring data in push mode in a timely manner. The data helps quickly locate network faults. In addition, telemetry provides a uniform data stream format, which simplifies data collection and analysis to solve problems in traditional network O&M.



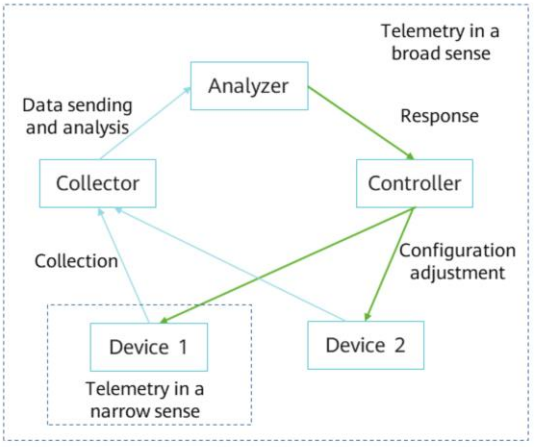
Refined monitoring
The collected data is of high precision and various types, which can fully reflect the network status.

Fast fault locating
Users want faults to be quickly located in seconds or even sub-seconds on a complex network.

Proactive data reporting
For telemetry, subscription needs to be performed only once. Then devices can continuously report data, reducing the pressure on devices to process query requests.

HUAWEI

Telemetry Network Model



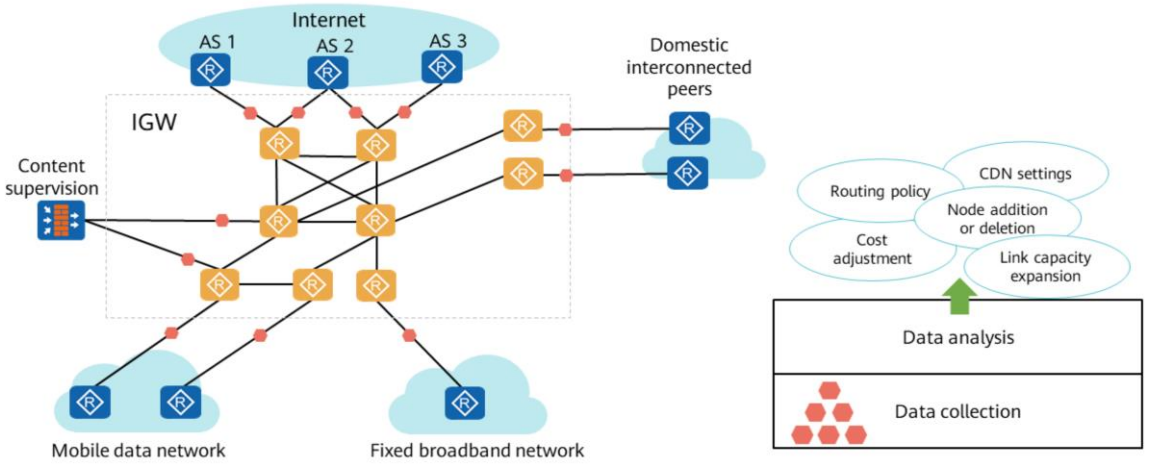
- The telemetry network model is defined in two ways:
  - In a broad sense, telemetry is a self-closed-loop system consisting of network devices, a collector, an analyzer, and a controller.
  - In a narrow sense, telemetry is the function of sending sampled device data to a collector.

Using telemetry, a collector can collect a large amount of device data and send the data to an analyzer for comprehensive analysis. The analyzer sends the results to a controller, which then adjusts device configurations accordingly, enabling the analyzer to determine whether the device status after adjustment meets expectations in real time.

HUAWEI

Telemetry Application Scenario: WAN

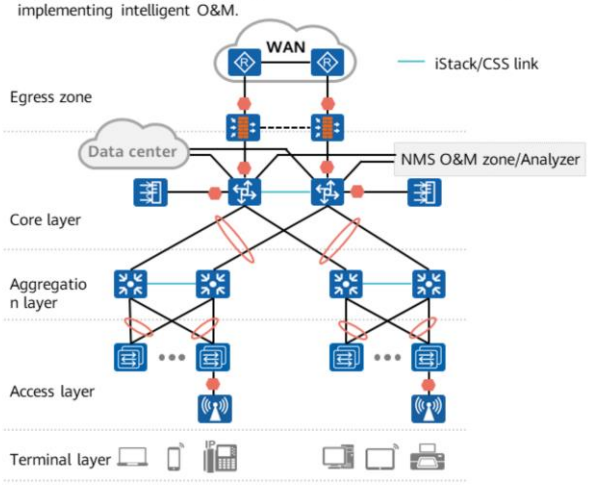
- Interface data of routers is reported in real time to help you gain insight into the traffic direction.



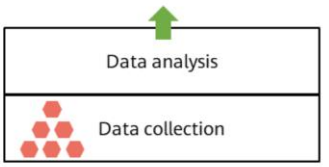
HUAWEI

Telemetry Application Scenario: Campus Network

- On a campus network, telemetry can be used to monitor and report device status and wired and wireless data to the analyzer at the same time, implementing intelligent O&M.



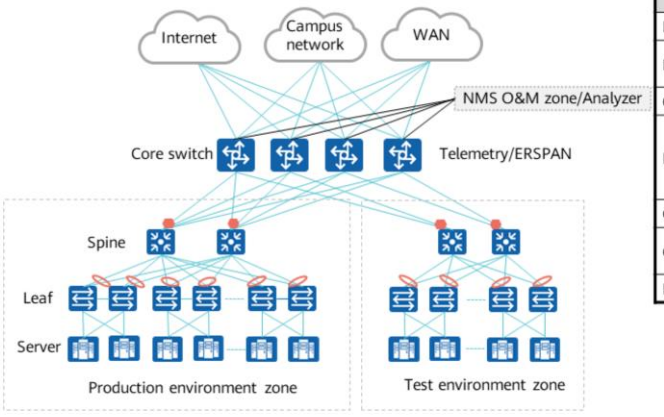
Measurement Object	Metric
AP	CPU usage, memory usage, and number of online clients
Radio	Number of online users, channel usage, noise, traffic, backpressure queue, interference rate, and power
User	RSSI, negotiated rate, packet loss rate, and latency
Device/Board	CPU and memory usages
Port	Numbers of sent and received broadcast, multicast, and unicast packets, numbers of discarded received and sent packets, and numbers of received and sent error packets



HUAWEI

Telemetry Application Scenario: Data Center

- Upon the receipt of data sent through telemetry, the analyzer calculates the forwarding path, forwarding latency, and link latency of packets. In addition, it analyzes the application interaction relationship and associates applications with network paths. The analyzer can collect statistics, analyze data, and display analysis results, implementing intelligent O&M.



Measurement Object	Metric
Device	CPU and memory usages
Board	CPU, memory, FIB entry, and MAC entry usages
Chip	TCAM usage
Port	Number of received/sent packets, number of bytes, number of lost packets, number of error packets, and numbers of broadcast, multicast, and unicast packets
Queue	Buffer size
Optical module	Receive/Transmit optical power, current, voltage, and temperature
Packet loss behavior	Detection of packet loss due to congestions

HUAWEI

- The collector in the data center collects device performance data through telemetry and collects device flow mirroring data through ERSPAN.

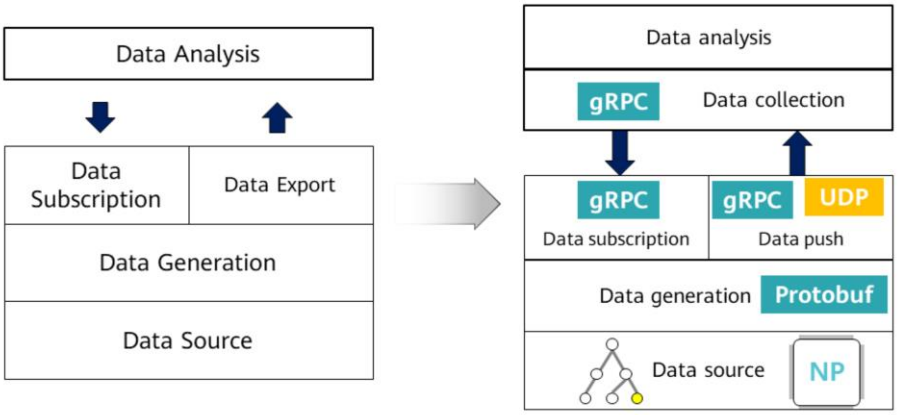
Contents

1. Introduction to Telemetry
2. **Telemetry Technical Fundamentals**
3. Telemetry Configuration and Practice

HUAWEI

Telemetry Framework

- The framework of telemetry in a narrow sense consists of four modules: data source, data generation, data subscription, and data push.

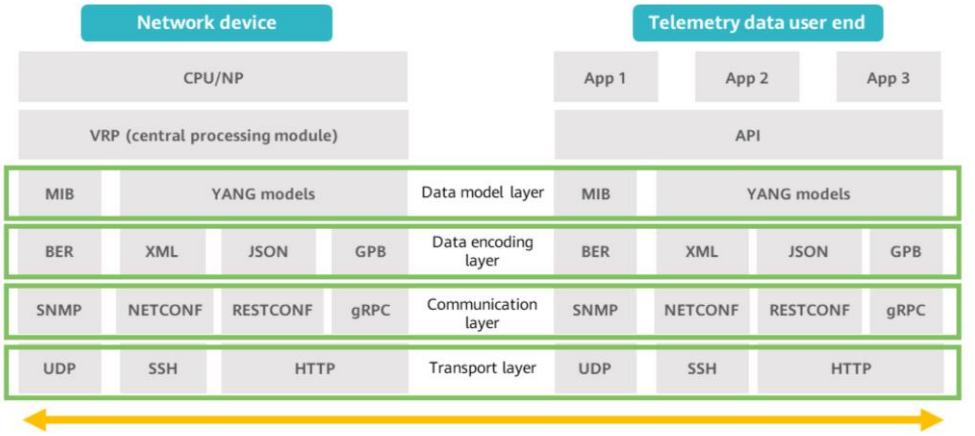


HUAWEI

- For details about the framework, see the corresponding RFC draft at https://tools.ietf.org/html/draft-song-ntf-02.
- Google Remote Procedure Call (gRPC) is an open-source remote procedure call (RPC) system developed by Google.
- User Datagram Protocol (UDP) provides a method for an application to send encapsulated IP packets without establishing a connection.
- Protocol buffers (Protobuf) is a mechanism for serializing structured data.

Telemetry Protocol Stack

- The telemetry protocol stack is divided into the transport layer, communication layer, data encoding layer, and data model layer.

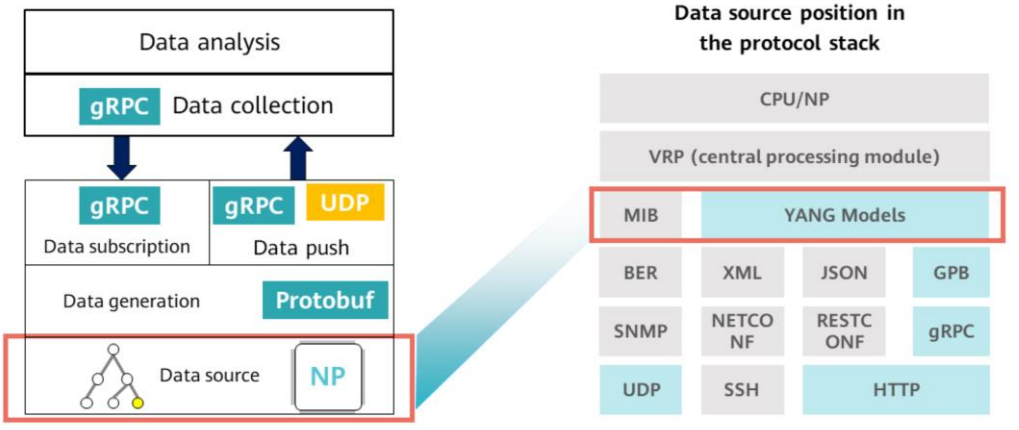


HUAWEI

- A YANG model is similar to a menu for a fast-food restaurant. If a customer wants to eat a hamburger or fried chicken, the customer writes an A4 paper purchase list with one hamburger and two fried chickens according to the menu, folds the list into a stamp-sized note, puts it in the GPB envelope, and gives the envelope to the messenger gRPC at the door. The gRPC then rides on an HTTP/2 electric motorcycle and goes to the fast food restaurant. The messenger gRPC gives the GPB envelope to the restaurant boss, and the restaurant manager opens the envelope to check whether the items that the customer orders are on the restaurant menu.

Telemetry Data Source

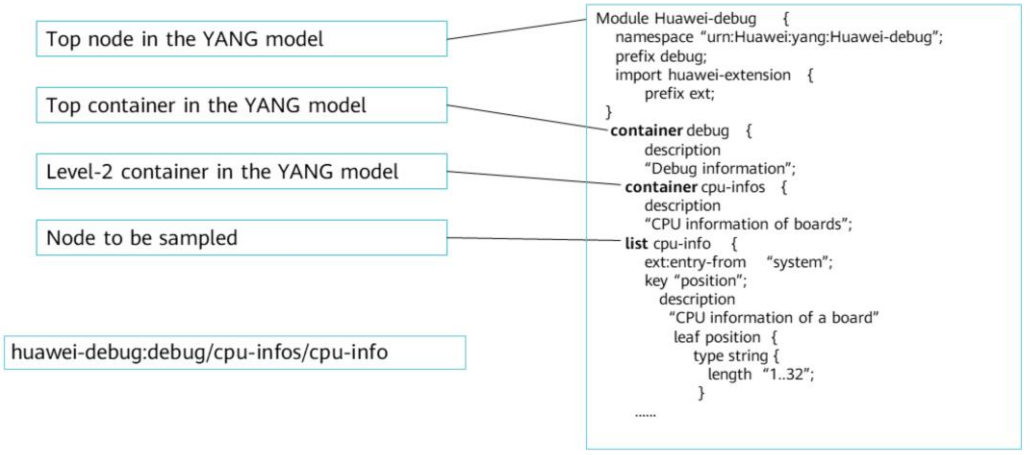
- The telemetry data source defines the data that can be obtained. Telemetry uses various types of YANG models to define data sources available to devices, such as Huawei-YANG, IETF-YANG, and OpenConfig-YANG.



HUAWEI

Data Source > Data Subscription > Data Generation > Data Push

YANG Model and Sampling Path

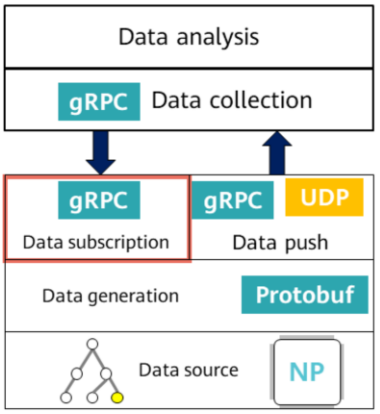


HUAWEI

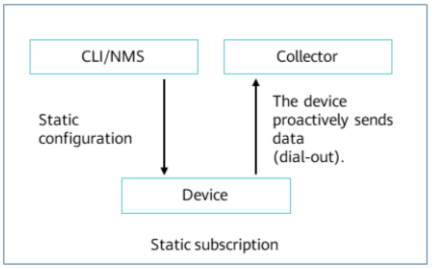
- For more information about the YANG model, visit https://datatracker.ietf.org/doc/rfc7895/.

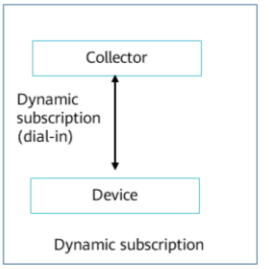
Telemetry Data Subscription

- Telemetry data subscription defines the data receive end and transmit end, as well as their interactions.

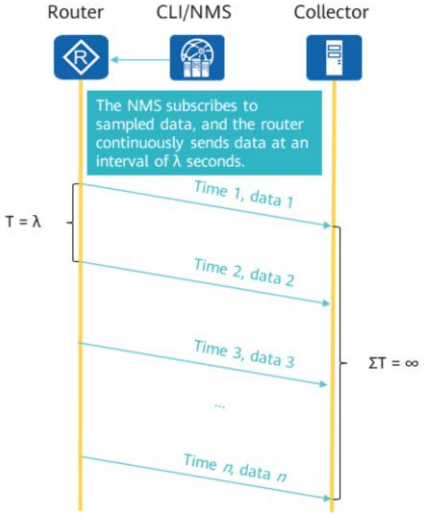


Huawei provides two subscription modes for telemetry: static and dynamic.
In static telemetry subscription, a device functions as a client and a collector functions as the server. The device proactively initiates a connection to the collector to collect and send data. This mode is applicable to long-term inspection.
In dynamic telemetry subscription, a collector functions as a client and initiates a connection to a device functioning as a server to collect data. This mode is applicable to short-term monitoring.





Static Telemetry Subscription

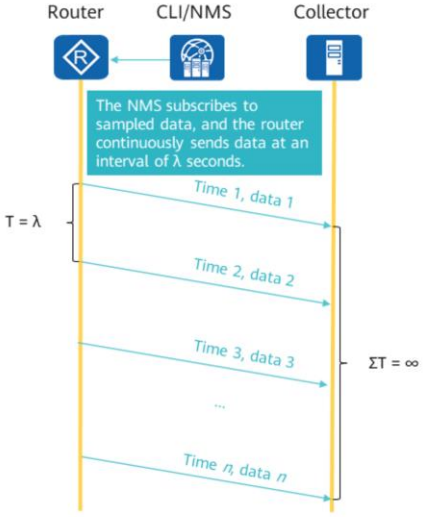


Static Telemetry Subscription

1. The NMS delivers static subscription configurations to the device, including the sampling path, sampling interval, destination to which the sampled data is reported, and data. The configurations can be delivered in the following modes:
   - Through the CLI
   - Through the NETCONF model **openconfig-telemetry.yang**
2. The device starts to sample data, establishes a connection with the collector, and pushes the sampled data to the collector in the following modes:
   - Through gRPC using the RPC method **dataPublish**
   - Through UDP

HUAWEI

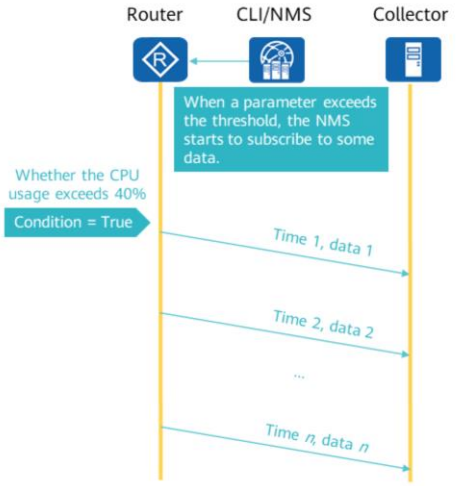
Static Telemetry Subscription



Configure the collector.	system-view Telemetry destination-group ipv4-address IP address, port number, protocol, and encryption mode for the collector
Configure the data to be sampled.	system-view telemetry sensor-group sensor-path Sensor sampling path
Create a subscription.	system-view telemetry subscription sensor-group Associate the sampling sensor group, and configure a sampling period, redundancy suppression, and a heartbeat interval.
Verify the configuration.	display telemetry sensor display telemetry destination display telemetry subscription display telemetry sensor-path

HUAWEI

Customized Event Reporting



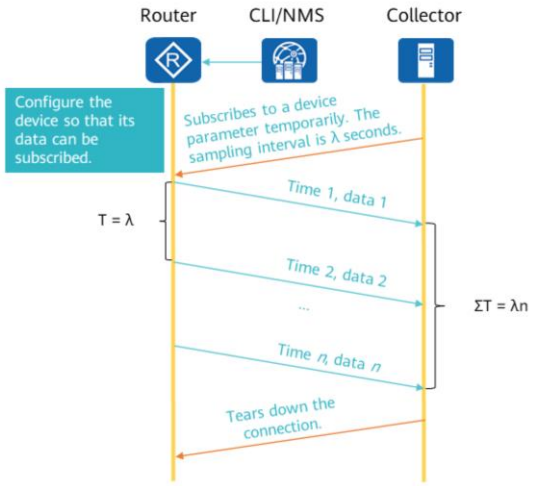
After a specific parameter exceeds the threshold, static subscription is triggered.

Configure a customized event in the static subscription.	system-view telemetry sensor-group Sensor-path //Configures a customized event. Example: sensor-path huawei-devm:devm/cpuInfos/cpuInfo condition express op-field systemCpuUsage op-type gt op-value 40

HUAWEI

Data Source > Data Subscription > Data Generation > Data Push

Dynamic Telemetry Subscription Using gRPC (1)



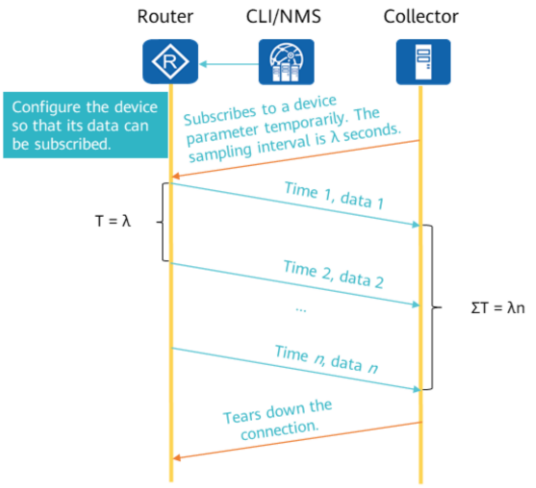
Dynamic Telemetry Subscription

1. The collector initiates a gRPC connection request to the router.
2. The collector initiates a dynamic subscription request to the router.
   - RPC method: Subscribe
   - Specify the sampling path and sampling interval.
3. The router starts to sample data, sends the sampled data (as the response to the dynamic subscription request) to the collector.

HUAWEI

- gRPC is a RPC system developed by Google.

Dynamic Telemetry Subscription Using gRPC (2)



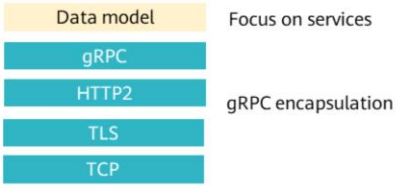
Parameter-specific data collection and on-demand data push

Configure the collector.	system-view grpc grpc server // Enters the gRPC server view. source-ip // Specifies a source IP address to be listened on. server-port // Specifies the number of a port to be listened on. service enable // Enables the gRPC service. (Optional) acl (Optional) idle-timeout (Optional) ssl-policy (Optional) ssl-verify peer commit
Verify the configuration.	display telemetry dynamic-subscription

HUAWEI

gRPC

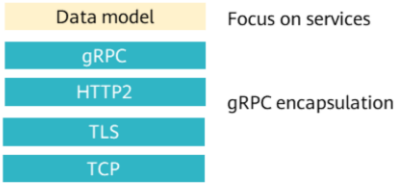
- gRPC, developed by Google, is a language-neutral, platform-neutral, and open-source remote procedure call (RPC) system. It can be used to perform secondary development for both communication parties (NMS and devices). This development focuses on services and shortens interconnection development periods.
- gRPC supports the following languages: C++, Node.js, Python, Ruby, Objective-C, PHP, C#, Java, and Go.
- gRPC is based on HTTP/2, which is better than HTTP/1 in performance. HTTP/2 features include bidirectional streaming, flow control, header compression, and multiplexing request of a single connection.



HUAWEI

gRPC Protocol Stack

- The gRPC protocol stack consists of the following layers:
  - TCP layer: This layer provides connection-oriented and reliable data links.
  - Transport Layer Security (TLS) layer: This layer ensures secure communication between devices and collectors using the TLS protocol, which is optional.
  - HTTP/2 application layer: gRPC is based on the HTTP/2 protocol. To be specific, gRPC uses HTTP/2 features such as bidirectional streams, flow control, and multiplexing.
  - gRPC layer: This layer defines the protocol interaction format for RPCs. Public RPC methods are defined in common .proto files, for example, huawei-grpc-dialout.proto.
  - Data model layer: This layer carries encoded service data. The data can be encoded in Google Protocol Buffers (GPB), XML, and JSON.



HUAWEI

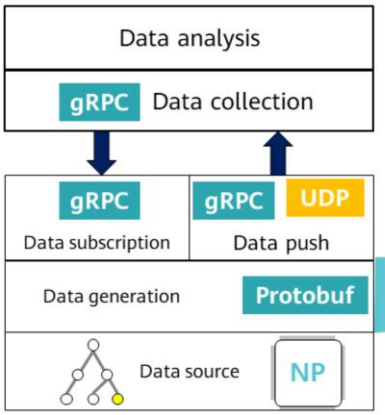
gRPC Service Methods

- gRPC defines four service methods (distinguished by streams).
  - Unary RPCs: The client sends a single request to the server and gets a single response back.
    - Example: rpc Cancel(CancelArgs) returns(CancelReply) {};
  - Server streaming RPCs: The client sends a request to the server, and the server continuously sends a stream back.
    - Example: rpc Subscribe(SubsArgs) returns(**stream** SubsReply) {};
  - Client streaming RPCs: The client continuously sends a stream to the server and waits for a response from the server.
    - Example: rpc LotsOfGreetings(**stream** HelloRequest) returns (HelloResponse) {};
  - Bidirectional streaming RPCs: Both the client and server send a stream.
    - Example: rpc dataPublish(**stream** serviceArgs) returns(**stream** serviceArgs) {};

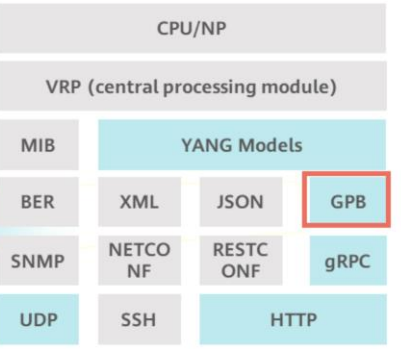
HUAWEI

Telemetry Data Generation

- With telemetry, data is encoded using GPB, which provides a mechanism for serializing structured data flexibly, efficiently, and automatically. Similar to XML and JSON, GPB is a binary encoding mode with better performance.



GBP position in the protocol stack



HUAWEI

Data Source > Data Subscription > Data Generation > Data Push

GPB Encoding

- GPB uses a .proto file to describe the message format and uses digits to replace flag names (keyword names).

Before GPB Encoding	huawei-telemetry.proto	After GPB Decoding

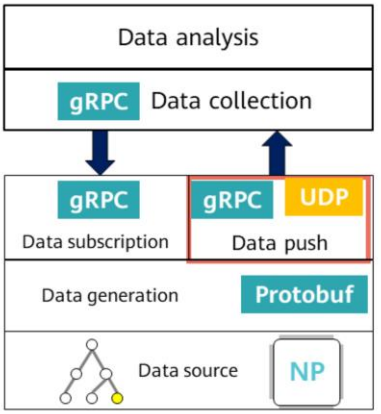
{
1:"HUAWEI"
2:"s4"
3:"huawei-ifm:ifm/interfaces/interface"
4:46
5:1515727243419
6:1515727243514
7{
1[{
1: 1515727243419
2 {
5{
1[{
5:1
16:2
25:"Eth-Trunk1"
}]
}
}
}]
}
}
	syntax = "proto3";
	package telemetry;
	
	message Telemetry {
	string node_id_str = 1;
	string subscription_id_str = 2;
	string sensor_path = 3;
	string proto_path = 13;
	uint64 collection_id = 4;
	uint64 collection_start_time = 5;
	uint64 msg_timestamp = 6;
	TelemetryGPBTable data_gpb = 7;
	uint64 collection_end_time = 8;
	uint32 current_period = 9;
	string except_desc = 10;
	string product_name = 11;
	Encoding encoding =12;
	string data_str = 14;
	
	}
		{
		"node_id_str":"HUAWEI",
		"subscription_id_str":"s4",
		"sensor_path":"huawei-ifm:ifm/interfaces/interface",
		"collection_id":46,
		"collection_start_time":"2018/1/12 11:20:43.419",
		"msg_timestamp":"2018/1/12 11:20:43.514",
		"data_gpb":{
		"row":[{
		"timestamp":"2018/1/12 11:20:43.419",
		"content":{
		"interfaces":{
		"interface":[{
		"ifAdminStatus":1,
		"ifIndex":2,
		"ifName":"Eth-Trunk1"
		}]
		}
		}
		}]
		}

HUAWEI

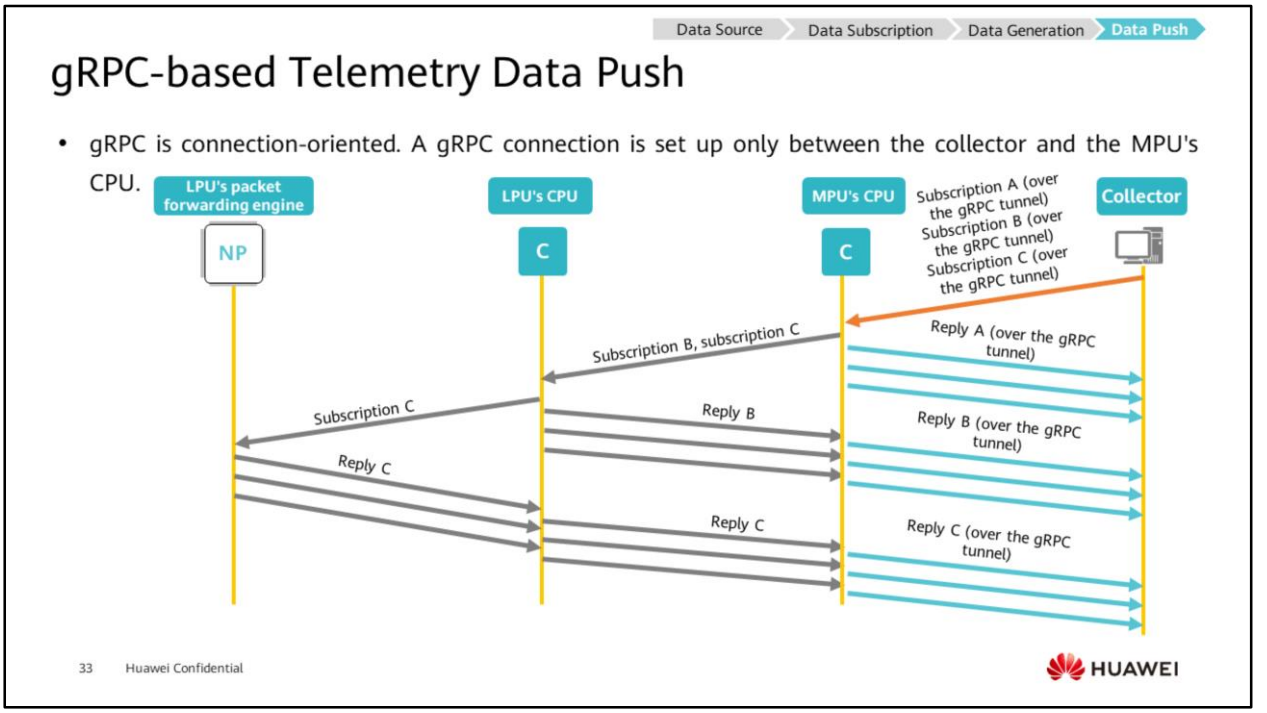
- GPB transmits data in binary mode with a small number of bytes for each transmission, and therefore stands out from other encoding methods, such as XML and JSON, in terms of transmission efficiency. Data collection efficiency is a key concern of Telemetry.
- For more information, see https://developers.google.com/protocol-buffers/.

Telemetry Data Push

- Telemetry data can be pushed in gRPC or UDP mode.



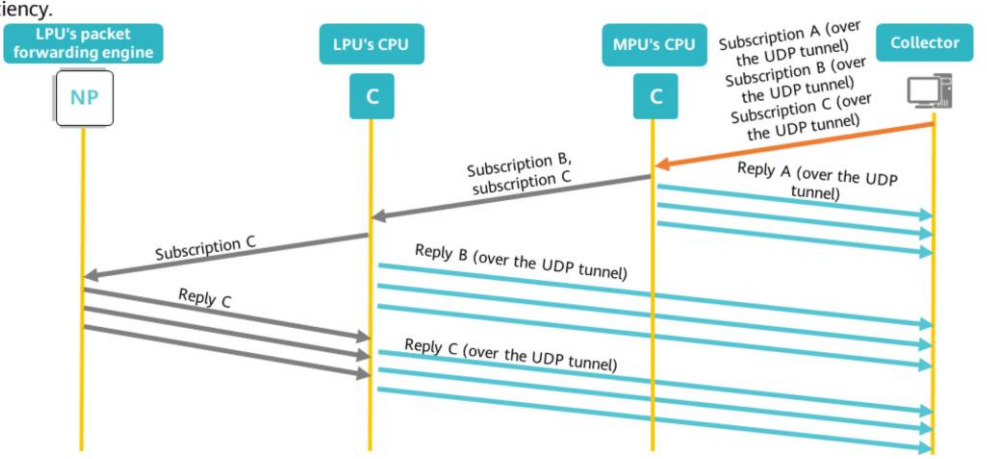
HUAWEI



- The collector functions as the gRPC client, and the device functions as the gRPC server.
- The collector constructs data in GPB or JSON format based on the subscribed event, compiles a .proto file through Protocol Buffers, establishes a gRPC channel with the device, and sends a request message to the device using gRPC.
- After receiving the request, the device parses the .proto file using Protocol Buffers to restore the data for processing.
- After data sorting is complete, the device re-compiles the data using Protocol Buffers and sends a response to the collector using gRPC.
- The collector receives the response message. So far, the gRPC interaction ends.

