Enterprise Bearer WAN Solution

HUAWEI

Foreword

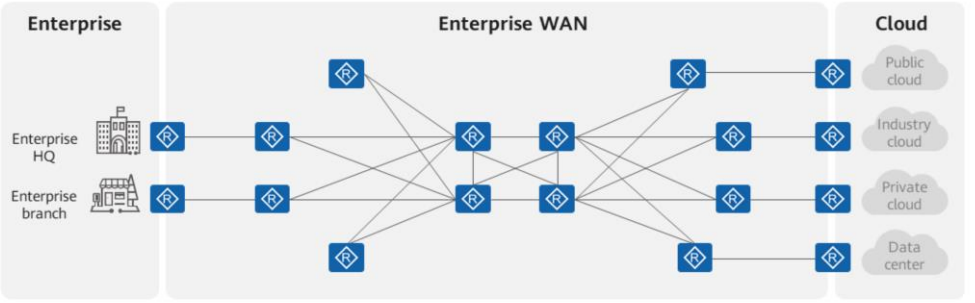
- A wide area network (WAN) is a computer network that connects local area networks (LANs) or metropolitan area networks (MANs) in different regions. A WAN allows information and network resources to be shared in a large scope.
- An enterprise IP bearer WAN is a backbone WAN used to implement cross-region 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.

Objectives

- Upon 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.

Contents

Enterprise WAN Overview



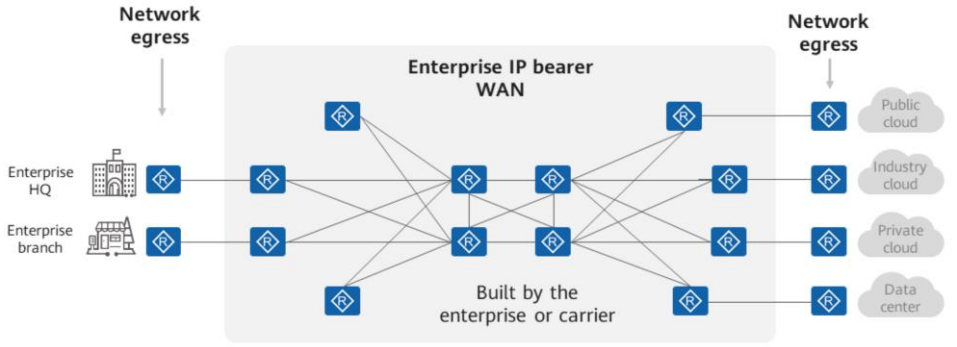
Enterprise WAN definition

- 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

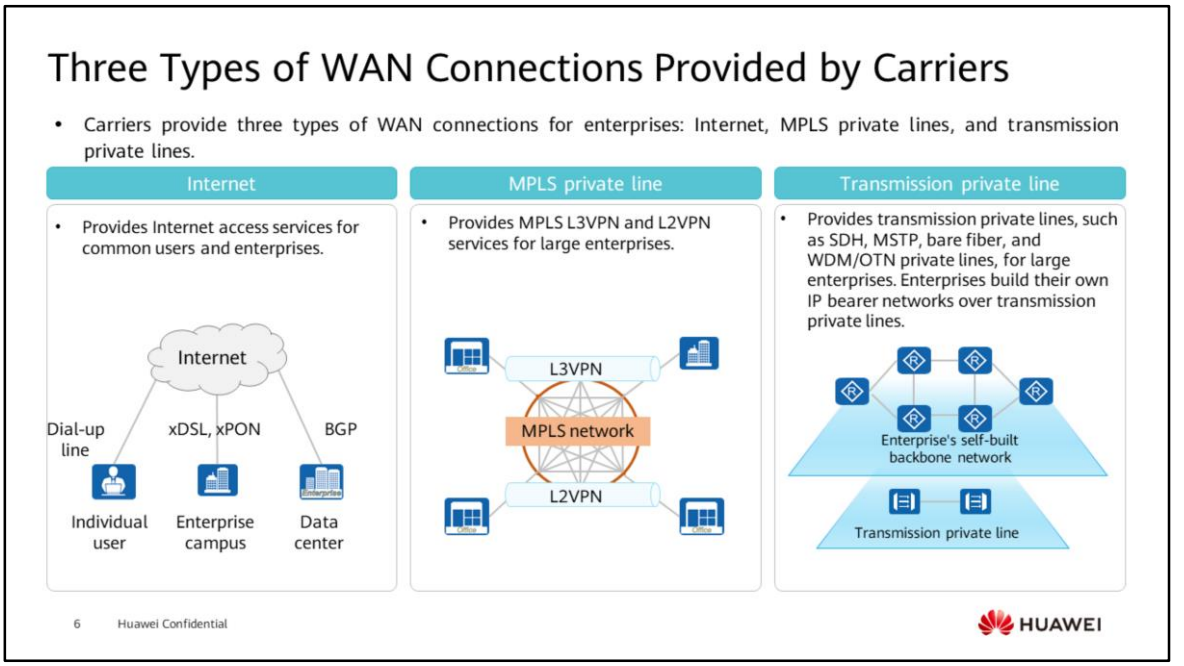
- Self-built, for internal use
- Self-built, for external use

Enterprise IP bearer WAN



Enterprise WAN = Enterprise network egress + enterprise IP bearer WAN (built by the enterprise or carrier)

HUAWEI

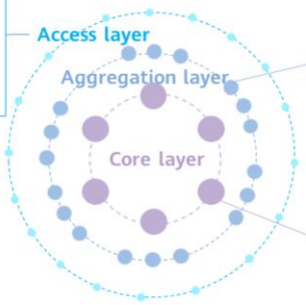


- Internet:
  - Site-to-Internet private line: Ethernet private line/xPON private line/xDSL private line. The access is restricted by geographic locations. It applies to inter-enterprise communication over Internet-based encrypted tunnels.
  - Dial-up connection: low bandwidth and low tariff. The access 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

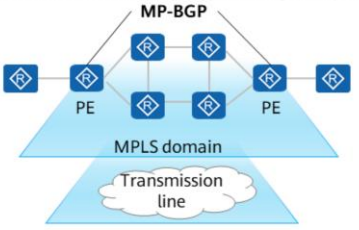
Typical Bearer Technologies of the Enterprise IP Bearer WAN

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.

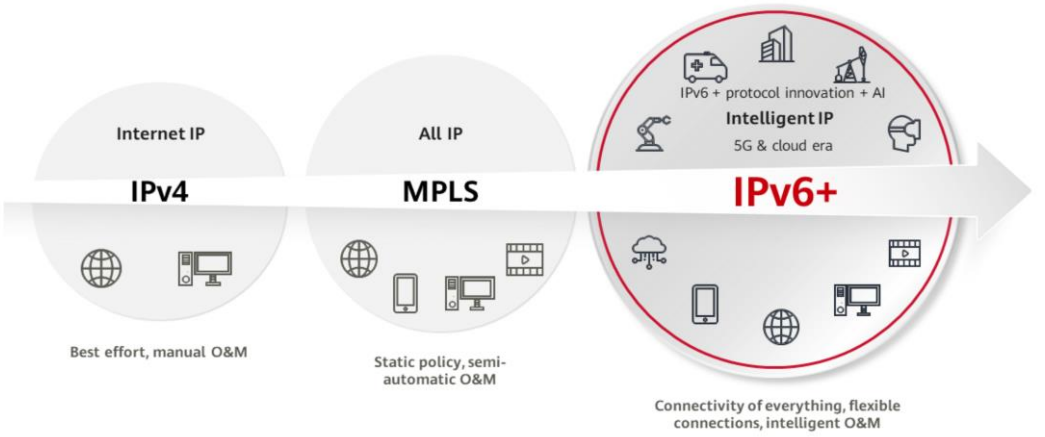
MPLS bearer

- 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.



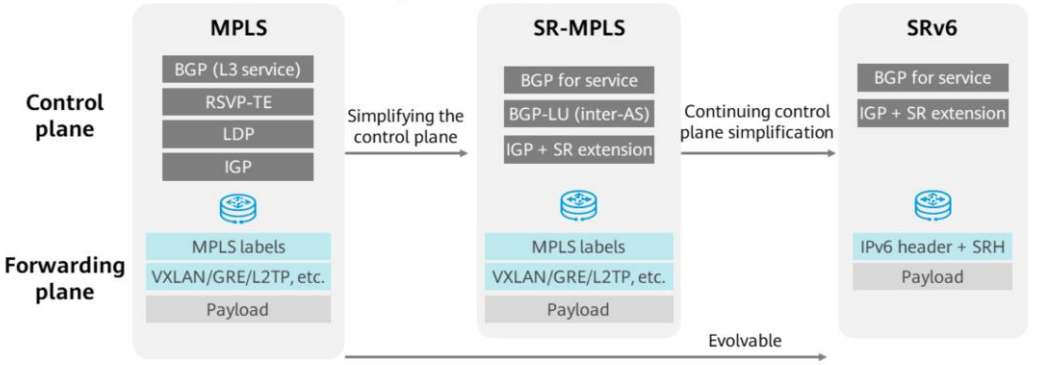
HUAWEI

Enterprise IP Bearer WAN Moving Towards the IPv6+ Era



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.

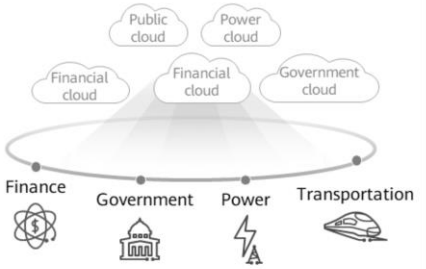
Contents

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**

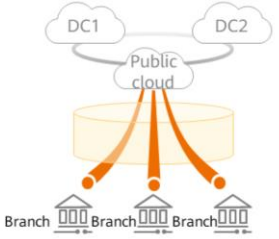
Train control and dispatching
Manual monitoring > Over-the-horizon monitoring
- **Bandwidth > 100 Mbit/s**
- **Clock precision < 3 µs**

Fuel pipe safety detection
Manual inspection > UAV inspection
- **Any access, flexible connection**

3 Challenges Facing the WAN in the Cloud Era

Cloudification of millions of enterprises

How can networks be as agile as clouds?



Service deployment time

Dozens of days >>> < 1 hour

Production service bearer

Can IP provide deterministic experience?

Relay protection

10 ms One-way delay

200 µs Two-way delay variation

A failure to meet these requirements may mean a protection failure, incurring accidents.

Train control

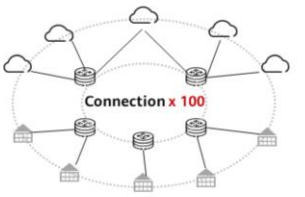
10 ms One-way delay

50 ms Failover duration

A failure to meet these requirements may mean train out-of-control.

Connection scale x 100 ↑

How can O&M be simpler and networks more reliable?

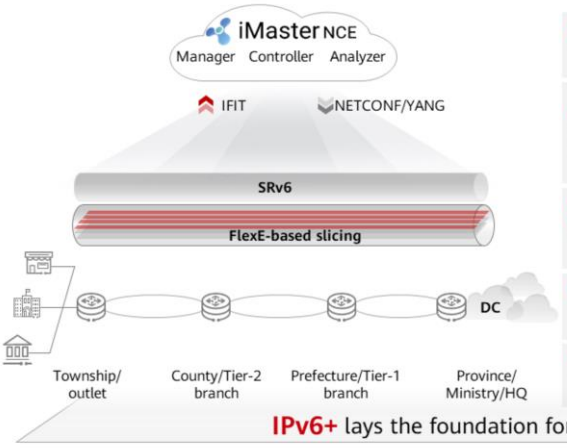


0 artifact, **0** frame freezing

24/7 online

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

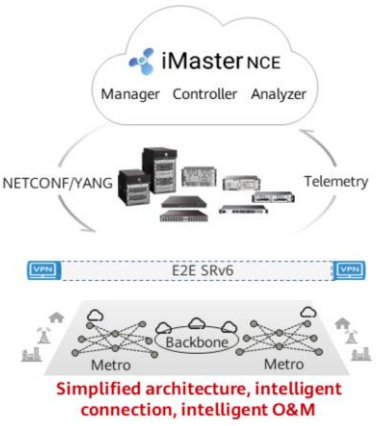
NetEngine 8000 M6

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

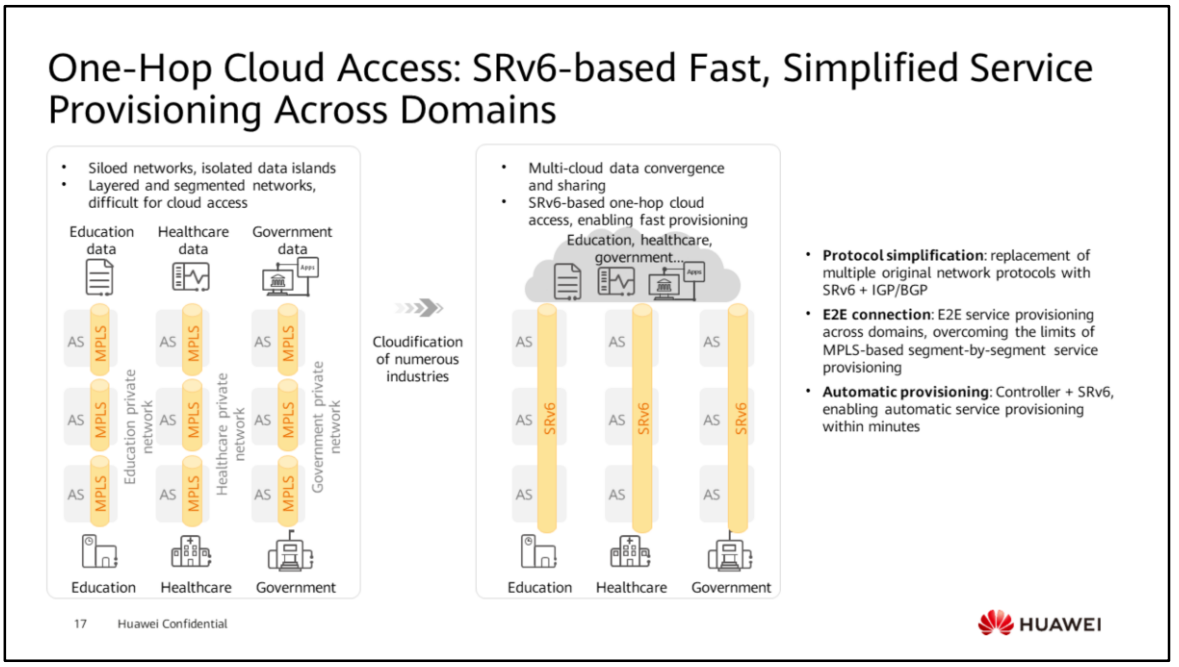
- **NE management:** topology, alarm, configuration, and inventory management
- **Service management**: tunnel and VPN service management

Controller

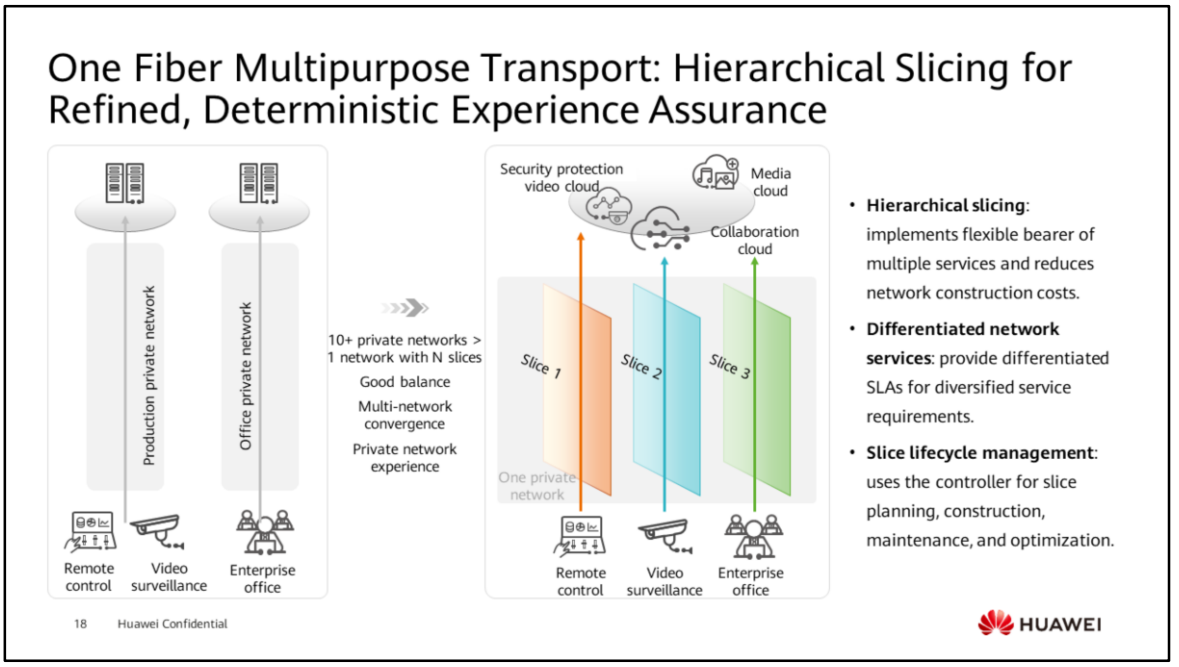
- **Centralized path computation**: path computation based on multiple constraints
- **Logical topologies**: cost, delay, and bandwidth topologies
- **Network optimization**: service path adjustment and optimization

Analysis

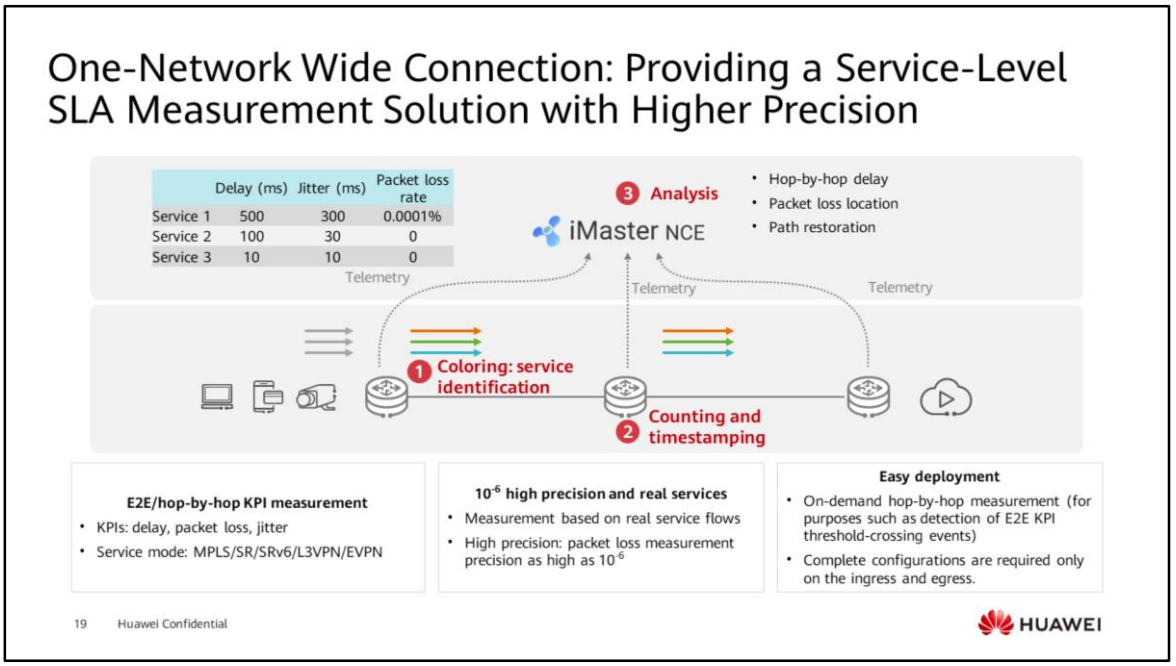
- **Basic network analysis**: display and analysis of performance, traffic, and quality
- **Analysis-based prediction**: traffic, fault, and exception prediction



- 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.