UDP-based Telemetry Data Push

- Compared with gRPC, data push based on UDP is optimized. In addition to the UDP tunnel with the MPU's CPU, the collector establishes a UDP tunnel with the LPU's CPU. This reduces the pressure on the MPU's CPU and improves efficiency.



HUAWEI

Format of Pushed Data (Based on gRPC)

- The gRPC encapsulation layer is provided by the gRPC open-source software. The following figure shows the protocol stack for pushing data based on gRPC.

Hierarchy		Description
TCP layer		Underlying communication protocol, which is based on TCP connections.
TLS layer		This layer is optional. It is based on the TLS 1.2–encrypted channel and bidirectional certificate authentication.
HTTP/2 layer		gRPC is based on HTTP/2, which is better than HTTP/1 in performance. HTTP/2 features include bidirectional streaming, flow control, header compression, and multiplexing request of a single connection.
gRPC layer		Defines the protocol interaction format for RPCs.
Data model layer	RPC layer	The device proactively initiates an RPC request to the collector, which is defined in the **huawei-grpc-dialout.proto** file. A collector proactively initiates an RPC request for dynamic subscription to a device, which is defined in the **huawei-grpc-dialin.proto** file.
	Telemetry layer	This is defined in the **huawei-telemetry.proto** file.
	Service data layer	Service data at this layer needs to be encoded by the .proto file of the corresponding service, and the NMS needs to decode the data using the .proto file of the corresponding service as well. For example, if the sampling path is **huawei-debug:debug/cpu-infos/cpu-info**, the **huawei-debug.proto** file needs to be used for data encoding and decoding.

HUAWEI

Format of Pushed Data (Based on UDP)

- The format of the data pushed based on UDP is as follows.

Hierarchy		Description
UDP layer		Underlying communication protocol, which is connectionless based on UDP.
DTLS layer		This layer is optional. It is based on the TLS 1.2–encrypted channel and bidirectional certificate authentication.
Message header layer		This layer describes the message header format of pushed telemetry data. The IETF definition is available at https://datatracker.ietf.org/doc/draft-ietf-netconf-udp-pub-channel/01/?include_text=1 The draft has six versions ranging from 00 to 05. Currently, telemetry is implemented based on the 01 version, as shown in the following figure.
Notification message layer	Telemetry layer	This is defined in the **huawei-telemetry.proto** file.
	Service data layer	Service data at this layer needs to be encoded by the .proto file of the corresponding service, and the NMS needs to decode the data using the .proto file of the corresponding service as well. For example, if the sampling path is **huawei-debug:debug/cpu-infos/cpu-info**, the **huawei-debug.proto** file needs to be used for data encoding and decoding.

HUAWEI

Contents

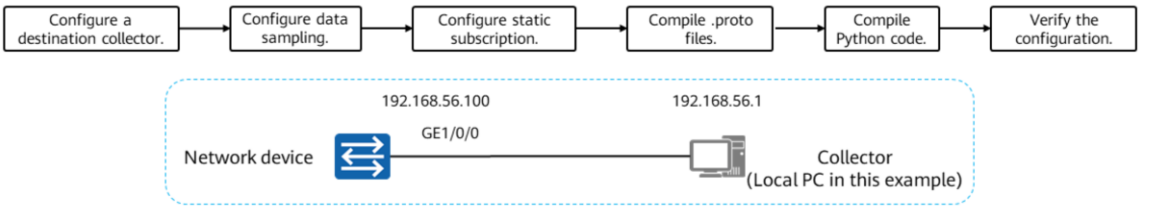
1. Introduction to Telemetry
2. Telemetry Technical Fundamentals
3. **Telemetry Configuration and Practice**

HUAWEI

Case: Configuring Static Telemetry Subscription

- Description:
  - A company has deployed a CE12800 switch with the management IP address 192.168.56.100. To better collect device performance data, it is required that the device push CPU information to the collector in static telemetry subscription mode.

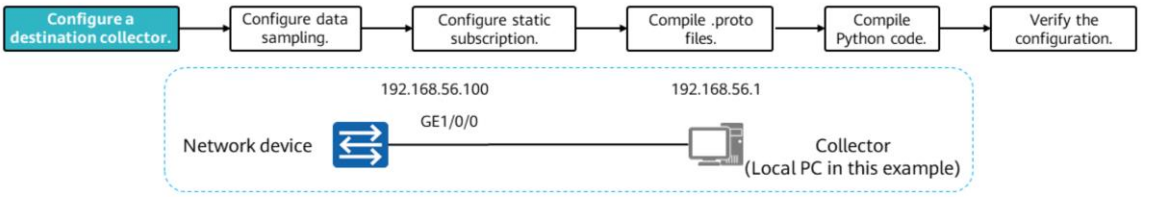
Configuration Roadmap



- Device configuration:
  - Enter the telemetry view.
  - Configure a destination collector for receiving sampled data.
  - Configure data sampling.
  - Configure static telemetry subscription.
- Collector configuration:
  - Prepare and compile .proto files.
  - Compile Python code: Import modules, create a gRPC server, obtain telemetry data, and decode telemetry data.

HUAWEI

Configuring a Destination Collector



1. Enter the telemetry view.

```
<CE1> system-view immediately
[CE1] telemetry
[CE1-telemetry]
```

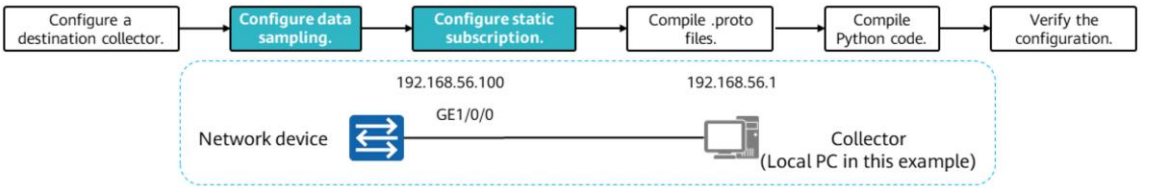
2. Configure a destination collector for receiving sampled data.

   Create a destination group where a destination collector belongs. In this example, create a destination group named **Dest1** and set the IP address and port number of the destination collector to 192.168.56.1 and 20000, respectively.

```
[CE1-telemetry] destination-group Dest1
[CE1-telemetry-destination-group-Dest1] ipv4-address
192.168.56.1 port 20000 protocol grpc no-tls
```


HUAWEI

Configuring Data Sampling and Static Subscription



3. Configure data sampling.

   When configuring static subscription to collect sampled data, create a sampling sensor group and specify a sampling path. In this example, create a sampling group named **Sensor1** and configure a sampling path of CPU information.

```
[CE1-telemetry] sensor-group Sensor1
[CE1-telemetry-sensor-group-Sensor1] sensor-path
huawei-devm:devm/cpuInfos/cpuInfo
```

4. Configure static subscription.

   Create a subscription by associating the destination group and the sampling sensor group. In this example, the destination group **Dest1** is associated with the sampling sensor group **Sensor1**, and the sampling interval is set to 1000 ms. After the configuration is complete, the device continuously pushes data to the destination collector.

```
[CE1-telemetry]subscription Sub1
[CE1-telemetry-subscription-Sub1]destination-group Dest1
[CE1-telemetry-subscription-Sub1]sensor-group Sensor1
sample-interval 1000
```


HUAWEI

Compiling .proto Files for the Collector (1)

5. Compile .proto files.
   - After the device pushes the sampled data to the collector using telemetry, the collector must connect to the device to obtain the sampled data. gRPC and GPB knowledge and development experience are required for interconnection.
   - The .proto files corresponding to the device are released. The collector needs to perform secondary development based on these .proto files to connect to the device. You can download the Huawei device .proto files from the Huawei official website.
     - Common .proto file:
       - **huawei-grpc-dialin.proto**: The collector functions as the client to send an RPC request to the device to subscribe to sampled data for dynamic telemetry subscription.
       - **huawei-grpc-dialout.proto**: The device functions as the client to initiate an RPC request to the collector to push data for static telemetry subscription.
       - **huawei-telemetry.proto**: After sampling service data, the device uses the telemetry header to encapsulate the data, facilitating interconnection with the collector.
     - Service .proto files:
       - **huawei-xxx.proto**: After sampling service data, the device encodes the sampled data in GPB method using this file, and the NMS decodes data using this file as well. Example: **huawei-mac.proto** and **huawei-mpls.proto**.

HUAWEI

Compiling .proto Files for the Collector (2)

- Compile .proto files to obtain the method invoked by the server. You can create a Python script, for example, **run_codegen.py**, and compile .proto files with the following codes:

```
from grpc_tools import protoc
protoc.main(
    (
        '',
        '-I./protos',
        '--python_out=.',
        '--grpc_python_out=.',
        './protos/huawei-grpc-dialout.proto', #dialout file path
    )
)
```

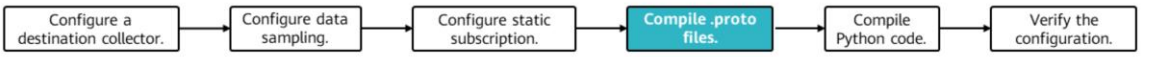
```
protoc.main(
    (
        '',
        '-I./protos',
        '--python_out=.',
        '--grpc_python_out=.',
        './protos/huawei-telemetry.proto', # telemetry file path
    )
)
```

```
protoc.main(
    (
        '',
        '-I./protos',
        '--python_out=.',
        '--grpc_python_out=.',
        './protos/huawei-devm.proto', # devm file path
    )
)
```


HUAWEI

- After the files are compiled successfully, multiple Python files are generated in the current folder.

Compiling .proto Files for the Collector (3)



- Run the run_codegen.py script file to obtain related files generated in the current directory after .proto files are compiled in Python.

.py files generated after .proto files are compiled

HUAWEI

Compiling Python Code (1)

Configure a destination collector. → Configure data sampling. → Configure static subscription. → Compile .proto files. → **Compile Python code.** → Verify the configuration.

6. Compile Python code to obtain device information.

Import modules. --

```
from concurrent import futures
import time
import importlib
import gRPC
import huawei_gRPC_dialout_pb2_gRPC
import huawei_telemetry_pb2
```

```
_ONE_DAY_IN_SECONDS = 60 * 60 * 24

def serve():
server = gRPC.server(futures.ThreadPoolExecutor(max_workers=10))
huawei_gRPC_dialout_pb2_gRPC.add_gRPCDataserviceServicer_to_server(Telemetry_CPU_Info(), server)
server.add_insecure_port('192.168.56.1:20000')
server.start()
try:
    while True:
      time.sleep(_ONE_DAY_IN_SECONDS)
except KeyboardInterrupt:
    server.stop(0)
```

Create a gRPC server object. --

Register the data listening service. --

Set the socket listening port. --

Start the gRPC server. --

Infinite loop listening. --

HUAWEI

- The gRPC module is installed by running the **pip install grpc** command.
- The huawei_grpc_dialout_pb2_grpc and huawei_telemetry_pb2 modules are generated after the .proto files are compiled.

Compiling Python Code (2)

Create a class to inherit the Servicer method. --

Define the initialization method. --

Inherit the dataPublish method. --

Dynamically load a module for obtaining data through telemetry. --

Dynamically load a module for obtaining data through telemetry. --

Decode and print data. --

```
class Telemetry_CPU_Info(huawei_gRPC_dialout_pb2_gRPC.gRPCDataserviceServicer):
    def __init__(self):
        return
    def dataPublish(self, request_iterator, context):
        for i in request_iterator:
            print ('############ start ############\n')
            telemetry_data = huawei_telemetry_pb2.Telemetry.FromString(i.data)
            print (telemetry_data)

            for row_data in  telemetry_data.data_gpb.row:
                print ('----------------')
                print ('The proto path is :'+telemetry_data.proto_path)
                print ('----------------')
                module_name = telemetry_data.proto_path.split('.')[0]
                root_class  = telemetry_data.proto_path.split('.')[1]

                decode_module = importlib.import_module(  module_name+'_pb2')
                print (decode_module)
                decode_func = getattr(decode_module,root_class).FromString
                print ('----------  content is ----------\n')
                print (decode_func(row_data.content))
                print ('----------  done ----------')
if __name__ == '__main__':
  serve()
```


HUAWEI

Verifying the Configuration

7. Verify the configuration.
   - Run the following script to check whether the CPU information is successfully obtained.

```
---------- content is ----------

cpuInfos {
  cpuInfo {
    entIndex: 16842753
    interval: 8
    ovloadThreshold: 90
    position: "1"
    systemCpuUsage: 6
    unovloadThreshold: 75
  }
}

---------- done ----------
```


HUAWEI

Quiz

1. (Multiple-answer question) Which of the following methods can be used to push data using telemetry? ( )
   A. TCP
   B. UDP
   C. gRPC
2. (Short-answer question) What are telemetry data subscription modes?

HUAWEI

1. BC
2. Static subscription and dynamic subscription.

Summary

- Telemetry remotely collects data from physical or virtual devices at a high speed.
- In a narrow sense, the telemetry framework consists of four modules: data source, data generation, data subscription, and data push.
- The data source uses YANG models to define devices. Data is generated using the GPB encoding method. Data subscription can be implemented in static or dynamic mode. Sampled data can be pushed based on gRPC or UDP.
- Python scripts can be run to subscribe to sampled device data through telemetry.

HUAWEI

More Information

- Protobuf official website
  - https://developers.google.com/protocol-buffers/
- gRPC official website
  - https://www.grpc.io/

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

OPS Fundamentals and Practices

HUAWEI

Foreword

- A conventional network device is a closed system that provides fixed functions and services and is not dynamic or flexible. With the rapid development and popularization of networks, diversified and differentiated requirements emerge one after another. Conventional network devices cannot meet these requirements. Some customers require devices with specific openness so that they can develop their own functions and deploy proprietary management policies to implement automatic management and maintenance, lowering management costs.
- Huawei has launched the open programmability system (OPS) to provide openness and programmability capabilities for network devices and enable users to perform secondary development, fully unleashing the potential of devices.
- This course introduces the OPS, describes its principles and application scenarios, and provides a case.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe basic OPS concepts.
  - Describe basic OPS functions.
  - Understand the OPS working principles.
  - Implement the basic OPS configuration.

Contents

1. **OPS Overview**
2. OPS Fundamentals
3. OPS Application Scenarios
4. OPS Configuration Practices

HUAWEI

OPS Overview (1)

- In the following scenarios:
  - During network deployment, services need to be configured for a large number of devices after hardware installation.
  - The network administrator wants devices to automatically take actions when alarms are generated.
  - The network is disconnected when the network administrator wants to remotely deliver commands and periodically collect device information through compiled scripts.

Administrator

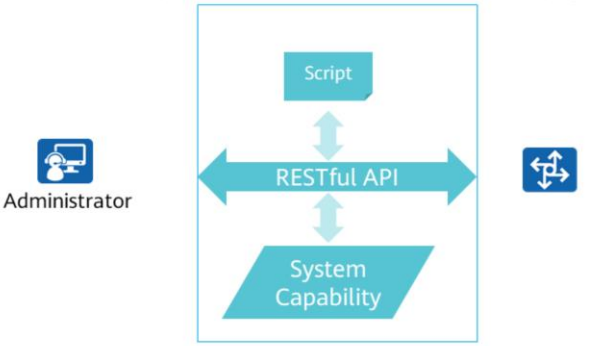
Conventional network devices cannot effectively solve the problems in these scenarios and meet the requirements.

HUAWEI

- Traditional network devices are relatively closed and cannot meet flexible and differentiated network management requirements.

OPS Overview (2)

- Huawei offers the OPS.
- The OPS allows users and third-party developers to develop and deploy network management policies using open RESTful APIs. It implements rapid service expansion, automatic function deployment, and intelligent device management, helping reduce network operation and maintenance costs and simplify network operations.



HUAWEI

- With the OPS, you can compile scripts based on their requirements and import the scripts to network devices for running, which is flexible and efficient.

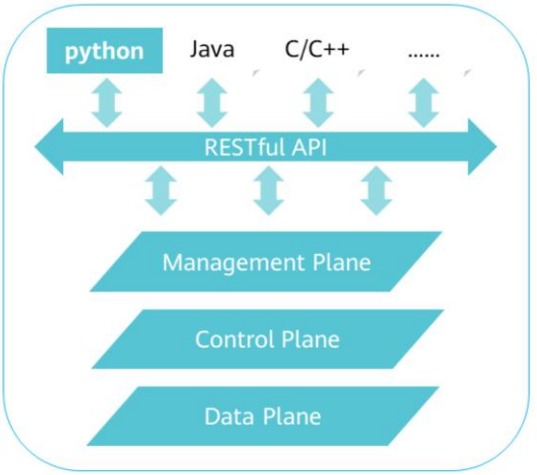
Contents

1. OPS Overview
2. **OPS Fundamentals**
3. OPS Application Scenarios
4. OPS Configuration Practices

HUAWEI

OPS Architecture

- The OPS is developed on Huawei Versatile Routing Platform (VRP). It uses open RESTful APIs to work with VRP's management plane, control plane, and data plane, allowing for function expansion on a device.
- The OPS provides managed objects (MOs) to open devices. Each MO is uniquely identified by a Uniform Resource Identifier (URI). The client can perform operations on objects using standard HTTP methods, such as GET, PUT, POST, and DELETE.



HUAWEI

- The VRP system is developed by Huawei based on years of research and network application experience and its intellectual property rights is owned by Huawei.
- Managed object (MO): an object that can be used to manage network devices by invoking RESTful APIs, such as CPU information, system information, and interface information.
- Uniform Resource Identifier (URI): identifies a specific resource. In the OPS, URIs are used to identify MOs. For example, the URI of the CPU information is /devm/cpuInfos/cpuInfo, which uniquely identifies the CPU information.
- Uniform resource locator (URL): A URL is a URI that can be used to present a resource and specify how to locate the resource, for example, http://www.ietf.org/rfc/rfc2396.txt and ftp://ftp.is.co.za/rfc/rfc1808.txt.
- Huawei network devices that support the OPS provides a running environment for Python scripts. Scripts in Java and C/C++ languages are not supported.

Introduction to RESTful APIs

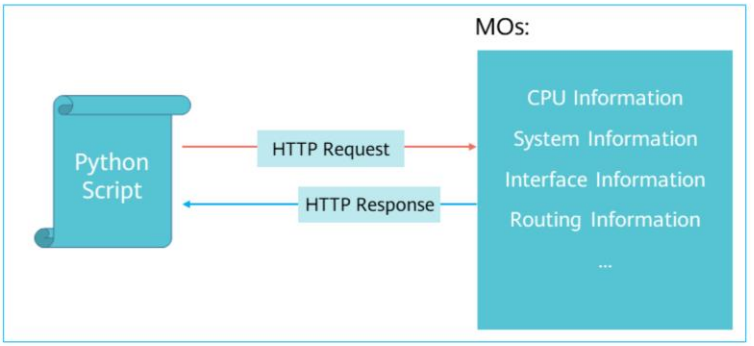
- Representational State Transfer (REST) is a style of software architecture. The design concepts and principles are as follows:
  - All elements on networks can be abstracted as resources.
  - Each resource has a unique resource ID. Operations on resources do not change the resource IDs.
  - Standard methods are used to operate resources. The core operations are GET, PUT, POST and DELETE defined in HTTP.
  - All operations are stateless.
- RESTful API: application programming interfaces (APIs) that comply with the REST architecture.
- The OPS defines a set of RESTful APIs and uses URIs to identify open MOs. You can access MOs using standard HTTP methods (GET, PUT, POST, and DELETE).

HUAWEI

- An API is a particular set of rules and specifications that are used for communication between software programs.
- For more information about RESTful, see the HCIP Programming and Automation Course — RESTful Fundamentals and Practices.

OPS RESTful API Interaction Example

- With RESTful APIs defined by the OPS, you can write Python scripts to send HTTP requests to perform operations on MOs of network devices. Network devices will return HTTP response messages based on the operation results.

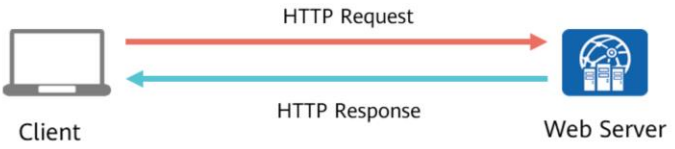


HUAWEI

- The OPS allows you to compile Python scripts, install the scripts on network devices, and send HTTP requests when the scripts are running to manage network devices.

Overview of HTTP

- Hyper Text Transfer Protocol (HTTP) is the most widely used network transfer protocol on the Internet. It is used to transfer hypertext from the World Wide Web server to the local browser.
- HTTP uses the client/server (C/S) architecture model and uses TCP/IP to transfer data, such as HTML files, image files, and query results. HTTP uses the TCP port 80 by default.

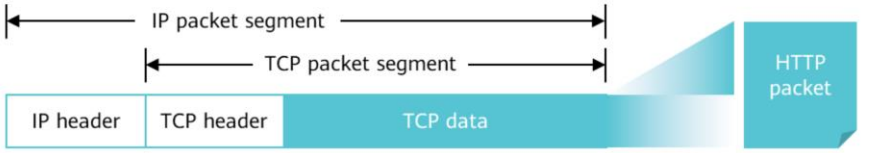


HUAWEI

- Currently, the implementation of RESTful APIs uses the HTTP standard specifications. Therefore, this section briefly describes HTTP.

HTTP Packet Format (1)

- HTTP packets are carried over TCP. The following figure shows the position of HTTP packets in TCP packets.



- An HTTP packet (request message and response message) consists of the following:
  - Start line: describes operations to be performed in a request packet and the execution result in a response packet.
  - Header field: describes additional information about the request and response packets. For example, the header field Host in a request packet specifies the Internet host of the requested resource, and the header field Content-Type in the response packet header indicates the type of the content in the packet body.
  - Body: The body of a request packet is the data to be sent by the client to the web server. The body of a response packet is the data to be returned by the web server to the client, such as HTML texts, images, and videos.

HUAWEI

HTTP Packet Format (2)

- The syntax of HTTP packets is as follows:

	Request Packet Format	Response Packet Format
Start line	<method> <request-URL> <version>	<version> <status-code> <reason-phrase>
Header field	<headers>	<headers>
Body	<entity-body>	<entity-body>

The request and response packets differ only in the syntax of the start line.

- method: action that the client expects the server to perform on a resource. It is a single word, for example, GET or POST.
- request-URL: path of the requested resource on the server.
- version: HTTP version used by packets. The format is as follows: HTTP/<major version No.>.<minor version No.>.
- status-code: HTTP status code, which indicates the execution result of a request operation. The status code is a three-digit number.
- reason-phrase: reason phrase, which explains the status code in human-readable mode.

HUAWEI

- <headers> and <entity-body> are the header field and body of the packet on the previous page.

HTTP Packet Format (3)

- HTTP packet instance:

	Request Packet	Response Packet
Start line	GET / HTTP/1.1	HTTP/1.1 200 OK
Header field	Host: www.example.com User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) Accept: text/html,application/xhtml+xml Accept-Language: zh-CN,zh;q=0.9	Date: Mon, YY-MM-DD 10:21:47 GMT Server: VWebServer Last-Modified: Mon, 15 Jun 2020 10:11:12 GMT ETag: "e211-5a81ca574d000" Accept-Ranges: bytes Vary: Accept-Encoding Content-Length: 13358 Content-Type: text/html
Body	The GET method is usually used to request the server to send a resource whose body is empty.	<HTML> <HEAD> <TITLE>Example Homepage</TITLE> </HEAD> <BODY> <!– Body goes here --> </BODY> </HTML>

HUAWEI

- Header field:
  - Host: contains host name and port number of the web server.
  - User-Agent: contains information about the user agent originating the request.
  - Accept: specifies response media types that are acceptable.
  - Accept-Language: indicate the set of natural languages that are preferred in the response.
  - Date: represents the date and time at which the message was originated.
  - Server: contains information about the software used by the origin server to handle the request.
  - Last-Modified: last modified date for the requested object.
  - ETag: specifies an identifier for a specific version of a resource, often a message digest.
  - Accept-Ranges: allows a server to indicate that it supports range requests for the target resource.
  - Vary: describes what parts of a request message, aside from the method, Host header field, and request target, might influence the origin server's process for selecting and representing this response.
  - Content-Length: indicates length of the response body in octets.
  - Content-Type: indicates Multipurpose Internet Mail Extensions (MIME) type of content.
- For details about the header fields, see RFC 2616.

OPS RESTful API Description Example

- The HTTP requests sent by a user must be constructed based on the RESTful APIs defined by the OPS. The API for obtaining system information is described as follows:

Request example including the HTTP header (XML format):

```
GET /system/systemInfo HTTP/1.1
 Host: localhost
 Accept-Encoding: identity
 Content-Length: 66
 Content-type: application/xml
 Accept: application/xml

 <?xml version="1.0" encoding="UTF-8"?>
 <systemInfo>
</systemInfo>
```

Response example including the HTTP header (XML format):

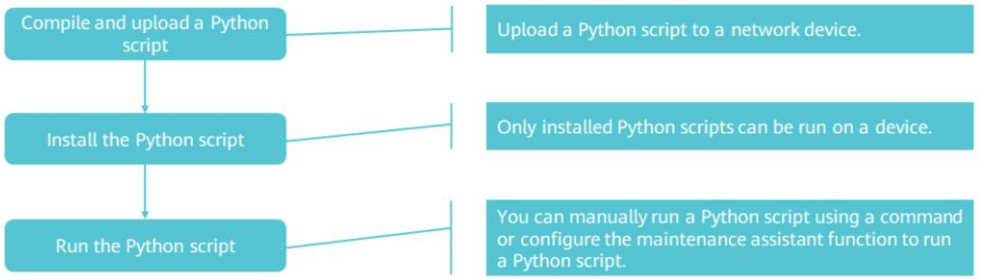
```
HTTP/1.1 200 OK
 Cache-Control: no-cache
 Content-Type: application/xml

 <?xml version="1.0" encoding="UTF-8"?> <systemInfo>
  <sysName>HUAWEI</sysName>
  <platformName>VRP</platformName>
  <platformVer>v800r007c00</platformVer>
  <productName>Device</productName>
  <productVer>V800R007C00</productVer>
  <patchVer></patchVer>
  <esn>123456789</esn>
  <mac>00e0-fc34-1234</mac>
</systemInfo>
```


HUAWEI

- The formats of the OPS RESTful API request and response packets are similar to those of the HTTP request and response packets described in the previous slide.
- Extensible Markup Language (XML) is designed to transmit and store data.
- Currently the OPS RESTful APIs use the XML format to transmit data. In a later version, the APIs can use the JavaScript Object Notation (JSON) format to transmit data. Therefore, the body of the OPS RESTful API request and response packets is in XML format.
- You can download RESTful API Reference on the network device page of http://support.huawei.com.

Usage of Python Scripts

- Currently, the OPS RESTful APIs can be used only in the Embedded Running Environment (ERE). You can compile Python scripts based on OPS APIs and install them on the devices that provide a Python running system. When a Python script is running, an HTTP request is sent to the system to manage devices.
- The following figure shows the Python script execution process.



HUAWEI

- The maintenance assistant is a function of Huawei network devices. You can set the trigger conditions and the Python script to be executed when the conditions are met. The system monitors device running in real time. When the specified trigger condition is met, the network device system automatically executes the Python script to complete the actions defined in the script. For more information about the maintenance assistant, see the Huawei network device product documentation.

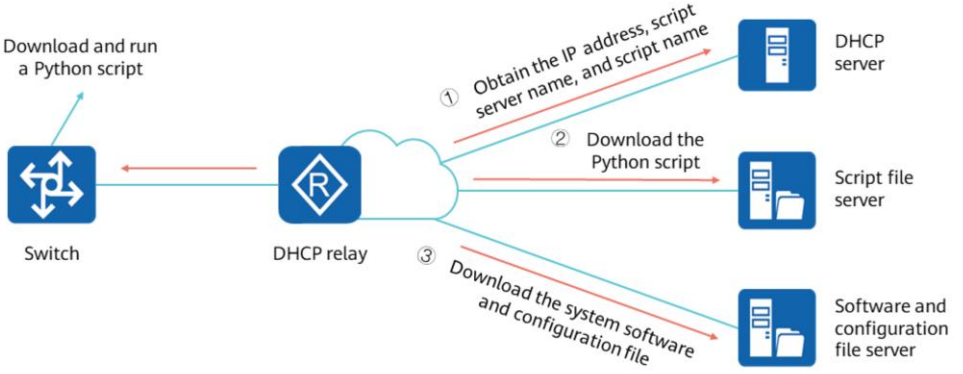
Contents

1. OPS Overview
2. OPS Fundamentals
3. **OPS Application Scenarios**
4. OPS Configuration Practices

HUAWEI

Scenario 1: Automatic Deployment of Unconfigured Devices

- The OPS can implement automatic deployment of unconfigured devices without the need of on-site installation, reducing labor costs and improving deployment efficiency.



HUAWEI

- DHCP server: allocates the temporary IP address, default gateway, and script file server address to the device to be automatically deployed.
- DHCP relay agent: forwards packets exchanged between the device to be automatically deployed and the DHCP server when they are located on different network segments.
- Script file server: stores scripts (Python) required for automatic network device deployment. By running the script files, a network device can obtain information such as the IP address of the software and configuration file server, version file, and configuration file.
- Software and configuration file server: stores system software, configuration files, and patch files required for automatic network device deployment.

Scenario 2: Automatic Health Check

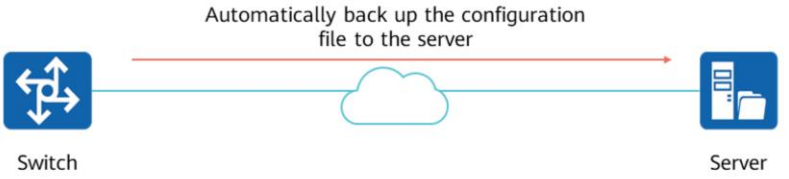
- During conventional device health check, you need to log in to a device and run commands to check the hardware and service running status.
- With the OPS function, a device can automatically run the health check commands, periodically collects health check results, and sends the results to a server for analysis. This function reduces maintenance workload. These commands are delivered through the Python scripts installed in the network device system instead of being remotely delivered. Therefore, you do not need to worry about network disconnection.

HUAWEI

- A Python script can be compiled to deliver commands. When the network is disconnected, the execution result is temporarily stored on the device. After the network is recovered, the execution result is transmitted to the server. Therefore, the impact of network disconnection can be mitigated.

Scenario 3: Automatic Backup of Configuration Files

- A network device subscribes to the configuration file saving events through the maintenance assistant. After the configuration is saved, the device automatically runs the Python script and sends its configuration file to the server for backup, reducing the manual backup workload.



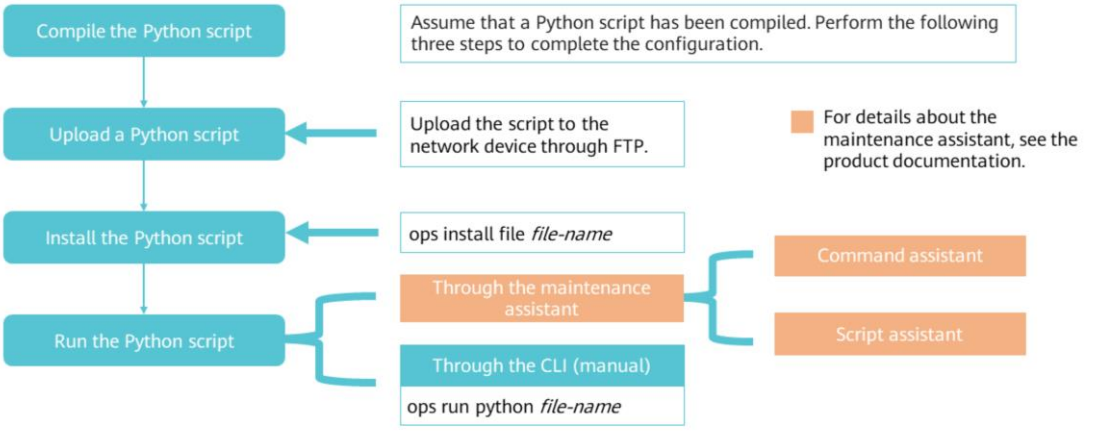
Contents

1. OPS Overview
2. OPS Fundamentals
3. OPS Application Scenarios
4. **OPS Configuration Practices**

HUAWEI

Configuration Roadmap

- The configuration roadmap is as follows: The core is to compile Python scripts, which will be detailed later. For the operation modes and involved commands of the other three steps, see the following figure.



HUAWEI

Compiling a Python Script (1)

- Python scripts are used to deliver HTTP requests to the system for management. Huawei provides Python script templates. You only need to search for RESTful APIs based on the functions to be implemented and modify code.
- Example of a Python script template:

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-

import traceback
import httplib
import string

class OPSConnection(object):
    """Make an OPS connection instance."""

#: an initialization class that creates an HTTP connection
 def __init__(self, host, port = 80):
        self.host = host
        self.port = port
        self.headers = {
(See the next slide for the rest of code.)
```

- The import statement is used to import third-party modules required in the Python script template.
- The OPSConnection class is used to invoke RESTful APIs. This class defines methods for establishing HTTP connections and does not need to be modified.

HUAWEI

Compiling a Python Script (2)

```
            "Content-type": "text/xml",
            "Accept":       "text/xml"
            }
        self.conn = None

# Disable an HTTP connection.
    def close(self):
        """Close the connection"""
        self.conn.close()

    # Create device resources.
    def create(self, uri, req_data):
        """Create operation"""
        ret = self.rest_call("POST", uri, req_data)
        return ret

    # Delete device resources.
    def delete(self, uri, req_data):
        """Delete operation"""
        ret = self.rest_call("DELETE", uri, req_data)
        return ret
(See the next slide for the rest of code.)
```

- Methods defined in the OPSConnection class:
  - def _init_(): is an initialization method that creates an HTTP connection.
  - def close(): shuts down an HTTP connection.
  - def create(): creates device resources.
  - def delete(): deletes device resources.

HUAWEI

Compiling a Python Script (3)

```
# Query device resources.
def get(self, uri, req_data = None):
    """Get operation"""
    ret = self.rest_call("GET", uri, req_data)
    return ret

# Modify device resources.
def set(self, uri, req_data):
    """Set operation"""
    ret = self.rest_call("PUT", uri, req_data)
    return ret

# Invoke classes internally.
def rest_call(self, method, uri, req_data):
    """REST call"""
    print('|---------- request: ----------|')
    print('%s %s HTTP/1.1\n' % (method, uri))
    if req_data == None:
        body = ""
    else:
        body = req_data
```

- Methods defined in the OPSConnection class:
  - def get(): queries device resources.
  - def set(): modifies device resources.
  - def rest_call(): indicates a class method that is invoked by a method mentioned above (create, delete, get, or set).
- You do not need to modify the methods defined in the OPSConnection class.

HUAWEI

Compiling a Python Script (4)

```
            print(body)
        if self.conn:
            self.conn.close()
        self.conn = httplib.HTTPConnection(self.host, self.port)

        self.conn.request(method, uri, body, self.headers)
        response = self.conn.getresponse()
        response.status = httplib.OK    # stub code
        ret = (response.status, response.reason, response.read())
        print('|----------------------------------  response: ----------------------------------|')
        print('HTTP/1.1 %s %s\n\n%s' % ret)
        print('|-------------------------------------------------------------------------------|')
        return ret

# Define the function for obtaining system startup information.
def get_startup_info(ops_conn):

    uri = "/cfg/startupInfos/startupInfo“

    req_data = \
'''<?xml version="1.0" encoding="UTF-8"?>
```

- You need to compile a method to invoke the RESTful API through an instance of the OPSConnection class to manage devices. For example, the def get_startup_info(ops_conn) method in this document is invoked to obtain system startup information. You can specify the method name.
- uri = "/cfg/startupInfos/startupInfo" is the URI of the system startup information. Different MOs have different URIs. You need to change the URI as required.

The bold code needs to be modified based on the site requirements when you compile a Python script.

HUAWEI

Compiling a Python Script (5)

```
<startupInfo>
</startupInfo>
'''
# The response data of the system startup information is similar to the following:
   '''
   <?xml version="1.0" encoding="UTF-8"?>
   <rpc-reply>
     <data>
       <cfg xmlns="http://www.huawei.com/netconf/vrp" format-version="1.0"
content-version="1.0">
         <startupInfos>
           <startupInfo>
             <position>6</position>
             <nextStartupFile>flash:/vrpcfg.cfg</nextStartupFile>
             <configedSysSoft>flash:/system-software.cc</configedSysSoft>
             <curSysSoft>flash:/system-software.cc</curSysSoft>
             <nextSysSoft>flash:/system-software.cc</nextSysSoft>
             <curStartupFile>flash:/vrpcfg.cfg</curStartupFile>
             <curPatchFile>NULL</curPatchFile>
             <nextPatchFile>NULL</nextPatchFile>
           </startupInfo>
```

- You can obtain the values of uri and req_data from *RESTful API Reference* based on the site requirements.
- The system startup information in this document is described in the *RESTful API Reference* as follows:

Operation	URI	Description
GET	/cfg/startupinfos/startupinfo	Obtain system startup information

Request example

<?xml version="1.0" encoding="UTF-8"?>

<startupInfo>

</startupInfo>

HUAWEI

Compiling a Python Script (6)

```
        </startupInfos>
      </cfg>
    </data>
  </rpc-reply>
  '''

  # Execute a GET operation request.
  ret, _, rsp_data = ops_conn.get(uri, req_data)
  if ret != httplib.OK:
      return None

  return rsp_data

# main() function defines the operations to be performed during script running. You can modify the function according to service requirements.
def main():
  """The main function."""
  host = "localhost"
  try:
      # Establish an HTTP connection.
      ops_conn = OPSConnection(host)
      # Invoke the function for obtaining system startup information.
```

- According to the *RESTful API Reference*, the HTTP operation corresponding to the RESTful API for obtaining system startup information is GET. Therefore, the get method of the OPSConnection class instance is invoked.
- You can change the get method to the create, delete, or set method of the OPSConnection class based on the device management function to be implemented.
- req_data indicates the request content, and rsp_data indicates the response message after the network device performs the request operation.
- In the main() method, host indicates the loop address. Currently, RESTful APIs can be invoked only inside the device, that is, the value is localhost.

HUAWEI

Compiling a Python Script (7)

```
        rsp_data = get_startup_info(ops_conn)
        # Disable an HTTP connection.
        ops_conn.close()
        return

    except:
        errinfo = traceback.format_exc()
        print(errinfo)
        return

if __name__ == "__main__":
    main()
```

- You can invoke the method compiled based on the RESTful API in the main() method, for example, the get_startup_info method in this document, to complete the desired function.
- The main() method can flexibly implement various management functions of a network device by using various RESTful APIs and structures of Python such as loops and if-else.
- The last line of the script indicates that the main() method is executed.

HUAWEI

Case: Obtaining CPU Information of a Device

- Description: A network administrator compiles a Python script and uploads the script to a network device to obtain CPU information by using the OPS function of the device. The networking is as follows.

HUAWEI

Compiling a Python Script (1)

- Based on the function of obtaining device CPU information to be implemented, obtain the URI, request message, and response message by referring to the RESTful API Reference. The result is as follows.

URI information:

Operation	URI	Description
GET	/devm/cpuInfos/cpuInfo	Query a board's CPU usage

Request message:

```
<?xml version='1.0' encoding='UTF-8'?>
  <cpuInfo>
    <position></position>
    <entIndex></entIdex>
    <systemCpuUsage></systemCpuUsage>
    <ovloadThreshold><ovloadTreshold>
    <unovloadThreshold></unovloadThreshold>
</cpuInfo>
```

Response message:

```
<?xml version="1.0" encoding="UTF-8"?>
<cpuInfo>
   <position chassis="1" slot="17">17</position>
   <entIndex>17891329</entIndex>
   <systemCpuUsage>5</systemCpuUsage>
   <ovloadThreshold>95</ovloadThreshold>
   <unovloadThreshold>75</unovloadThreshold>
</cpuInfo>
```


HUAWEI

- After knowing the format of the response message, you can parse the response message in the Python script. In this case, the response message is only displayed. You can try to parse the response message to implement more complex functions.

Compiling a Python Script (2)

- Modify the related code in the Python script template based on the URI and request content. Other code in the template does not need to be modified.

Define the get_cpu_info method.

```
def get_cpu_info(ops_conn):
    uri = "/devm/cpuInfos/cpuInfo"          #URI corresponding to the CPU information
    req_data = \                            #Set the content of the request for obtaining CPU information.
      '''<?xml version='1.0' encoding='UTF-8'?>
   <cpuInfo>
     <position></position>
     <entIndex></entIndex>
     <systemCpuUsage></systemCpuUsage>
     <ovloadThreshold></ovloadThreshold>
     <unovloadThreshold></unovloadThreshold>
   </cpuInfo>
      '''
   ret, _, rsp_data = ops_conn.get(uri, req_data)   #Execute the get operation.
   if ret != httplib.OK:
      return None
   return rsp_data
```

Invoke the get_cpu_info method in the main method.

```
def main():
   """The main function."""

   host = "localhost"
   try:
      # Establish an HTTP connection.
      ops_conn = OPSConnection(host)
      # Invoke the function for obtaining CPU information.
      rsp_data = get_cpu_info(ops_conn)
      # Disable an HTTP connection.
      ops_conn.close()
      return

   except:
      errinfo = traceback.format_exc()
      print(errinfo)
      return
```


HUAWEI

Uploading a Python Script

- After the Python script is compiled, save it as cpu_demo.py. The network administrator enables the FTP server on the local PC, and the network device functions as the FTP client to download the Python script.

```
<CE1>ftp 192.168.56.1                                          # Connect to the FTP server.
Trying 192.168.56.1 ...
Press CTRL + K to abort
Connected to 192.168.56.1.
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new user
User(192.168.56.1:(none)):1                                    # Enter the user name.
331 Give me your password, please
Enter password:                                                # Enter the password.
230 Logged in successfully
[ftp]get cpu_demo.py                                           # Obtain the cpu_demo.py script file.
Warning: The file may not transfer correctly in ASCII mode.
500 Unidentified command SIZE cpu_demo.py
200 PORT command okay
150 "D:\FTPSERVER\cpu_demo.py" file ready to send (4605 bytes) in ASCII mode
..
226 Transfer finished successfully.                            # FTP transmission is complete.

FTP: 4605 byte(s) received in 1.135 second(s) 3.962Kbyte(s)/sec.
```


HUAWEI

- For details about how to enable the FTP server on the local PC, you can easily search the way from a search engine.

Compiling a Python Script

- After the Python script is uploaded, run the ops install file file-name command to install the script.

```
<CE1>ops install file cpu_demo.py                                    # Install the Python script.
```

- Verify the installation result.

```
<CE1>display ops script                                              # Check the installed Python
script.
-------------------------------------------------------------------------------
 Index      Size(Byte)  Filename
-------------------------------------------------------------------------------
     0            4605  cpu_demo.py
-------------------------------------------------------------------------------
```

The **cpu_demo.py** script has been successfully installed.

HUAWEI

Running the Python Script (1)

- Run the **ops run python** *file-name* command to run the Python script.

```
<CE1>ops run python cpu_demo.py
|---------------------------------- request: ----------------------------------|          #  Content of the get operation
GET /devm/cpuInfos/cpuInfo HTTP/1.1

<?xml version='1.0' encoding='UTF-8'?>
   <cpuInfo>
     <position></position>
     <entIndex></entIndex>
     <systemCpuUsage></systemCpuUsage>
     <ovloadThreshold></ovloadThreshold>
     <unovloadThreshold></unovloadThreshold>
  </cpuInfo>

|---------------------------------- response: ----------------------------------|          # Content of the response message
HTTP/1.1 200 OK

<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply>
  <data>
    <devm xmlns="http://www.huawei.com/netconf/vrp" format-version="1.0" content
-version="1.0">

(See the next slide for the rest of code.)
```


HUAWEI

Running the Python Script (2)

```
    <cpuInfos>                                                    # Returned CPU information. In this case, the network device has
      <cpuInfo>                                                     two cpus.
        <position chassis="1" slot="1">1</position>               # position indicates the CPU position information.
        <entIndex>16842753</entIndex>                             # entIndex indicates the CPU index information.
        <systemCpuUsage>4</systemCpuUsage>                        # systemCpuUsage indicates the CPU usage.
        <ovloadThreshold>90</ovloadThreshold>                     # ovloadThreshold indicates the CPU overload clearance threshold.
        <unovloadThreshold>75</unovloadThreshold>                 # unovloadThreshold indicates the CPU non-overload threshold.
      </cpuInfo>
      <cpuInfo>
        <position chassis="1" slot="17">17</position>
        <entIndex>17891329</entIndex>
        <systemCpuUsage>4</systemCpuUsage>
        <ovloadThreshold>90</ovloadThreshold>
        <unovloadThreshold>75</unovloadThreshold>
      </cpuInfo>
    </cpuInfos>
  </devm>
 </data>
</rpc-reply>