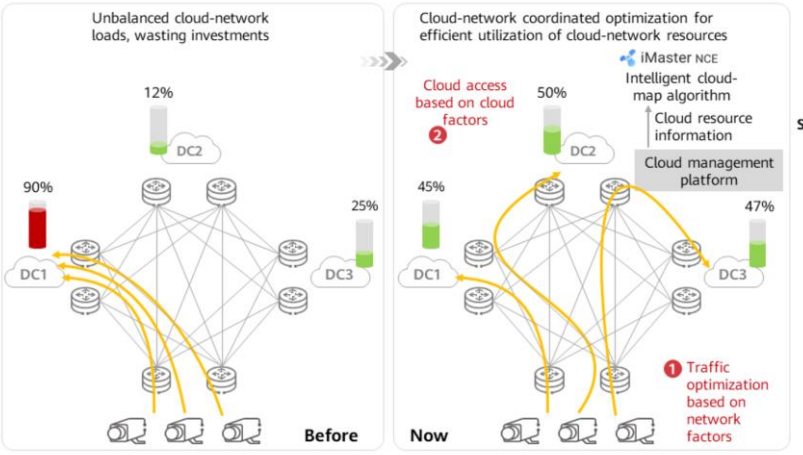


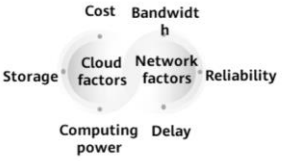
- Huawei uses the hierarchical slicing technology to power IP-based production networks, ensuring deterministic SLAs for production services.
- In the past, production and office services were carried over multiple independent private networks. Repeated network construction resulted in high investment 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

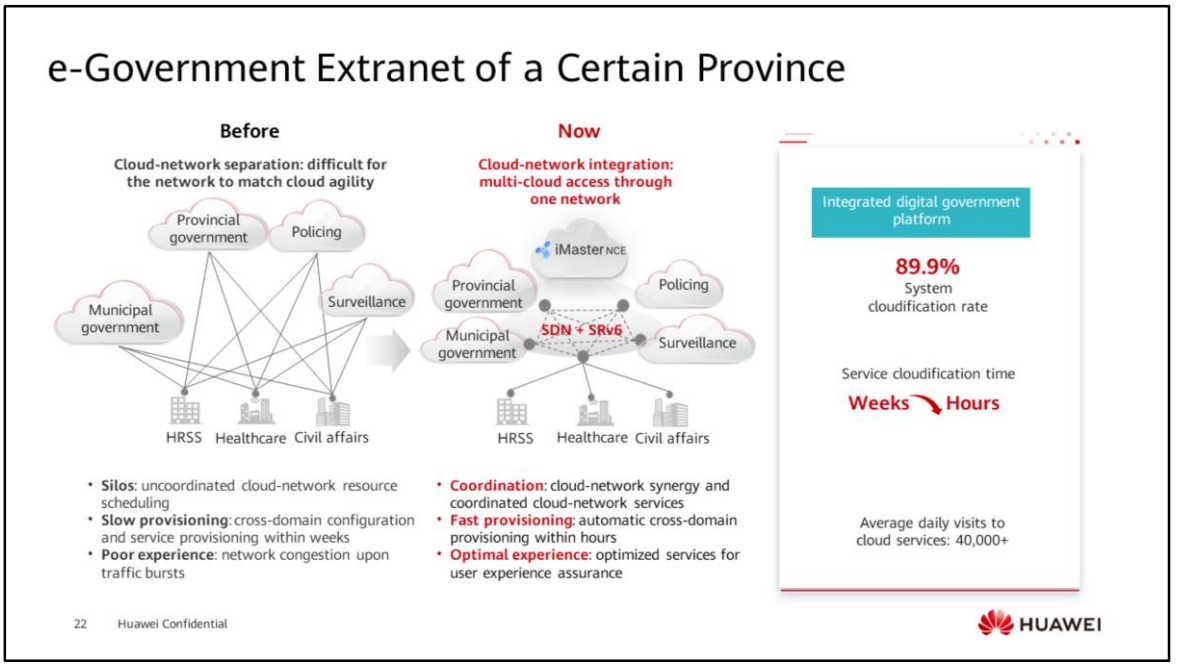




- The intelligent cloud-map algorithm combines the cloud and network factors.
- Load balancing is performed based on service conditions and cloud loads. Services are intelligently scheduled to the most appropriate cloud based on SRv6 and SDN technologies.

HUAWEI

Contents

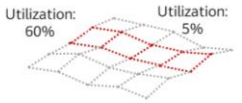


- Quick Network Adjustment upon Cloud Changes, Integrated Service Provisioning, Cloud-based Data Sharing.

Intelligent Traffic Optimization on an Enterprise WAN

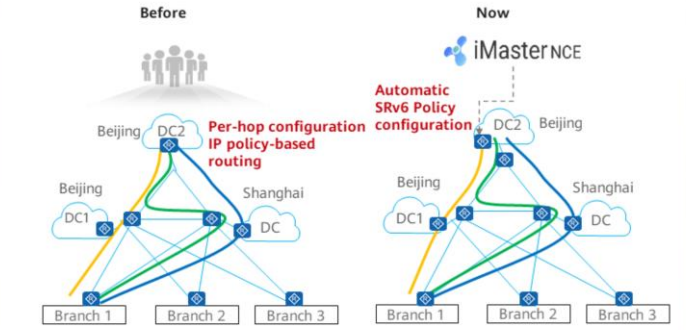
Service scenarios

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

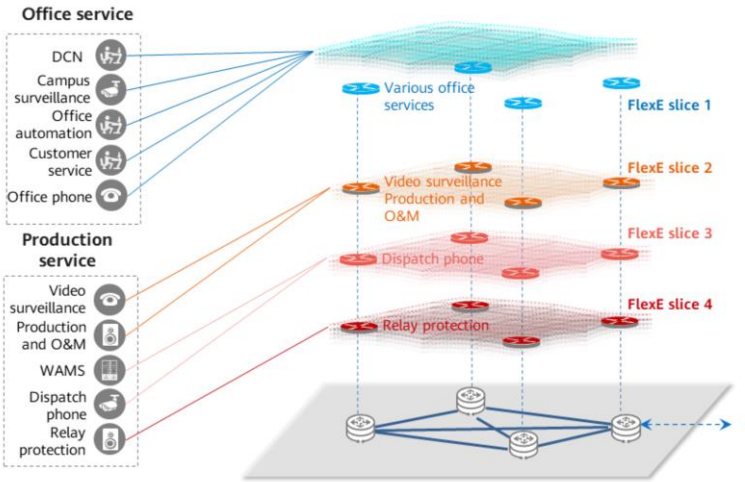
Solution



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

- 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.

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.

Enterprise Bearer WAN Architecture and Key Technologies

HUAWEI

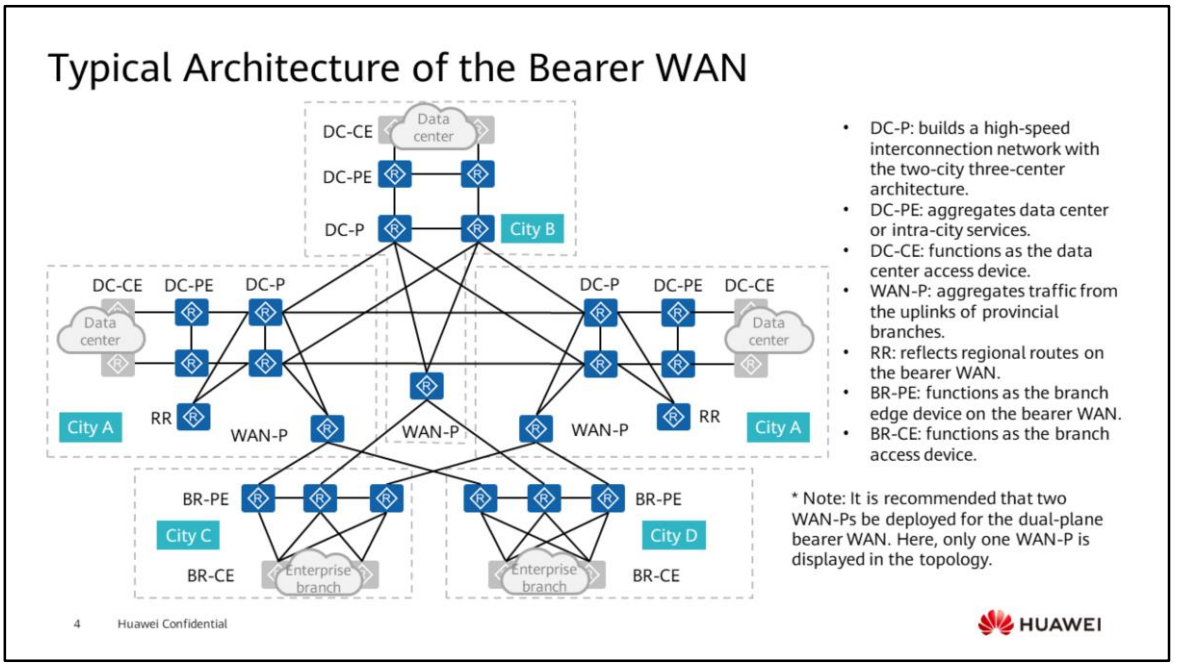
Foreword

- 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.

Objectives

- Upon 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.

Contents

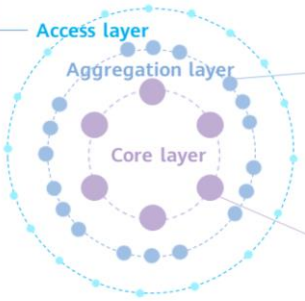


- The Ps for intra-city data center are directly connected using WDM or bare optical fibers. The link bandwidth can reach 10 Gbit/s. To reduce costs, consider connecting local data centers with remote data centers through carriers' MSTP links.

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.

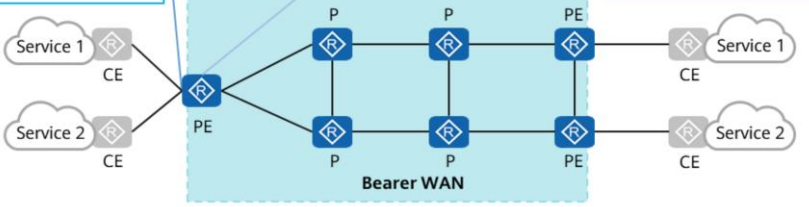
Core layer (P):
- Generally adopts the full-mesh + dual-plane architecture.
- Provides stable, reliable high-speed traffic forwarding between regions.

Technologies Used by the Typical Bearer WAN Architecture

For CEs accessing the same PE, VPNs are used to isolate services:
- L2VPN: Layer 2 communication
- L3VPN: Layer 3 communication

Service tunnels are established between PEs for VPN service recursion. For example:
- MPLS LDP tunnels
- MPLS TE tunnels
- SR-MPLS BE/TE tunnels
- SRv6 BE tunnels

Data forwarding:
- MPLS or IPv6 forwarding

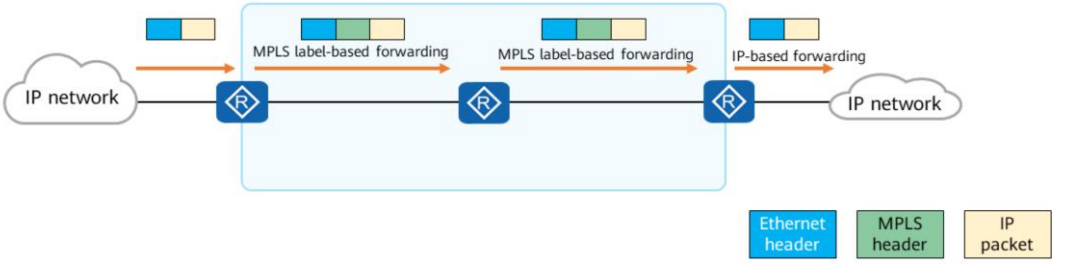


HUAWEI

Contents

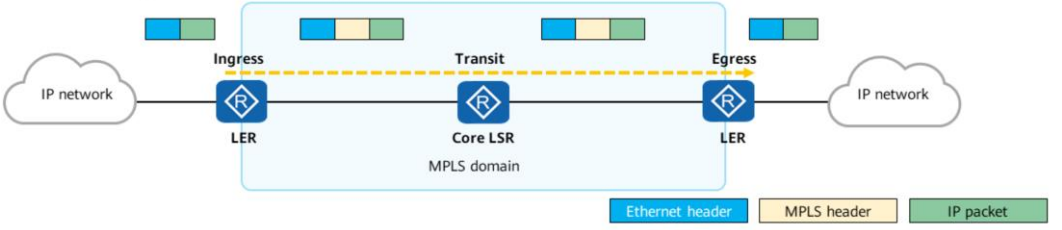
MPLS Overview

- 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.



MPLS Terms (1)

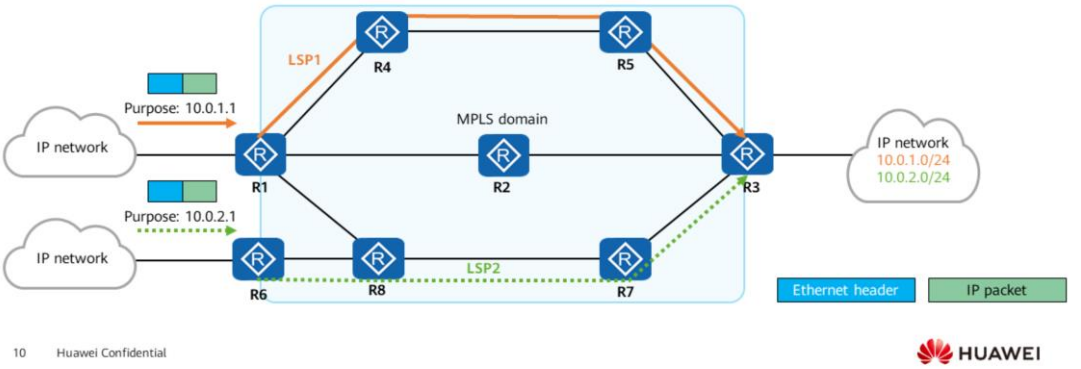
- MPLS domain: consists of a series of consecutive network devices that run MPLS.
- Label switching router (LSR): a routing device, such as a router or switch, that runs MPLS. An LSR that resides at the edge of an MPLS domain and connects to a non-MPLS network is called a label edge router (LER). An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

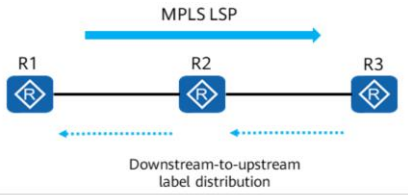
- For more information about MPLS, see HCIP-Datacom-Advanced Routing & Switching Technology.

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).

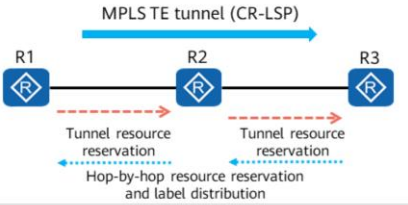
LDP: used to establish common LSPs

- 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.



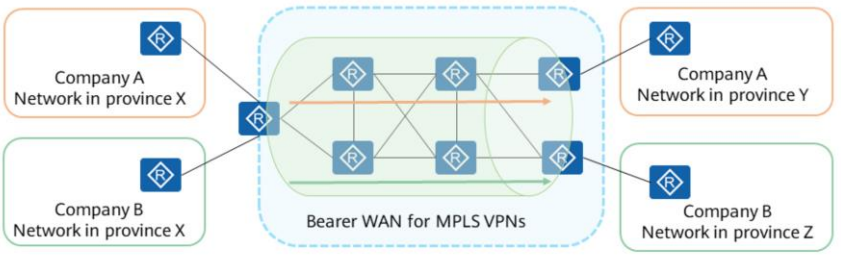
RSVP-TE: used to establish CR-LSPs

- 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).



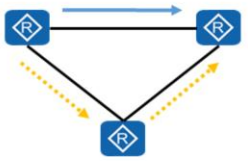
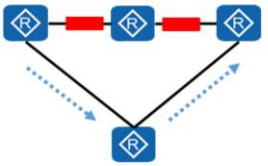
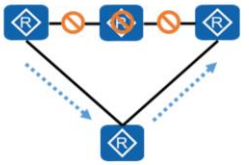
MPLS LDP Overview

- 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.



MPLS TE Overview

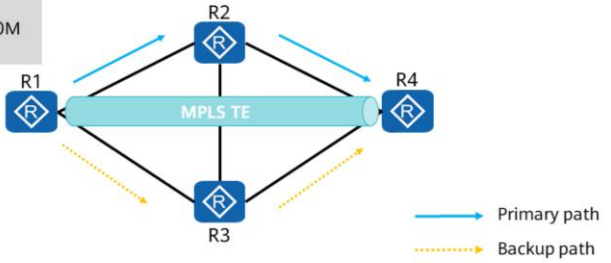
- MPLS TE, as its name suggests, is a combination of MPLS and TE. It provides functions such as path planning, traffic optimization, and fault protection for MPLS VPN services.
- Compared with MPLS LDP, MPLS TE enhances VPN traffic control and protection.

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.

MPLS TE Tunnel

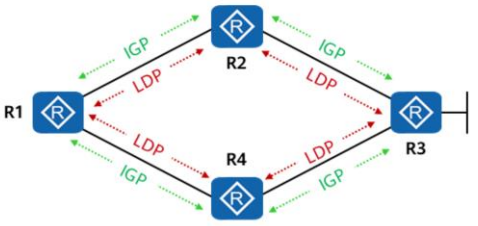
- MPLS TE often associates multiple LSPs with a virtual tunnel interface, and such a group of LSPs is called an MPLS TE tunnel.
- An MPLS TE tunnel provides SLA assurance, but requires complex configuration and manual planning.

```
[R1] interface tunnel1
[R1-Tunnel1] ip address ...
[R1-Tunnel1] tunnel-protocol mpls te
[R1-Tunnel1] destination R4
[R1-Tunnel1] mpls te bandwidth 50M
...
```



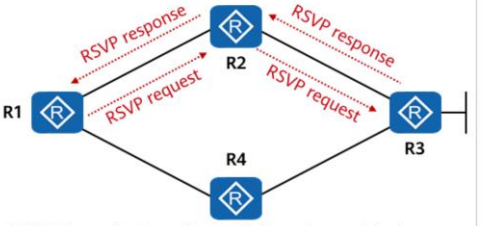
Issues with MPLS LDP and RSVP-TE

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

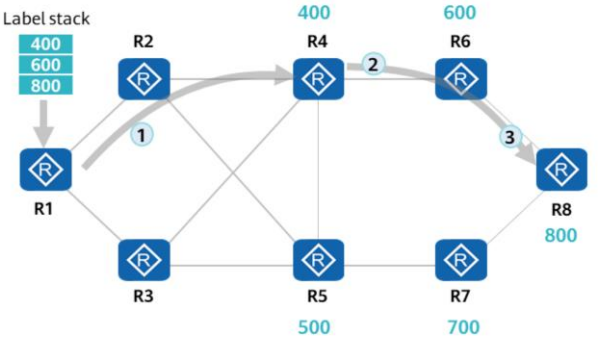


- RSVP-TE is complex to configure and does not support load balancing.
- To implement TE, devices need to exchange a large number of RSVP packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- RSVP-TE uses a distributed architecture, so that each device only knows its own state and needs to exchange signaling packets with other devices.

How can we simplify the control plane?

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.



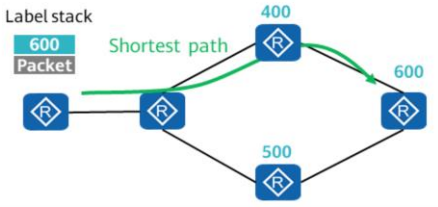
- Source routing:
  - The source node selects a path and pushes an ordered label stack into the packet.
  - Other nodes on the network forward the packet according to the label stack encapsulated into the packet.
- SR has the following characteristics:
  - Extends existing protocols (e.g. IGP) to facilitate 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.

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.

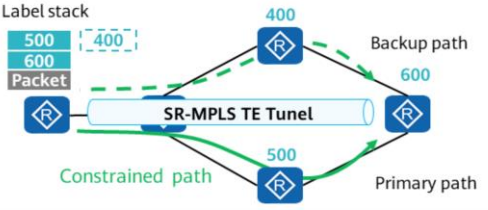
SR-MPLS BE

- 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.



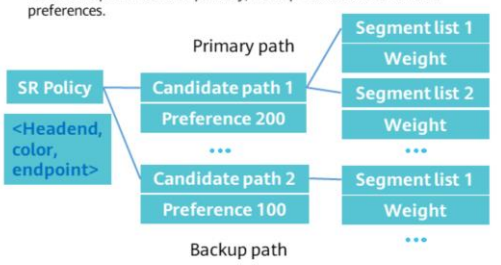
HUAWEI

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.

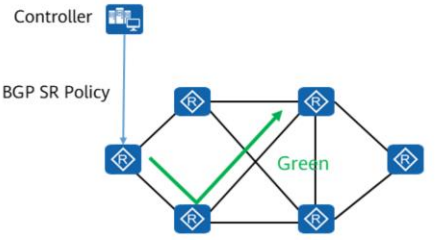
SR Policy model

- 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.

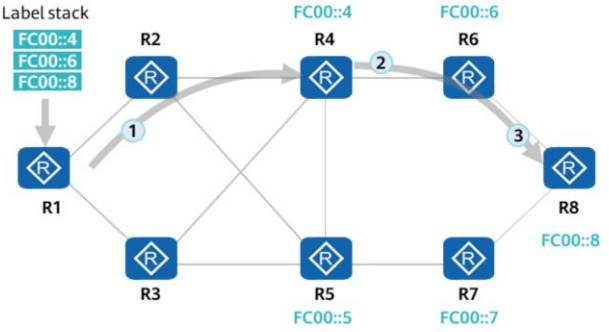


HUAWEI

- 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.

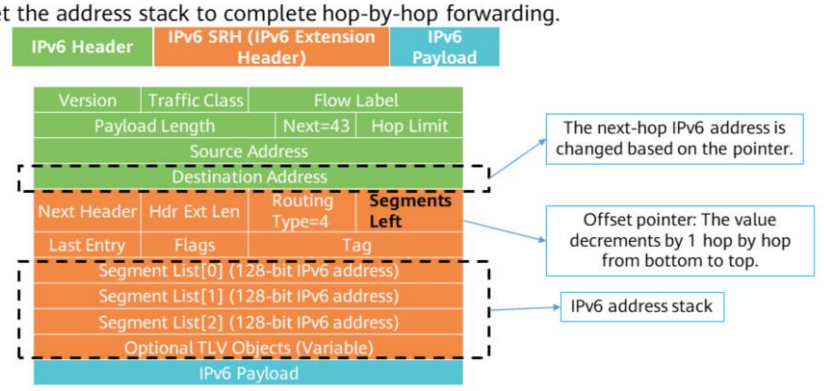


Source routing:

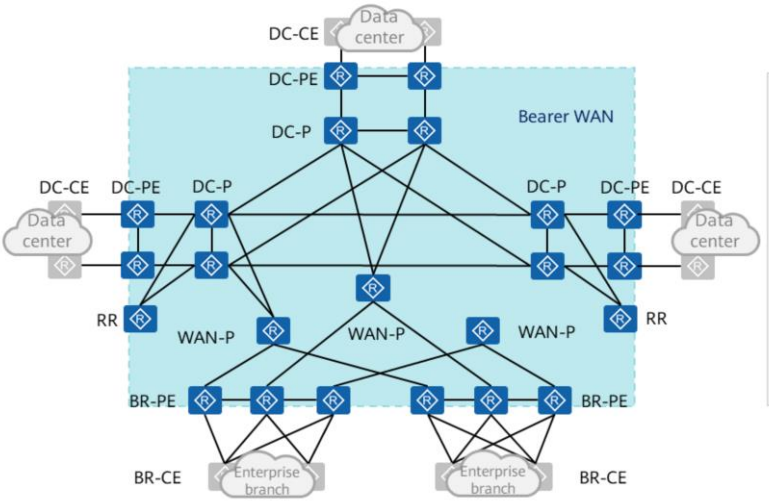
- The source node selects a path and pushes an ordered label stack into the packet.
- Other nodes on the network forward the packet according to the label stack encapsulated into the packet.

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.



Summary: WAN Bearer Technologies



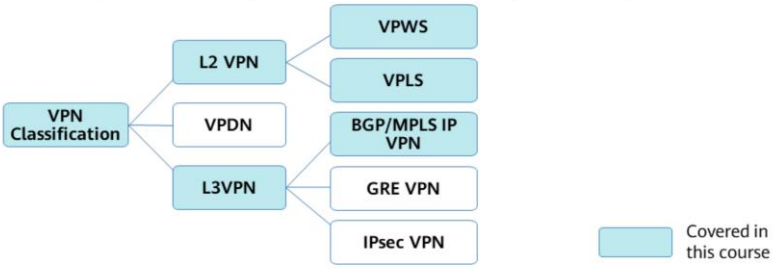
Enterprise bearer WANs can be roughly classified into two types: MPLS network and IPv6 network.

- MPLS network:
  - MPLS LDP
  - MPLS TE
  - SR-MPLS
- IPv6 network:
  - SRv6

Contents

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

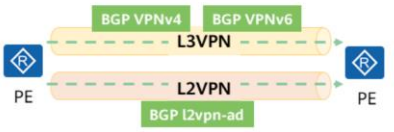
- Traditional switching networks, such as asynchronous transfer mode (ATM) and frame relay (FR) networks, are integrated with IP or MPLS networks. As a result, Layer 2 virtual private network (L2VPN) emerges. L2VPN includes Virtual Pseudo Wire Service (VPWS) and Virtual Private LAN Service (VPLS):
  - 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

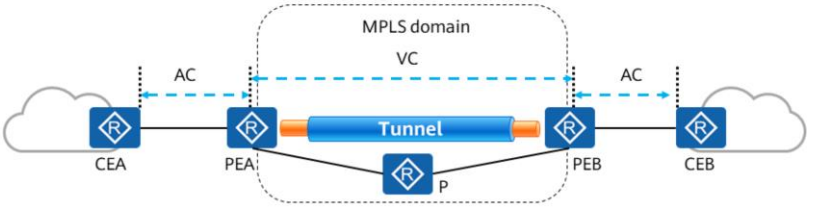
- EVPN provides both Layer 2 and Layer 3 capabilities and functions as a control plane protocol to transmit Layer 2 and Layer 3 information (MAC/IP addresses). EVPN can be used together with the traditional VPN technology.
- EVPN supports EVPN VPLS, EVPN VPWS, EVPN L3VPN, etc.

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.

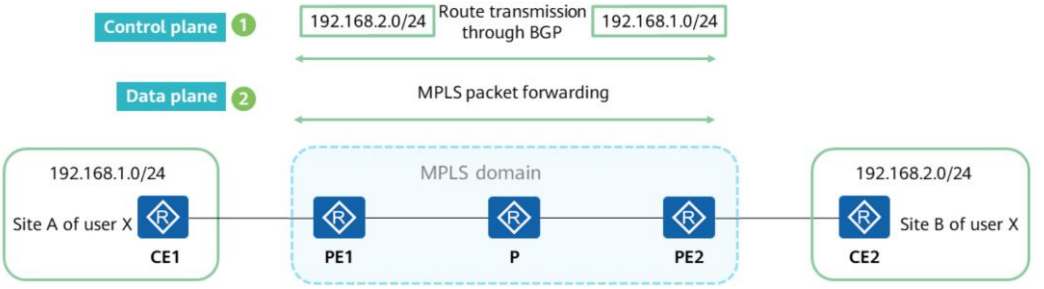


HUAWEI

- VCs are also called pseudo wires (PWs) in some documents.

Traditional WAN L3VPN Overview

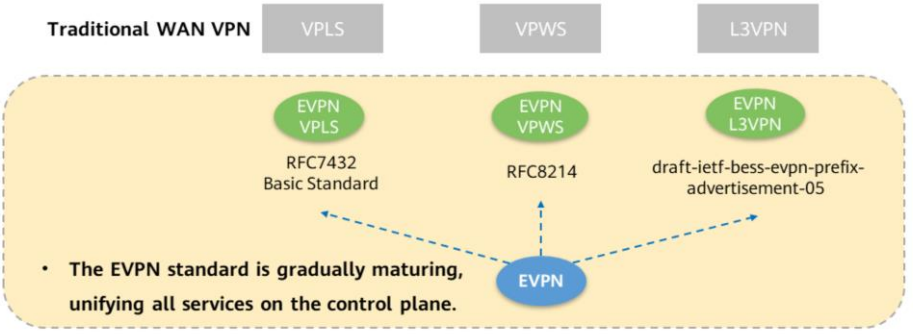
- Traditional WAN L3VPN generally refers to BGP/MPLS L3VPN.
- It distinguishes the control plane from the data plane. The control plane uses MP-BGP to advertise VPN routes, and the data plane uses MPLS LSPs to forward VPN packets.
- BGP/MPLS L3VPN is widely used on carrier networks to provide MPLS Layer 3 private line services for enterprises.



HUAWEI

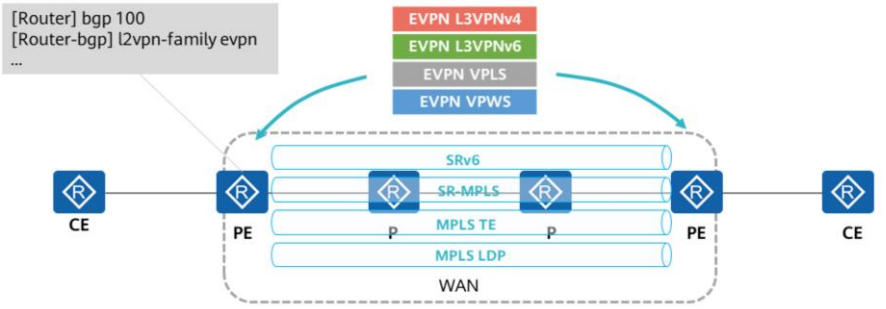
WAN EVPN Overview

- 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.



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

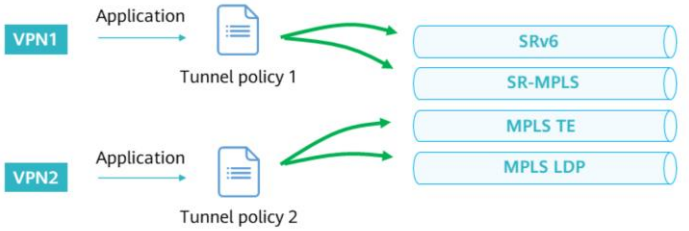
Summary: WAN VPN

- 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.

Contents

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.



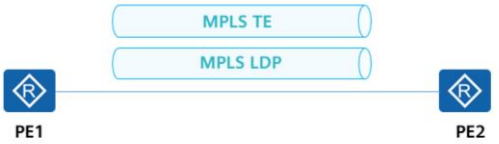
HUAWEI

- 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
```



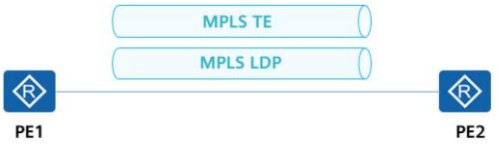
HUAWEI

- 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
```

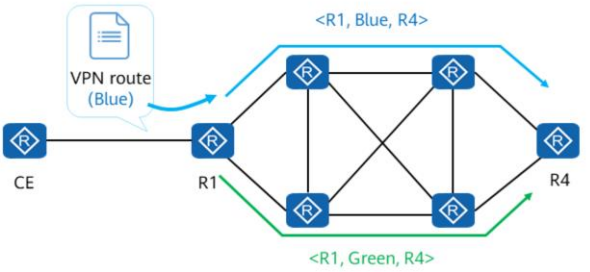


HUAWEI

- 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.

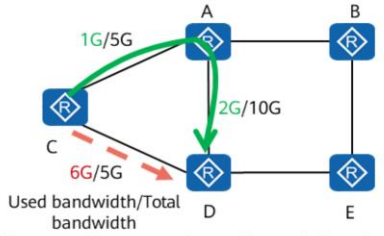
HUAWEI

- 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

Network Congestion Background

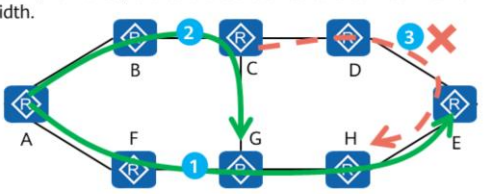
Drawback of fixed bandwidth-based route selection and related solution



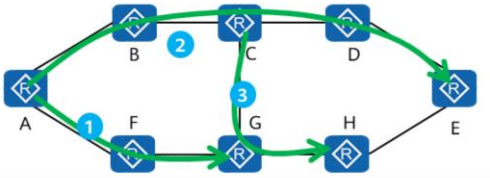
The network computes forwarding paths based on bandwidth. The link from Router C to Router D is the shortest forwarding path. The rate of service traffic from Router C to Router D exceeds the link bandwidth, and packet loss occurs. Although other links are idle, the algorithm still selects the shortest 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:



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

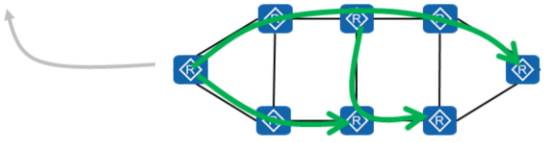
The controller computes paths based on the optimization target. Optimization targets include:

- Least path cost
- Shortest path delay
- Maximum link bandwidth utilization
- ...

3. Optimization result delivery

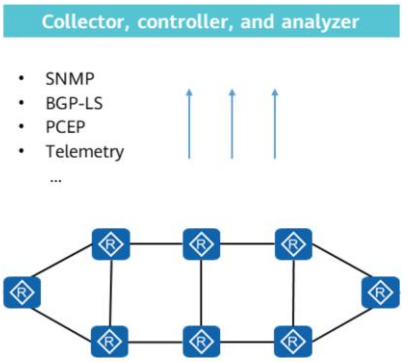
The controller delivers the computation result to network devices in any of the following modes:

- NETCONF
- PCEP
- BGP SR Policy
- ...



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.

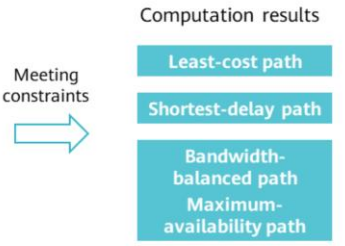
HUAWEI

- 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.

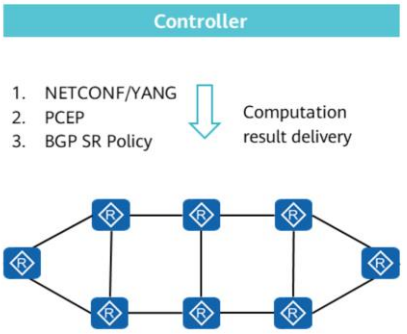


HUAWEI

- 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.

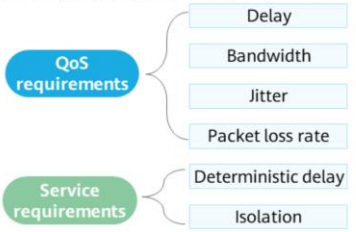
3 BGP SR Policy

- 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.