|-----------------------------------------------------------------------------|
```


HUAWEI

Quiz

1. (Multiple-answer question) The OPS function uses standard HTTP methods to access managed objects (MOs). Which of the following methods are supported? ( )

   A. GET

   B. PUT

   C. POST

   D. DELETE

HUAWEI

1. ABCD

Summary

- The OPS is a unique function of Huawei network devices. It uses RESTful APIs and standard HTTP methods to access MOs on devices, helping users easily cope with various network O&M requirements.
- Huawei provides Python script templates. You only need to search for RESTful APIs based on the functions to be implemented and modify code, which is easy to use.

HUAWEI

More Information

- For more REST information, see the paper written by Dr. Roy Fielding.
  - https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm
- For more information about HTTP, see RFC 2616.
- Python official website: https://www.python.org
- RESTful API reference: https://support.huawei.com

HUAWEI

- The REST software architecture was first mentioned by Roy Fielding in his doctoral paper. Roy Fielding is one of the major authors of the HTTP specifications.

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

RESTful Fundamentals and Practice

HUAWEI

Foreword

- The open ecosystem of the computing industry brings booming development of multiple fields. The network industry is seeking transformation and development, where Software Defined Networking (SDN) is a concept that attracts great attention.
- In the network communications field, network openness is an inevitable trend of industry development. In a fully connected, intelligent world, opening more interfaces means quick connections with other networks. Network openness not only transforms networks, but also further segments the industry chain, bringing new opportunities for industry development. Nowadays, REST is selected and preferred by a growing number of companies, and likewise, RESTful application programming interfaces (APIs) are growing in popularity across the entire network.
- This course aims to help development engineers understand the concepts and functions of SDN, REST and RESTful, working principles of HTTP, and typical practices in invoking RESTful APIs.

HUAWEI

Objectives

- On completion of this course, you will be able to:
  - Describe the development of SDN.
  - Have full knowledge of the background, concepts, and highlights of REST.
  - Understand the relationship between REST, RESTful, and RESTCONF.
  - Master the HTTP packet format and field descriptions.
  - Understand the relationship between HTTP, HTTPS, and HTTP/2.
  - Grasp basic capabilities for invoking RESTful APIs.

HUAWEI

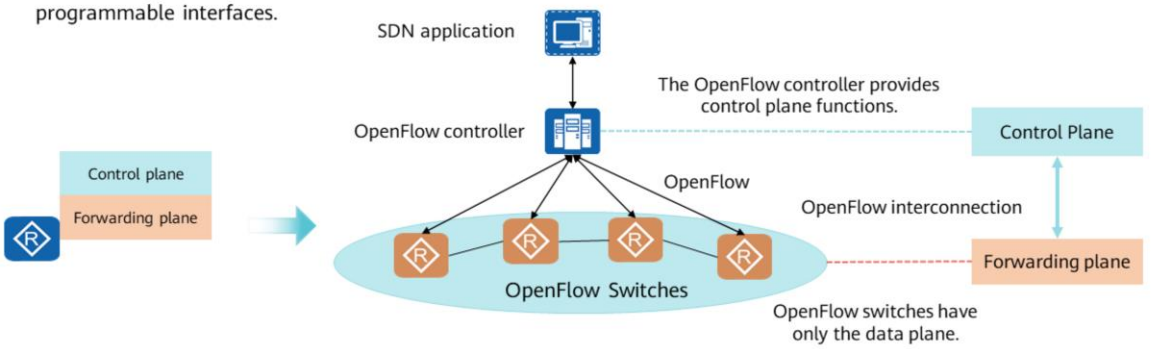
Contents

1. **SDN Overview**
2. REST and RESTful
3. Working Principle of HTTP
4. Practices in Invoking RESTful APIs

HUAWEI

SDN Origin

- SDN was developed by the Clean Slate Program at Stanford University as an innovative new network architecture. The core of SDN is to separate the control plane from the data plane of network devices to implement centralized control of the network control plane and provide good support for network application innovation.
- SDN has three characteristics in initial phase: forwarding-control separation, centralized control, and open programmable interfaces.



HUAWEI

- OpenFlow was defined in the initial phase of SDN. With technology development, many other southbound interface (SBI) protocols are defined between the controller and network devices.

Essential Requirements of SDN

- The essence of SDN is to make networks more open, flexible, and simple. It builds a core brain for a network and implements fast service deployment, traffic optimization, or network service openness through centralized control in the global view.
- Highlights:
  - Provides centralized management, simplifying network management and O&M.
  - Shields technical differences, simplifying network configuration and reducing O&M costs.
  - Offers automatic optimization, improving network utilization.
  - Deploys services rapidly, shortening the service provisioning time.
  - Builds an open network, supporting open and programmable third-party applications.

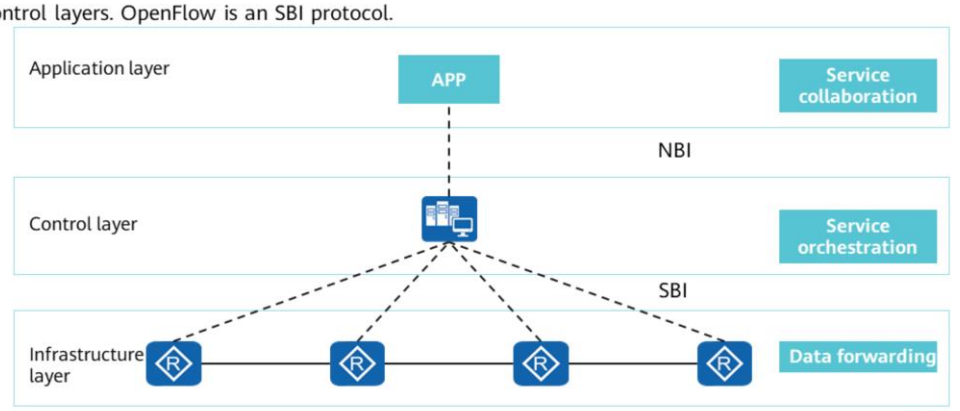
SDN transforms the network architecture.

HUAWEI

- SDN is a broader concept, not limited to OpenFlow. Separation between the control and data planes is a method rather than the essence of SDN.

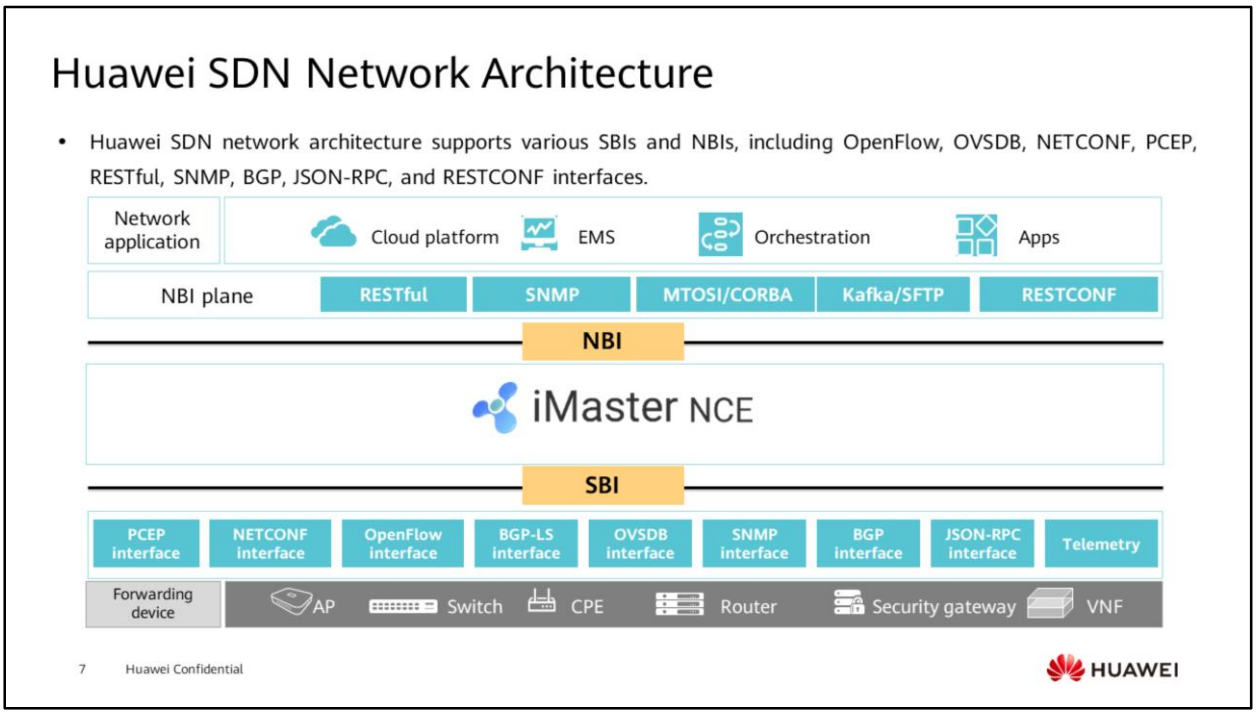
SDN Architecture

- An SDN architecture consists of the application layer, control layer, and infrastructure layer. To communicate between these layers, SDN uses open interfaces where the northbound interfaces (NBIs) communicate between the infrastructure and control layers and the southbound interfaces (SBIs) communicates between the application and the control layers. OpenFlow is an SBI protocol.



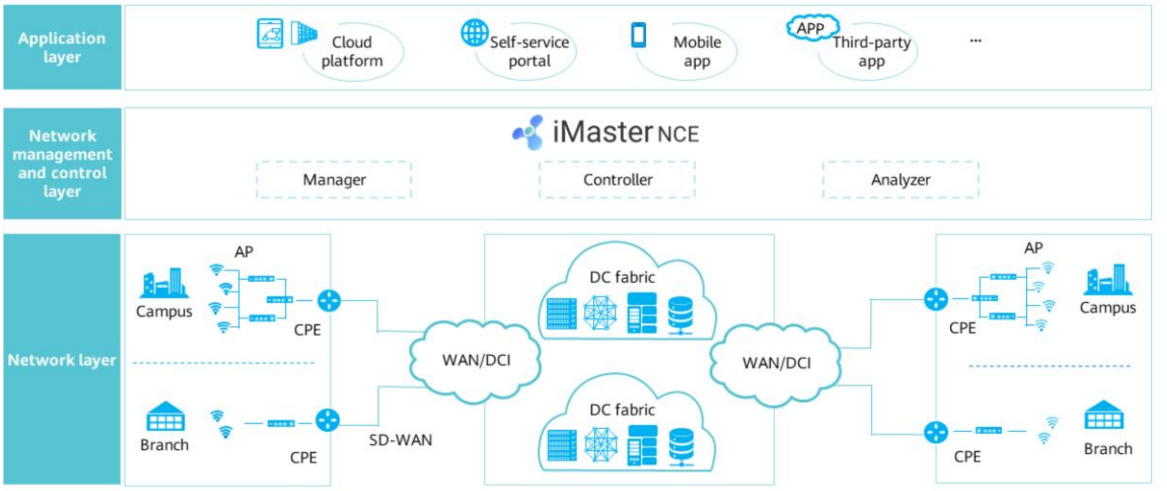
HUAWEI

- Application layer: provides various upper-layer applications for service intents, such as OSS and OpenStack. The OSS is responsible for service orchestration of the entire network, and OpenStack is used for service orchestration of network, compute, and storage resources in a DC. There are also other applications at this layer. For example, a user deploys a security app. This app invokes NBIs of the controller, such as Block (Source IP，DestIP), regardless of the device locations. Then the controller delivers different instructions to network devices based on different southbound protocols.
- Control layer: The SDN controller is deployed at this layer and is the core of the SDN network architecture. The control layer is the brain of the SDN system and implements network service orchestration.
- Infrastructure layer: A network device receives instructions from the controller and performs data forwarding.
- NBI: NBIs, mainly RESTful APIs, are used by the controller to interconnect with the application layer.
- SBI: SBIs are used by the controller to interact with devices through protocols such as NETCONF, SNMP, OpenFlow, and OVSDB.



- Cloud platform: resource management platform in a cloud DC. The cloud platform manages network, compute, and storage resources. OpenStack is the most mainstream open-source cloud platform.
- The Element Management System (EMS) manages one or more telecommunication network elements (NEs) of a specific type.
- Container-based orchestration: The container-based orchestration tool can also provide the network service orchestration function. Kubernetes is a mainstream tool.
- MTOSI or CORBA is used to interconnect with the BSS or OSS. Kafka or SFTP can be used to connect to a big data platform.

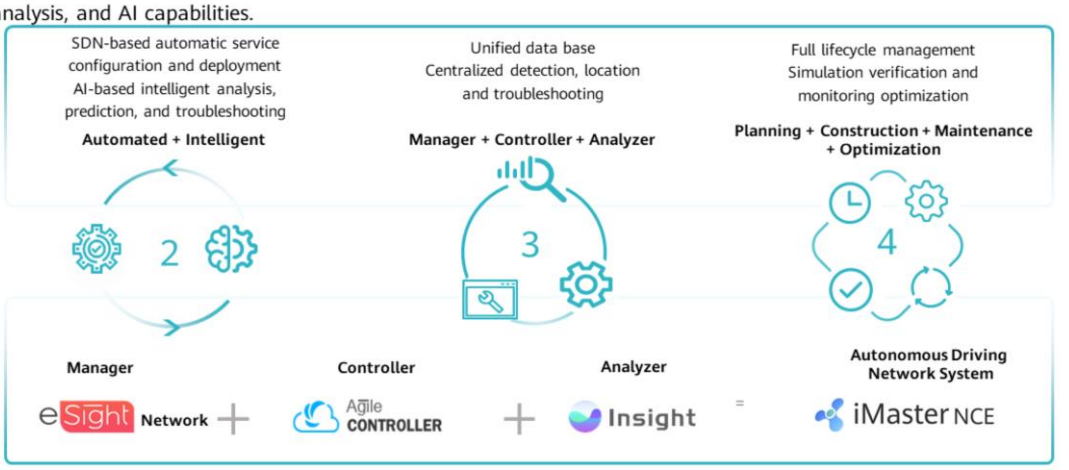
Huawei SDN Solution - Integrating Management, Control, and Analysis to Build an Intent-Driven Network



HUAWEI

Introduction to iMaster NCE

- Huawei iMaster NCE is an industry intelligent network automation platform that integrates management, control, analysis, and AI capabilities.



HUAWEI

- iMaster NCE effectively connects physical networks with business intents and implements centralized management, control, and analysis of global networks. It enables resource cloudification, full lifecycle automation, and data analytics-driven intelligent closed-loop management according to business and service intents and provides open network APIs for rapid integration with IT systems.
- Huawei iMaster NCE can be used in the enterprise data center network (DCN), enterprise campus, and enterprise branch interconnection (SD-WAN) scenarios to make enterprise networks simpler, smarter, open, and secure, accelerating enterprise service transformation and innovation.

iMaster NCE Application

iMaster NCE

Data center	iMaster NCE-Fabric *
Enterprise campus	iMaster NCE-Campus *
SD-WAN	iMaster NCE-WAN
IP WAN	iMaster NCE-IP
WAN transmission	iMaster NCE-T

* Introduced in this document

HUAWEI

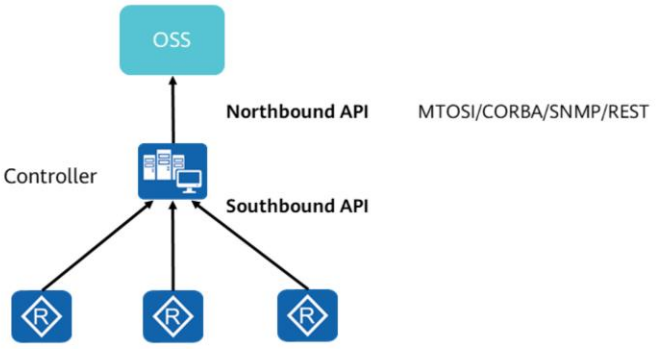
Contents

1. SDN Overview
2. **REST and RESTful**
   - Background of REST
   - Overview of REST
3. Working Principle of HTTP
4. Practices in Invoking RESTful APIs

HUAWEI

Basic Concepts of Southbound and Northbound

- Southbound and northbound are relative concepts. Generally, the interfaces provided by an upper-layer system for a lower-layer system are called southbound APIs, and the interfaces provided by a lower-layer system for an upper-layer system are called northbound APIs.
- APIs provide the interaction function to implement data transmission.

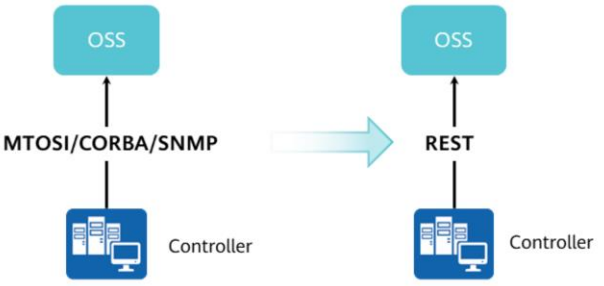


HUAWEI

- The operation support system (OSS) is a necessary support platform for telecom services.
- MTOSI: Multi-Technology Operations System Interface
- Common Object Request Broker Architecture (CORBA) is a standard object-oriented application program system specification formulated by Object Management Group (OMG).

Development of Northbound APIs

- There are many northbound API standards, for example, MTOSI and CORBA. Vendors' controllers need to open northbound APIs compliant with specific standards to easily interconnect with carriers' integrated network management platform.
- Devices may use diverse northbound APIs such as SNMP, CORBA, and REST. Nowadays, these APIs are gradually unified into the **REST** style.



HUAWEI

API Evolution: Unified URI Naming Conventions

- REST standardizes universal resource identifier (URI) naming conventions. Resource-oriented URI names are easy to understand.
- A URI identifies the location of a resource.

Without REST: URIs were randomly named.	With REST: URIs are independent and methods are clear.
URIs without constraints on naming conventions: • Reserving a meeting room: **/reserve/**meetingroom/B25R • Canceling a reserved meeting room: **/cancel/**meetingroom/B25R • Querying the meeting room status: /meetingroom/B25R?**method=query**	URIs with naming conventions of REST: • Reserving a meeting room: /meetingroom/B25R, **with the calling method being POST** • Canceling a reserved meeting room: /meetingroom/B25R, **with the calling method being DELETE** • Querying the meeting room status: /meetingroom/B25R, **with the calling method being GET**

HUAWEI

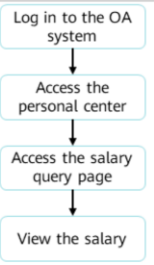
API Evolution: Stateless Design

- REST uses a stateless design to enhance system scalability.
- Stateful: The server needs to save and maintain the state information of previous requests. Each subsequent state depends on the previous state.
- Stateless: The server sends back the same response for invocation of the same request, function, or method, without depending on other requests. The server does not need to maintain state information, facilitating expansion. At least one URI is available to locate a resource.

Without REST: both stateful and stateless used

A client can obtain the required information only after multiple requests. The figure on the right shows an example of logging in to the office automation (OA) system to view the salary. The process is complex, affecting user experience.

The required information cannot be directly located using only one URI.



With REST: stateless

In the REST architecture, the required result can be obtained with only one request. For example, the salary can be queried using the following URIs:

Salary of Tom: http://oa.company.com/salary/tom

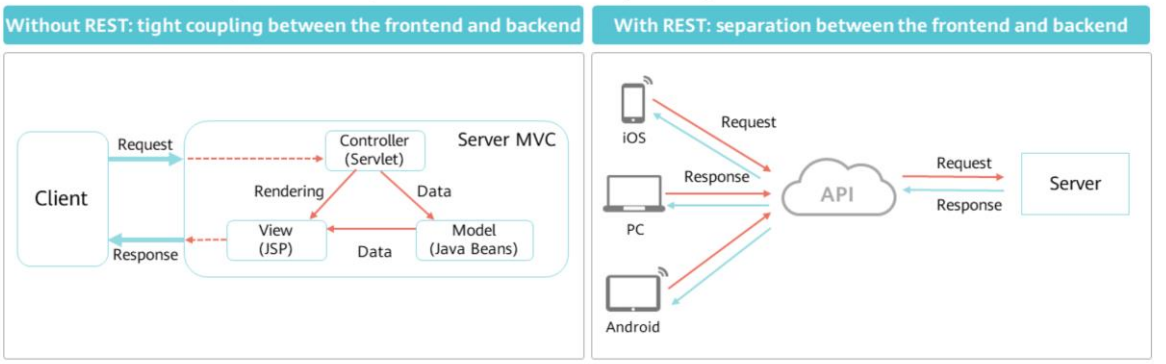
Salary of Jerry: http://oa.company.com/salary/jerry

HUAWEI

- Stateful request: A server generally needs to save and maintain the status information of previous requests. Each request can use information about the previous requests by default.
- Stateless request: The processing result on the server must be based on the information carried in the same request.

API Evolution: Separation Between Frontend and Backend (1)

- REST provides good cross-platform compatibility and separates the frontend from the backend.
  - The frontend mainly involves the pages that can be viewed by users.
  - The backend mainly involves the logical functions related to data processing.

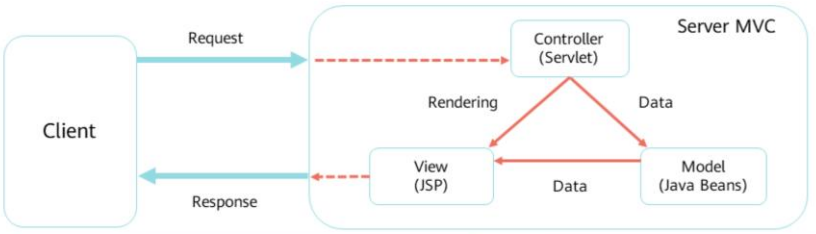


HUAWEI

- An API is a set of predefined functions or methods for connecting different components of a software system. It is a set of routines that can be accessed by applications and developers based on software or hardware without having to access the source code or understand the details of the internal working mechanism. For example, if a computer needs to invoke information in a mobile phone, we simply need to connect the computer and the mobile phone by using a data cable. In this example, the interfaces on the computer and mobile phone at both ends of the data cable.

API Evolution: Separation Between Frontend and Backend (2)

- In early web projects, rendering is generally performed on the server. A client directly obtains and interacts with the rendered view on the server.



Advantage: The entire page is sent back with only one request, featuring a high loading speed.

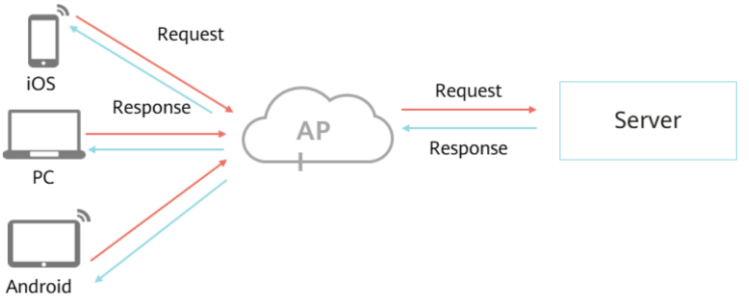
Disadvantage: Components are closely associated with each other. If a view is modified on the frontend, the logical modules of the backend must be modified accordingly.

HUAWEI

- Rendering refers to the process of transforming views such as HTML into visual images that can be seen by human eyes.
- For web applications of early days, a view is a graphical user interface (GUI) composed of HTML elements. For today's web applications, the GUI incorporates new elements such as Adobe Flash, XHTML, XML/XSL, and WML.
- The Model-View-Controller (MVC) is a pattern for creating web applications and consists of three parts: controller, model, and view.
    - The controller is responsible for processing user interaction in an application. Generally, the controller reads data from the view, controls user input, and sends data to the model.
    - The model is responsible for processing the data logic of an application. Generally, the model stores and reads data from the database.
    - The view is responsible for processing data display. Generally, a view is created based on model data.

API Evolution: Separation Between Frontend and Backend (3)

- REST develops rapidly with the emergence of the mobile Internet. The server cannot render a GUI for each type of terminal. Therefore, rendering on the server cannot meet the requirements of mobile Internet. To maintain compatibility with various types of terminals, REST emerges and becomes pervasive.
- With REST, the server only processes data, and the client is responsible for GUI rendering. In this way, the frontend and backend are separated.
- REST-compliant RESTful APIs provide a unified communication method for various types of terminals, which is beyond the reach of traditional APIs.



HUAWEI

- An API is a set of predefined functions or methods for connecting different components of a software system. It is a set of routines that can be accessed by applications and developers based on software or hardware without having to access the source code or understand the details of the internal working mechanism.
- Separation between the frontend and backend has become an industry standard for the Internet projects in the industry. It lays a solid foundation for the large-scale distributed architecture, elastic computing architecture, microservice architecture, and multi-terminal services (such as browsers, vehicle-mounted terminals, Android, and iOS). The key to separation between the frontend and backend is that the frontend page invokes the RESTful API of the backend for data interaction.

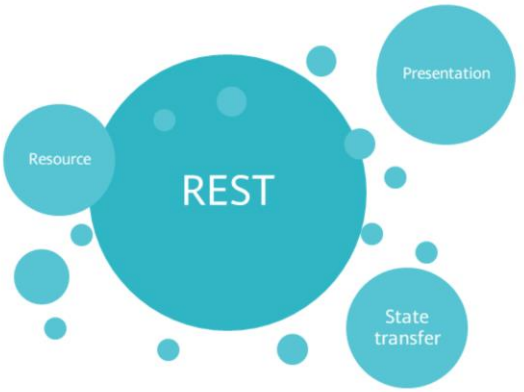
Contents

1. SDN Overview
2. **REST and RESTful**
   - Background of REST
   - Overview of REST
3. Working Principle of HTTP
4. Practices in Invoking RESTful APIs

HUAWEI

Overview of REST

- Representational State Transfer (REST) was proposed by Roy Thomas Fielding, HTTP's leading designer, in his doctoral dissertation in 2000. In short, REST is a design style.



REST uses the client-server interaction framework and involves the following concepts:

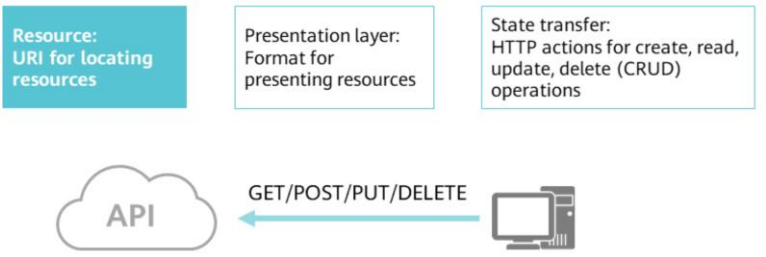
- **Resource**
- **Presentation**
- **State transfer**

HUAWEI

- Abstract of Roy's doctoral dissertation Architectural Styles and the Design of Network-based Software Architectures:
  - This dissertation explores a junction on the frontiers of two research disciplines in computer science: software and networking. Software research has long been concerned with the categorization of software designs and the development of design methodologies, but has rarely been able to objectively evaluate the impact of various design choices on system behavior. Networking research, in contrast, is focused on the details of generic communication behavior between systems and improving the performance of particular communication techniques, often ignoring the fact that changing the interaction style of an application can have more impact on performance than the communication protocols used for that interaction. My work is motivated by the desire to understand and evaluate the architectural design of network-based application software through principled use of architectural constraints, thereby obtaining the functional, performance, and social properties desired of an architecture.

REST Concepts - Resource

- A resource is an information entity. All things on the network can be abstracted as resources. Each resource has a URI. For example, ports, network elements (NEs), boards, and equipment rooms are resources.
- Resources and operations are the core of REST. The URI is used to locate resources, and HTTP actions (GET, POST, PUT, and DELETE) are used to describe operations.

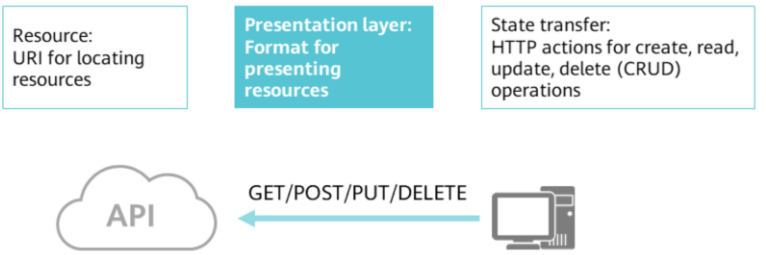


HUAWEI

- REST is short for Representational State Transfer, in which the main entity — resource — is not presented.

REST Concepts - Presentation Layer

- Representation is the way a resource is presented.
- For example, a text can be presented in TXT, HTML, XML, JSON, or even binary format; an image can be presented in JPG, PNG, or other formats.

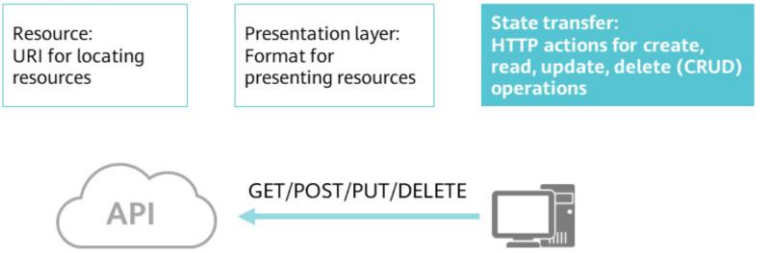


HUAWEI

- A URI represents only a resource entity but not its presentation.

REST Concepts - State Transfer

- Accessing a URI represents an interaction between the client and the server. In this process, data and status changes are involved.
- Hypertext Transfer Protocol (HTTP) is a stateless protocol. This means that all states are stored on the server side, not on the client side. Therefore, the client performs operations on the server by using a certain method, so that a state transfer occurs on the server. Such a state transfer depends on the presentation layer, so it is called "Representational State Transfer".
- The method used by the client is **HTTP**.



HUAWEI

RESTful and RESTCONF

- RESTful APIs comply with the **REST design style**. There are no mandatory requirements on RESTful APIs, and therefore RESTful APIs can be defined freely.
- RESTCONF APIs also comply with the **REST design style**. Unlike RESTful APIs, RESTCONF APIs must comply with **RFC 8040** defined by Internet Engineering Task Force (IETF). RFC 8040 defines RESTCONF APIs and their specifications. RESTCONF APIs are based on HTTP and are used to access data defined in Yet Another Next Generation (YANG). RESTCONF allows web applications to access configuration data, status data, and event notifications of network devices in a modular and scalable manner.

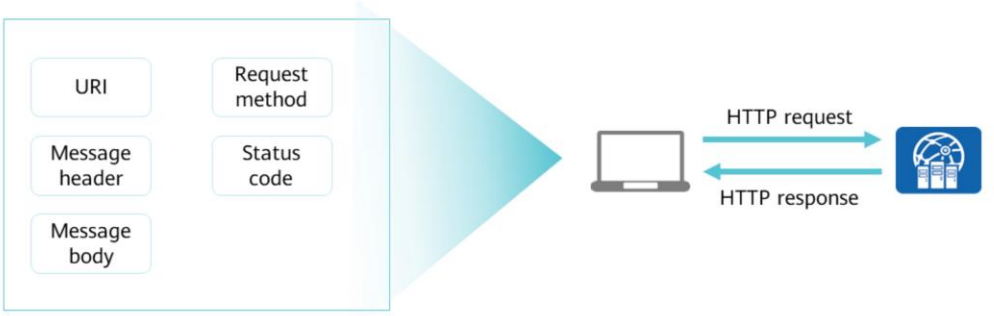
REST + **Conf**iguration

HUAWEI

- YANG defines the storage content and configuration of data.

Defining RESTful APIs

- What is the content defined when we define RESTful APIs?



All the preceding fields are contained in HTTP packets. Therefore, REST makes full use or heavily relies on HTTP.

HUAWEI

- Relationship between the URI and URL.
  - The URL is a subset of the URI. The former must be an absolute path, while the latter can be an absolute path or a relative path. For example, http://127.0.01:8080/AppName/rest/product/1 is a URL, and AppName/rest/product/1 is a URI.

Contents

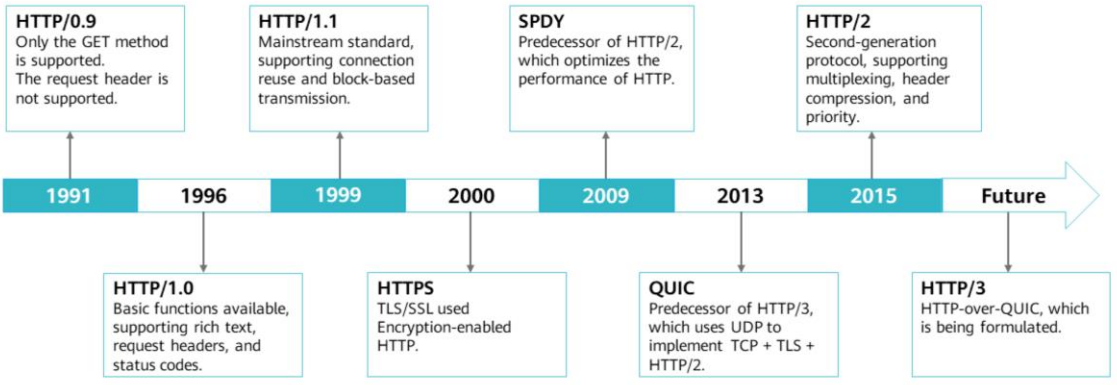
1. SDN Overview
2. REST and RESTful
3. **Working Principle of HTTP**
   - Overview of HTTP
   - HTTP/1.1
   - HTTPS and HTTP/2
4. Practices in Invoking RESTful APIs

HUAWEI

- As mentioned earlier, REST makes full use or heavily relies on HTTP. Next, we will move on to HTTP.

Development of HTTP

- HTTP was proposed by Tim Berners-Lee in 1990. After the WWW Alliance was established, the IETF working group further optimized and released the HTTP protocol. Over the evolution from HTTP/0.9 to HTTP/3, HTTP offers increasingly high performance.



HUAWEI

- SPeeDY (SPDY) is a TCP-based application-layer protocol developed by Google. Its objective is to optimize the performance of HTTP and shorten the loading time of web pages and improve security by using technologies such as compression, multiplexing, and priority. The core idea of SPDY is to minimize the number of TCP connections. SPDY is an enhancement to HTTP, instead of a protocol for replacing HTTP.
- Quick UDP Internet Connection (QUIC) is a UDP-based low-delay Internet transport layer protocol developed by Google. In November 2016, the IETF convened the first meeting of the QUIC working group, which attracted wide attention from the industry. This means that QUIC starts its standardization process and will become a next-generation transport layer protocol.

Overview of HTTP

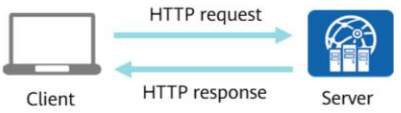
- Hypertext Transfer Protocol (HTTP) is an application layer protocol for distributed, collaborative, and hypermedia information systems. HTTP has been used by the World Wide Web since its inception.
- It uses TCP/IP to transmit data and uses TCP port 80 by default. Websites starting with http:// are standard HTTP services.

HUAWEI

- The data transmitted using HTTP can be HTML, images, texts, and so on.

Working Principle of HTTP

- HTTP is based on the client/server (C/S) architecture. The HTTP request and response process is as follows:
  - A client establishes a TCP connection with the server.
  - The client sends an HTTP request, which consists of a request line, request header, empty line, and request data.
  - The server receives the request and sends back an HTTP response, which consists of a status line, response header, empty line, and response body.
  - The TCP connection is released.
  - The client parses the status line, response header, and response body in sequence and displays the response packet. If the body contains HTML data, the client formats the data based on the HTML syntax and displays the data.

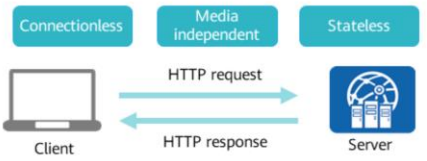


HUAWEI

- An HTTP client is usually a browser. A web server can be an Apache server or an Internet Information Services (IIS) server.
- When a TCP connection is released, if the value of the Connection field in the packet header is close, the server proactively closes the TCP connection, and the client passively closes and releases the TCP connection. If the value of Connection is keepalive, the connection lasts for a period of time and can continue to receive requests.

Features of HTTP

- HTTP has the following features:
  - Connectionless: Only one request is processed for each connection. After processing the client's request, the server disconnects from the client.
  - Media independent: Any type of data can be sent by HTTP as long as both the client and the server know how to handle the data content. It is required for the client as well as the server to specify the content type using appropriate MIME-type.
  - Stateless: The server and client are aware of each other only during a current request. Afterwards, both of them forget about each other. This facilitates quick processing of a large number of transactions and enhances the protocol scalability.



HUAWEI

- The browser differentiates the displayed content such as HTML, XML, GIF, and flash based on MIME-type.
- Advantages of the connectionless feature: This mode saves the transmission time and improves the concurrent performance. No persistent connection is established. Instead, one response is made to each request. However, if a connection is repeatedly established and torn down, the efficiency is affected. In HTTP/1.1, a TCP connection is maintained between the browser and the server for a period of time and will not be disconnected immediately after a request ends.
- Stateless means that, if the processing of subsequent packets requires the previously exchanged information, the information must be retransmitted. Although HTTP/1.1 is a stateless protocol, cookies are introduced to implement the function of maintaining status information.
- A cookie is a text file stored on a client. This file is associated with a specific web page and saves the information about the web page accessed by the client.

Contents

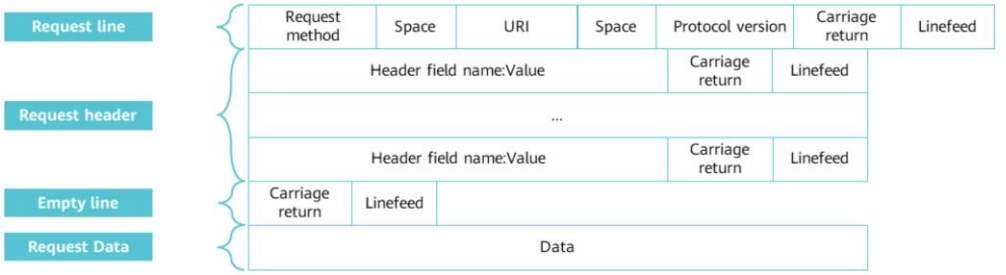
1. SDN Overview
2. REST and RESTful
3. **Working Principle of HTTP**
   - Overview of HTTP
   - HTTP/1.1
   - HTTPS and HTTP/2
4. Practices in Invoking RESTful APIs

HUAWEI

- HTTP/1.1 has been widely used since it was proposed in 1999 and has become a mainstream standard for more than 20 years. In the following part, we will introduce HTTP packets, which are based on HTTP/1.1.

Client Request Message

- An HTTP request message sent by a client consists of a request line, request header, empty line, and request data. The following figure shows the format of a request message.

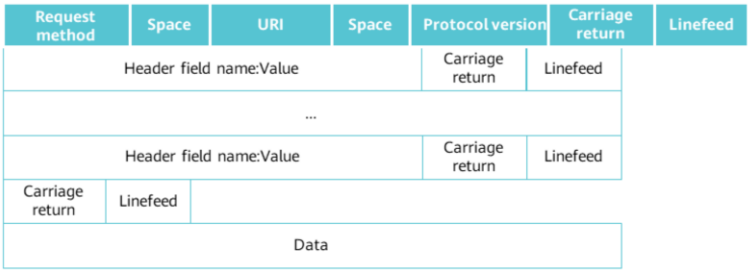


HUAWEI

Client Request Message - Request Line

- A request line consists of the request method field, URI field, and HTTP version field.
  - Request method: HTTP request method, for example, GET and POST. An HTTP client (for example, a browser) must specify the request type when sending a request to the server.
  - URI: A URI identifies the resource involved in a request.
  - Protocol version: The protocol version allows the sender to indicate the format of a message and its capacity for understanding further HTTP communication.

Example of a request line: GET http://www.w3.org/pub/WWW/TheProject.html HTTP/1.1



HUAWEI

HTTP Request Method

- HTTP requests can use multiple request methods. HTTP 1.0 defines three request methods: GET, POST, and HEAD. HTTP 1.1 provides six new request methods: OPTIONS, PUT, PATCH, DELETE, TRACE, and CONNECT.

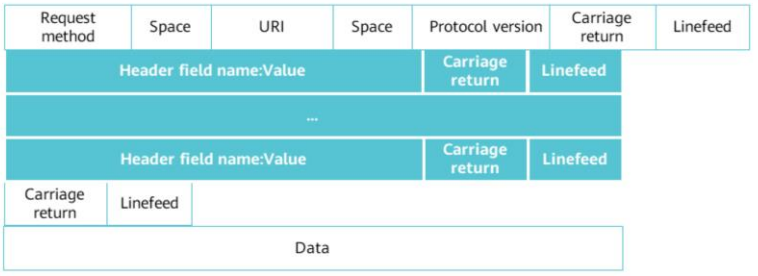
Method	Function
GET	Requests the specified page information. The server returns the specific data.
POST	Submits data, for example, a form.
HEAD	Similar to the GET method. However, the response does not contain any specific data. HEAD is used to obtain the header.
PUT	Updates and modifies data.
DELETE	Deletes a specified page.
CONNECT	Implements HTTP proxy.
OPTIONS	Allows the client to check the server performance.
TRACE	Echoes back the request received by the server. This is used for testing or diagnosis.
PATCH	Partially updates known resources.

HUAWEI

Client Request Message - Request Header

- The request header allows the client to send additional information about the request to the server. These fields act as request modifiers whose semantics are equivalent to the parameters invoked in the programming language method.

Example of a request header:
Accept: text/html
Accept-Encoding: gzip, deflate, br
Connection: keep-alive



HUAWEI

Request Line > Request Header > Request Data

Request Header Fields

Request Header Field	Description
Accept	Media types which are acceptable for the response.
Accept-Encoding	Data encoding format that can be used by the client for decoding, for example, gzip.
Accept-Language	Natural languages that are preferred as a response to the request. This field is mandatory if the server can provide more than one language.
Authorization	Authorization information, which is usually contained in the response to the WWW-Authenticate header sent by the server.
Content-Type	Media type of the data. The default value is text/plain.
Content-Length	Length of the request message body.
Host	Name of the host to be accessed by the client.
Referer	Resource from which the client accesses the server. This field contains a URI, indicating that the client accesses the requested page from the page represented by the URI.
User-Agent	Information about the user who sends the request and the client type.
Cookie	Used by the client to save the data returned by the server. Generally, the user identity information is saved.
Connection	Whether to disconnect the connection after the request is processed.

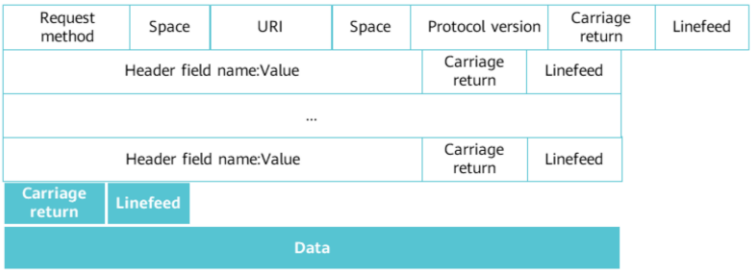
HUAWEI

- In HTTP 1.0, each connection involves only one request and response and is closed after the request is processed. HTTP 1.0 does not have the Host field. In HTTP 1.1, multiple requests and responses can be transmitted in the same connection, and multiple requests can be processed concurrently.
- WWW-Authenticate is a simple and effective user identity authentication technology in the early stage.
- The browser differentiates the displayed content such as HTML, XML, GIF, and flash based on MIME-type.
- For more information, refer to RFC HTTP 1.1 at https://www.ietf.org/rfc/rfc2616.html.

Client Request Message - Request Data

- Empty line: The server is notified of the end of a request header through an empty line.
- Request data: If the GET method is used, the request data is empty. If the POST method is used, the request data is the data to be submitted.

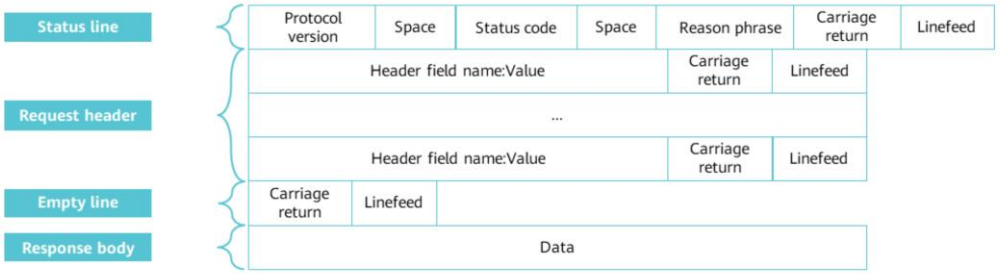
Example of request data:
user=admin&password=123456



HUAWEI

Server Response Message

- An HTTP response consists of four parts: status line, response header, empty line, and response body.

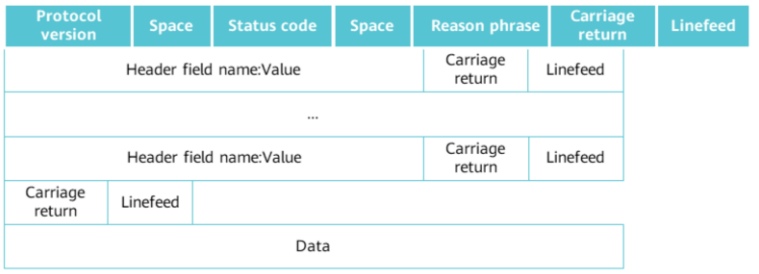


HUAWEI

Server Response Message - Status Line

- The first line of the response message is the status line, which consists of the protocol version, status code, and reason phrase, separated by a space character.
  - Protocol version: The protocol version allows the sender to indicate the format of a message and its capacity for understanding further HTTP communication.
  - Status code: a three-digit result code, which is used to return the operation result to the client.
  - Reason phrase: a short text description of the status code, which facilitates understanding.

Example of a status line: HTTP1.1 200 OK



HUAWEI

Status Code

- An HTTP status code is a three-digit number indicating the status of a response from the server. It is used to return the operation result to the client.

Status Code		Description	
1XX	(Informational) The request is received.	100 Continue	If the request is received, go to the next step.
2XX	(Successful) The request is successful.	200 OK	Success. The response body is available.
		201 Created	Resource created successfully.
		204 No Content	Success. No response body is available.
3XX	(Redirection) Further operations need to be performed.	301 Moved Permanently	A new URI is allocated to the target resource, and all future resources will be associated with the new URI.
4XX	(Client Error) Request error.	400 Bad Request	The request body is incorrect and carries error information.
		401 Unauthorized	Authorization failed. For example, the certificate does not match.
		403 Forbidden	Access denied. The possible cause is that the user attempts to perform operations beyond the permission or the login user name or password is incorrect.
		404 Not Found	The requested resource cannot be found.
5XX	(Server Error) Server error.	500 Internal Server Error	The request cannot be executed due to an internal server error. The user needs to resend the request later.
		501 Not Implemented	The function has not been implemented.

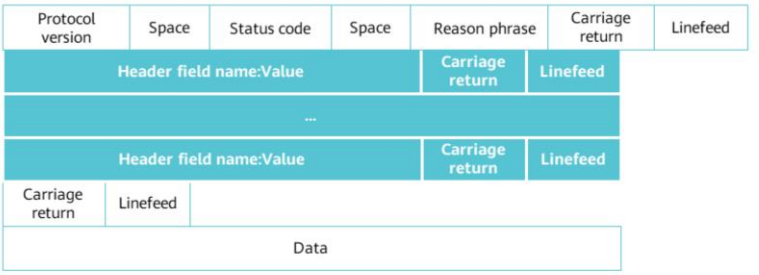
HUAWEI

Status line > Response header > Response body

Server Response Message - Response Header

- A response header allows the server to transfer additional information about the response. These header fields provide information about the server and the resource identified by the URI.

Example of a response header: Server: JSP3/2.0.14



HUAWEI

- The response header describes the basic information about the server and data. The server uses the response header to notify the client of how to process the data that it replies to.

Status line | Response header | Response body

Common Fields in the Response Header

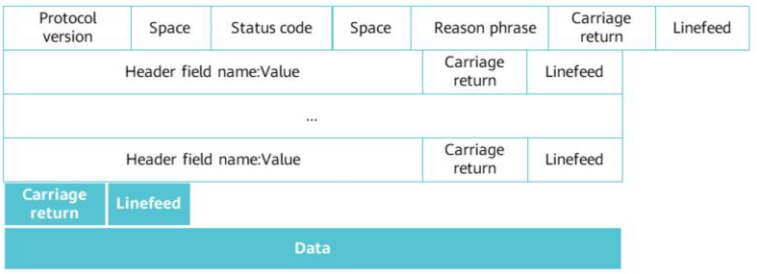
Response Header Field	Description
Allow	Request methods, for example, GET and POST, supported by the server.
Content-Encoding	Encoding method of a document. The content type specified by Content-Type can be obtained only after decoding.
Content-Length	Content length. This field is required only when the client uses a persistent HTTP connection.
Content-Type	Media type of the data. The default value is text/plain.
Date	Current Greenwich Mean Time (GMT) time.
Location	Location from which the client obtains resources. This field is used together with status code 302 to specify a new URI as the recipient.
Server	Type of a server.
Set-Cookie	Cookie associated with the page.
Transfer-Encoding	Data transfer format.
WWW-Authenticate	Type of authorization information that the client should provide in the Authorization header. This field is mandatory in a response that contains the 401 (Unauthorized) status line.

HUAWEI

- The HTTP response header is often combined with the status code. For example, the status code 302 (indicating that the location has changed) is usually used together with the Location header, and the status code 401 (Unauthorized) must be used together with a WWW-Authenticate header. The response header can be used to set the cookie, specify the date, instruct the client to refresh the page at the specified interval, and so on.

Server Response Message - Response Body

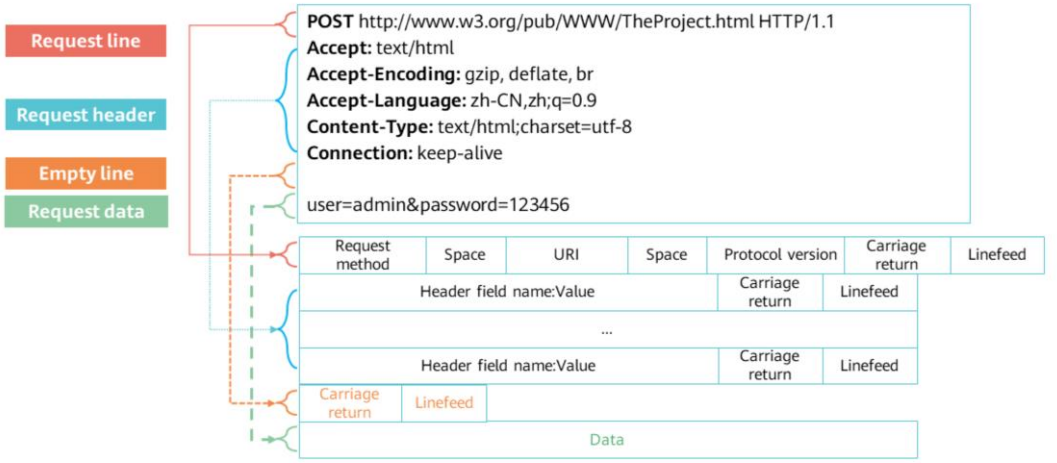
- Empty line: It is used to inform the server of the end of the response header.
- Response body: message body of the response. If data of the pure data type is requested by the client, pure data is returned. If an HTML page is requested, the HTML code is returned.



HUAWEI

Client Request Message - Example

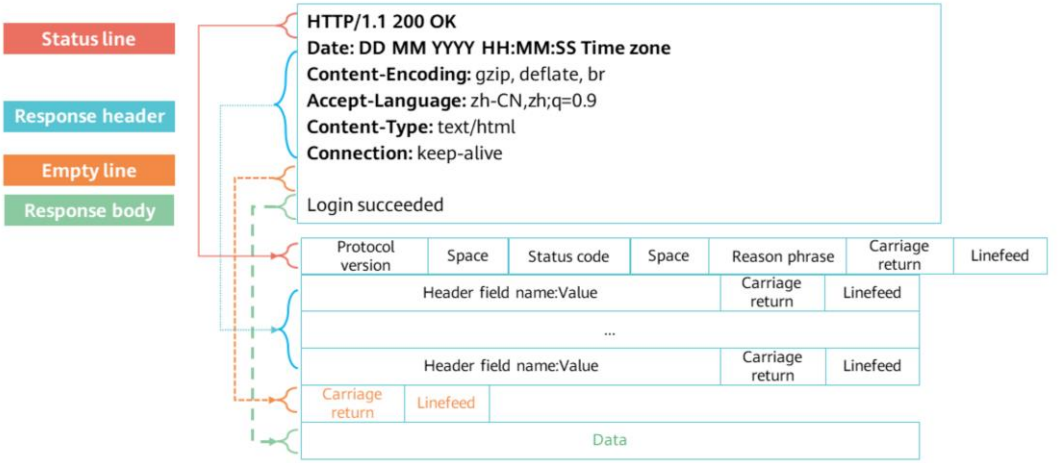
- A client sends a request containing the user name and password to the server for login authentication.



HUAWEI

Server Response Message - Example

- The server sends a response to the client, indicating that the authentication is successful.



HUAWEI

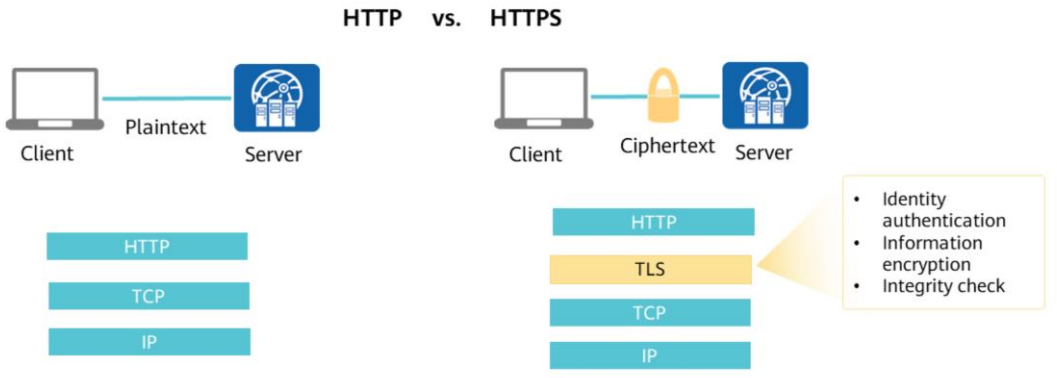
Contents

1. SDN Overview
2. REST and RESTful
3. **Working Principle of HTTP**
   - Overview of HTTP
   - HTTP/1.1
   - HTTPS and HTTP/2
4. Practices in Invoking RESTful APIs

HUAWEI

Overview of HTTPS

- Based on HTTP, Hypertext Transfer Protocol over Secure Sockets Layer (HTTPS) uses SSL/TLS for encryption, enhancing security.



HUAWEI

- HTTP transmits information in plaintext, which may pose risks of information interception, tampering, and hijacking. Transport Layer Security (TLS) provides identity authentication, information encryption, and integrity check functions, and therefore can prevent such problems.

Overview of SSL/TLS

- Secure Sockets Layer (SSL) and Transport Layer Security (TLS) are security protocols that ensure security and data integrity for network communication.
- TLS is the successor of SSL. SSL and TLS encrypt data transmitted on network connections between the transport layer and application layer without affecting the TCP and HTTP protocols. Therefore, if HTTPS is used, it does not require much modification on HTTP pages.

Risks of HTTP communication without using SSL/TLS

1. Eavesdropping: **HTTP transmits information in plaintext**, and third parties can obtain communication data.
2. Tampering: Third parties can tamper with the communication content.
3. Pretending: Third parties can pretend others to participate in communications.

Features of HTTPS communication using SSL/TLS

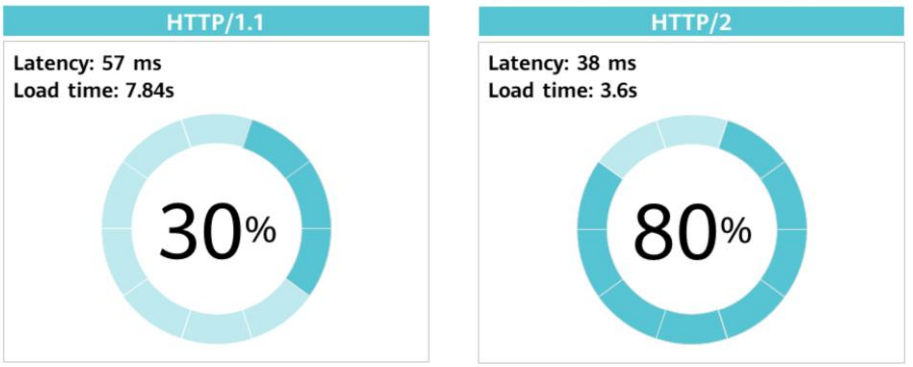
1. Data integrity: HTTPS uses SSL/TLS to **encrypt and transmit all HTTP packets.** All information is encrypted and cannot be intercepted or tampered with by third parties.
2. Data verification: Communication data is verified. If data is tampered with, the communication parties can immediately detect the tampering.
3. Trusted data source: Identity certificates are used to prevent identity spoofing.

HUAWEI

- For more information, refer to the RFC document at https://www.ietf.org/rfc/rfc5246.html.

Overview of HTTP/2

- HTTP/2 is the second major version of the HTTP protocol. It was initially named HTTP 2.0 and is based on the SPeeDY (SPDY) protocol.
- HTTP/2 greatly improves web performance without changing HTTP semantics, methods, status codes, URIs, and header fields.

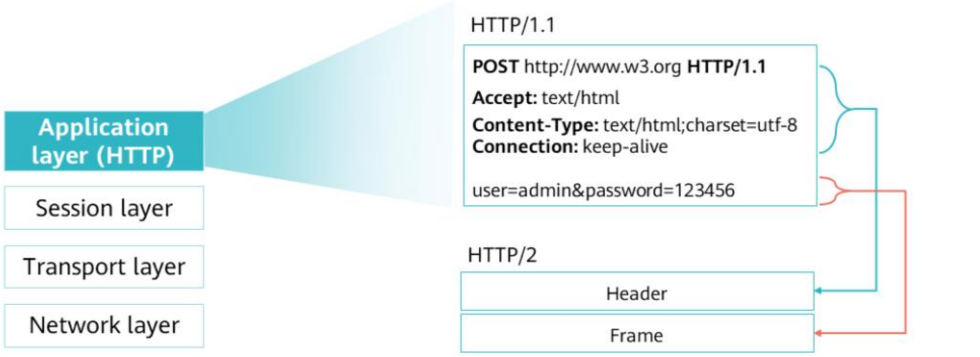


HUAWEI

- SPDY is a TCP-based application-layer protocol developed by Google. SPDY aims to optimize the performance of HTTP and shorten the loading time of web pages and improve security by using technologies such as compression, multiplexing, and priority. The core idea of SPDY is to minimize the number of TCP connections. SPDY is an enhancement to HTTP, instead of a protocol for replacing HTTP.

Core of HTTP/2 - Binary Transmission

- The core of performance enhancement in HTTP/2 is binary transmission.
- In HTTP/1, data is transmitted in text mode, which has some defects. This is because various scenarios need to be considered to achieve robustness, due to the diversity of text formats. On the contrary, the binary format involves only 0s and 1s, featuring significant convenience and robustness.



HUAWEI

- Enhancements to HTTP/2:
- Header compression: The HPACK algorithm is used to compress headers to reduce the header size and improve performance.
- Multiplexing: A request message can be divided into frames, which are sent in sequence and are reassembled at the other end. In HTTP/1.1, when a client sends multiple requests through a TCP connection, the server can only respond to the requests in sequence. Subsequent requests may be blocked.
- Resource pushing: In addition to responding to client requests, the server can push additional resources to clients.
- Priority: HTTP/2 defines complex priority rules. A browser can request multiple resources at a time and specify priorities to help the server determine how to process these resources, avoiding resource competition.

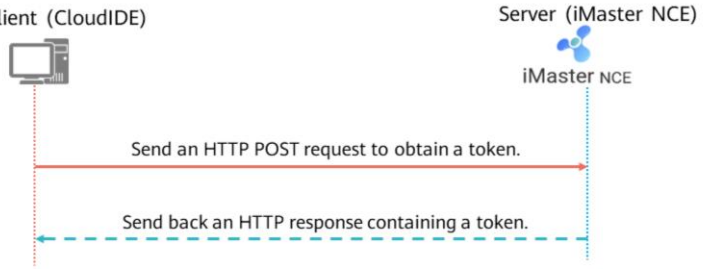
Contents

1. SDN Overview
2. REST and RESTful
3. Working Principle of HTTP
4. **Practices in Invoking RESTful APIs**

HUAWEI

Case: Invoking RESTful APIs to Obtain Tokens (1)

- Description:
  - Huawei iMaster NCE is a network automation and intelligence platform that integrates management, control, analysis, and artificial intelligence (AI) functions. In the SDN network architecture, iMaster NCE centrally manages network devices. You can implement intent-driven networks based on open northbound APIs of iMaster NCE.
  - This case describes an example of how to use the Python Requests module to log in to iMaster NCE and invoke service interfaces.

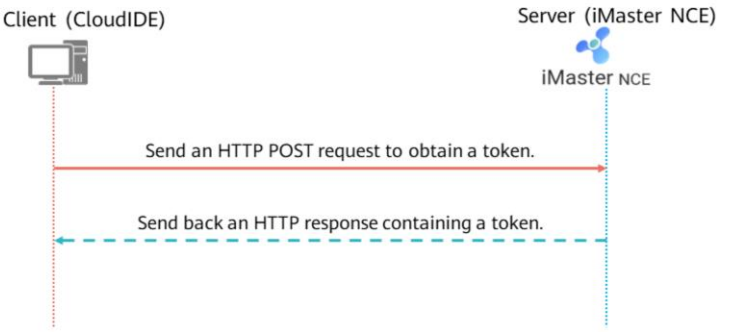


HUAWEI

- In this case, two objects are involved in the networking: CloudIDE and iMaster NCE.
- CloudIDE is a cloud-based development environment provided by HUAWEI CLOUD. The local environment can also be used to write code.
- A sandbox environment is available to iMaster NCE in the datacom developer community.
- A token is a character string used for authentication on APIs.
- For details, see HCIP-Datacom-Northbound Openness Lab Guide.

Case: Invoking RESTful APIs to Obtain Tokens (2)

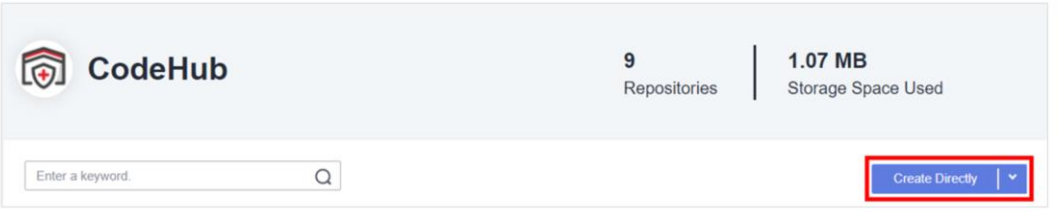
- Configuration roadmap:
  - Prepare the environment, including the IDE and sandbox environment.
  - Compile the code for invoking Python RESTful APIs.
  - Verify the configuration.



HUAWEI

Procedure: Create a Code Repository

- Access the CodeHub console and create a project and a code repository on the cloud.

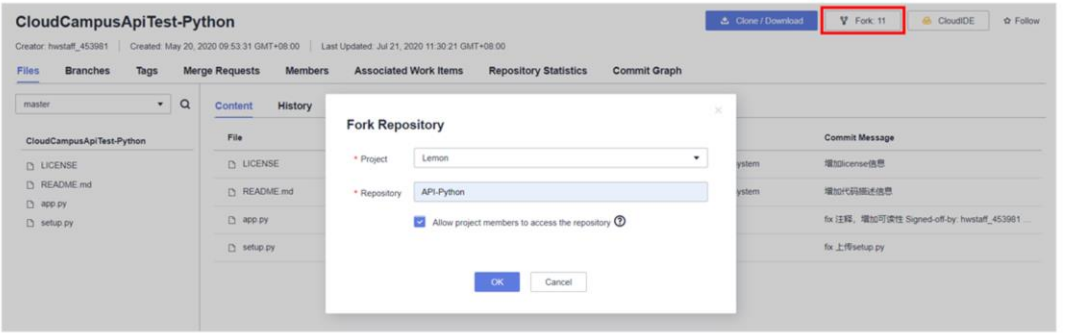


HUAWEI

- For more operations, see the CodeHub guide at https://support.huaweicloud.com/codehub/index.html.

Procedure: Fork the Complete Code

- Fork the complete code to the repository. In this example, the complete code has been uploaded to DevCloud on HUAWEI CLOUD. You can use Huawei CloudIDE to download the code to the local host. (Alternatively, you can perform online programming.)



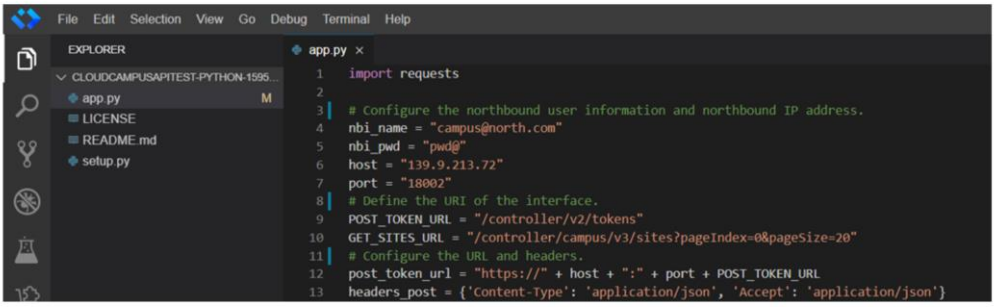
HUAWEI

- Code in DevCloud on HUAWEI CLOUD: https://devcloud.cn-north-4.huaweicloud.com/codehub/project/68494a8ad06b4eea9a1c3f18be115161/codehub/620653/home

Procedure: Start the CloudIDE

- Click **CloudIDE** to open the code.

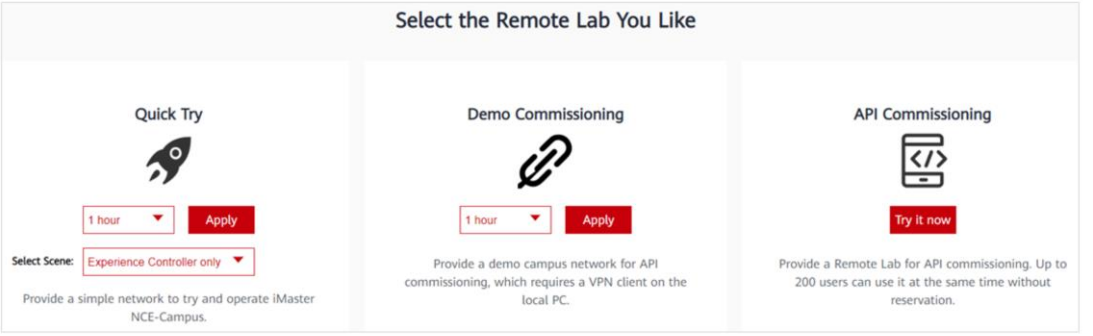
- Check the code in CloudIDE.



HUAWEI

Procedure: Reserve a Sandbox Environment

- Reserve a sandbox environment.
  - The developer community provides the sandbox experience. The intent-driven campus network sandbox environment is used as an example.
  - You can experience the sandbox environment in three ways: **Quick Try, Demo Commissioning, and API Commissioning**. Here, **Quick Try** in CloudIDE is used as an examxple.

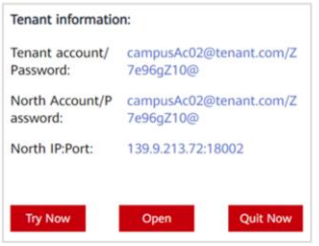


HUAWEI

- Website for reserving a sandbox environment: https://devzone.huawei.com/openecosystem/experienceView/campus.html

Procedure: Check Sandbox Information

- Check sandbox information.
  - **Tenant account:** account for logging in to iMaster NCE.
  - **North Account:** account used by a location application to interact with iMaster NCE.
  - **Northbound IP:Port:** URL for logging in to iMaster NCE. You can click **Try Now** to log in to iMaster NCE.
  - **Password:** password for logging in to the sandbox environment.

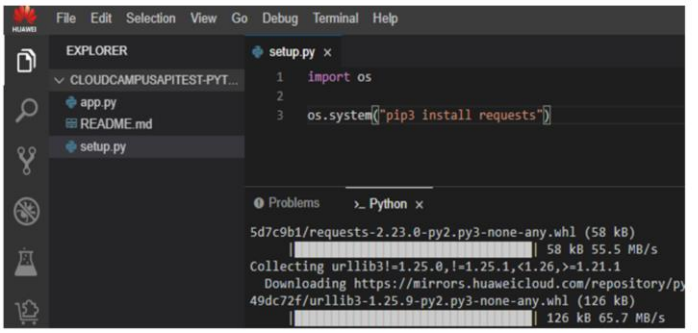


HUAWEI

Procedure: Install the Requests Module

- Install the Requests module.
  - The Python Requests module is required in this experiment. Create and run the setup.py script in CloudIDE to install the Requests module. Alternatively, run pip3 install requests on the CloudIDE Python CLI to install the Requests module.