Contents

New Requirements for Bearer Network SLA Assurance

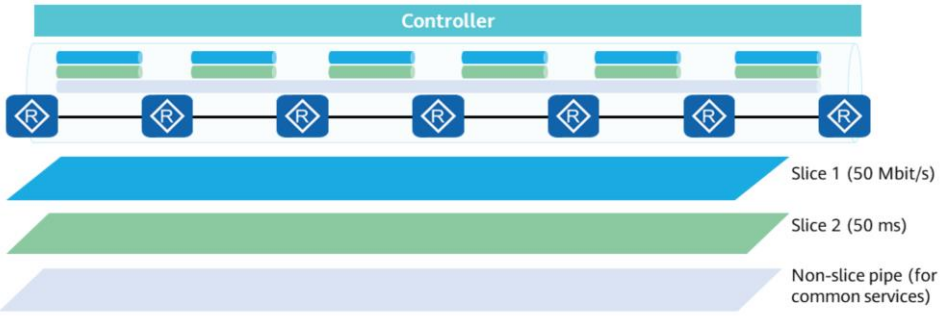
- An SLA is a formal commitment between a service provider and a customer. In the WAN service field, in addition to basic connectivity requirements, the SLA also focuses on deterministic delay, bandwidth, reliability, and isolation (security).
- Generally, a bearer path carries the traffic of multiple 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

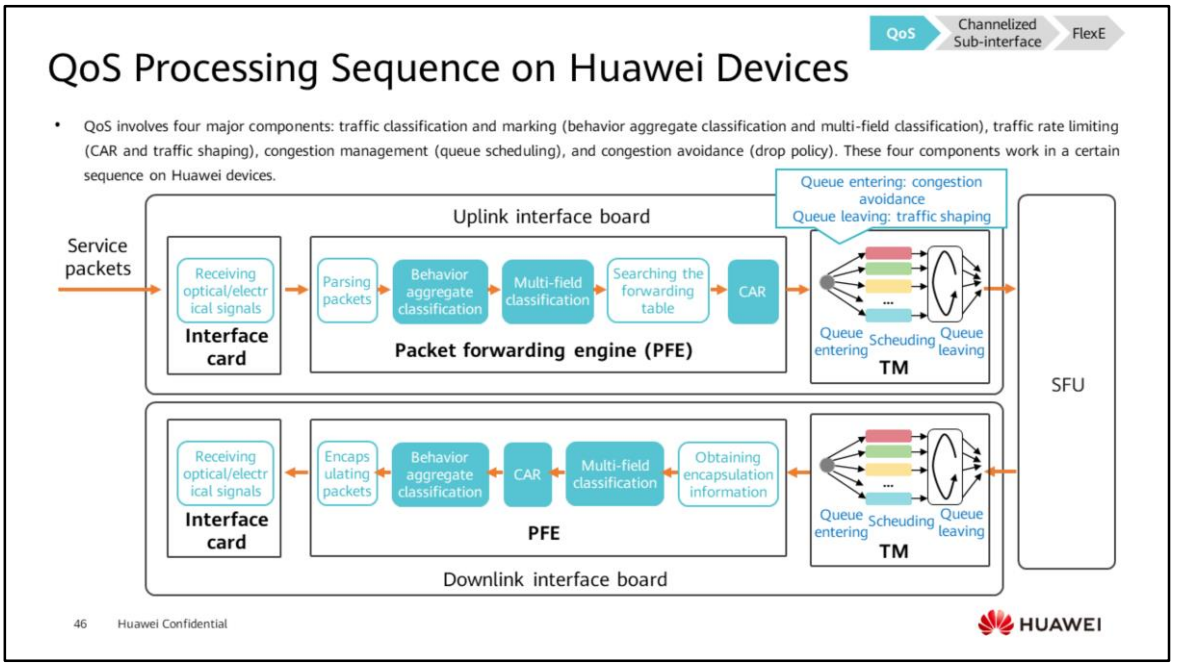
- An application needs to apply for specific services before sending packets.
- The network reserves network resources for the application using a protocol such as RSVP.

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.

DiffServ Model-based QoS Components

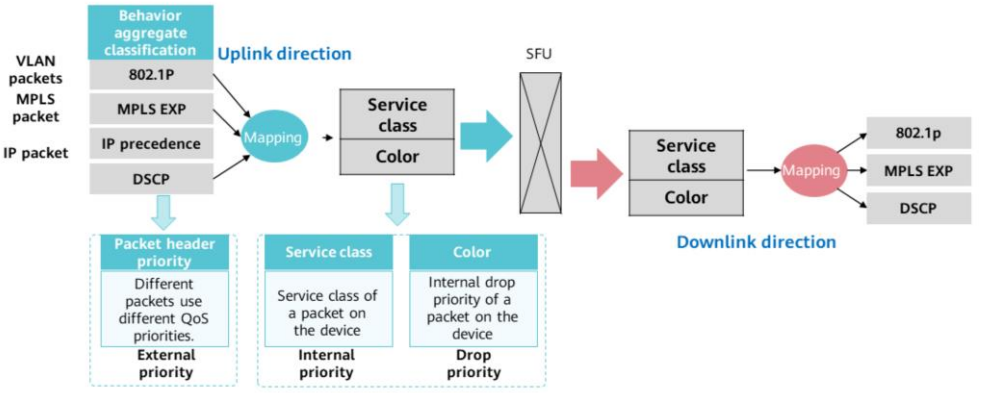
- DiffServ model-based QoS consists of four components: traffic classification and marking, traffic policing and shaping, congestion management, and congestion avoidance.
  - Traffic classification and marking: Dividing data packets into different classes or configuring different priorities for packets 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.



- 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.



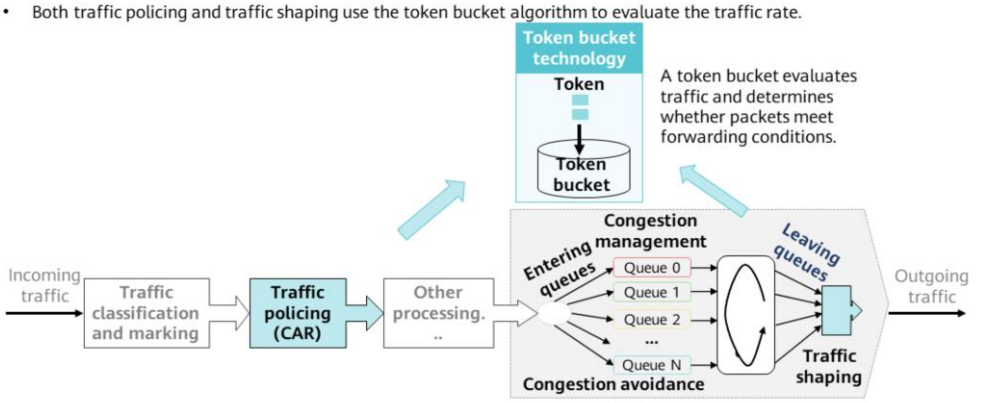
DSCP/IP Precedence/ 802.1p/EXP Value Mapping

Priorities in descending order

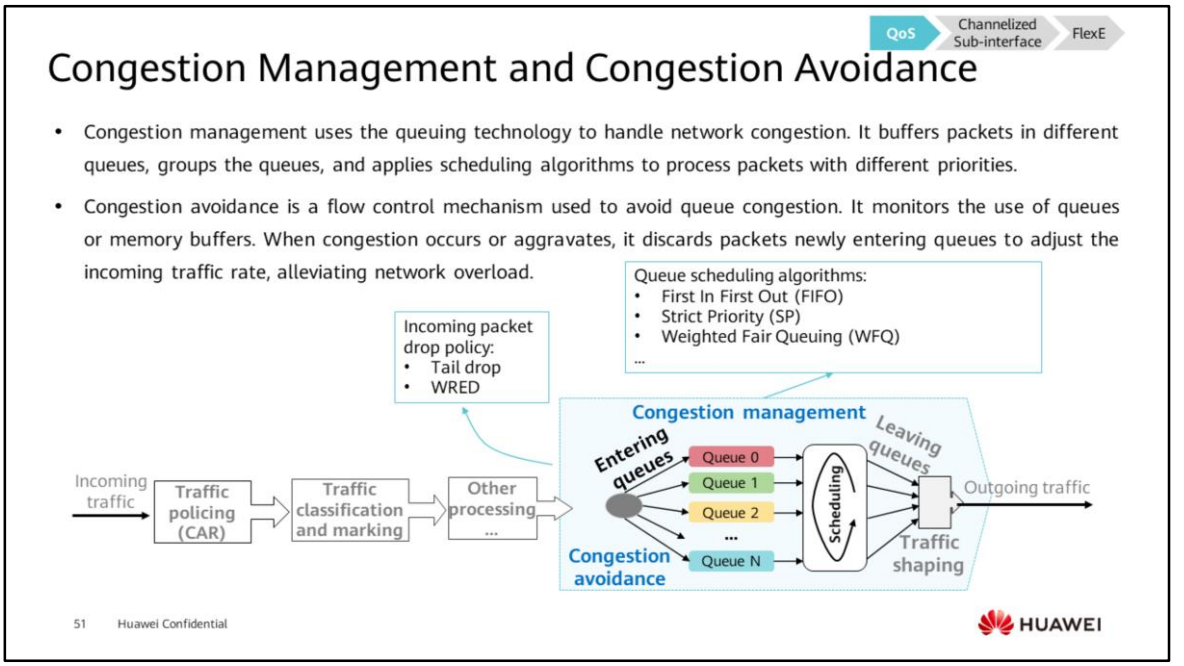
802.1p	MPLS EXP	IP Precedence	DSCP	DSCP Name				
7	7	7	56-63	CS	CS7			
6	6	6	48-55		CS6			
5	5	5	40-47	EF	EF			
4	4	4	32-39	AF	AF4	AF41	AF42	AF43
3	3	3	24-31		AF3	AF31	AF32	AF33
2	2	2	16-23		AF2	AF21	AF22	AF23
1	1	1	8-15		AF1	AF11	AF12	AF13
0	0	0	0-7	BE	BE			

Traffic Policing and Traffic Shaping

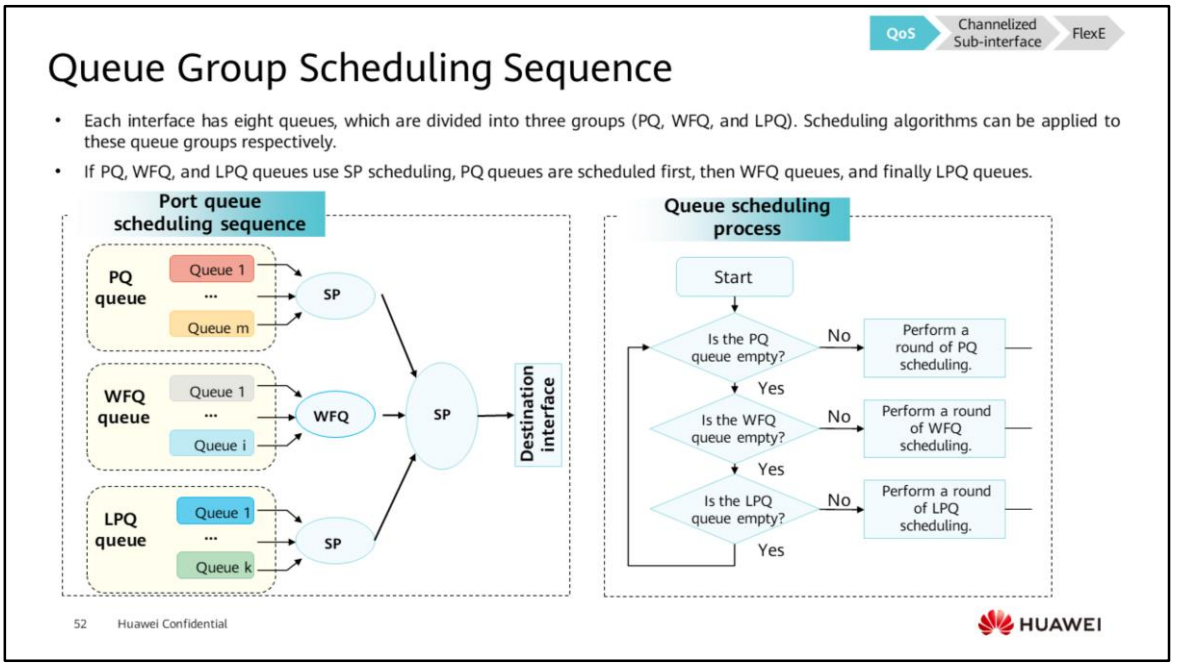
- Both traffic policing and traffic shaping are traffic rate limiting technologies. The former monitors the incoming traffic of a device and limits the incoming traffic rate to a permitted range. If the traffic rate is too high, excess packets are discarded or the packet priorities are re-set. Traffic 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.



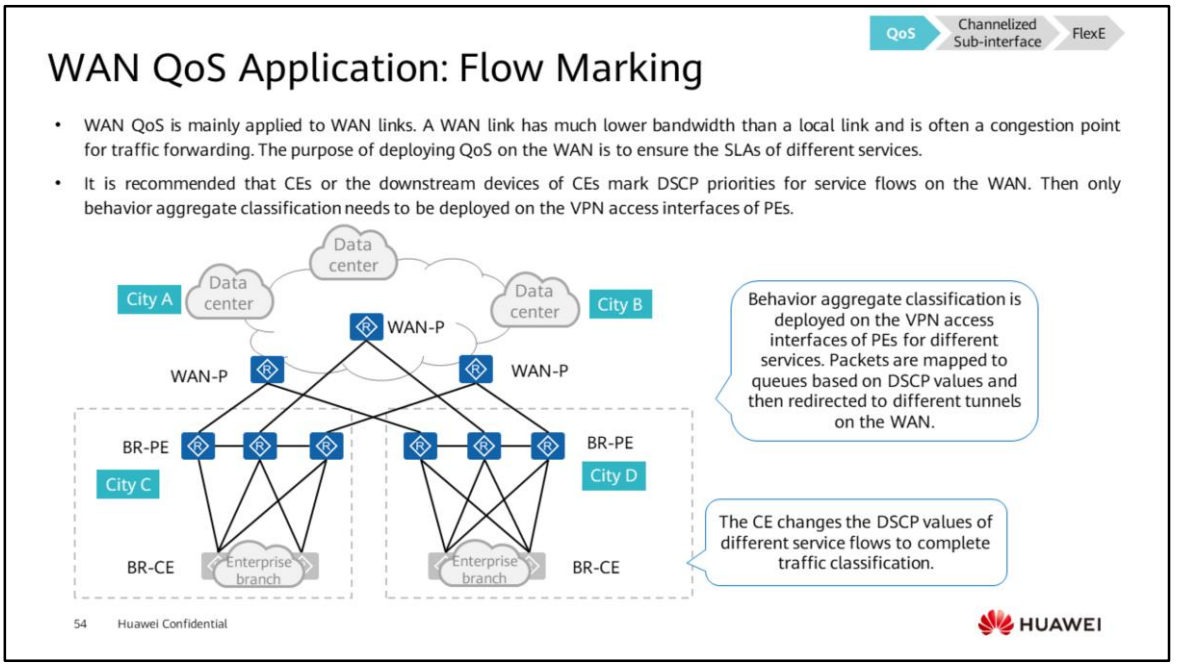
HUAWEI



- 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.



- 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

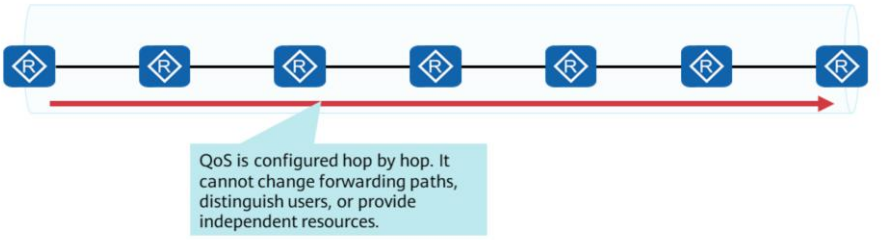
- The WAN's forwarding bottleneck lies in the egress WAN link in each region. For example, remote sites are interconnected through a carrier's MSTP private line, which has limited bandwidth. To prevent service flows from being discarded due to rate limiting after entering WAN links on the carrier network, 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

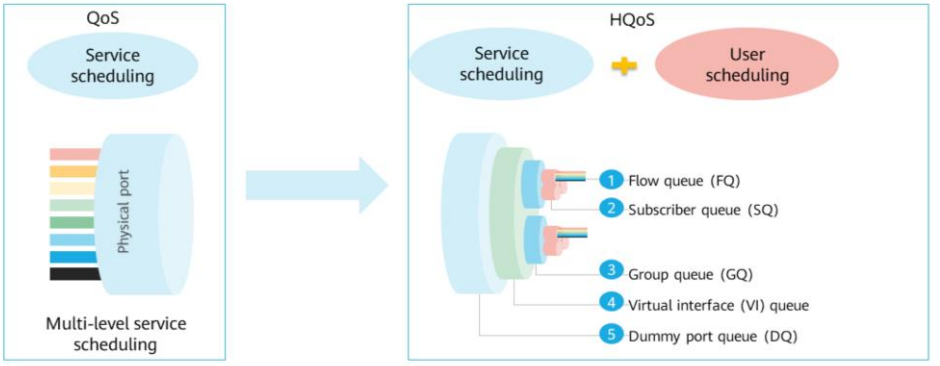
QoS Limitations

- 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.



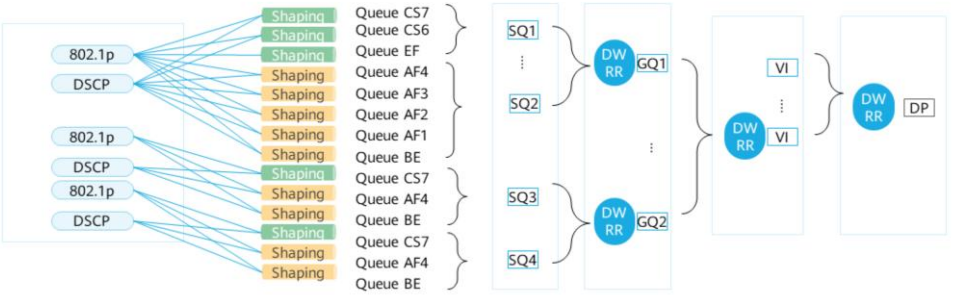
HQoS: Providing Scheduling and Bandwidth Guarantee for Users

- 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.



HQoS: Providing User-Level Bandwidth Guarantee When the CIR of Each Scheduler Does Not Exceed the Upper Limit

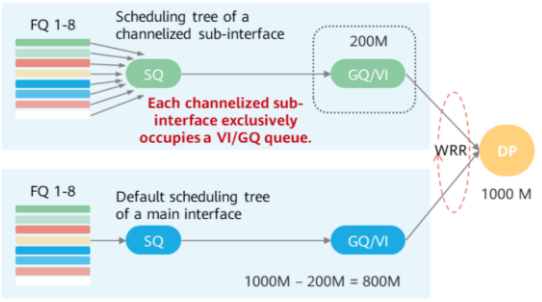
- HQoS adds hierarchical schedulers such as SQ, GQ, and VI. Each scheduler has two attributes (CIR and PIR) and uses SP scheduling for flows whose rates are between the CIR and PIR. The CIR is preferentially guaranteed.
- HQoS guarantees bandwidth through strict CIR deduction and shaping. The sum of 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

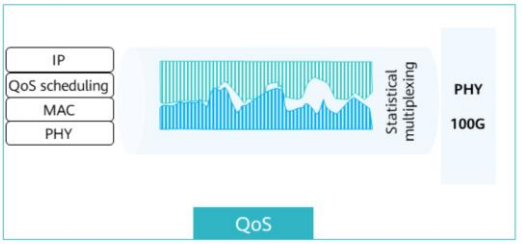
- Bandwidth guarantee: Channelized sub-interfaces use HQoS-based hierarchical scheduling to implement flexible and refined resource management. Each channelized scheduling tree has independent buffer resources and bandwidth resources, providing bandwidth guarantee.
- Delay guarantee: Because the resources for channelized sub-interface are strictly guaranteed, the delay can be guaranteed within a certain range.