```
import os

os.system("pip3 install requests")
```



HUAWEI

- To run the code, right-click setup.py and choose Run Python File in Terminal from the shortcut menu.

Procedure: Compile a Python Script

- Compile a Python script.

Configure the northbound user information and northbound address.

Define the URI of an interface.

Configure the URL.

Configure the request header.

Initiate a request, whose body is in JSON format.

Parse the returned data to obtain the token.

```
import requests

nbi_name = "campus02@north.com"
nbi_pwd = "5aAZ4lh44@"
host = "139.9.213.72"
port = "18002"

POST_TOKEN_URL = "/controller/v2/tokens"

post_token_url = "https://" + host + ":" + port + POST_TOKEN_URL
headers_post = {'Content-Type': 'application/json', 'Accept': 'application/json'}

r = requests.post(post_token_url, headers=headers_post, json={"userName":
nbi_name, "password": nbi_pwd}, verify=False)

token_id = r.json()['data']['token_id']
print("1.[Get Token Id]")
print("[post_token_url]:"+post_token_url)
print("[token_id]:"+token_id)
```


HUAWEI

Procedure: Verify the Configuration

- The token (specified by token_id) is obtained by invoking the RESTful API.

```
[Get Token Id]
[Post_token_url]: https://139.9.213.72:18002/controller/v2/tokens
[Token_id]: x-
rs5jvy7z47vsum042r2qc9vtaq44o9hgqohf3uca7z9clio5vvvz86al3s5dvs0aphpck5defw45uq
pi3xdgddmn482kgajutf7vlfs649rw6k5c6r6l0a5i1g4b5j9c
```


HUAWEI

Quiz

1. (Single-answer question) Which of the following HTTP request methods does not need to contain request data in an HTTP request? ( )

   A. POST

   B. GET

   C. PUT

   D. DELETE

HUAWEI

1. B

Summary

- SDN innovates the network architecture. It uses a controller to build a more open, flexible, and simplified network.
- REST is an architectural or design style rather than a standard. It provides only design principles and constraints.
- REST makes full use of HTTP. The core of REST is to perform operations (such as GET, POST, PUT, and DELETE) on resources in multiple forms.
- By invoking RESTful APIs, users can write Python scripts to implement automation as required.

HUAWEI

More Information

- REST dissertation
  - https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm
- RFC HTTP/1.1
  - https://www.ietf.org/rfc/rfc2616.html
- RFC HTTP/2
  - https://www.ietf.org/rfc/rfc7540.html
- RFC HTTPS
  - https://www.ietf.org/rfc/rfc2818.html

HUAWEI

Thank you.

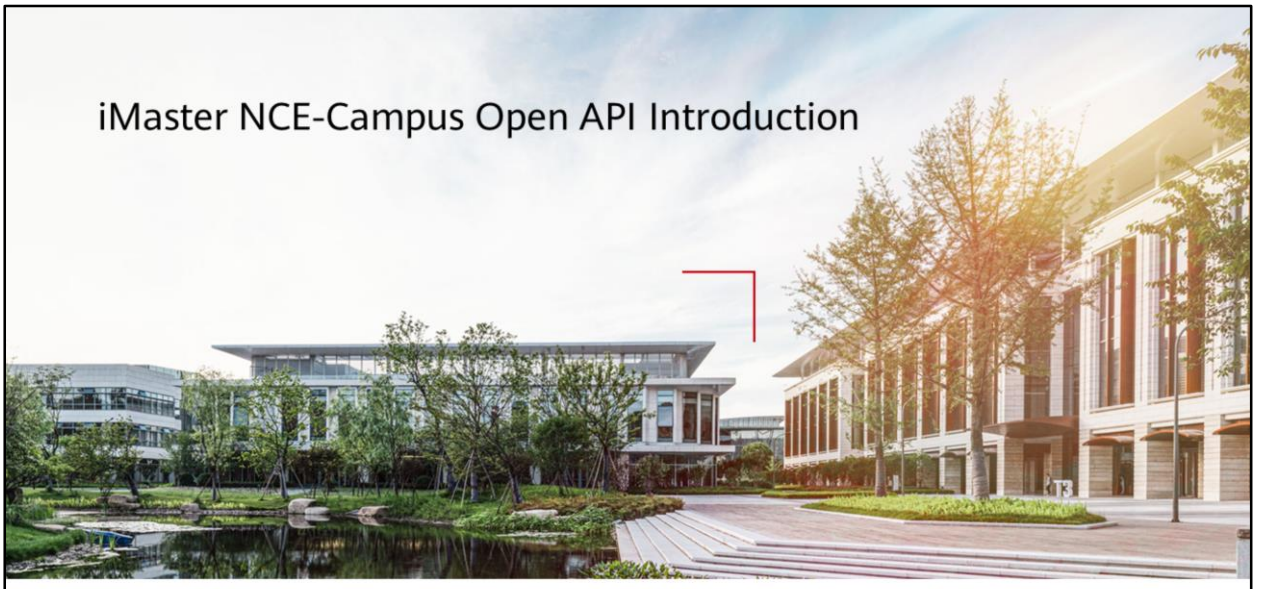
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

iMaster NCE-Campus Open API Introduction



HUAWEI

Foreword

- iMaster NCE-Campus is a next-generation SDN controller for enterprise campuses and is a core component of Huawei CloudCampus Solution.
- iMaster NCE-Campus centrally manages and controls enterprise campuses and provides automatic deployment, security defense, and visualized O&M. It helps customers build service-centric dynamic scheduling capabilities of network services.
- iMaster NCE-Campus provides standard-compliant open RESTful APIs for partners to develop innovative applications and continuously create value for enterprise customers.

HUAWEI

Objectives

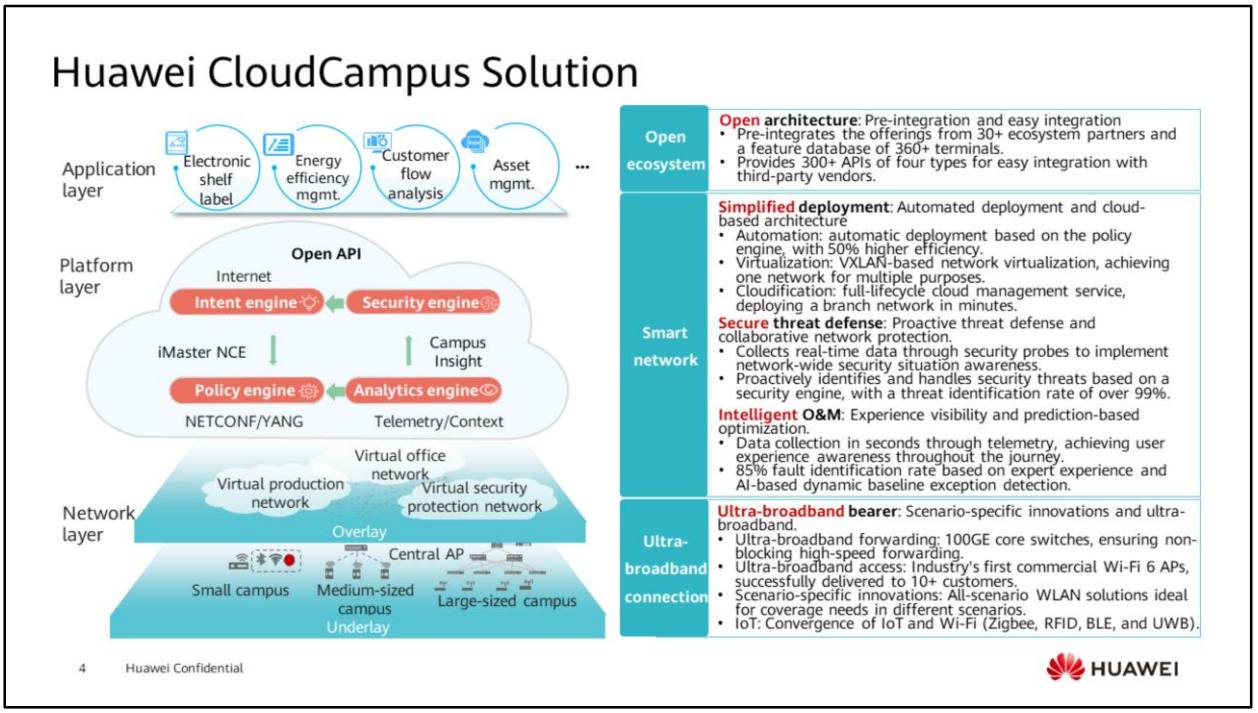
- On completion of this course, you will be able to:
  - Describe Huawei CloudCampus Solution.
  - Describe main open APIs of iMaster NCE-Campus.
  - Describe typical application scenarios of iMaster NCE-Campus.

HUAWEI

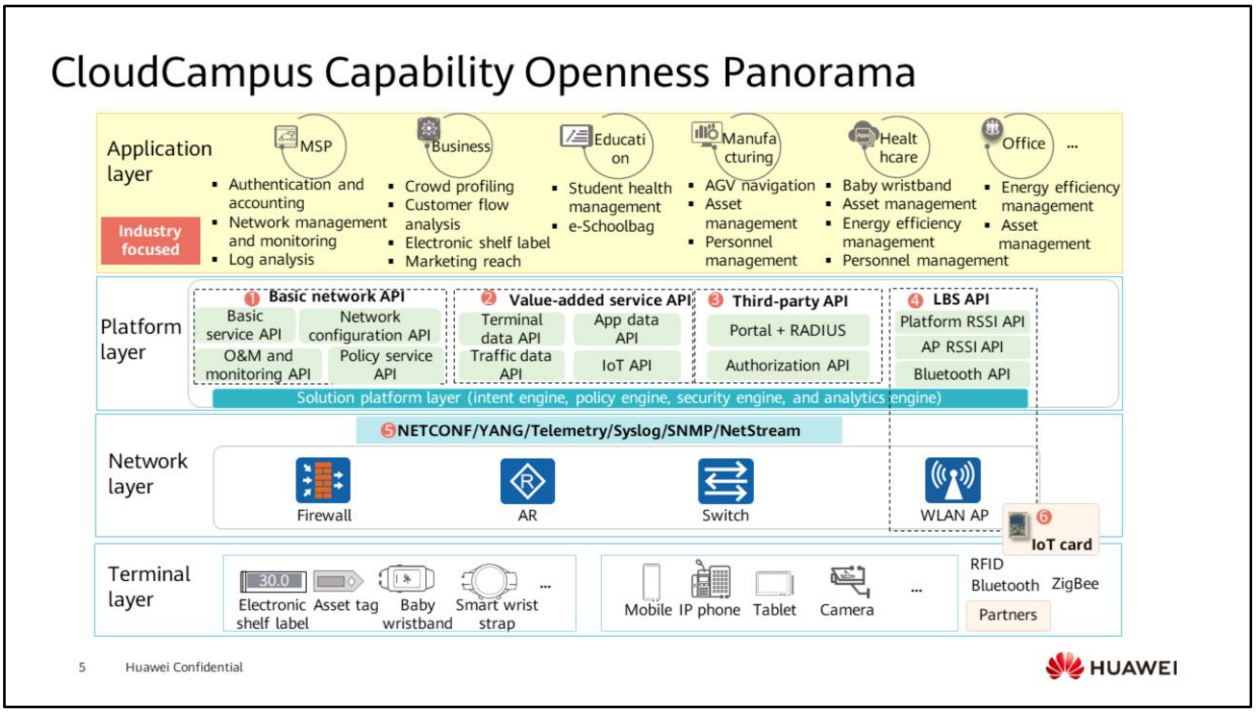
Contents

1. **Solution Overview**
2. Typical Scenarios and Open Capabilities
3. Openness and Cooperation Cases

HUAWEI

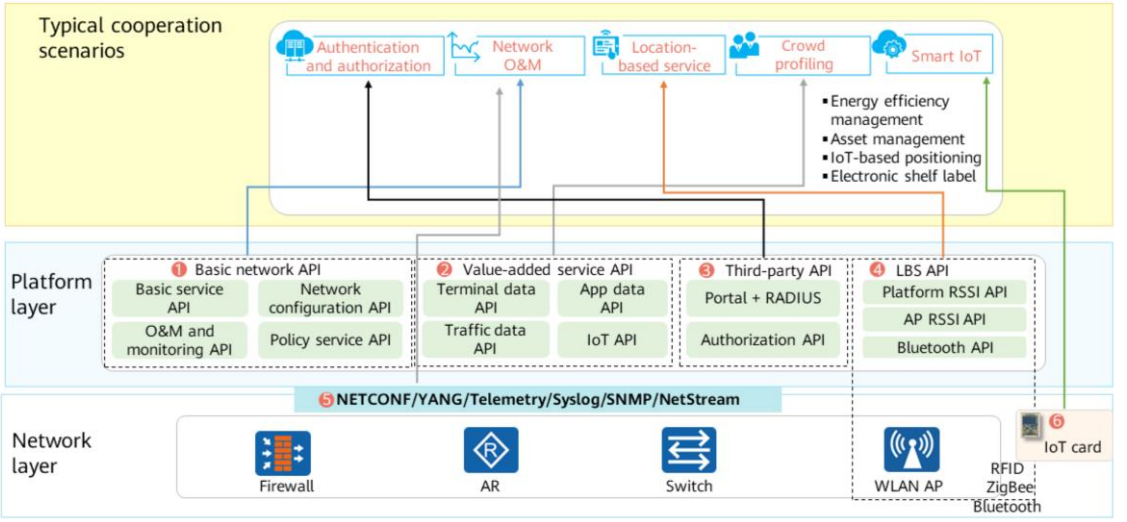


- iMaster NCE-Campus is Huawei's next-generation autonomous driving network management and control system for campus networks. It is a network automation and intelligent platform integrating manager, controller, analyzer, and AI functions. As such, this platform drives enterprise cloudification and digital transformation, and creates a shortcut to more automated network management and intelligent network O&M.



- Application layer: It focuses on mainstream scenarios and provides industry-wide applications. At this layer, network service data is shared, meeting value-added data and operation requirements.
- Platform layer: It provides four types of APIs and supports industry-standard network interconnection protocols.
- Network layer: It provides various open interfaces, such as NETCONF, YANG, and Telemetry, improving device manageability. APs are compatible with third-party IoT cards to implement IoT.
- Terminal layer: It supports access of IoT terminals (such as ZigBee, RFID, and BLE), and access of wired and wireless terminals (such as mobile phones, IP phones, tablets, and cameras).

Typical Cooperation Scenarios and Open APIs



HUAWEI

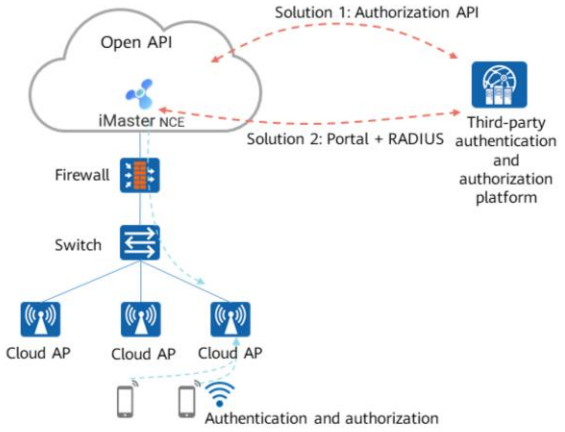
Contents

1. Solution Overview
2. **Typical Scenarios and Open Capabilities**
   - Authentication and Authorization
   - LBS
   - Crowd Profiling
   - Network O&M
   - Smart IoT
3. Openness and Cooperation Cases

HUAWEI

Introduction to Authentication and Authorization

- Authentication and authorization: A third-party authentication and authorization platform interworks with iMaster NCE-Campus to perform network access authentication and accounting for terminals that access Huawei intent-driven campus networks.

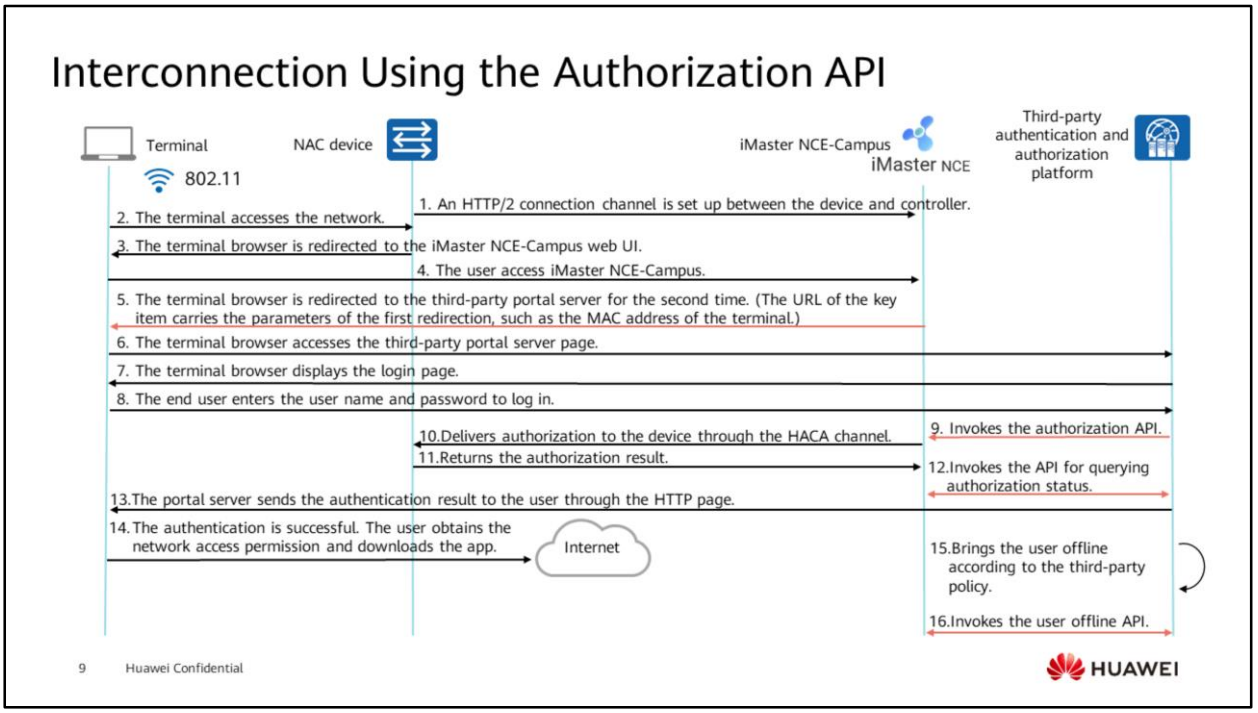


- iMaster NCE-Campus provides two integration solutions for authentication and authorization using the third-party authentication and authorization platform. Tenants can select a solution based on the site requirements.

Integration Solution	Suggestions
Portal + RADIUS	iMaster NCE-Campus functions as a RADIUS client to interact with the third-party authentication and authorization platform for user authentication and authorization. Accounting is recommended in this solution.
Authorization API	After users are authenticated, the third-party authentication and authorization platform invokes the authorization API to instruct iMaster NCE-Campus to authorize the users. Accounting is not recommended in this solution.

HUAWEI

- The tenant or MSP wants to use an existing or third-party authentication platform to authenticate user identities and authorize users for network access authentication through the web page (authentication portal). For example, an MSP provides a unified access authentication page for tenants.



- To access the Internet, a user connects to the SSID of a Wi-Fi network and logs in to the portal pushed by a developer app. The developer app calls the authorization API of Huawei iMaster NCE-Campus to deliver the user's Wi-Fi access permission to the AP. The user then can access the Internet.
- NAC is short for network access control.
- Huawei Agile Cloud Authentication (HACA) is based on the mobile Internet protocol HTTP/2.

Interconnection Using the Authorization API

Sample Code (Windows)

The API uses HTTPS and the port 18008.
This is a RESTful API, and **Content-Type** is **application/json**.
The following shows only a part of the API URL. For example, if the address of iMaster NCE-Campus is **https://10.186.40.148:18008/**, the complete URL of the authorization API is **https://10.186.40.148:18008/controller/cloud/v2/northbound/accessuser/haca/authorization**.

Request

```
POST
/controller/cloud/v2/northbound/accessuser/haca/authorization
{
   "deviceMac":"AA:AA:AA:00:00:89",
   "deviceEsn":"AA50082935AAAA000088",
   "ssid":"YWFh",
   "policyName":"dfdsffd",
   "terminalIpV4":"10.11.11.11",
   "terminalMac":"ff-ff-ff-ff-ff-f1",
   "username":"liao",
   "nodeIp":"10.186.40.148",
   "apMac":"AA:AA:AA:00:00:89"
}
```

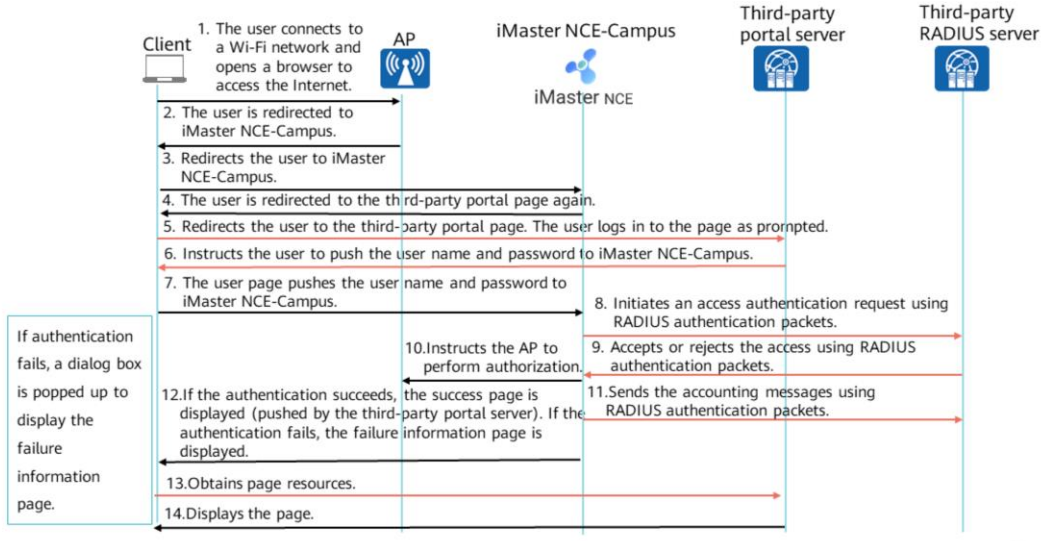
Response

```
STATUS CODE 200
{
  "errcode": "",
  "errmsg": "",
    "psessionid":
    "846d65a4e1bb38c350a280f46e6d2f6ab0a688df7edbb72b77bc5d03e3
    926cd6"
}
```


HUAWEI

- For more information about API-based authentication and authorization, visit https://devzone.huawei.com/cn/enterprise/campus/apiSolution.html.

Interconnection Using Portal and RADIUS



HUAWEI

- For details about RADIUS-based authentication, visit https://devzone.huawei.com/cn/enterprise/campus/radiusSolution.html.

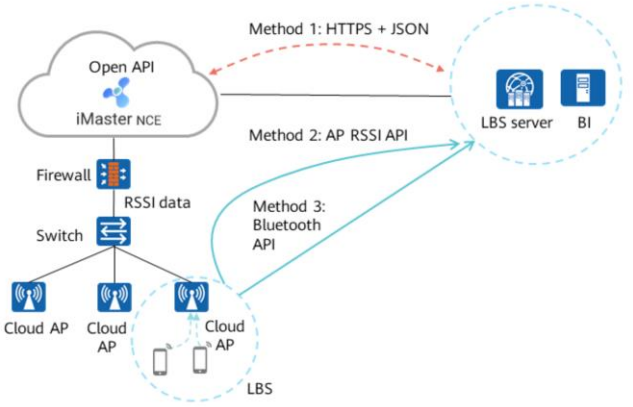
Contents

1. Solution Overview
2. **Typical Scenarios and Open Capabilities**
   - Authentication and Authorization
   - LBS
   - Crowd Profiling
   - Network O&M
   - Smart IoT
3. Openness and Cooperation Cases

HUAWEI

LBS Introduction

- A location-based service (LBS) is deployed on a third-party LBS application platform and used to detect and locate terminals managed on Huawei intent-driven campus networks. The LBS provides the customer flow analysis, Wi-Fi marketing, and navigation based on the terminal locations.



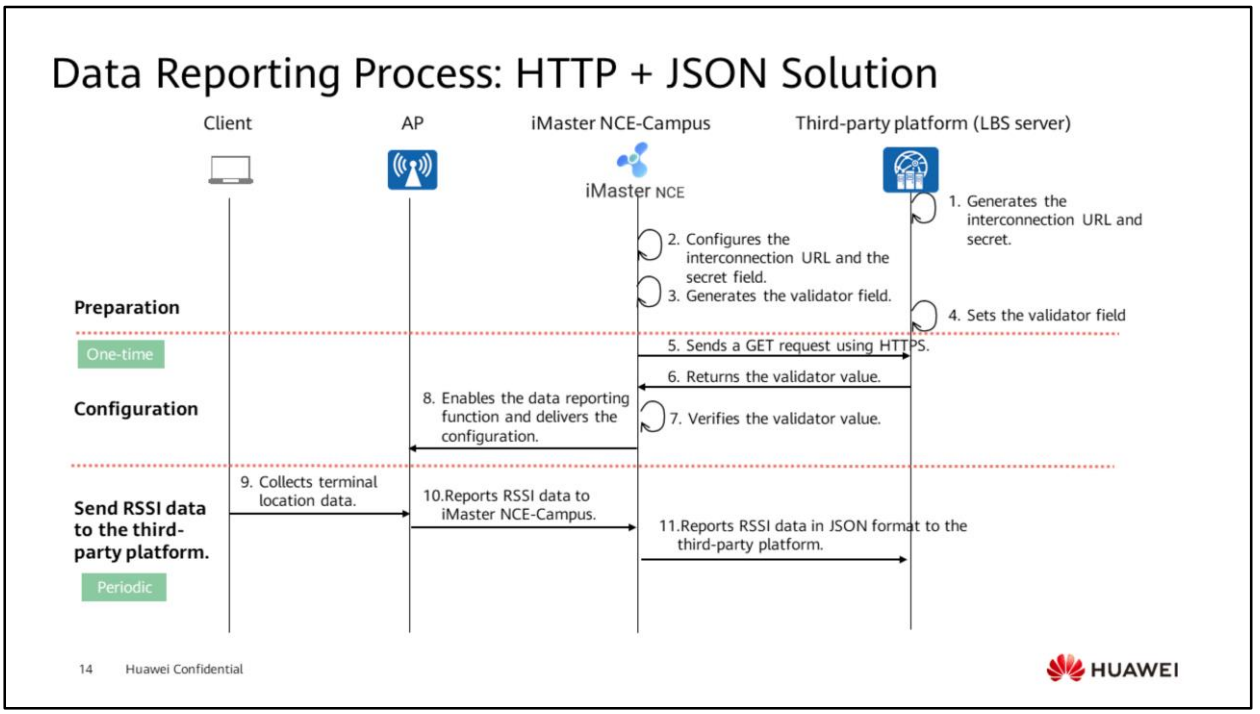
Application Scenario

- Huawei provides location data of Wi-Fi and Bluetooth terminals to a third-party partner. The partner parses the location data and provides VAS applications, such as heat map, track tracing, and customer flow analysis, for end customers.

Integration Solution	Suggestions
HTTPS + JSON	Applicable to the interconnection with an LBS system deployed based on software as a service (SaaS).
AP RSSI API	Applicable to the on-premise scenario.
Bluetooth API	Applicable to locating Bluetooth terminals and reporting terminal data.

HUAWEI

- Location-based service (LBS) uses various locating technologies to obtain the current locations of devices and pushes information and basic service for these devices through mobile Internet.
- iMaster NCE-Campus aggregates the terminal location data collected by cloud APs and periodically sends the data to the third-party LBS platform. After parsing and analyzing the location data with a series of algorithms, the LBS platform provides VASs, such as heatmap, tracking, and customer flow analysis, for customers.
- Remarks: Partners need to meet related standards based on application scenarios, such as EU General Data Protection Regulation (GDPR).



- iMaster NCE-Campus can directly report terminal location data to a third-party LBS platform. In this solution, iMaster NCE-Campus function as a relay agent.
- For details about this process, see "Wi-Fi Terminal Location Practice in Huawei CloudCampus Solution" in the HCIP-Datacom-NCE Northbound Openness Lab Guide.

Data Example: HTTP + JSON Solution

Sample Code (Windows)