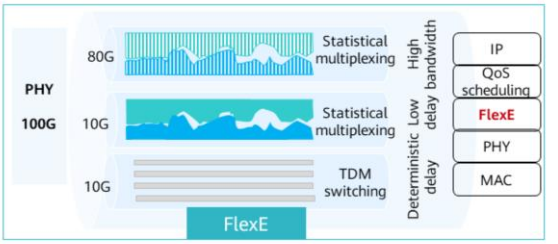


- Channelized sub-interfaces use the HQoS mechanism and exclusively occupy the HQoS VI/GQ scheduling trees and bandwidth to implement strict scheduling isolation.
- Remaining bandwidth of a main interface = Total interface bandwidth – Total bandwidth of all channelized sub-interfaces. Bandwidth is automatically 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.

FlexE: Delivering Isolation and Deterministic Delay Assurance Based on Independent Resources

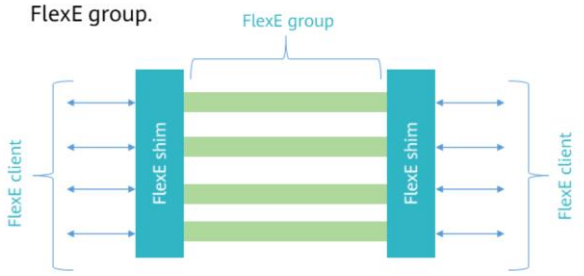
- Flexible Ethernet (FlexE) is a new technology introduced between the MAC and PHY layers. It is a lightweight enhanced technology for IP networks and is compatible with the existing Ethernet standard and QoS capabilities.
- It provides isolated FlexE interface links, enabling one network to carry different types of services.
- It provides an independent P2P time division multiplexing (TDM) tunnel for each service, meeting isolation and deterministic delay requirements.





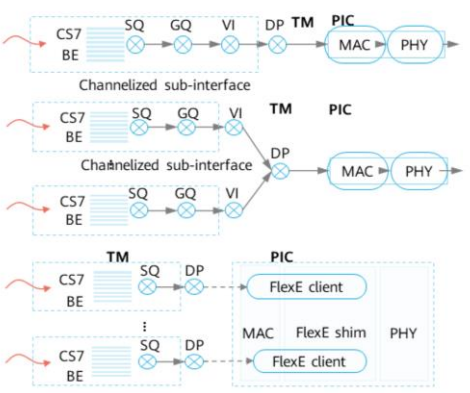
Basic FlexE Architecture

- 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.

Comparison of Slicing Technologies



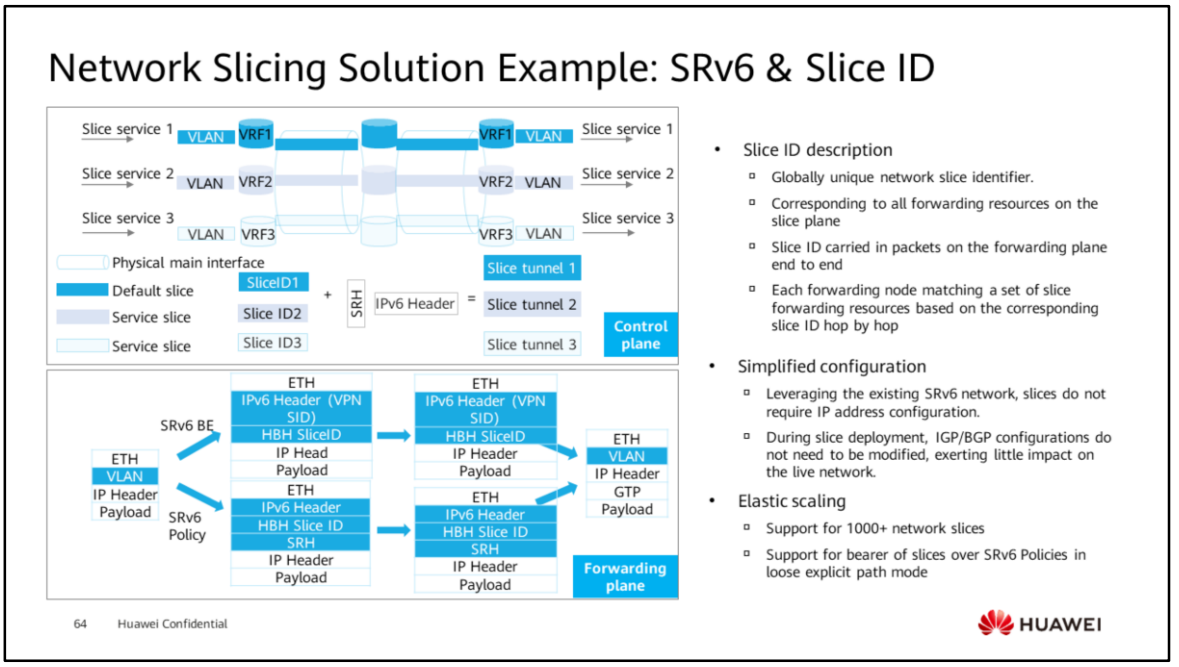
QoS: All traffic shares eight queues. QoS schedules resources in a unified manner to maximize the statistical multiplexing capability. It cannot differentiate users, and so cannot provide independent resource reservation for different users.

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.

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.



- 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

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.

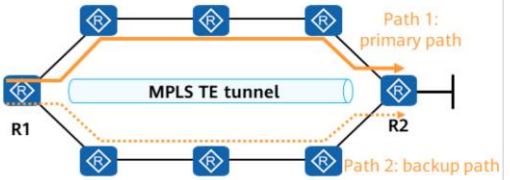
- Because different WAN VPN technologies use different terms, this section briefly describes various protection mechanisms, but does not describe specific protection technologies.

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.

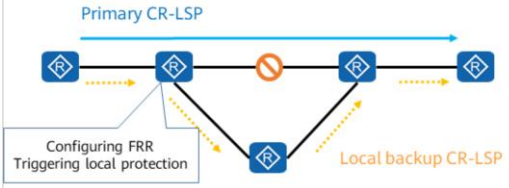
E2E protection

- 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

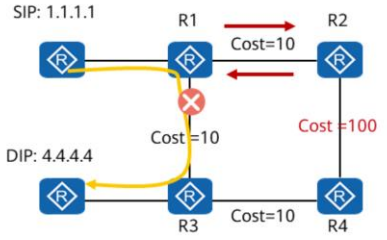
- MPLS TE E2E protection is classified into HSB and ordinary backup. In HSB protection, the backup path and primary path are created at the same time.
- Segment Routing adopts Topology Independent-Loop Free Alternate (TI-LFA), an enhancement of FRR, for local protection.
- Fast detection mechanism: Fast 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.

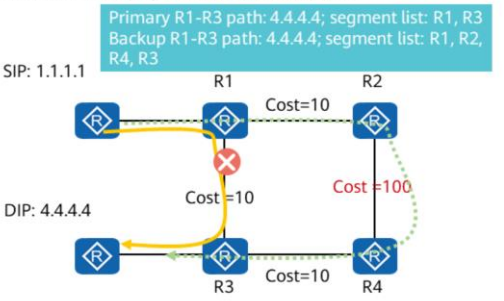
Limitations of the traditional LFA algorithm

- 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.

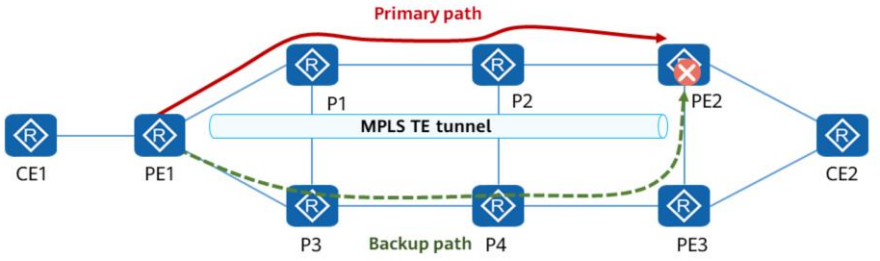


HUAWEI

- 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

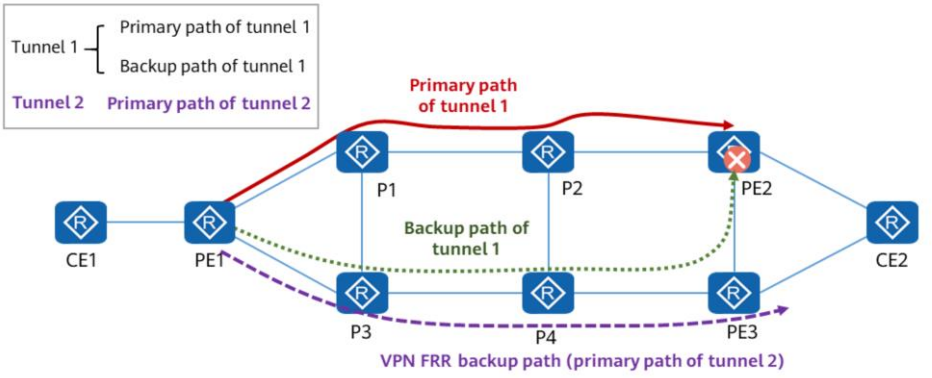
- Technologies such as HSB can protect E2E paths, but cannot rectify faults on PEs.
- In this example, a TE tunnel with both primary and backup paths is established between PE1 and PE2. If PE2 is faulty, the backup path cannot solve the problem.



HUAWEI

VPN FRR

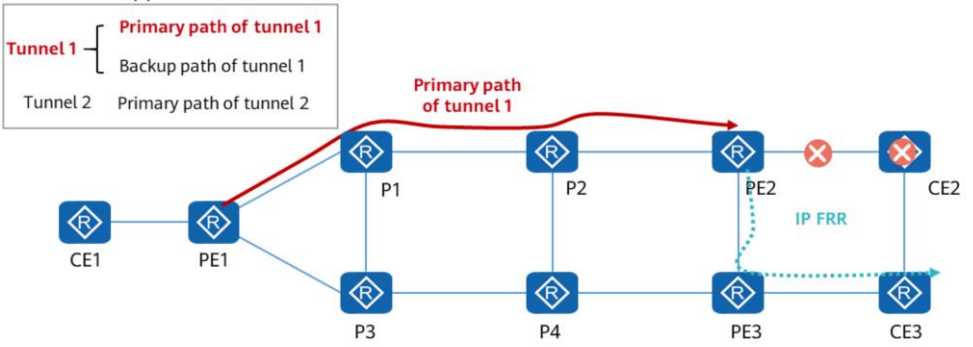
- VPN FRR sets the primary and backup forwarding paths pointing to the active and standby PEs on the remote PE in advance. It works with fast PE fault detection to accelerate fault-triggered E2E service convergence in scenarios where a CE is dual-homed to two PEs.



HUAWEI

IP FRR

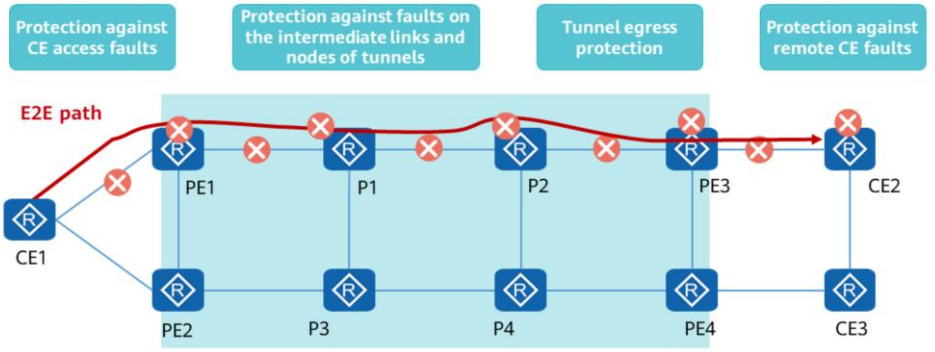
- If the link between PE2 and CE2 fails but PE2 still functions properly, the tunnel between PE1 and PE2 is still available. In this case, E2E tunnel switching is not required. PE2 selects PE3 as the backup next hop. If the link between PE2 and CE2 or CE2 fails, IP FRR is implemented to rapidly switch IP traffic.
- IP FRR is applicable to the IP network between CEs and PEs.



HUAWEI

Summary: Multi-Level Network Protection

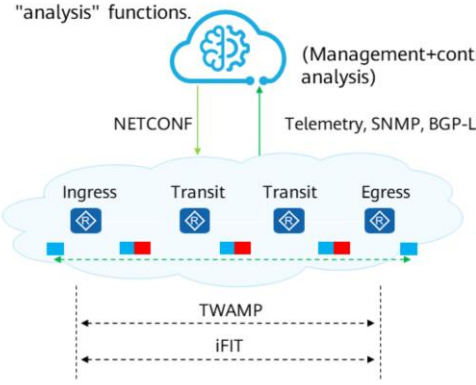
- To sum up, different protection measures are taken to ensure tunnel reliability based on the locations of faults on E2E paths.
- From the perspective of E2E forwarding paths, it is recommended that multi-level protection be used:



Contents

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.

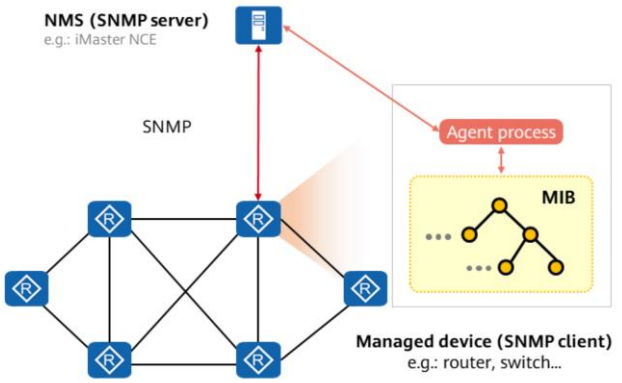
Network measurement

- 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.

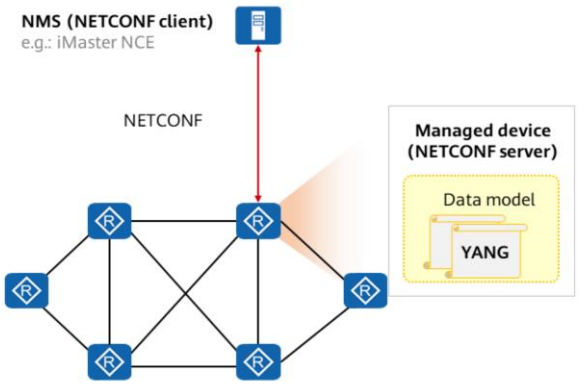
SNMP



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.

NETCONF



SNMP's drawbacks

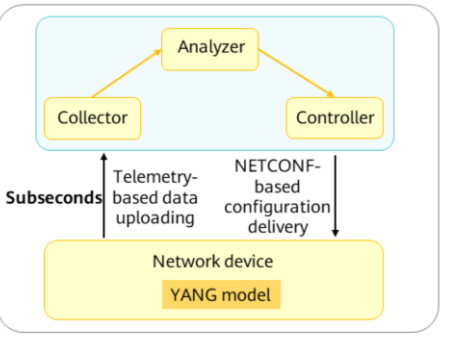
SNMP is not a configuration-oriented protocol. On a large-sized network with a complex topology, SNMP cannot meet network management requirements, especially the configuration management requirements.

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.

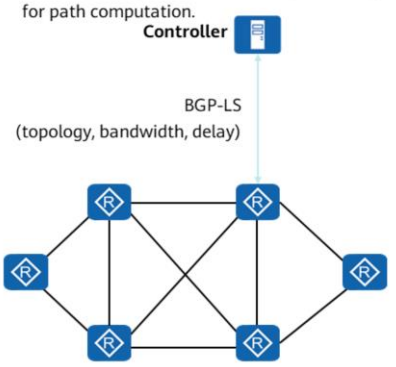
Telemetry

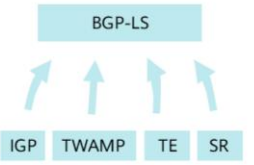
- Telemetry, also known as network telemetry, is mainly used to monitor networks, including packet check and analysis, intrusion and attack detection, intelligent data collection, and application performance management. Generally, it is used together with NETCONF. The analyzer analyzes the data collected 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.



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.



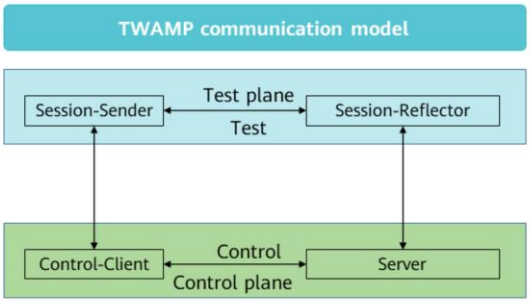


BGP-LS aggregates information collected by various network layer protocols, including:

- 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

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.

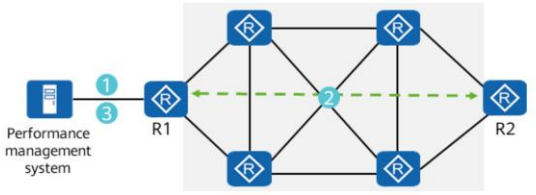
1. The Controller sends test packets to the Responder, and the Responder functions as the Session-Reflector to reflect the test packets.
2. The Session-Reflector does not need to learn Session status. After receiving the test packets, the Session-Reflector copies the necessary information, generates 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

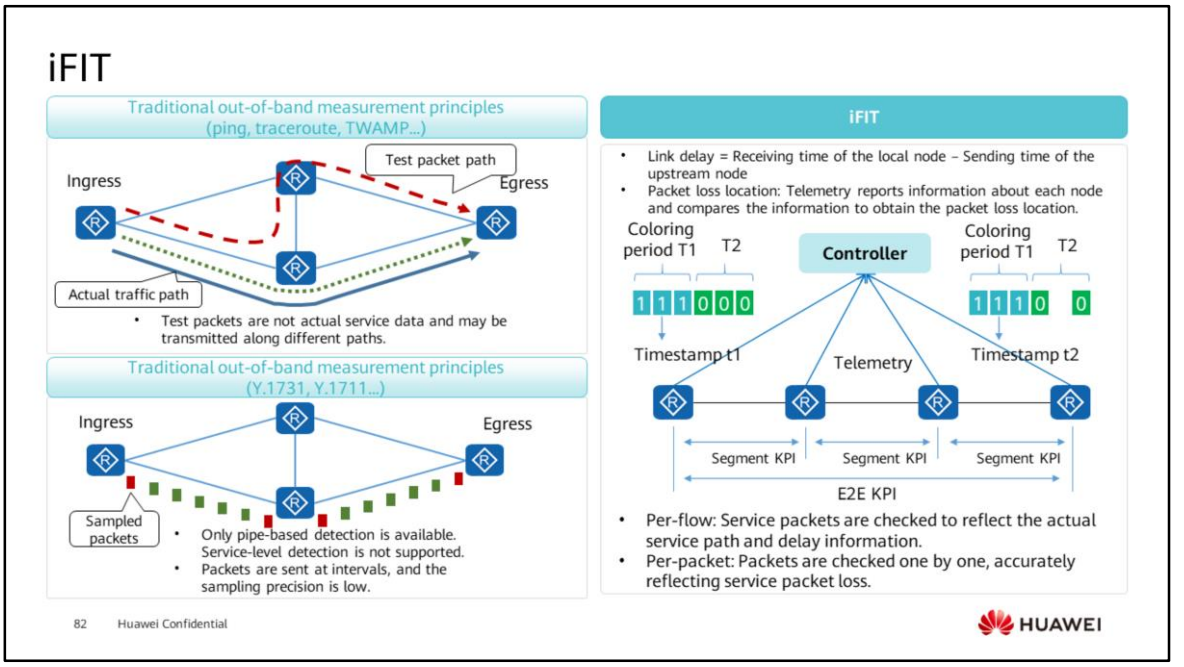
- The standard TWAMP version uses TCP for control plane negotiation, and test packets are based on UDP. The reflector needs to know the session status so that devices of different vendors can communicate with each other.
- TWAMP Light does not involve control plane negotiation, and test packets are also 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.