```
{
    "data":[
        {
            "apMac":"4C:FA:CA:D8:23:A0",
            "terminallist":[
                {
                    "terminalMac":"88:19:08:F1:88:45",
                    "rssi":-68,
                    "timestamp":1557460789000
                },
                {
                    "terminalMac":"90:2E:1C:6A:2A:57",
                    "rssi":-57,
                    "timestamp":1557460789000
                }
                ]
        }
    ],
    "secret":"Test@1234",
    "type":"Aplocation"
}
```

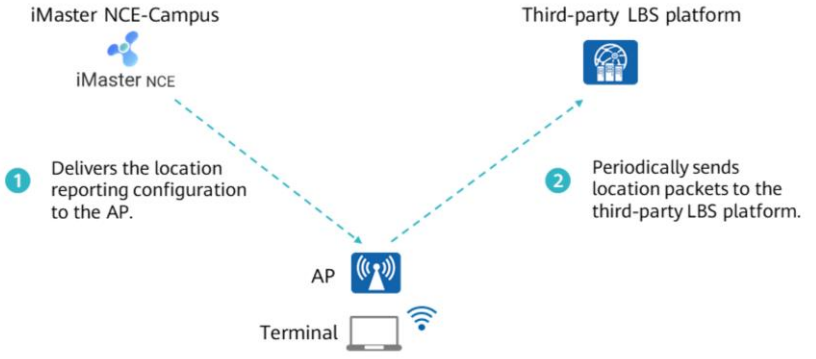
- secret: Used by a third-party LBS platform to verify the data sent by iMaster NCE-Campus. If the secret does not match that associated with the reported URL, the data is discarded. The value is generated by the third-party LBS platform, and is a string of 32 hexadecimal characters.
- validator: Before sending data, iMaster NCE-Campus checks whether the validator returned by the peer end is correct. If the validator does not match that associated with the reported URL, iMaster NCE-Campus does not send Wi-Fi terminal location data to the third-party LBS platform.

HUAWEI

- The validator value is in UUID format and is generated by iMaster NCE-Campus.
- For more examples, visit https://developer.huaweicloud.com/techfield/network.html#CloudCampus.

Data Reporting Process: AP RSSI API Solution

- iMaster NCE-Campus allows WLAN APs to report terminal location data directly to a third-party LBS platform. iMaster NCE-Campus delivers the location reporting configuration to APs, and then the APs periodically send location packets to the third-party LBS platform.

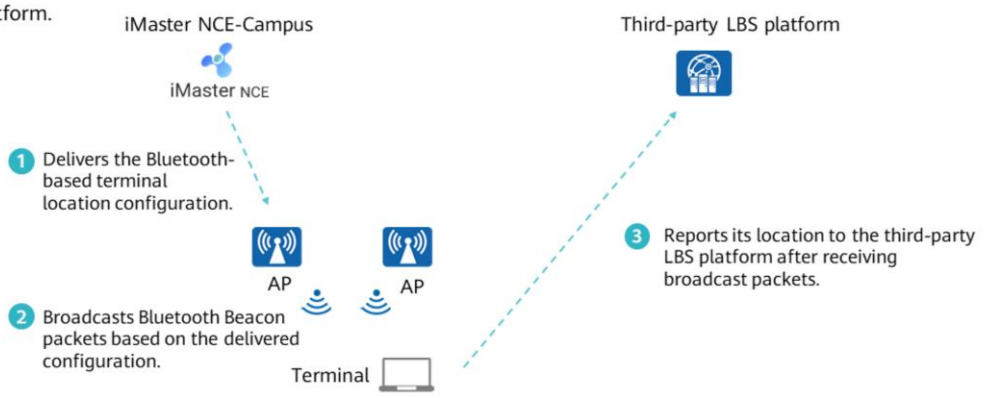


HUAWEI

- For more about the AP location reporting solution, see https://devzone.huawei.com/cn/enterprise/campus/lbsWiFiSolution.html#Wi-Fi Terminal Data Reporting Process.

Data Reporting Process: Bluetooth API Solution

- iMaster NCE-Campus allows WLAN APs to report terminal location data through Bluetooth Beacon packets. iMaster NCE-Campus delivers the Bluetooth-based terminal location configuration to APs. The APs then broadcast Bluetooth Beacon packets. After receiving the broadcast packets, terminals report their locations to the third-party LBS platform.



HUAWEI

- For more about the Bluetooth API solution, see https://devzone.huawei.com/cn/enterprise/campus/lbsBluetoothSolution.html.

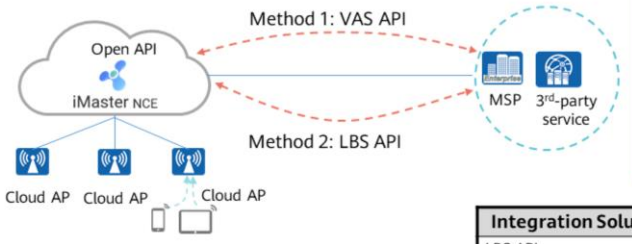
Contents

1. Solution Overview
2. **Typical Scenarios and Open Capabilities**
   - Authentication and Authorization
   - LBS
   - Crowd Profiling
   - Network O&M
   - Smart IoT
3. Openness and Cooperation Cases

HUAWEI

Crowd Profiling

- Crowd profiling: A third-party big data analysis platform analyzes the locations and status of terminals detected and managed on Huawei intent-driven campus networks to obtain the characteristics of crowd profiles in an area, so that target content can be pushed to the customers to provide personalized services.



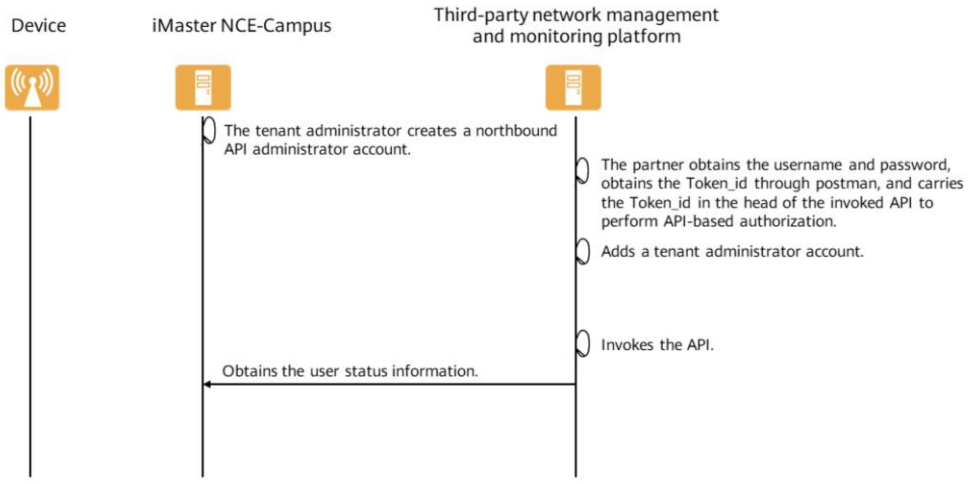
Application Scenario

- The business industry has the requirements for precision marketing. Direct marketing can tag specific people based on crowd profiles, and push target content to provide personalized services based on these tags. Crowd profiling requires two types of data: one is the data left online; the other type of data is left in brick-and-mortar stores (offline data), which can be collected using Wi-Fi probes and from network access data of customers.

Integration Solution	Suggestions
LBS API	Terminal location information is provided.
VAS API	Terminal login and logout information, network application information, and network traffic information are provided.

HUAWEI

Invoking a VAS API



HUAWEI

- For more about VAS APIs, visit https://devzone.huawei.com/cn/enterprise/campus/valueAddedApi.html.

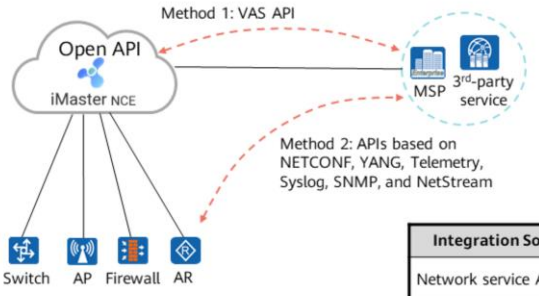
Contents

1. Solution Overview
2. **Typical Scenarios and Open Capabilities**
   - Authentication and Authorization
   - LBS
   - Crowd Profiling
   - Network O&M
   - Smart IoT
3. Openness and Cooperation Cases

HUAWEI

Network O&M Overview

- Network O&M: Third-party network management platforms (such as MSP-owned or mainstream network management and monitoring platforms) can manage or monitor devices managed by iMaster NCE-Campus.



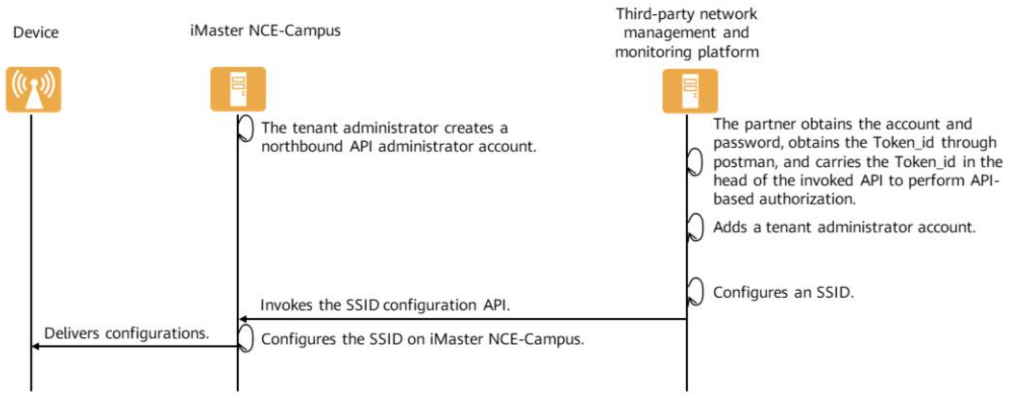
Application Scenario

- MSPs and customers use their existing network management platform to manage or monitor devices that are supposed to be managed by iMaster NCE-Campus, such as creating tenant administrator accounts, managing devices, configuring networks for specified devices, and monitoring device status and alarms.
- Currently, two methods are available to implement network O&M:
  - Network service APIs: The third-party network management platform manages and monitors devices through RESTful APIs.
  - Traditional device APIs based on NETCONF, YANG, Telemetry, and SNMP: Network management platforms can directly configure, manage, and maintain devices through these APIs.

Integration Solution	Suggestions
Network service API	Applicable to interconnection with various third-party private SaaS platforms.
APIs based on NETCONF, YANG, Telemetry, and SNMP	Applicable to interconnection with traditional network management platforms.

HUAWEI

Interconnecting with a Third-Party Network Management and Monitoring Platform



HUAWEI

- For more basic network solutions, visit https://devzone.huawei.com/cn/enterprise/cloudcampus/quickStart.html#network.

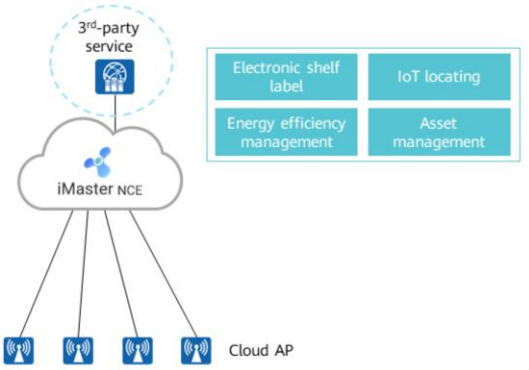
Contents

1. Solution Overview
2. **Typical Scenarios and Open Capabilities**
   - Authentication and Authorization
   - LBS
   - Crowd Profiling
   - Network O&M
   - Smart IoT
3. Openness and Cooperation Cases

HUAWEI

Introduction to Smart IoT

- Smart IoT: A third party expands an AP into an IoT base station through the open card capability of the AP to implement functions such as energy efficiency management, asset management, IoT locating, and electronic shelf label.



Application Scenario

- When partners want to deploy IoT applications (such as electronic shelf label, IoT locating, energy efficiency management, and asset management), they can use Huawei's network infrastructure to provide IoT signal coverage (ZigBee, Bluetooth, and RFID), without the need to deploy a secondary IoT network.

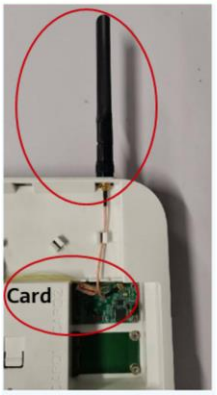
- Based on Huawei Cloud Campus Solution, partners can develop IoT cards and IoT service software on the cards to provide IoT services based on the network infrastructure.
  - Huawei provides the network infrastructure, open AP hardware, and basic IoT card management and monitoring.
  - Partners develop IoT cards, card software, and IoT service software.

HUAWEI

Smart IoT: Open Cards

Built-in card

IoT slot: allows IoT cards to be installed to provide extended functions.



AP4050DN-E

External card

MT800

HUAWEI

- For more about the smart IoT solution, visit https://devzone.huawei.com/cn/enterprise/cloudcampus/quickStart.html#iot.

Contents

1. Solution Overview
2. Typical Scenarios and Open Capabilities
3. **Openness and Cooperation Cases**

HUAWEI

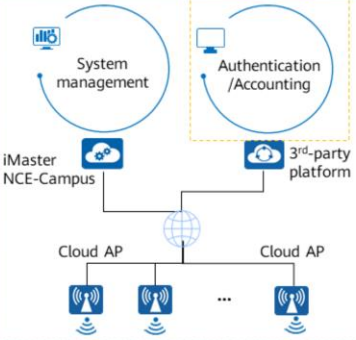
Unified Cloud-based Wi-Fi Management in XX

Background

- XX provides Wi-Fi services over more than 800 cities and has 1.5 million users.

Challenges

- How to ensure Wi-Fi signal quality and user access experience in diversified tenant environments.
- **O&M for massive devices.** A large number of tenants have devices scattered in difference places, resulting in high installation and O&M costs.



Solution Implementation

iMaster NCE-Campus implements plug-and-play and fast deployment of devices and provides full-lifecycle cloud-based and centralized management, simplifying O&M. It provides open APIs to seamlessly interconnect with third-party authentication and accounting management platforms, implementing secure network management.

Cooperation Scenario and API

Authentication and authorization: Portal + RADIUS API

Customer Benefits

- **Network coverage**: High-quality Wi-Fi coverage
- **Cloud-based mgmt**: 80% lower OPEX
- **Smooth upgrade**: Securer customer investment

HUAWEI

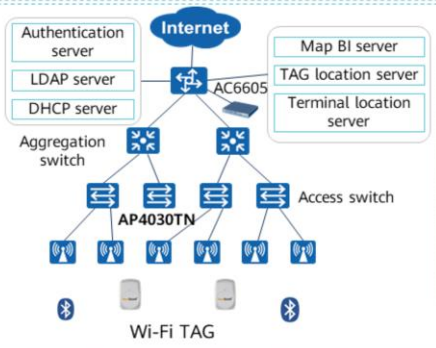
XX Builds the Future Business

Background

- XXX is a large city complex that integrates industry R&D, apartments, hotels, and businesses.
- It is a landmark building group in XXX. Wireless networks are an important infrastructure.

Challenges

- **Consumer experience**: The customer requires high-density, high-quality wireless network access to provide precise indoor navigation in more than 160 stores and 30,000 parking spaces.
- **Smart operations**: The customer wants to use big data analytics technology to describe consumer behavior and visualize staff locations in support of marketing and operational decision-making.



Solution Implementation

Three-RF APs have an independent radio to provide LBS services for Wi-Fi terminals and tags. Bluetooth beacons in the underground garage achieve a locating precision of 1 m to 3 m. Open APIs connect to app platform to implement customer flow analysis, indoor navigation, and precision marketing, laying a solid foundation for operational decision-making.

Cooperation Scenario and API

LBS: AP RSSI API

Customer Benefits

- **Network quality**: High-quality Wi-Fi network
- **Customer flow analysis**: Big data-based precision marketing
- **Indoor navigation**: Improved customers' shopping experience

HUAWEI

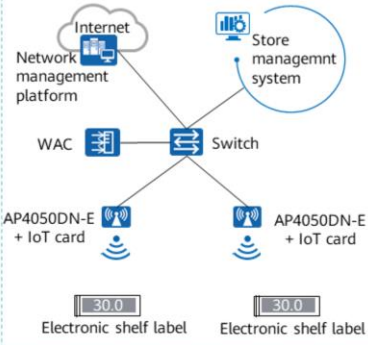
Digital Transformation of the XX Supermarket

Background

- XX is a well-known chain supermarket in China.
- One of China's top 500 enterprises and leading enterprise in China's national merchandising and agricultural industrialization

Challenges

- **Shopping experience**: The customer wants to provide high-quality wireless network access to improve customers' shopping experience.
- **Manpower**: The customer wants to reduce manpower and O&M costs, and improve the accuracy of supermarket services.



Dual-network convergence

Huawei AP4050DN-E provides built-in slots for IoT modules to converge Wi-Fi and IoT networks, achieving unified planning of the two networks.

Electronic shelf label

Electronic shelf labels interwork with the supermarket management system and customer's Enterprise Resource Planning (ERP) system to display prices in real-time and provide interactive functions such as code scanning, coupon claiming, and out-of-stock warning.

Wireless Internet service

Huawei AP4050DN-E helps build a high-quality wireless network for customers to scan QR codes for shopping or entertainment.

Cooperation Scenario and API

Smart IoT: IoT card

Customer Benefits

- **Converged mgmt.**: Converged high-quality Wi-Fi and IoT networks
- **Smart operations**: ESL-based smart operation supermarket
- **Wireless Internet access**: Improved customers' shopping experience

HUAWEI

Quiz

1. (Multiple-answer question) Which of the following layers are available for openness on Huawei intent-driven campus networks? ( )
   A. Application layer
   B. Platform layer
   C. Network layer
   D. Terminal layer

HUAWEI

1. ABCD

Summary

- This course describes the architecture, cooperation scenarios, and API packet structure involved in Huawei CloudCampus Solution to help readers understand the basic ecosystem openness capabilities of Huawei CloudCampus Solution and provide guides for interconnection between iMaster NCE and third-party applications based on these capabilities. It also helps MSPs, VAS application partners, and customers quickly implement system interconnection and service convergence, and helps customers and partners leverage network values.

HUAWEI

Thank you.

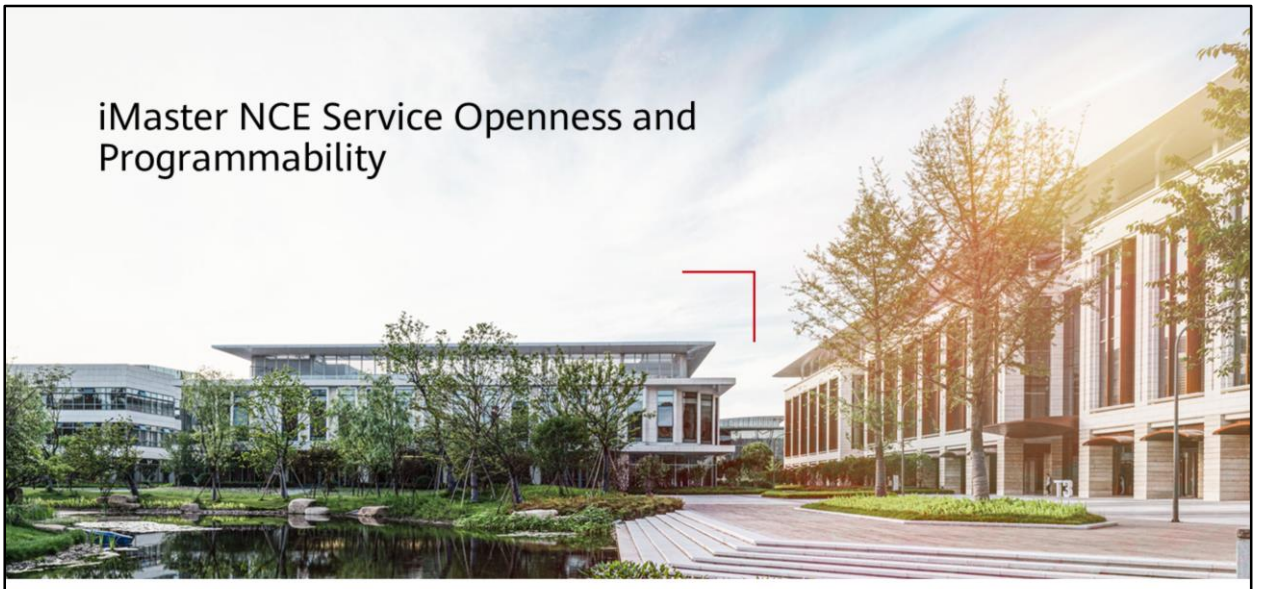
把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

iMaster NCE Service Openness and Programmability



HUAWEI

Foreword

- As digital network transformation moves forward, new services and technologies are burgeoning, and network requirements become more complex. Traditional closed networks cannot meet the requirements for rapid service innovation. It is an inevitable trend to improve network openness using existing assets. More agile planning of new services and network service orchestration based on customized customer intents are core capabilities.
- NCE service openness and programmability are based on the open architecture driven by the YANG model. Huawei enables openness and programmability at the NE and network layers in the form of Specific NE Driver (SND) and Specific Service Plugin (SSP) packages, implementing quick interconnection between new devices and fast rollout of new network services.
- This course introduces the background of service openness and programmability and Huawei's practices.

HUAWEI

Objectives

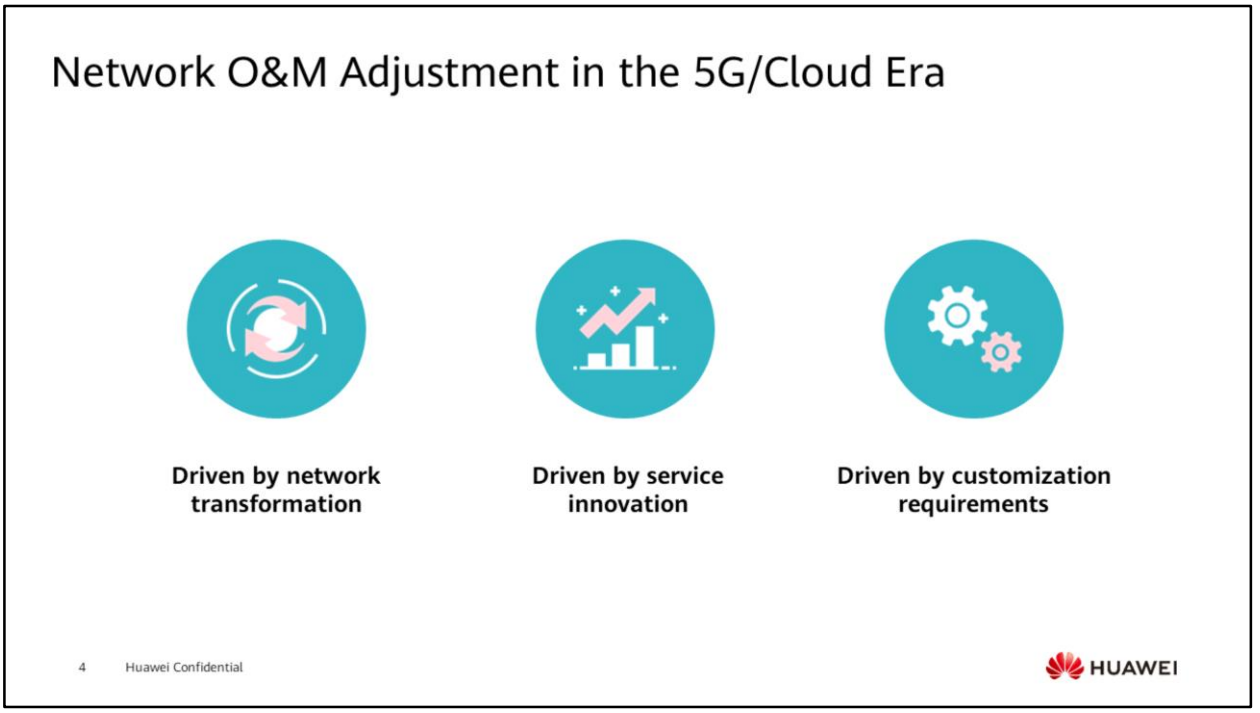
- On completion of this course, you will be able to:
  - Describe the basic concepts of Huawei NCE service openness and programmability.
  - Describe the open and programmable architecture and working principles of NCE services.
  - Describe the application scenarios of NCE service openness and programmability.
  - Use the iMaster NCE open programmability system (OPS) to interconnect new devices and roll out new services.

HUAWEI

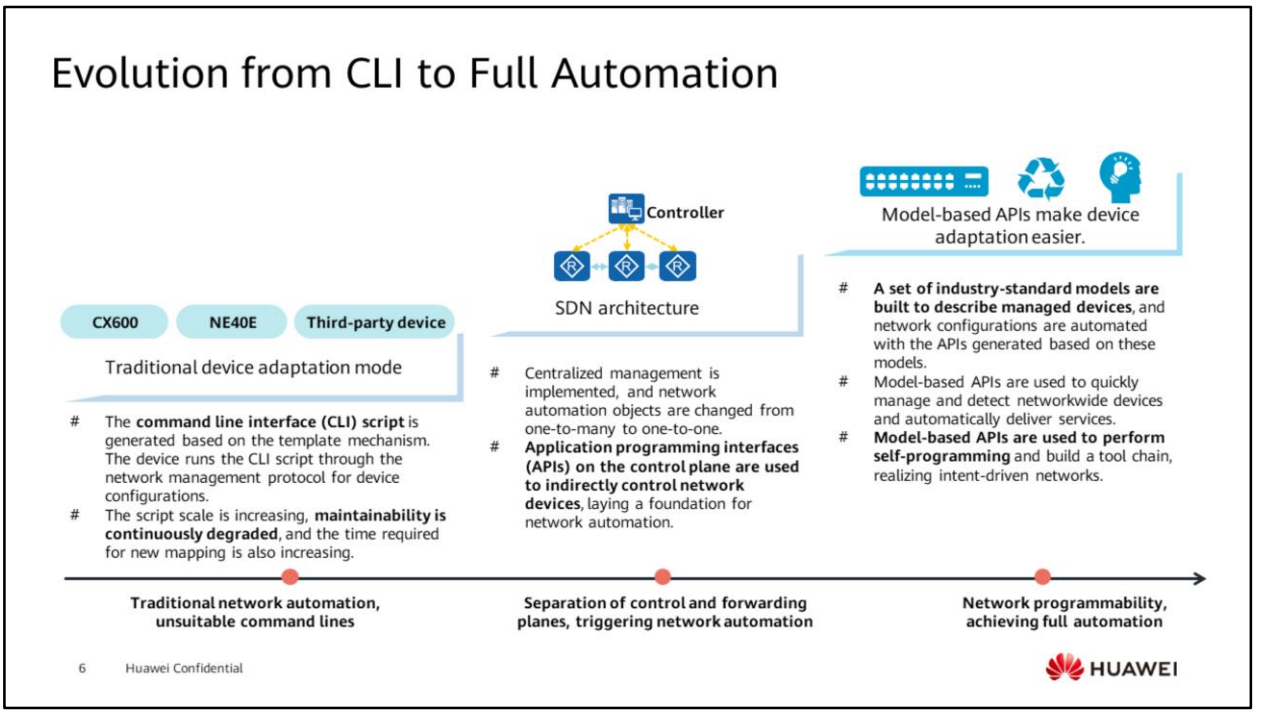
Contents

1. **Background**
2. Introduction to NCE Service Openness and Programmability
3. Key Capabilities
4. Related Concepts
5. Practice Cases

HUAWEI



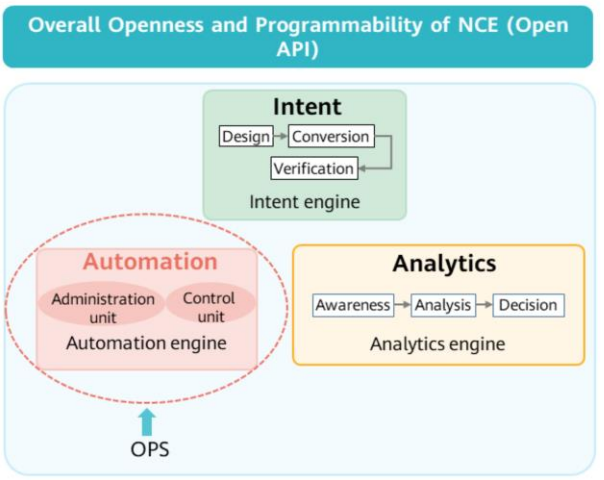
- First, let's take carriers as an example. Globally, most carriers face the problems of revenue decrease and OPEX increase. Moreover, as OTT providers continue to preempt market shares, more and more carriers take OTT providers as their competitors. These factors drive carriers to transform their networks. In this case, carriers are faced with the following problems: how to implement multi-network convergence, multi-vendor collaboration, and fast and efficient management of converged networks.
- In the 5G era, everyone predicts that 5G will lead to new businesses and services. However, carriers raise requirements for the rollout of new services, and device vendors implement the requirements. The rollout period is half a year or several years. It takes only a few months for OTT providers to launch new services, which makes it impossible for carriers and OTT providers to compete equally. There are many reasons for slow service rollout. One of the reasons is that there is a gap between carriers and vendors. That is, carriers do not understand devices, and vendors do not understand carrier services. It is an urgent issue to eliminate the impact of this gap and enable carriers and vendors to play their roles in the fields they are familiar with and quickly provision new services.
- Finally, the products provided by vendors are universal, that is, they are applicable to most operators. Carriers want systems to match their service requirements and enterprise cultures. Therefore, they have customization requirements. For example, a carrier writes the customization capability into its bidding document or customizes enterprise specifications. From the perspective of vendors, customization requirements of customers generate high costs. Therefore, the best solution is to provide the customization capability and let customers complete customization by themselves.



- On traditional networks, network automation refers to the process of generating command line scripts based on the template mechanism and enabling devices to run the received command line scripts through the network management protocol. It does not change the way it interacts with network devices. During device adaptation, network management engineers use Python or Perl to compile a specific function with a narrow application scope to implement a series of automatic operations, or use automation tools such as Ansible and Puppet to implement more complex automation tasks. Network management engineers need to adapt to network devices to be supported one by one, regardless of whether they write scripts or use automation tools. As the script scale becomes larger and larger, script maintainability decreases continuously, and the time required for adding a new version increases accordingly. With the advent of the Internet of Everything (IoE) era, the time to market (TTM) of new services has become a core indicator for enterprises to survive.
- With the great success of the commercialization of cloud computing, the concept of software-defined networking (SDN, sometimes referred to as “software-driven network”), which was popular only in the academic circle, has begun to flourish. On an SDN network, the separation between the control and forwarding planes is highly recommended. In an ideal SDN network, a centralized controller becomes an indispensable basis. As the brain of the entire network, it collects information about the network topology, calculates an optimal path globally based on service requirements, and notifies devices along the path. When receiving a service packet, these devices forward the packet according to a path determined by the centralized controller.

Huawei's NCE Service Openness and Programmability Practice

- iMaster NCE (NCE for short) is an innovative network cloudification engine of Huawei. The overall openness and programmability of NCE include automation, analytics, and intent. The goal is to build an open programmable architecture throughout the lifecycle to satisfy customers' requirements for openness and programmability.
- **Huawei open programmability system (OPS)**, a closed-loop component of the **automation engine**, are crucial to the entire open programming system of iMaster NCE.



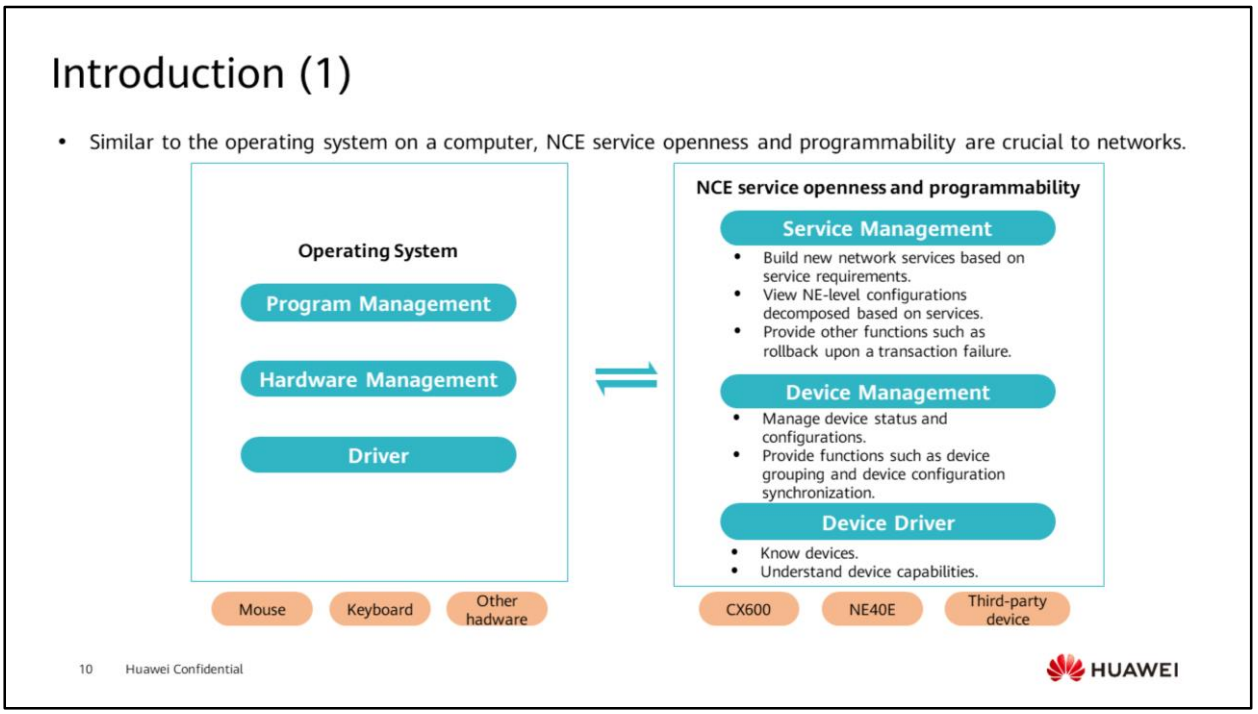
HUAWEI

- iMaster NCE is an innovative network cloudification engine developed by Huawei. Positioned as the brain of future cloud-based networks, NCE integrates functions such as network management, service control, and network analysis. It is the core enablement system for network resource pooling, network connection automation, and O&M automation. NCE aims to build an intent-driven network (IDN) that is first automated, then self-adapting, and finally autonomous.
- The overall openness and programmability of NCE include automation, analytics, and intent. The goal is to build a full-lifecycle open and programmable architecture to satisfy customer needs. The OPS, as a part of the automation engine, are crucial for the entire open programming system of NCE to form a closed loop. Equivalent to the limbs of the human body, the OPS is an executor, which needs to be flexible to support the automatic closed-loop capability driven by the brain of an intent-driven network.

Contents

1. Background
2. **Introduction to NCE Service Openness and Programmability**
3. Key Capabilities
4. Related Concepts
5. Practice Cases

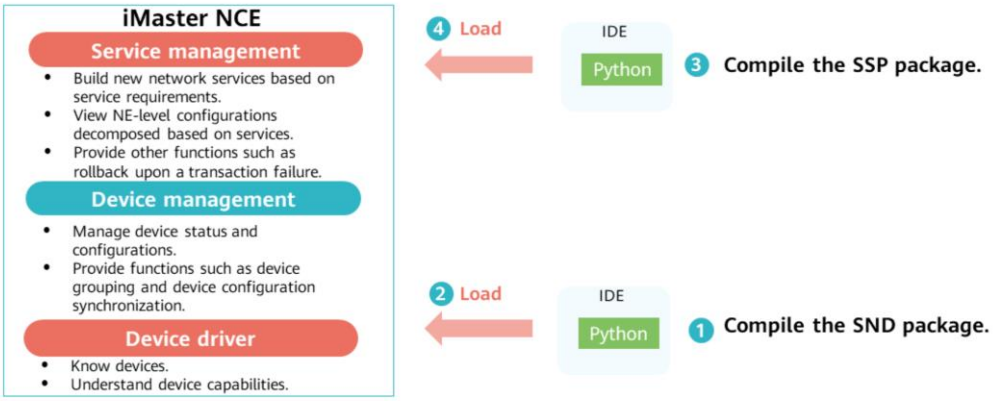
HUAWEI



- The open architectures of different industries are similar. Similar to the operating system on a computer, NCE service openness and programmability are crucial to networks.
- To connect the operating system to managed hardware, such as the mouse and keyboard, you need to install corresponding drivers. The drivers enable the operating system to recognize the hardware. NCE service openness and programmability have similar functions. The difference is that switches and routers are managed in the datacom industry. First, we need to understand and manage these switches and routers. That is, load device drivers first, and then add and understand the specific capabilities of the devices.
- At the upper layer, the operating system implements hardware management, specifically, managing the status of the mouse and keyboard. NCE service openness and programmability implement device management, such as managing device status and configurations.
- At the top layer, the operating system provides program management to manage various applications, such as Word and Excel. Note that the mouse and keyboard capabilities are required for using these programs. NCE service openness and programmability implement service management at the top layer, that is, building network service capabilities based on application scenarios. In addition, NCE provides capabilities such as rollback up on a transaction failure and automatic detection of device configuration changes to improve O&M.

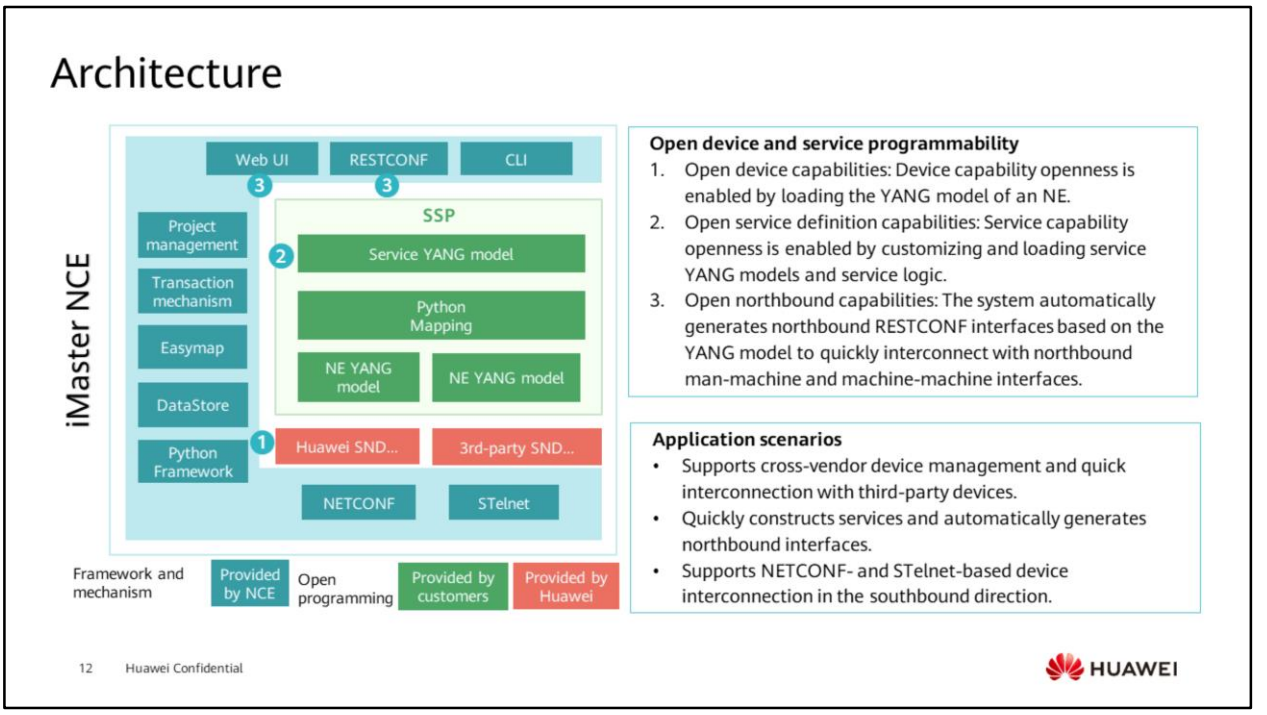
Introduction (2)

- NCE service openness and programmability are implemented based on two key software packages: Specific NE Driver (SND) and Specific Service Plugin (SSP). By compiling and loading two driver packages, you can quickly interconnect new devices and build new services.



HUAWEI

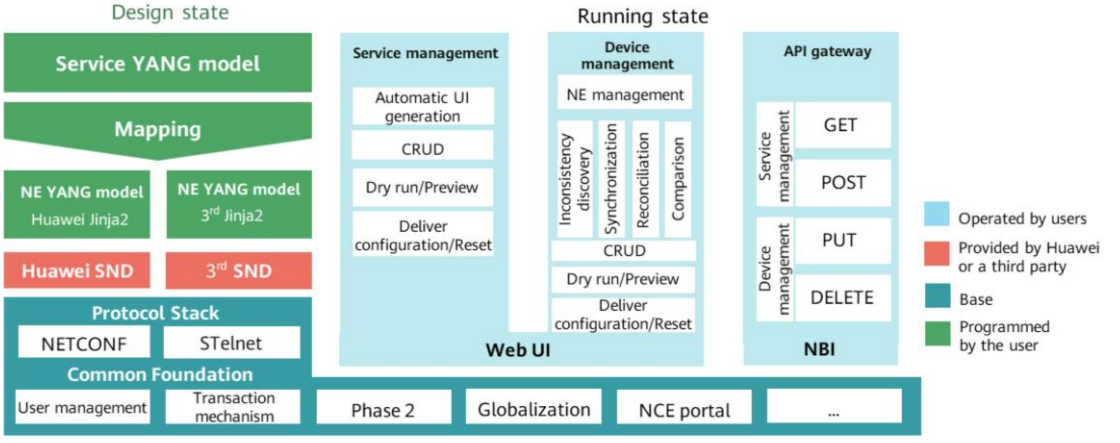
- NCE service openness and programmability depend on two software packages: SND and SSP.
  - Specific NE Driver (SND): provides a data model for the iMaster NCE OPS to interact with NEs.
  - Specific Service Plugin (SSP): defines a data model for completing network service configuration.
- Engineers compile SND packages and load them to iMaster NCE to quickly interconnect with new devices. Then, engineers compile SSP packages and load them to iMaster NCE to quickly construct new services.



- NE YANG model: YANG files generated by abstracting atomic capabilities (such as creating sub-interfaces) at the device layer. They are provided by device vendors.
- Service YANG model: YANG files generated by abstracting service models can be used to generate northbound interfaces and configuration GUIs.
- Easymap: a mapping logic algorithm that decomposes network-layer services into NE-layer services.

Principles

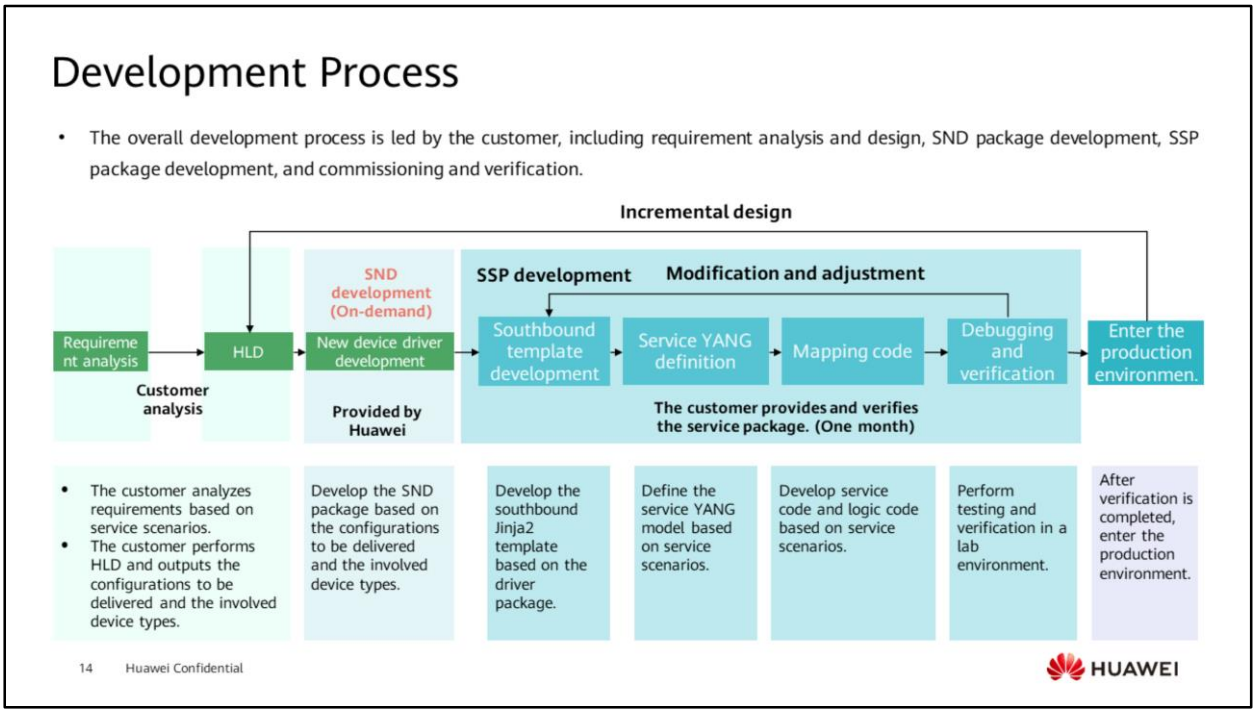
- NCE service openness and programmability consist of the design state and running state. The design state is used to design and develop service logic. The running state automatically generates service and device management GUIs and northbound interfaces based on the model defined in the design state. Users can configure devices and deliver services through the GUIs and northbound interfaces.



HUAWEI

- The design state is used to establish the mapping between the service YANG model and NE YANG model. The system provides the mapping logic algorithm to decompose network-layer services into NE-layer services. Currently, the NCE service openness and programmability framework supports two layers of mapping logic: 1. Mapping from the service model to the device model, which is processed by the SSP package. 2. Mapping from the device model to protocol packets, which is processed by the SND package.
- The running state uses the mappings established in the design state to manage devices and provision services. Specifically:
  - Service management automatically generates a service management GUI based on the service YANG model to add, delete, modify, and query services.
  - Device management automatically generates an NE management GUI based on the NE YANG model to add, delete, modify, and query NE resources, achieving functions such as difference comparison, data synchronization, and configuration reconciliation.
  - API gateway automatically generates northbound RESTCONF interfaces based on the service and device YANG models and works with the mapping between the two models to add, delete, modify, and query services and NE resources.
- The running state provides the dryRun function to help users preview the results of the current operation and the modification of related device configurations.
- Jinja2 is a Python template engine. NCE service openness and programmability use

Jinja2 to quickly complete the template-based processing of SSP packages.



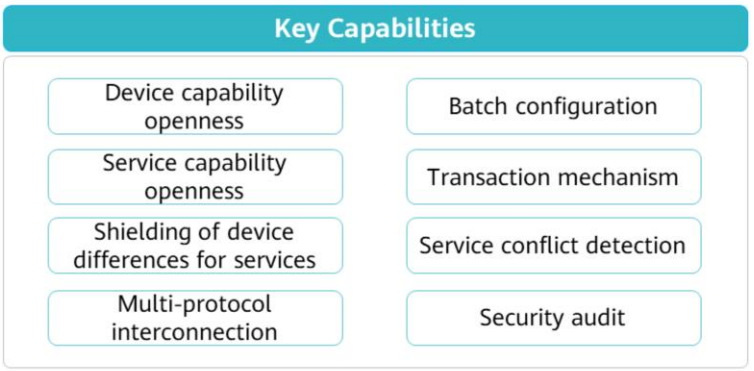
- The development process of NCE service openness and programmability is as follows:
  - First, analyze requirements based on service scenarios and output the high level design (HLD). In this phase, analyze the configuration commands to be delivered and the involved device types, and then start the development of a Specific NE Driver (SND) package. The SND package is developed as required. If the SND package of a device exists and the SND package to be delivered is supported, you do not need to develop the SND package again.
  - Then, develop a Specific Service Plugin (SSP) package. Step 1: Develop the southbound Jinja2 template. The southbound Jinja2 template can be considered as the tailoring of the open interfaces of the device. There are many open capabilities of devices. However, we only need to use some of them. Therefore, find and select the required ones. Step 2: Define the service YANG model and determine northbound input parameters. Step 3: Develop the service logic. This step is optional. If the service layer can directly map and use the southbound template, skip this step.
  - Finally, perform commissioning and verification. After the commissioning and verification are completed, the use in the production environment is formally started. If there are incremental requirements, follow the incremental design process and perform incremental development with reference to the preceding steps.

Contents

1. Background
2. Introduction to NCE Service Openness and Programmability
3. **Key Capabilities**
4. Related Concepts
5. Practice Cases

HUAWEI

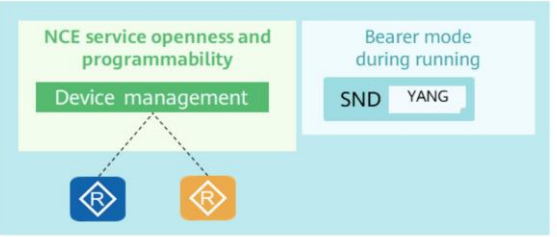
Overview of Key Capabilities



HUAWEI

Device Capability Openness

- Open atomic capabilities of devices. Driven by the device YANG model, the iMaster NCE OPS automatically generates northbound interfaces and configuration GUIs to quickly manage Huawei and third-party devices. It supports device data consistency comparison, configuration reconciliation, and configuration synchronization.



1. Quickly manage Huawei and third-party devices.
2. Open device configuration capabilities.

HUAWEI

Service Capability Openness

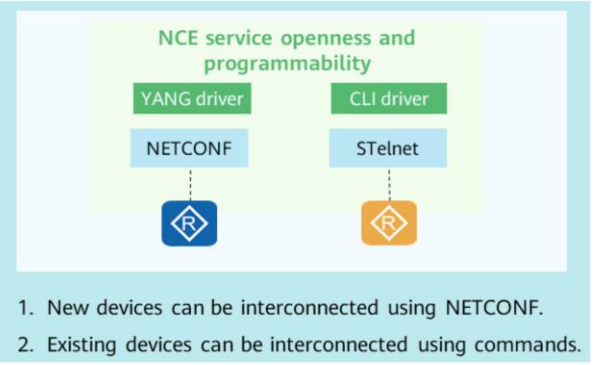
- Open service configuration mapping capabilities. Driven by the service YANG model, the iMaster NCE OPS automatically generates northbound and southbound interfaces to quickly build new services. The Easymap algorithm in the mechanism is used. Users only need to write the creation process. The update and deletion processes are calculated by the algorithm, thereby simplifying programming and reducing development difficulties.

HUAWEI

- Easymap: a mapping logic algorithm that decomposes network-layer services into NE-layer services. Currently, the NCE service openness and programmability framework supports two layers of mapping logic: 1. Mapping from the service model to the device model, which is processed by the SSP package. 2. Mapping from the device model to protocol packets, which is processed by the SND package.

Multi-Protocol Interconnection

- Support NETCONF- and STelnet-based device interconnection in the southbound direction. New devices can be interconnected using NETCONF, and existing devices can be interconnected using commands.

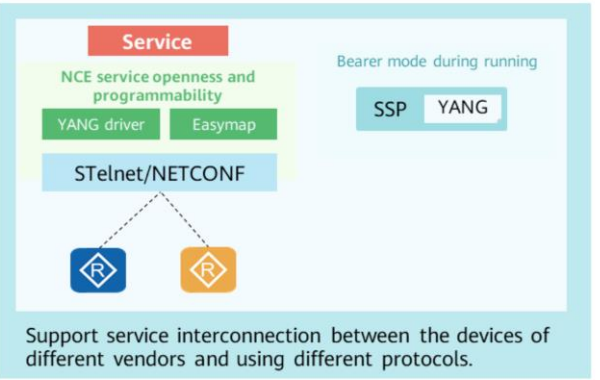


HUAWEI

- The SND package of the CLI driver is also a YANG file.

Shielding of Differences Between Underlying Devices at the Service Layer

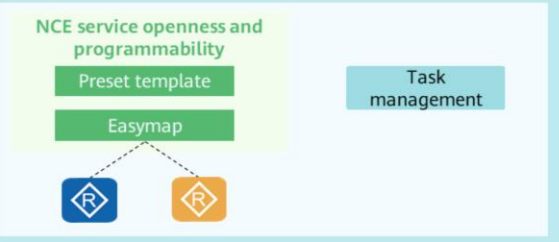
- Only specific services need to be focused on at the service layer. Devices of different vendors and using different protocols do not need to be interconnected separately. The OPS supports service interconnection between the devices of different vendors and using different protocols.



HUAWEI

Batch Configuration

- Provide batch configuration capabilities. Devices with the same configuration can be grouped. A template is preset in the system first, and then configurations are delivered in batches.
- Templates are applied in batches by device group to improve efficiency.



1. Support device grouping.
2. Preset templates, which are delivered in batches by device group.
3. Deliver asynchronous tasks in batches.

HUAWEI

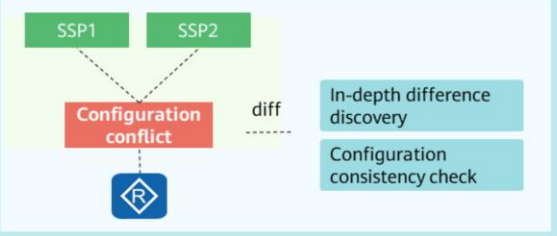
Transaction Mechanism

- Provide a transaction mechanism to ensure data consistency between the device and controller in both success and failure scenarios. The dryRun function enables the customer to check the correctness of configurations before delivery. The configurations can be repeatedly modified, and then the dryRun function is implemented until the configurations are correct. Then, the Commit operation can be started to commit the configurations to the device.

1. The transaction mechanism ensures data consistency between the device and controller.
2. Perform dryRun to check the correctness of configurations before delivery.
3. Perform the Commit operation to commit the configurations to the device. If the Commit operation fails, a rollback based on the transaction occurs automatically.

Service Conflict Detection

- Provide service conflict detection capabilities. The OPS supports source tracing of configured data sources. Configuration conflicts between different services can be detected before delivery. If the configurations of a device are modified by a third party, the device can automatically detect the modifications and re-deploy the configurations to restore the services interrupted by the modifications by the third party.
- Implement in-depth difference discovery. The controller directly compares the NE-level data decomposed by services with device data to check data consistency.

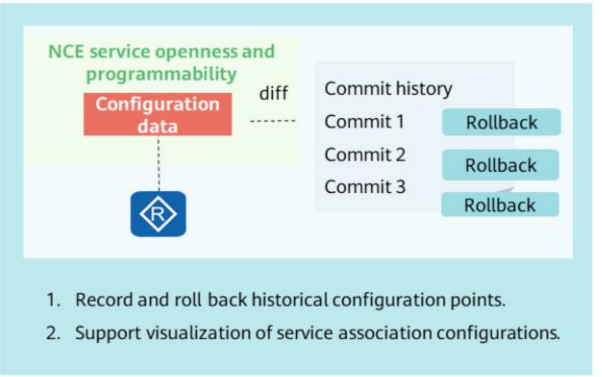


1. Perform data source tracing and conflict detection.
2. Support service and device data consistency check.
3. Automatically detect device configuration modifications by a third party and restore services.

HUAWEI

Security Audit

- Provide security audit capabilities. Historical configuration points can be recorded and rolled back, and service association configurations can be visualized.



HUAWEI

Contents

1. Background
2. Introduction to NCE Service Openness and Programmability
3. Key Capabilities
4. **Related Concepts**
5. Practice Cases

HUAWEI

Function Package

Specific NE Driver (SND)

A type of software package that provides a data model for interaction between the OPS and NEs.
The data model usually contains a **.py** file and a data model (YANG) of several features. Specifically:

- The former is used to define information related to an NE, such as device type, vendor, and connection information.
- The latter describes the data structure of NE-related features.

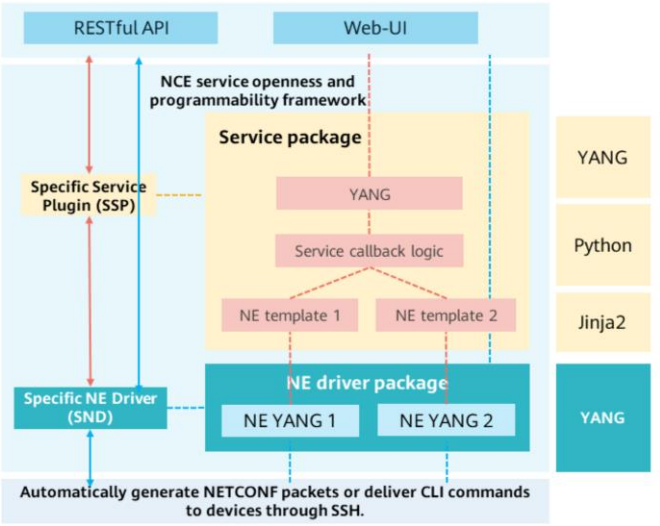
Specific Service Plugin (SSP)

A type of software package that defines a data model for completing network-level service configurations. The data model usually contains a Jinja2 template file, a Python mapping script, and a service YANG model. Specifically:

- The Jinja2 template describes the data structure of services and uses the Jinja2 syntax to perform operations such as interpolation, condition judgment, and recursion.
- The Python mapping script describes how to fill the data submitted by users into the template and map the data to the NE data structure.
- The service YANG model describes service parameters and is constructed based on service input.

HUAWEI

Package Mapping Mechanism



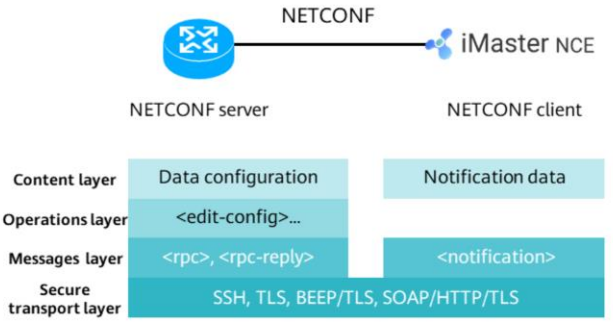
Packet mapping supports two layers of mapping logic: During the mapping from the service model to the device model, the logic is processed by the SSP package. During the mapping from the device model to protocol packets, the logic is processed by the SND package.

- NCE automatically generates northbound interfaces or configuration GUIs based on service models.
- Users send configuration requests to the service logic compiled by users through the interface provided by the service model.
- The service processing consists of two parts:
  - Python code processing, which processes service logic. For example, allocate resources and read NE or service information.
  - Jinja2 template processing, which processes vendor-related logic. A template is actually the data delivered to a device model. Devices from different vendors have different templates.

HUAWEI

- Currently, the NCE service openness and programmability framework supports two layers of mapping logic: 1. Mapping from the service model to the device model, which is processed by the SSP package. 2. Mapping from the device model to protocol packets, which is processed by the SND package.
- For SND package processing, if the device is a NETCONF device, NCE service openness and programmability automatically convert the model data into NETCONF packets.

NETCONF



NETCONF is a protocol defined by the IETF for installing, maintaining, and deleting configuration data on NEs. All NETCONF operations are implemented using XML-encoded RPCs. NCE service openness and programmability use NETCONF to communicate with NEs.

- Content layer: It needs to transmit configuration data and notification data.
- Operations layer: It defines an XML-encoded operation set for implementing RPC operations.
- Messages layer: It implements a simple framing mechanism based on the requirements of the transport protocol to encapsulate RPC messages or notification messages.
- Secure transport layer: It provides a communication path between the client and server. The NETCONF message hierarchy mechanism is implemented by using the communication paths that meet related conditions.

HUAWEI

- For more information about NETCONF, see NETCONF/YANG Principles and Practices.

YANG Model

- **YANG**, short for Yet Another Next Generation, is a standard-based and extensible data modeling language. It can be used to model the configuration and running status data of network devices, remote procedure calls, and server event notifications. YANG was originally designed to model network management data and provide a standardized content layer for the NETCONF model.

Module

A module is a collection of related definitions. Each module contains the module header, revision, and description statements.

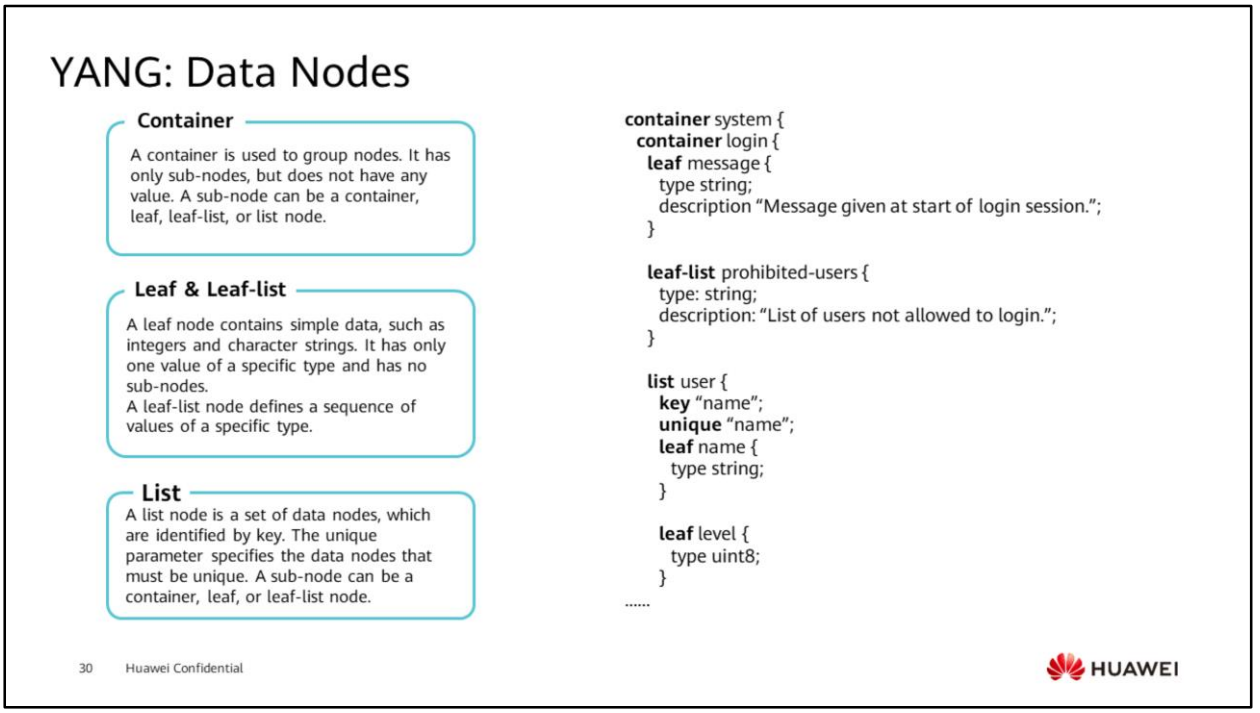
In a module, you can use the include statement to import submodules or use the import statement to import other modules. You can also use the augment statement in a module to extend a specified module.

A module may include the following four main types of data nodes, including Leaf, Leaf-list, Container, and List nodes.

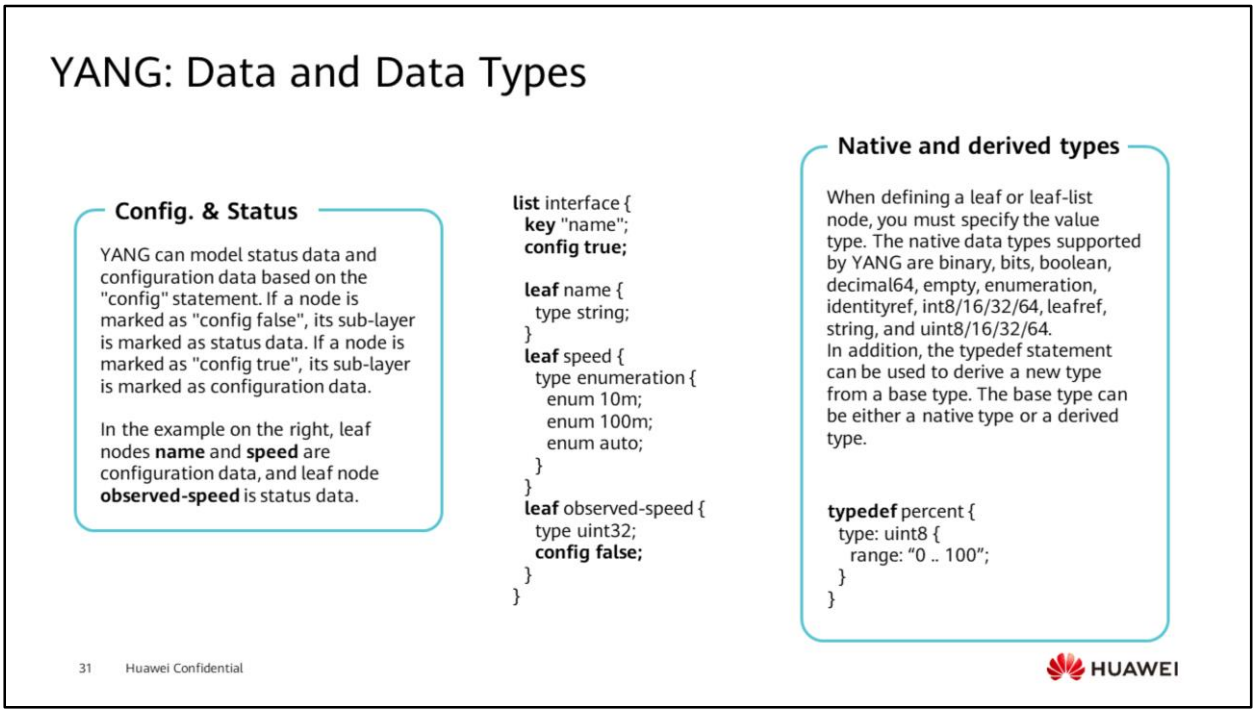
```
module hbng {
  namespace http://www.huawei.com/hbng;
  prefix hbng;