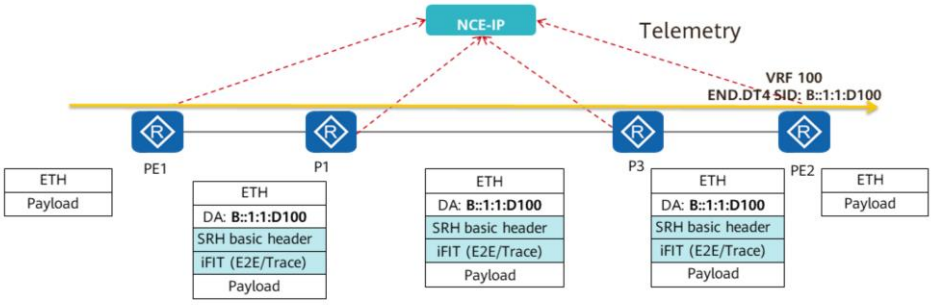
1. The Controller instructs network devices to establish performance measurement sessions. For example, R1 functions as the Control-Client to initiate IP performance measurement.
2. R1 and R2 initiate a TWAMP measurement session.
3. R1 reports the measurement result to the Controller.



- 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.

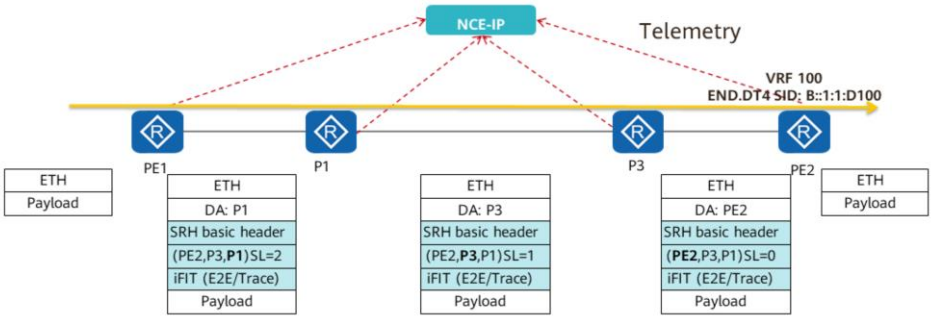


HUAWEI

- 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.

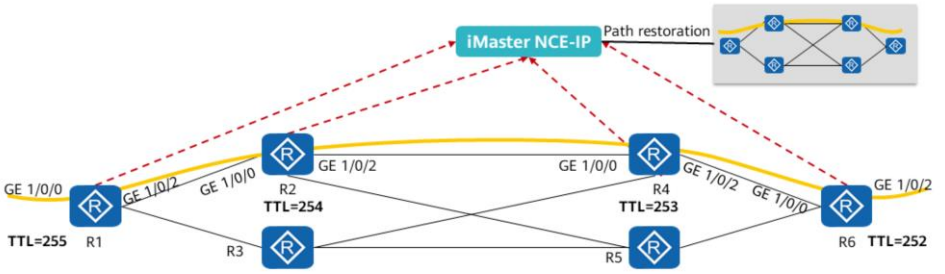
iFIT for SRv6 Policy Scenarios

- 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.



iFIT-based Service Path Display

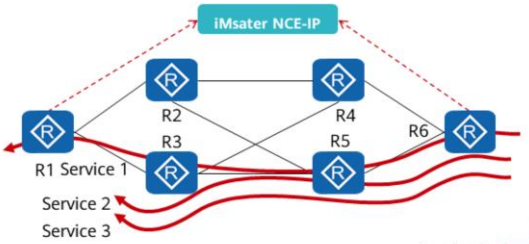
- In each measurement period, a device reports the flow direction, interface number, and TTL information when reporting packet statistics.
- iMaster NCE-IP restores the path information of the flow based on the information reported by each node.
- The implementation is independent of the tunnel type (SRv6/SRv6 Policy/SR-MPLS TE/SR-MPLS BE/MPLS...).



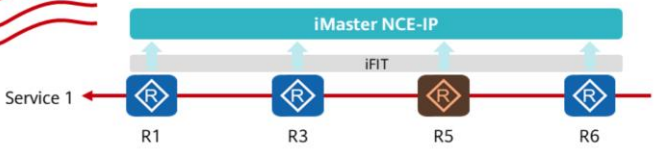
HUAWEI

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

- A
- 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.

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.

WAN VPN Technologies

HUAWEI

Foreword

- A virtual private network (VPN) is established using the WAN IP technology. A VPN, as its name suggests, has two features: virtual and private. Different VPNs share the underlying bearer network while logically isolating services.
- Before Segment Routing IPv6 (SRv6) came into place, a WAN VPN was generally carried over an MPLS network. And a WAN VPN carried over an MPLS network is called an MPLS VPN.
- This course introduces the models and classification of WAN VPN technologies, followed by the architecture fundamentals and evolution of these technologies.

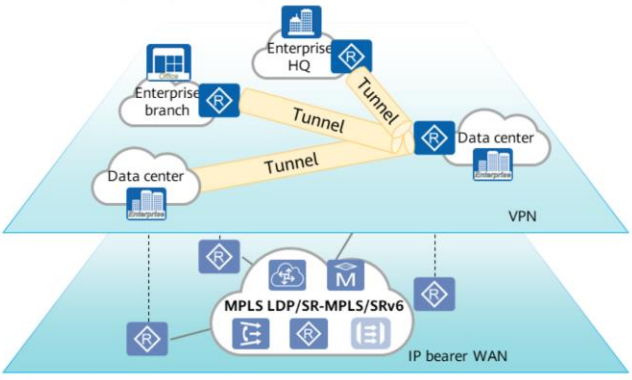
Objectives

- Upon completion of this course, you will be able to:
  - Describe the models and classification of WAN VPN technologies.
  - Describe the relationships between MP-BGP and WAN VPNs.
  - Describe the fundamentals of MPLS VPNs and SRv6 VPNs as well as differences between the two types of VPNs.
  - Describe the development trend of WAN VPN technologies.

Contents

WAN VPN Overview

- A virtual private network (VPN) is established using the WAN IP technology. A VPN, as its name suggests, has two features: virtual and private. Different VPNs share the underlying bearer network while logically isolating services.



Private

A VPN is a network dedicated to VPN users. For VPN users, a VPN provides the same service experience as a traditional private network. A VPN provides security assurance to protect VPN information against external threats. Different VPN services can be carried over the same underlying network while being isolated from each other.

Virtual

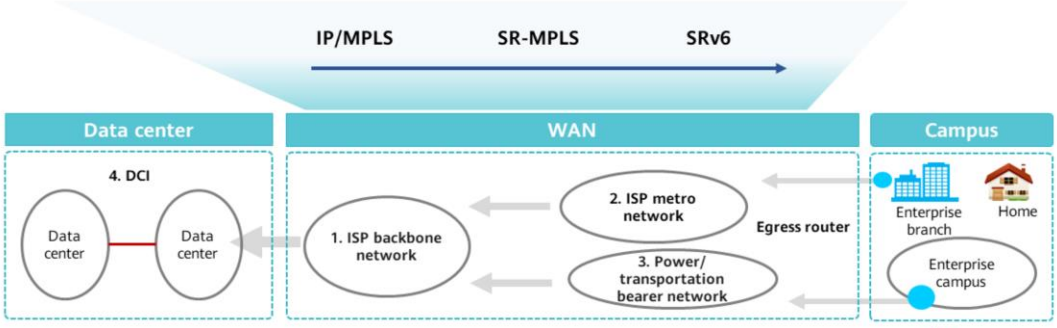
VPNs provide logical private networks over the same public physical bearer network through the virtualization technology. You can create multiple logical private networks based on service requirements.

HUAWEI

- In HCIP-Datacom-Core Technology > VPN Technology Overview, we learned that VPNs can be built over the Internet (e.g. IPsec VPN, L2TP VPN, SSL VPN, etc.) or private networks (MPLS VPN). In this course, we will learn the development and evolution of WAN VPN technologies.

Development of WAN IP Technologies

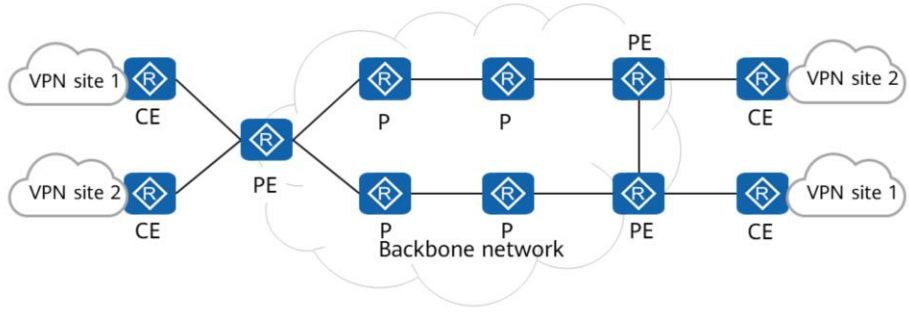
- WAN technologies are evolving towards all IP. For example, IP/MPLS replaces asynchronous transfer mode (ATM), and IEC 61850 enables IP-based power communication networks. WAN IP technologies are widely used in enterprise scenarios, such as power/transportation bearer networks, financial DCI networks, ISP backbone networks, and metro networks.
- WAN IP technologies are evolving from classic IP/MPLS to SR-MPLS and even SRv6.



HUAWEI

Basic WAN VPN Model

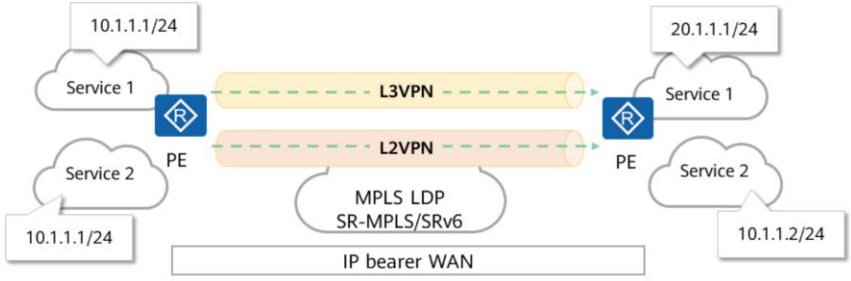
- A WAN VPN consists of three device roles: customer edge (CE), provider edge (PE), and provider (P).



- CE: user-side edge device that connects to service provider devices. A CE connects to one or more PEs for user access.
- PE: service provider edge device connecting to the CE. PEs are important network nodes that connect to both CEs and Ps.
- P: service provider device that does not connect to any CE.

Classification of Traditional WAN VPNs

- Depending on service types and network characteristics, VPNs are classified into L3VPNs and L2VPNs:
  - L3VPNs: carry Layer 3 services. L3VPNs use VPN instances to isolate services.
  - L2VPNs: carry Layer 2 services. L2VPNs use pseudo wires (PWs) to isolate services. Typical L2VPNs are VPWSs and VPLSs.



HUAWEI

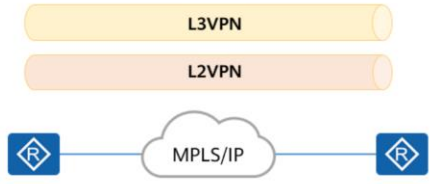
- As networks are moving towards all IP, most network services are Layer 3 IP services, and most VPNs are deployed as L3VPNs.
- Traditional L2VPN services include VPLS services and VPWS services. The VPLS service is a multipoint-to-multipoint L2VPN service, and the VPWS service is a point-to-point L2VPN service.

WAN VPN Classification Based on IP Bearer Technologies

- Depending on IP bearer technologies, WAN VPNs can be classified into MPLS VPNs and SRv6 VPNs.

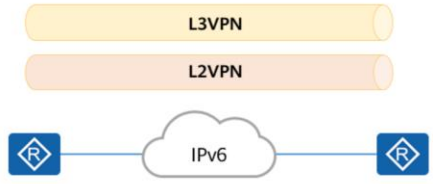
MPLS VPN

- The underlying bearer network of MPLS VPNs uses MPLS/IP on the forwarding plane. MPLS can carry L2VPN and L3VPN services.



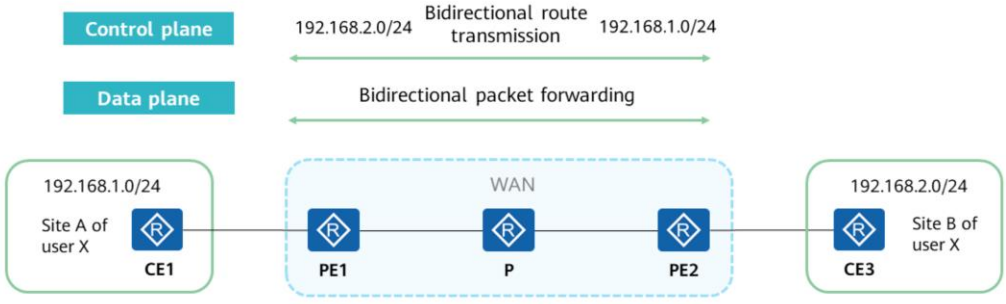
SRv6 VPN

- The underlying bearer network of SRv6 VPNs uses IPv6 on the forwarding plane.
- SRv6 is a segment routing technology that uses IPv6 addresses as segments and is widely considered as the evolution direction of next-generation WANs.



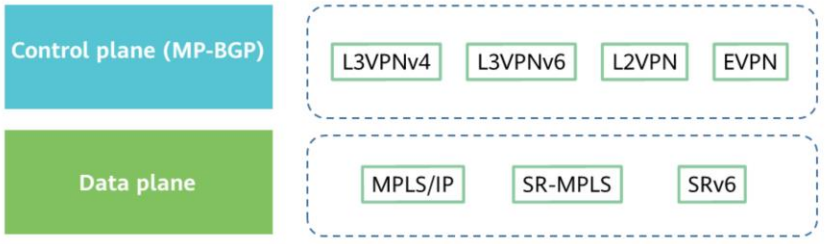
Hierarchical WAN VPN Technology Architecture

- The WAN VPN technology architecture can be simply divided into two planes: data plane and control plane.
- The control plane is responsible for transmitting CE-side route information, and the data plane is responsible for forwarding packets.



Overview of WAN VPN Technologies

- In the course of development, IP technologies are constantly evolving. There are multiple types of WAN VPN control and data plane technologies, and these technologies can be flexibly combined to meet the requirements of different service scenarios.
  - The control plane uses MP-BGP, which supports the use of different address families to transmit different route information.
  - The WAN data plane mainly uses MPLS/IP, SR-MPLS, or SRv6.

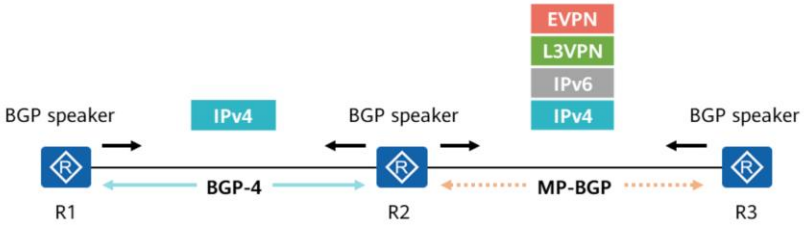


- L2VPN in the figure refers to the BGP L2VPN-AD address family on the Huawei device. It is used to exchange VPLS/VPWS information between MP-BGP peers.
- More data plane tunneling technologies, such as GRE, IPsec, and VXLAN, are used in other scenarios such as SD-WAN and DCN and are not described here.

Contents

MP-BGP Overview

- Multiprotocol Extensions for BGP-4 (MP-BGP), defined in RFC 4760, is used to extend BGP-4 to enable BGP to carry route information for multiple network layer protocols, such as IPv6, L3VPN, and EVPN.
- The extensions are backward compatible. That is, a router that supports MP-BGP can exchange route information with a router that supports only BGP-4.

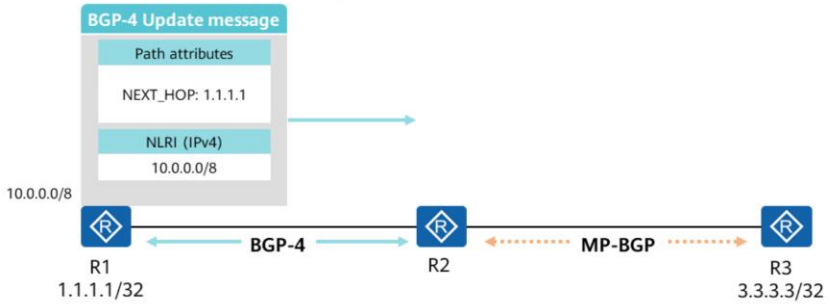


HUAWEI

- https://datatracker.ietf.org/doc/rfc4760/

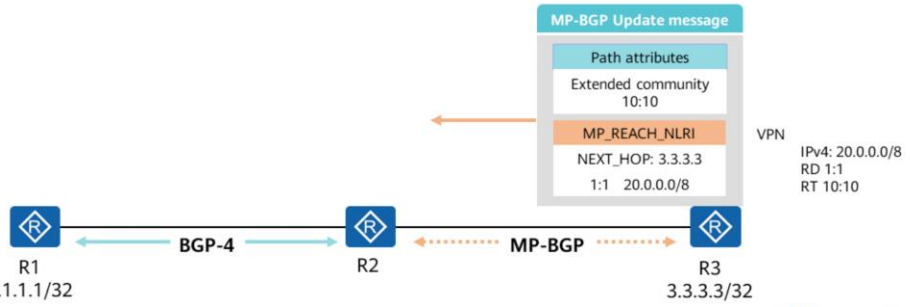
Example: BGP-4 Update Message

- BGP-4 has three attributes specific to the IPv4 address family: NEXT_HOP, AGGREGATOR, and IPv4 NLRI.
- On the network shown in the figure, R1 sends a BGP-4 Update message to R2. The advertised network segment is 10.0.0.0/8, and the next hop address is 1.1.1.1.
- In a BGP Update message, the NLRI field carries IPv4 network segment information, and the NEXT_HOP attribute in the Path attributes field carries the next hop address.