  import ietf-inet-types {
    prefix inet;
  }

  include foo-bar-types;

  description "This module can configure and manage the HBNGs (Huawei NE40-
X8)";

  revision 2018-04-20 {
    description "Initial revision";
  }

  augment "/app:applications" { ... }
}
```


HUAWEI

- In this example, the service YANG module hbng is customized.
- description describes the functions of the hbng module.
- revision is 2018-04-20, indicating the initial version of the hbng module.
- import and include introduce two modules for subsequent node definition.
- augment "/app:applications" { ... } indicates that the current module hbng is extended to the /app:applications directory of the app module.



- In this example, a container node named **system** is created, including the **login** container sub-node for recording login information.
- The container sub-node **login** contains the following:
  - A leaf node named **message**, which records the login prompt information.
  - A leaf-list node named **prohibited-users**, which records the blacklist of users who are not allowed to log in to the system.
  - A list node named **user**. In the list node, the unique key is defined as name and its type is character string; **level** is defined as user level and its type is number.



- In this example, the list interface is defined. **config true** indicates that the list is configuration data, and **config** false in **observed-speed** indicates that this leaf is status data.
- The leaf node **name** is a character string. The leaf node **speed** provides three options. **type** enumeration indicates that the enumerated values are **10m**, **100m**, and **auto**. The leaf node **observed-speed** is a positive integer of the **uint32** type.

YANG: Reuse & Selection

Reusable node group (grouping)

If a group of nodes can be reused in different modules or nodes, you can define this group of nodes as a reusable node group and use the **uses** keyword to reference modules or nodes.

Choices

YANG allows data models to use the choice and case statements to define a set of pattern nodes that cannot appear together.

```
grouping ip-port {
  leaf ip {
    type inet:ip-address;
    description "ip address";
  }

  leaf port {
    type inet:port-number;
    description "port number";
  }
}

container quadruple {
  container source {
    uses ip-port;
  }

  container destination {
    uses ip-port;
  }
}
```

```
container transfer-protocol {
  choice name {
    case a {
      leaf udp {
        type empty;
      }
    }

    case b {
      leaf tcp {
        type empty;
      }
    }
  }
}
```


HUAWEI

- In this example, a group node named **ip-port** is defined, including two leaf sub-nodes: **ip** and **port**.
- The container **quadruple** contains the source and destination information containers, both of which use the IP address and port information. The group node ip-port is reused.
- The container **transfer-protocol** is used to indicate the transmission protocol. The UDP and TCP protocols are provided. Either of them can be selected using the **choice** function. **case a** indicates that the UDP protocol is used, and **case b** indicates that the TCP protocol is used.

YANG: Operations

RPCs & Actions

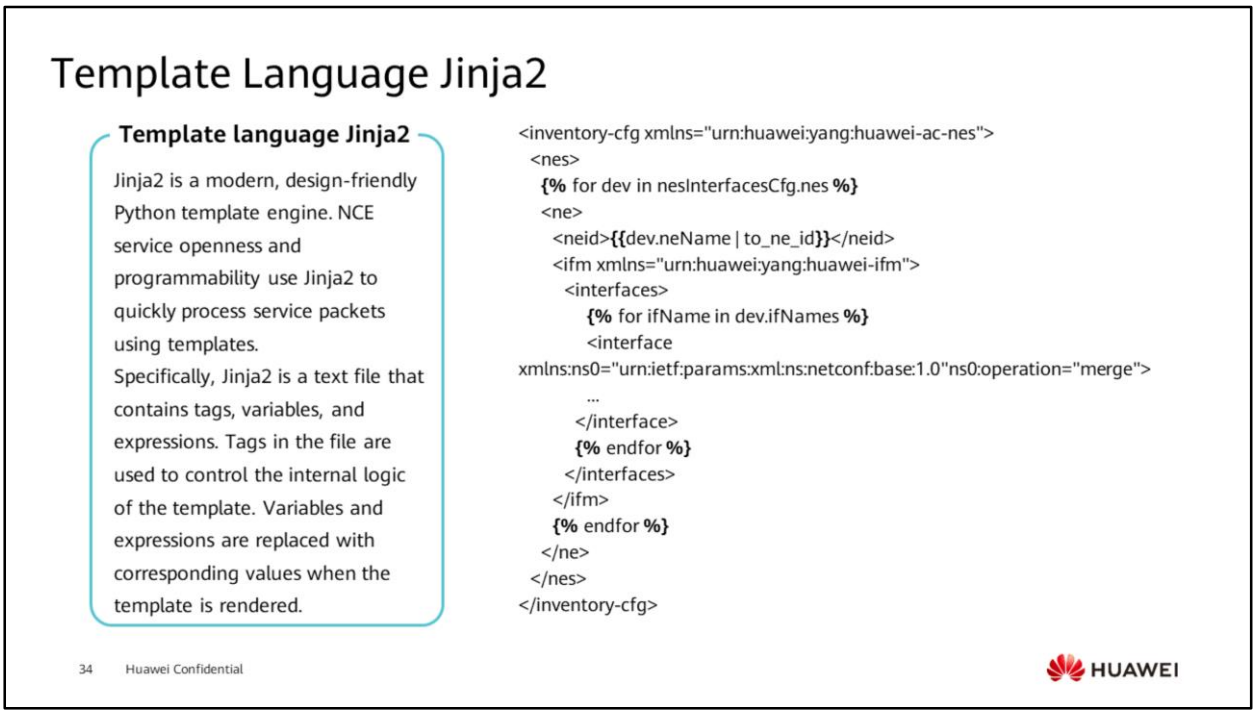
YANG allows the following operation definitions:

- Operations at the top layer of the module are defined using RPC statements. An RPC node usually consists of an input node and an output node. The input and output nodes can contain container, list, leaf, and leaf-list nodes.
- Operations in a container or list node are defined by action statements. Generally, an action node contains an input node and an output node.
- The difference between RPC and action is that RPC does not directly operate data, but action is directly bound to the node in the data storage.

```
module server-farm {
  yang-version: 1.1;
  namespace: "urn:example:server-farm";
  prefix: "sfarm";

  import ietf-yang-types {
    prefix: yang;
  }

  rpc reset-specified-servers {
    input {
      leaf-list servers {
        type: "inet:ip-address";
      }
    }
  }

  list servers {
    key: "name";
    leaf name {
      type: string;
    }
    leaf address {
      type: inet;
    }

    leaf location {
      type: string;
    }

    action reset {
      input {
        leaf reset-at {
          type: yang:date-and-time;
        }
      }

      output {
        leaf complete-at {
          type: yang:date-and-time;
        }
      }
    }
```


HUAWEI

- In this example, an RPC interface named **reset-specified-servers** is defined for resetting services. **input** indicates that the input parameter is the IP address of the server to be restarted. If **output** is not defined, the HTTP status is used to determine the returned result.
- The servers list node defines **action reset** to restart the corresponding service. **Input** defines the leaf node **reset-at**, which indicates that the input parameter is the restart time. **Output** defines the leaf node **complete-at**, which indicates that the returned result is the restart completion time.

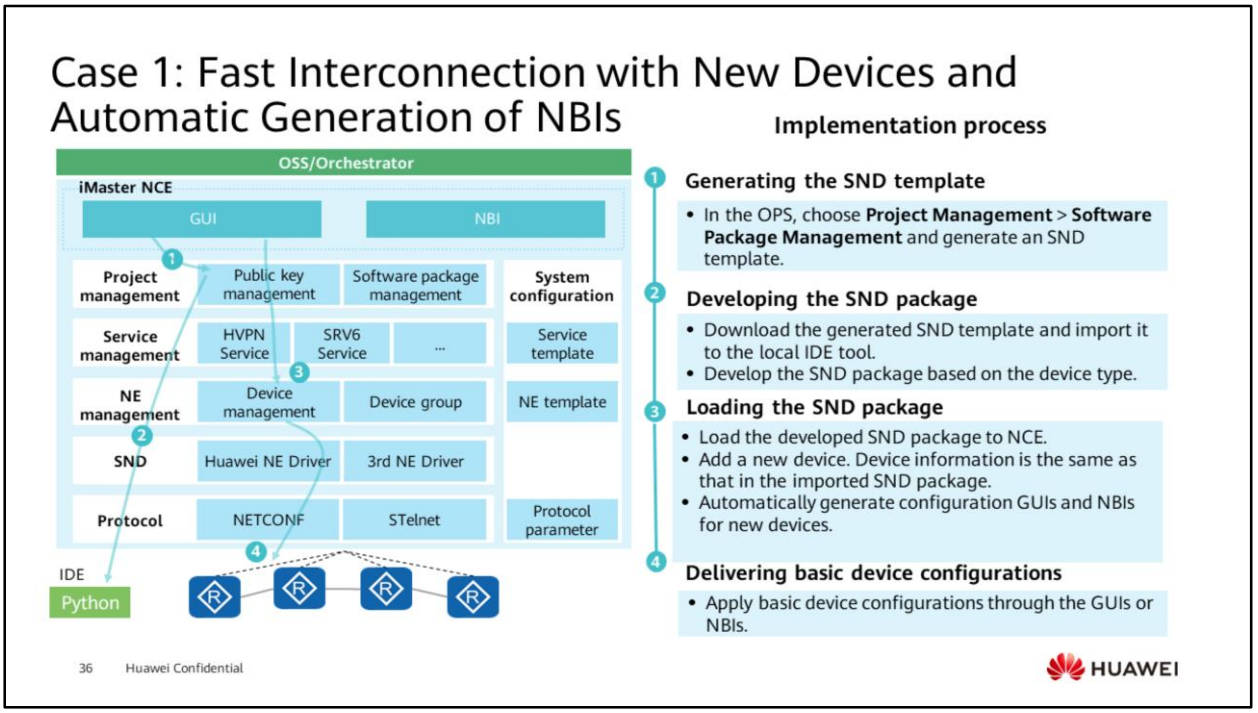


- The Jinja2 template is only a text file, which can be based on any text format (HTML, XML, CSV, etc.). In this example, the XML format is used.
- A template contains variables and expressions. The variables and expressions are converted to corresponding values when the template is used. It has the following common syntaxes:
  - {% ... **%}** contains Control Structures. In this example, **{% for dev in nesInterfacesCfg.nes %}** indicates that the for loop starts, and **{% endfor %}** indicates that the loop ends.
  - **{{...}}** contains an expression, which can be a constant, variable, mathematical formula, or logical statement.
  - {# ... **#}** indicates the comment.
- The variables in **{{...}}** can be modified using filters. Filters and variables are separated by vertical bars (|). For example, **{{ 'abc' | capitalize }}** indicates that the first letter is capitalized and the filtering result is **Abc**. In this example, **{{dev.neName | to_ne_id}}: to_ne_id** is a user-defined filter, indicating that the variable device name **dev.neName** is converted to the device ID.
- For more information, see the *Template Designer Document* at https://jinja.palletsprojects.com/en/2.11.x/.

Contents

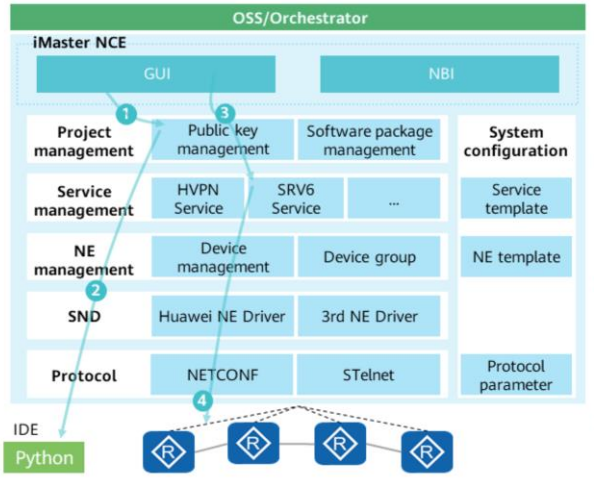
1. Background
2. Introduction to NCE Service Openness and Programmability
3. Key Capabilities
4. Related Concepts
5. **Practice Cases**

HUAWEI



- NCE uses the Specific NE Driver (SND) package to quickly interconnect and manage Huawei and third-party devices and open device configuration capabilities. To manage third-party devices, you need to obtain the YANG file of the device from the vendor's website. If third-party devices support only command lines and do not support NETCONF interconnection, Huawei can customize interconnection capabilities.
- Key capabilities:
    1. Quickly manage Huawei and third-party devices.
    2. Open device configuration capabilities.
    3. Automatically generate the configuration GUIs and northbound interfaces of new devices.

Case 2: Fast Building of New Services and Automatic Generation of NBIs



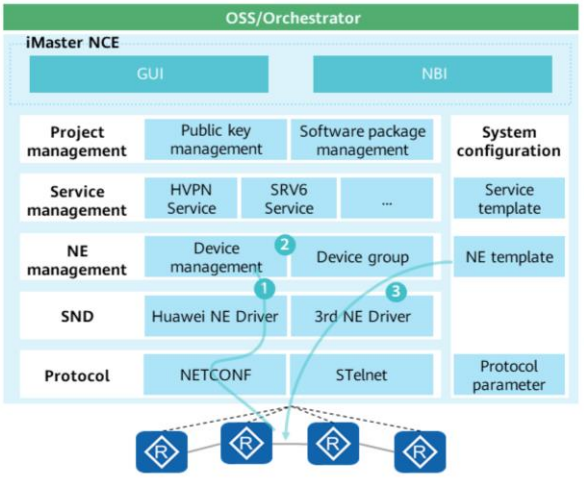
Implementation process

1. **Generating the SSP template**
   - In the OPS, choose **Project Management** > **Software Package Management** and generate an SSP template.
2. **Developing the SSP package**
   - Download the generated SSP template and import it to the local IDE tool.
   - Develop the SSP package based on the service scenario.
3. **Loading the SSP package**
   - Log in to the OPS and load the developed SSP package.
   - Automatically generate configuration GUIs and NBIs for new services.
4. **Delivering service configurations to devices**
   - Deliver service configurations to devices through the GUIs or NBIs.

HUAWEI

- In this example, the service openness capability is used. Similar to the device atomic capability openness, the system is developed based on the standard NETCONF protocol. The internal data model uses the YANG modeling language to automatically generate configuration GUIs and northbound interfaces based on the YANG model of services. In addition, the Easymap algorithm is provided for customers to write only the creation process, and the update and deletion are calculated by comparing algorithms. This simplifies customer programming.
- The service layer shields differences between devices, supports interconnection with different device types, and delivers configurations through different protocols. The maintenance personnel or upper-layer system only needs to view corresponding services. They do not need to know the specific vendor and protocol of the device. This feature improves interconnection efficiency and reduces the pressure on maintenance personnel.
- Key capabilities:
    1. Open service capabilities.
    2. Shield differences between underlying devices at the service layer.
    3. Automatically generate southbound and northbound interfaces based on the YANG model.

Case 3: Batch Configuration Delivery



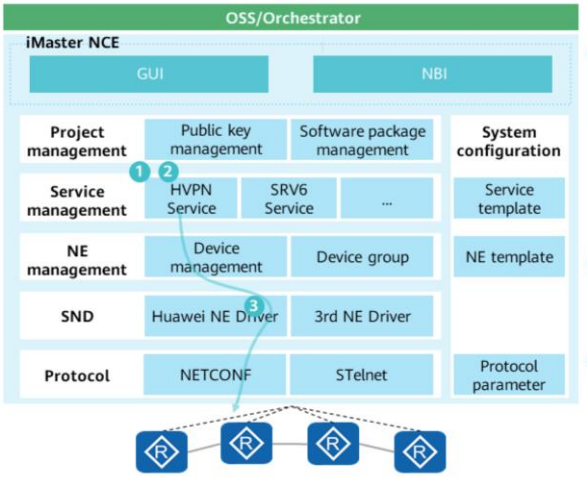
Implementation process

1. **Managing devices**
   - Choose NE Management > Device Management, add a device, and match the SND package based on the device NE type, software version, and vendor.
   - Synchronize device configurations to the OPS.
2. **Grouping devices on demand**
   - To apply one template to multiple devices, you can add these devices to the same device group.
3. **Delivering configurations in batches**
   - Some templates can be preset for common configurations (such as underlay), and device groups can be selected for batch delivery.

HUAWEI

- Devices with the same configuration can be grouped. A preset template can be applied to the system for batch configuration delivery. Currently, more than 60 templates are preset in the enterprise DCN for users to apply.

Case 4: Configuration Correctness Pre-check Using the dryRun Function



Implementation process

1. **Configuring services**

   Configure services on the **Service Management** page as required.
   - Create a service.
   - Select an existing service and modify the configuration parameters.
   - Select an existing service and delete it.

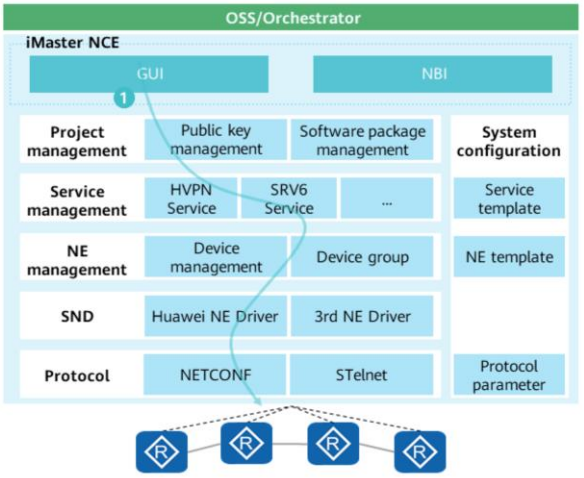
2. **Checking before service configuration delivery**
   - Perform dryRun and check whether the generated configurations are correct.

3. **Delivering service configurations to devices**
   - Perform the Commit operation to deliver configurations to the device.
   - View the delivered configurations through the associated data.

HUAWEI

- Before service configurations are delivered, the OPS provides the dryRun function to check the correctness of delivered configurations in advance. If an error occurs, modify the dryRun function. After the configurations are correct, commit the configurations again. The system provides a transaction mechanism to ensure data consistency between the device and controller. If the data fails to be synchronized, the system automatically rolls back the data to ensure that no residual data exists. For a service that is successfully delivered, you can view the delivered configurations of the associated device. In addition, you can view the delivered configurations in historical records. You can roll back the configurations based on the rollback point.
- Key capabilities:
  1. Use the dryRun function to check whether the delivered configurations are correct in advance.
  2. Provide a transaction mechanism to ensure data consistency between the device and controller. If a failure occurs, automatic rollback will be performed.
  3. Provide the visualized display of service association data and historical configurations.

Case 5: Automatic Detection of Service Configuration Conflicts



Implementation process

1 Automatic detection of service configuration conflicts

Before delivering configurations, the OPS automatically checks whether the configurations delivered to devices conflict with each other. For example, if service 1 delivers the description configuration of VLAN 2 to device 1 and service 2 also delivers the description configuration of VLAN 2 to device 1, this system checks whether the two values are the same. If they are different, a message indicating that the configurations conflict will be displayed.

On the **Service Management** page, deliver configurations. The system automatically checks whether service configurations conflict.

HUAWEI

Quiz

1. (Short-answer question) What are the core function packages of Huawei iMaster NCE service openness and programmability?
2. (Multiple-answer question) Which of the following are included in a Specific Service Plugin (SSP) package? ( )
   A. Service YANG model
   B. Device YANG model
   C. Jinaji2 template
   D. Python mapping script

HUAWEI

1. Specific NE Driver (SND) and Specific Service Plugin (SSP) packages.
2. ACD

Summary

- Network service openness and programmability pave the way for development of the network industry. Enabling the capabilities of network service openness and programmability is the cornerstone of building an open network ecosystem.
- Huawei OPS is a practice offered by Huawei in this field. Users can compile SND and SSP packages to quickly interconnect new devices and construct new services.
- To learn this chapter, you need to have a good understanding of Python, NETCONF, YANG, and Jinja2.

HUAWEI

More Information

- Log in to the HUAWEI CLOUD developer community and click the Datacom Network Openness and Programmability tab to obtain related tools and documents.
  - https://developer.huaweicloud.com/resource/network.html#AOC
- Key documents:
  - Datacom Network Openness and Programmability Development Guide: This document describes the development process and methods of related driver packages, and provides guidance for developers to develop SND and SSP packages, load the packages to the system, and automatically generate configuration GUIs and NBIs.
  - Datacom Network Openness and Programmability User Guide: This document describes operations related to the configuration GUIs and NBIs, and provides guidance for O&M personnel to deliver device and service configurations.

HUAWEI

Recommendations

- For more information about Jinja2, visit the following website:
  - https://jinja.palletsprojects.com/en/2.11.x/

HUAWEI

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

Copyright©2021 Huawei Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.