Example: MP-BGP Update Message

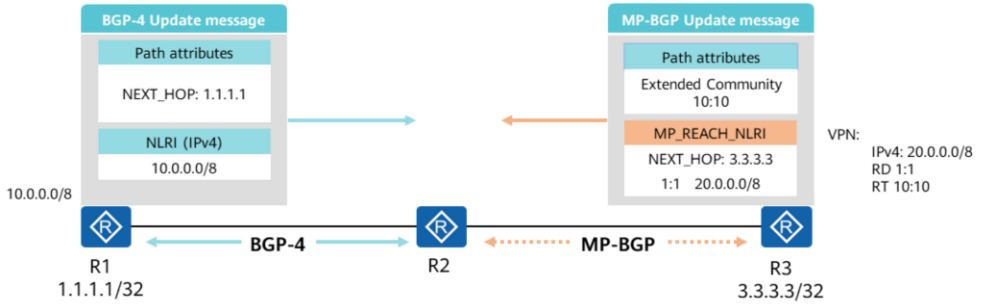
- To enable MP-BGP to support new address families and provide compatibility with IPv4, the IETF adds two more path attributes: MP_REACH_NLRI and MP_UNREACH_NLRI. The former indicates reachable destination information, and the latter indicates unreachable destination information.
- The structures of MP-BGP and BGP-4 Update messages are slightly different. On the network shown in the figure, R3 advertises the following VPNv4 route information to R2. The RD and IPv4 route information is stored in the new NLRI field, and the RT is stored in the community attribute of the Path attributes field.



HUAWEI

Comparison Between BGP-4 and MP-BGP Update Messages

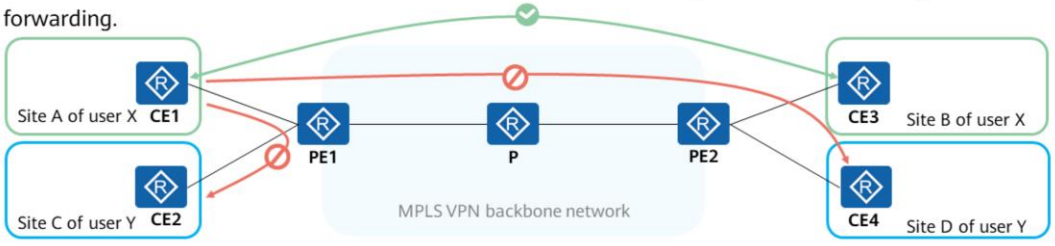
- MP-BGP defines new fields to carry route prefix and next hop information.
- The Update message structures of MP-BGP address families, such as IPv6, VPNv6, and EVPN, are the same.



Contents

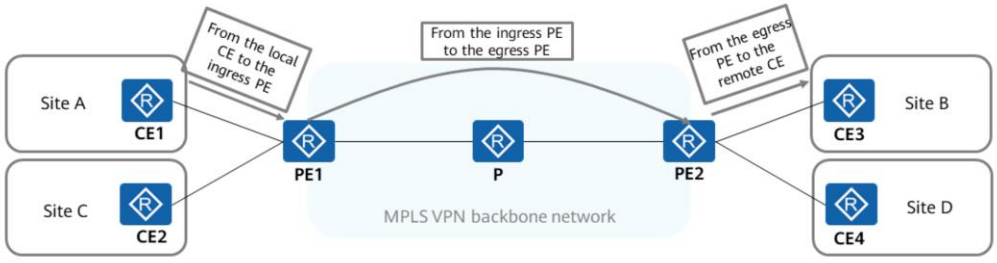
Common MPLS VPN Networking

- This section uses BGP/MPLS IP VPN as an example to describe MPLS VPN fundamentals.
- This example involves four CE sites, with two sites belonging to user X and two sites belonging to user Y. It is required that the sites of one user can communicate with each other but not with those of the other user.
- Communication between site A and site B of user X involves two phases: route exchange and data forwarding.



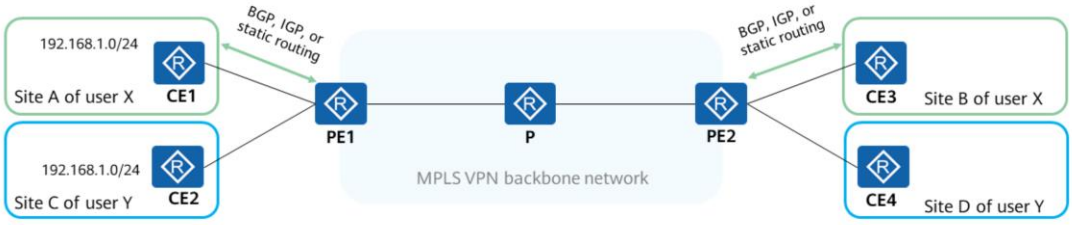
MPLS VPN Route Advertisement

- 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



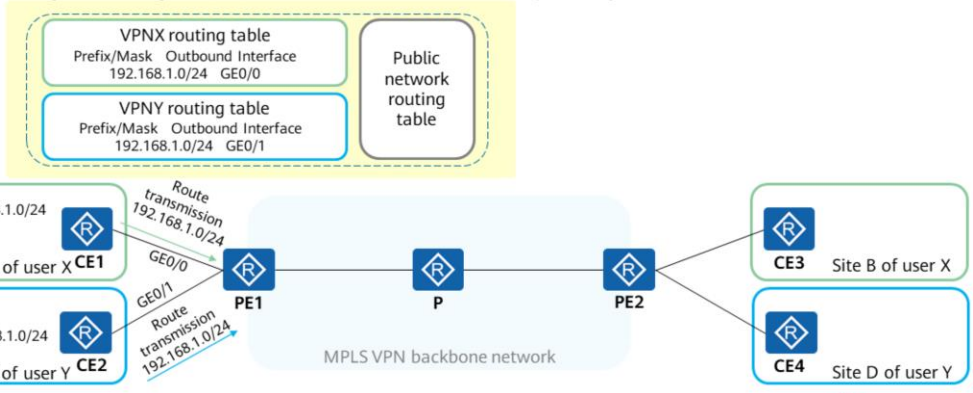
Route Exchange Between CEs and PEs

- The CEs and PEs can use static routing, 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 exchanges routes with the remote CEs.



Route Transmission from the Ingress PE to the Egress PE — VPN Local Isolation

- After receiving routes from CEs, PEs need to independently store routes of different VPNs and solve the problem caused by different customers using overlapping IP address spaces.
- The virtual routing and forwarding table (VRF), also called the VPN instance, is used to implement logical isolation between local VPNs.

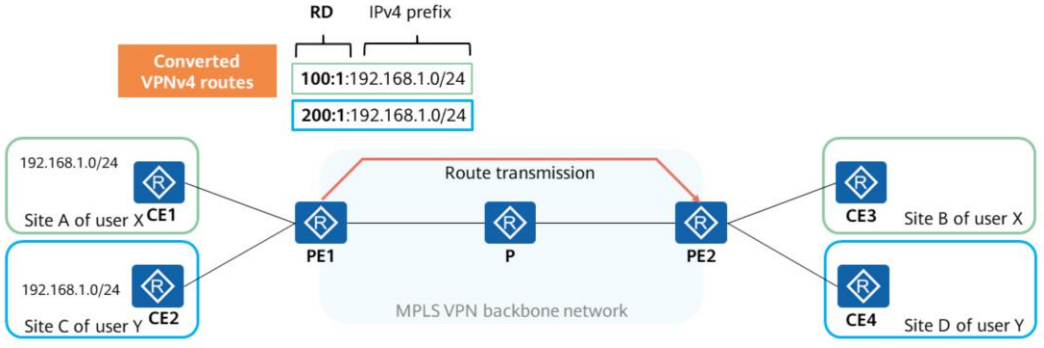


HUAWEI

Route Advertisement | Data Forwarding

RD: Distinguishing Routes During Route Distribution

- A VPN instance is only of local significance. It is used to isolate and distinguish routes on the local PE. When a local PE transmits route information to a remote PE, how can the PE distinguish routes with the same prefix? To solve this problem, route distinguishers (RDs) are introduced to distinguish routes with the same prefix and mask but different destinations.
- After receiving IPv4 routes from a CE, a PE adds RDs to these routes to convert them into globally unique VPN-IPv4 (VPNv4) routes.

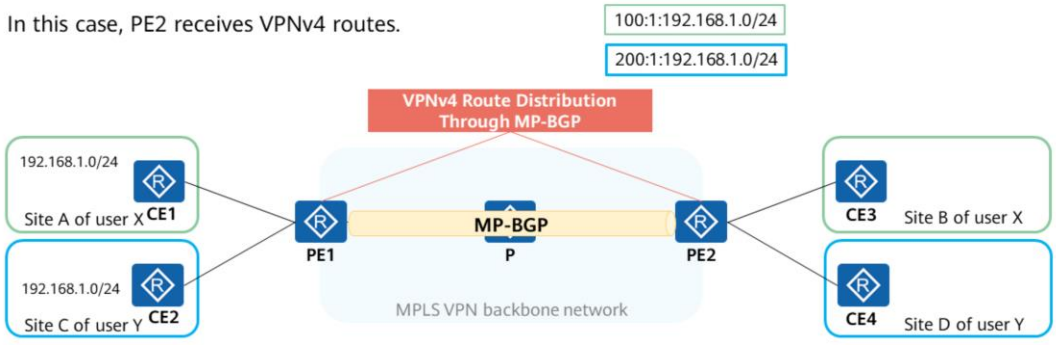


HUAWEI

- RD, 8 bytes long.
- The common RD formats are as follows:
  - 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)

VPNv4 Route Distribution Through MP-BGP

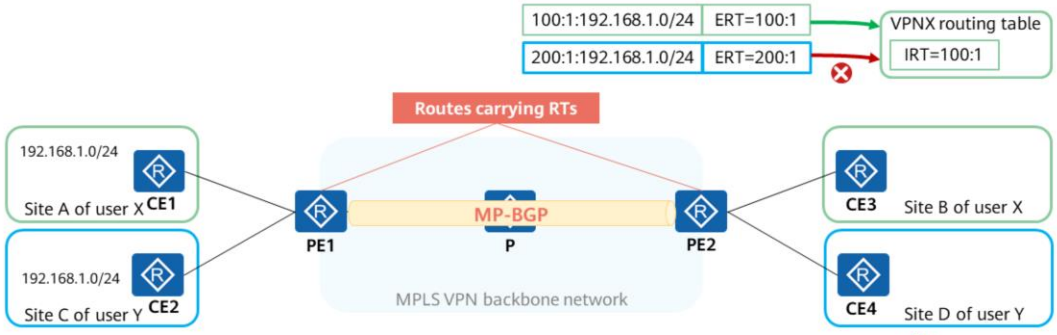
- Traditional BGP-4 cannot process VPNv4 routes.
- PEs need to establish MP-BGP peer relationships with each other and use MP-BGP to exchange VPNv4 routes.
- In this case, PE2 receives VPNv4 routes.



HUAWEI

RT: Controlling Route Import and Export

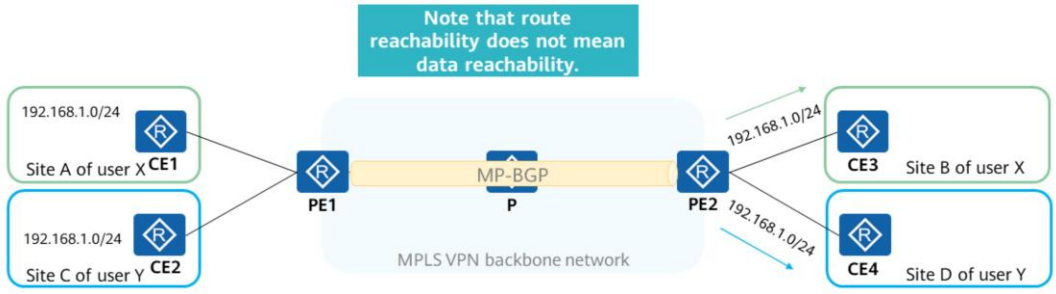
- In this case, PE2 receives two VPNv4 routes and has two VPN instances corresponding to sites B and D respectively. How does PE2 determine the import and export of VPN routes?
- The length of an RT is 32 bits. A VPNv4 route can carry one or more RTs. There are two types of RTs: import RT (IRT) and export RT (ERT). When creating a VRF on a PE, you need to specify the IRT and ERT. If an ERT carried in a VPNv4 route received by a PE is the same as an IRT of a local VRF on the PE, the PE imports the route into the VRF.



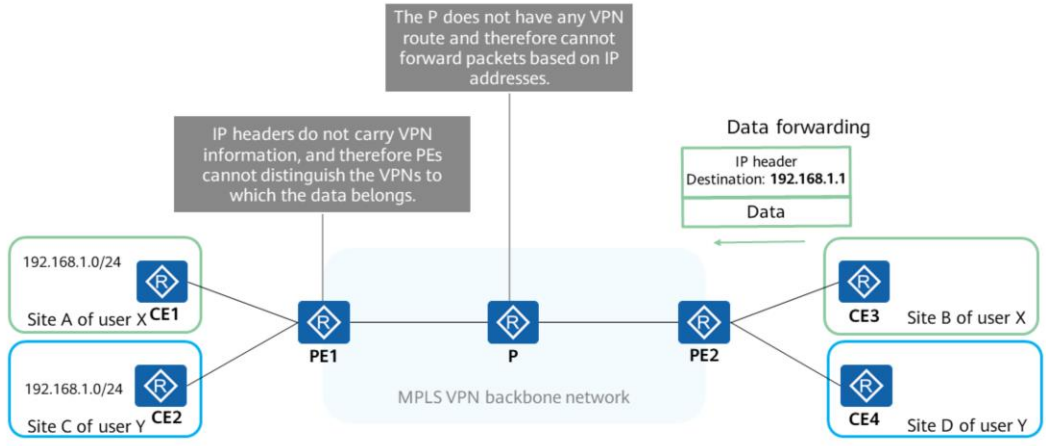
HUAWEI

Route Transmission from the Egress PE to the Remote CE

- PE2 removes the RD from VPNv4 routes and advertises them to the corresponding CEs.
- Similarly, other sites also advertise routes to each other. In this way, different sites of the same user are routable to each other.

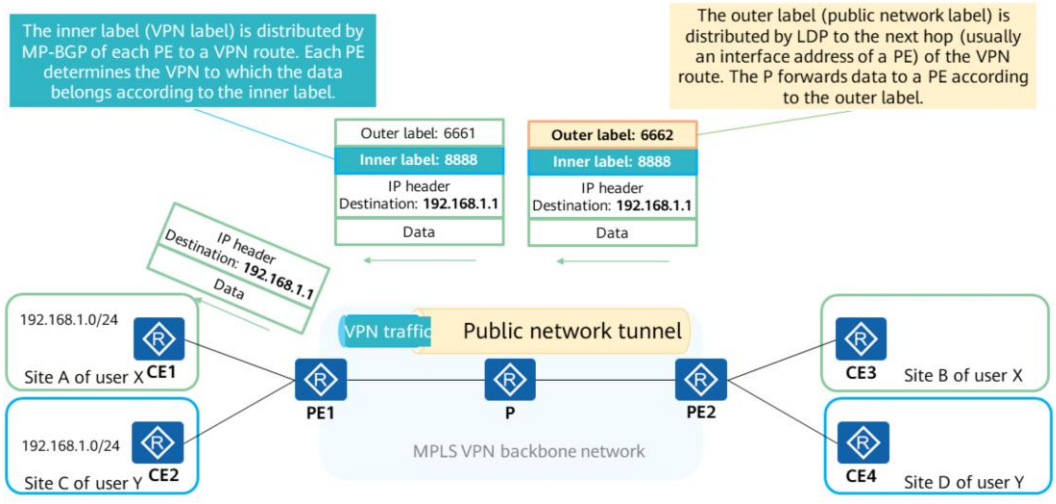


Problems Encountered During Data Forwarding

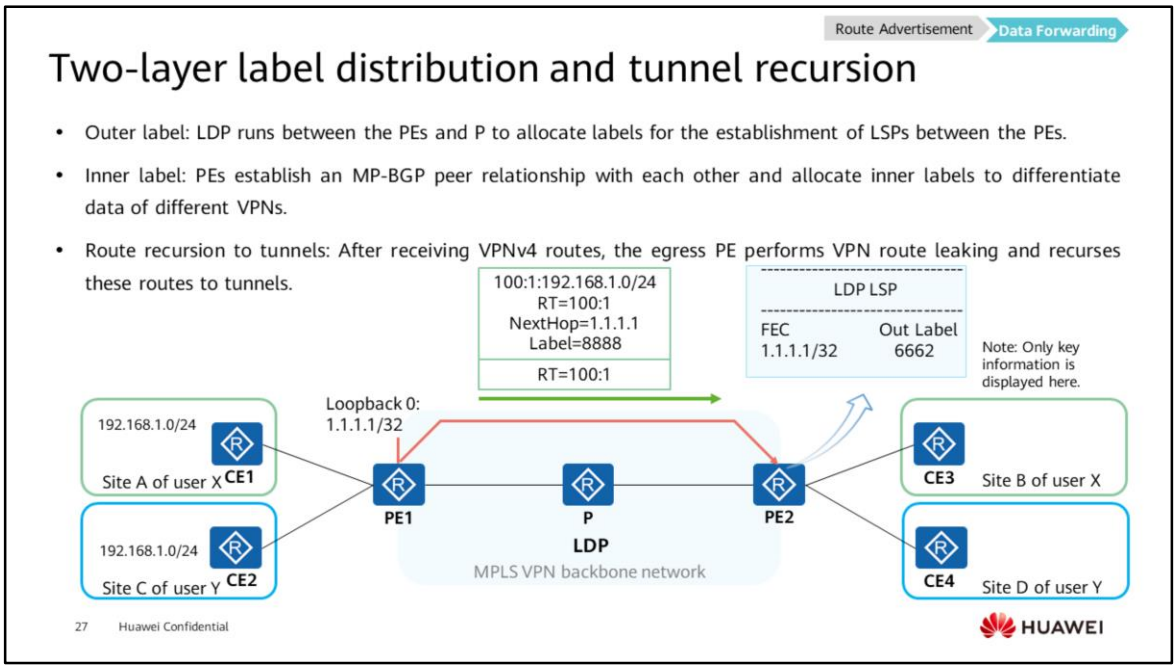


Route Advertisement | Data Forwarding

Problem Solving with Two Layers of Labels



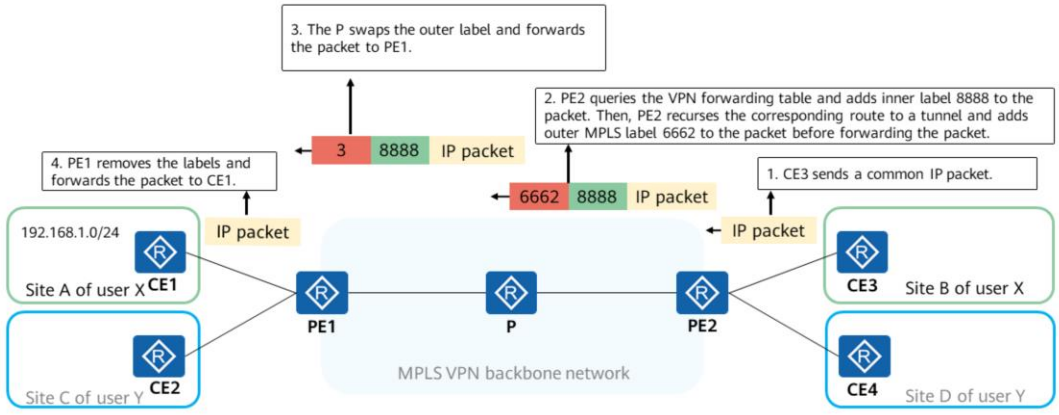
HUAWEI



- PEs can allocate VPN labels in either of the following ways:
  - Route-based MPLS label allocation: Each route in the VPN instance is assigned a label (one label per route). The disadvantage of this method is that when there are a large number of routes, more entries need to be maintained in the Incoming Label Map (ILM) table of the device, which increases the requirement on device capacity.
  - VPN instance-based MPLS label allocation: A label is allocated to the entire VPN instance. All routes in the VPN instance share the same label. The advantage of using this allocation method is that label resources are conserved.
- VPN route leaking: VPNv4 routes are matched against the VPN targets of the local VPN instance. After receiving a VPNv4 route, the PE directly imports the route to the local VPN instance without preferentially selecting the route or checking whether the tunnel to which the route recurses exists.
- Recursion to tunnels: To transmit VPN traffic to the other end through the public network, a public network tunnel is required to carry the VPN traffic. Therefore, after VPN routes are leaked, routes need to recurse to tunnels based on the destination IPv4 prefix. That is, the next hop of the IPv4 route has a corresponding LSP. The route is added to the routing table of the corresponding VPN instance only if the route can successfully recurse to a tunnel.

Data Forwarding Process

- The following uses packet forwarding from CE3 to CE1 as an example.



HUAWEI

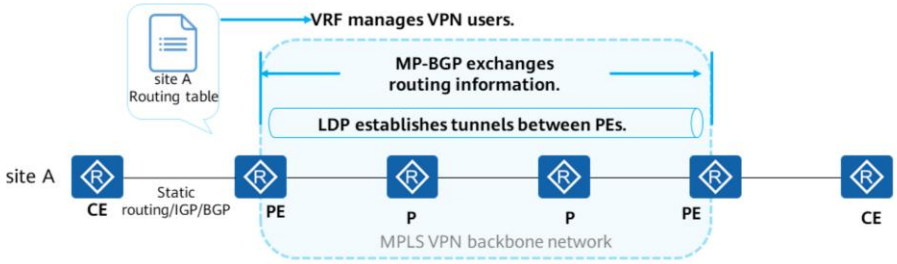
MPLS VPN Summary

- As one of the most typical WAN VPNs, the MPLS VPN with an architecture of separated forwarding and control planes is widely used in the industry.
- The control plane uses MP-BGP to transmit VPNv4 routes and inner labels.
- The data plane uses LDP to generate MPLS forwarding tunnels. During data forwarding, outer labels are generated upon tunnel recursion.
- Despite the development of WAN VPN technologies, the idea of separating the forwarding plane from the control plane is still used.

Contents

BGP/MPLS IP VPN Technology Architecture

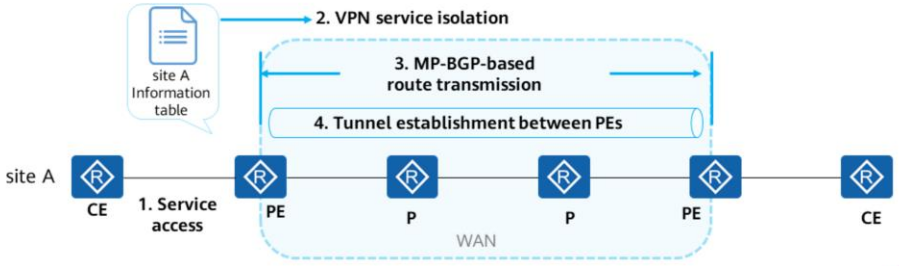
- BGP/MPLS IP VPN is not a single VPN technology. Rather, it is a comprehensive solution that combines multiple technologies, including:
  - MP-BGP: transmits site route information between PEs.
  - LDP: establishes tunnels between PEs.
  - VRF: manages VPN users on PEs.
  - Static routing, IGP, or BGP: exchanges route information between PEs and CEs.



HUAWEI

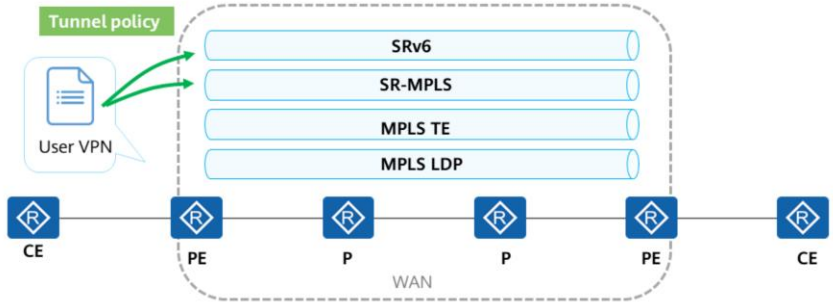
Abstracted WAN VPN Technology Architecture

- The WAN VPN technology architecture, which abstracts key BGP/MPLS IP VPN technologies, consists of four functional modules:
  1. Service access
  2. VPN service isolation
  3. MP-BGP-based route transmission
  4. Tunnel establishment between PEs



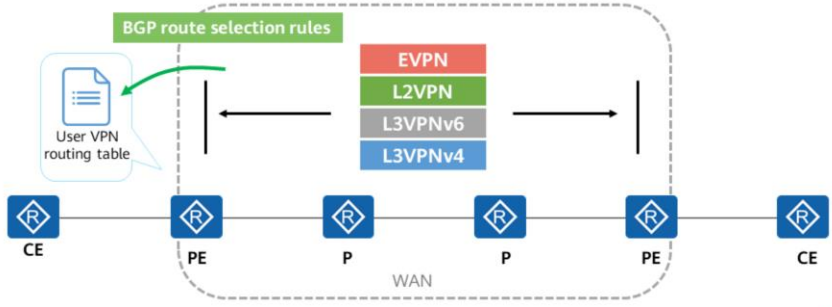
WAN Data Plane Development - SR

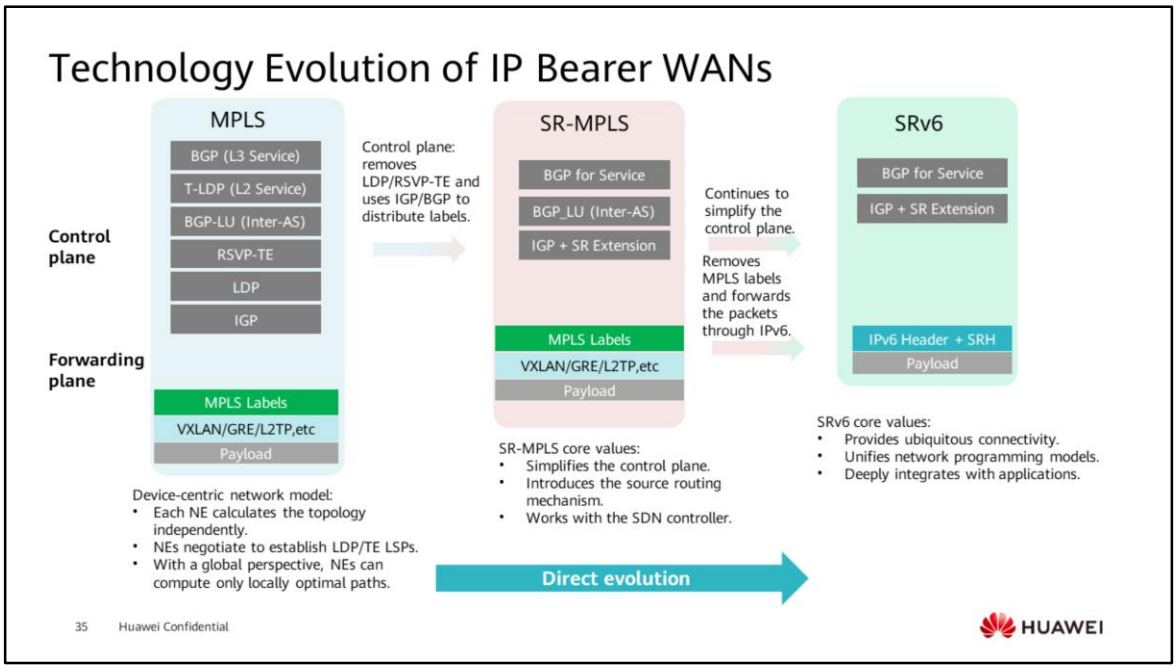
- In the WAN VPN, the data plane is presented as data forwarding tunnels.
- Common MPLS and SR forwarding tunnels include MPLS (LDP), MPLS TE, SR-MPLS, and SRv6 tunnels.
- Multiple tunneling technologies can be deployed on a WAN, and tunnel policies can be used to determine tunnels for route recursion.



WAN Control Plane Development - EVPN

- In the WAN VPN, the control plane uses MP-BGP.
- The MP-BGP address families gradually evolve from VPNv4, L2VPN, and other address families to the EVPN address family, with EVPN unifying the entire control plane.
- The control plane supports route exchange among multiple address families, but only the optimal routes are added to the VPN routing table.

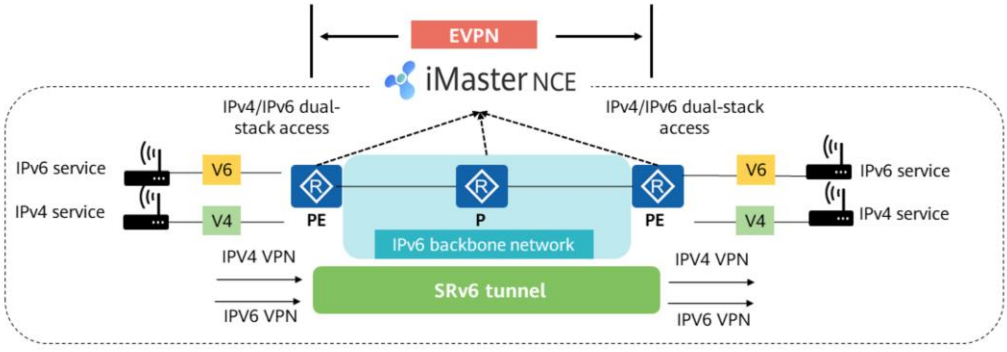




- Target LDP (T-LDP) is used for L2VPN services and distributes VC labels to L2VPN PWs.
- BGP Labeled Unicast (BGP-LU) (RFC 3017) is a routing protocol used in both inter-AS and intra-AS scenarios. It advertises MPLS paths between IGP areas or ASs. These routes may span one or more router hops.

Best Practices of IP Bearer WANs

- Huawei iMaster NCE manages WAN forwarders in a unified manner to enable the SRv6+EVPN-capable intent-driven IP WAN bearer network.



HUAWEI

Quiz

1. (Short-answer question) What is the relationship between EVPN and L2VPN?( )
2. (Single-answer question) In a BGP/MPLS IP VPN scenario, which of the following protocols is used to allocate outer labels?( )
   A. OSPF
   B. MP-BGP
   C. MPLS LDP
   D. RSVP-TE

- L2VPN (VPLS as an example) has many drawbacks, such as complex service deployment, limited network scale, and lack of dual-homing support. The IETF proposes EVPN (RFC 7432) to overcome these drawbacks. EVPN is originally designed as an L2VPN technology based on BGP extensions. However, EVPN also supports L3VPN with protocol extensions.
- C

Summary

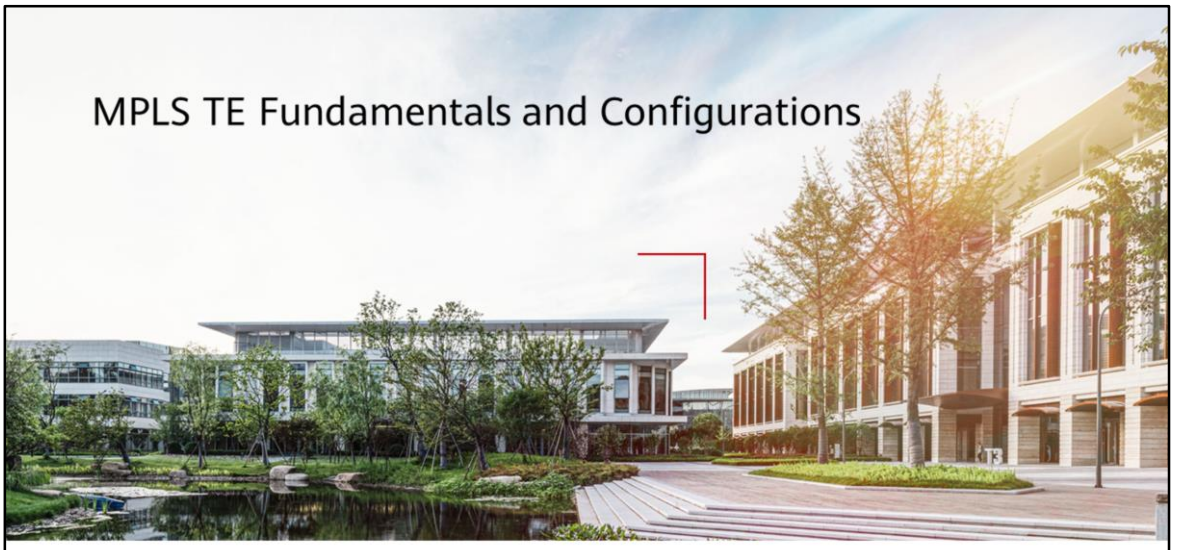
- Depending on IP bearer technologies, WAN VPNs are classified into MPLS VPNs and SRv6 VPNs. With the evolution of IPv6 networks, SRv6 and IPv6 naturally merge with each other. Therefore, SRv6 VPN is an inevitable trend.
- The WAN VPN control plane is converging. EVPN, which integrates L2VPN and L3VPN capabilities, is a best practice of WAN VPNs.
- The advantages of SR need to be demonstrated through the centralized path computation, global optimization, and other capabilities of the controller. For more information, refer to 05 Segment Routing.

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.

MPLS TE Fundamentals and Configurations



HUAWEI

Foreword

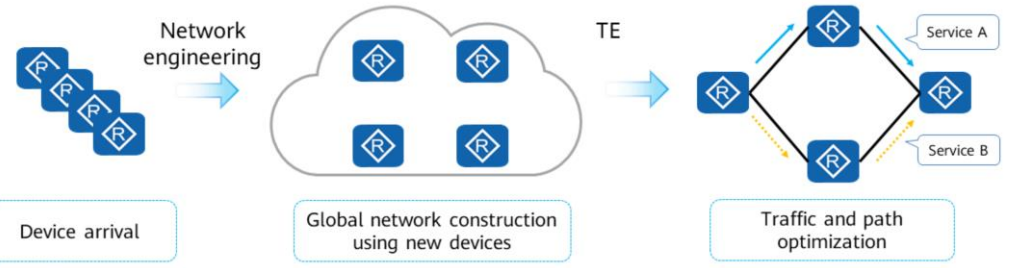
- Traffic engineering (TE) allows network nodes to establish data forwarding paths based on available resources on a network and reserve network bandwidth for critical traffic. By dynamically monitoring network traffic and device load and adjusting traffic management, routing, resource constraint, and other parameters in real time, traffic engineering optimizes network resource utilization and prevents traffic congestion arising from load imbalance.
- Short for MPLS traffic engineering, MPLS TE establishes constraint-based routed label switched paths (CR-LSPs) and diverts traffic to the CR-LSPs, allowing network traffic to be sent over specified paths.
- This course describes the fundamentals and features of MPLS TE.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the functions and implementation of MPLS TE.
  - Explain how MPLS TE works and its data forwarding modes.
  - Illustrate data protection and recovery of MPLS TE.
  - Explain advanced MPLS TE features.

Contents

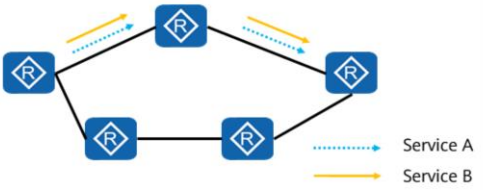
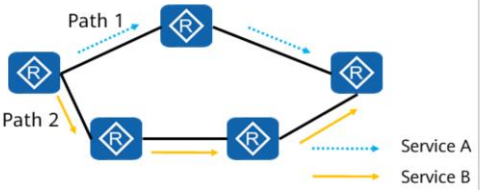
Concepts of TE



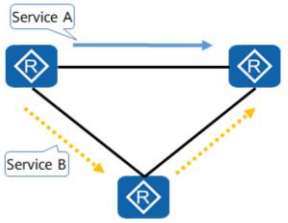
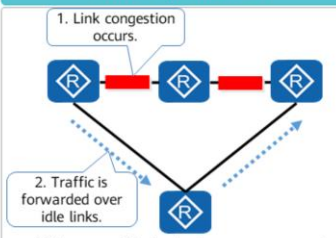
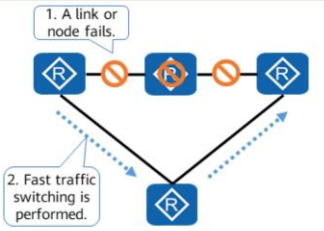
- Network engineering is a process of meeting traffic requirements through network design. In essence, it is a process of planning, designing, and deploying networks based on traffic requirements.
- TE allows network nodes to establish data forwarding paths based on available resources on a network and reserve network bandwidth for critical traffic. By dynamically monitoring network traffic and device loads and adjusting traffic management, routing, resource constraint, and other parameters in real time, TE optimizes network resource utilization and prevents traffic congestion arising from load imbalance.
- If network engineering is compared to road construction, reconstruction, and expansion, TE plays the roles of relieving traffic congestion and ensuring smooth traffic.

- The essence of network engineering is to understand network traffic requirements through surveys, and then design and deploy networks according to the actual requirements, which is a process of building new networks.
- TE is implemented on existing networks. It optimizes resource configuration and improves network performance through proper traffic planning.

Why Is Traffic Engineering Used?

Conventional IP network	Post-TE network
	
• On a conventional IP network, nodes select the shortest path as the optimal route regardless of factors such as bandwidth. This easily causes the shortest path to be congested with traffic, whereas other available paths are idle.	• By establishing CR-LSPs, TE allows specific traffic to be transmitted along specified paths based on traffic characteristics when traffic arrives at the ingress of CR-LSPs. • TE allows network administrators to deploy different forwarding paths to properly allocate network resources and prevent network congestion.

Major Functions of TE

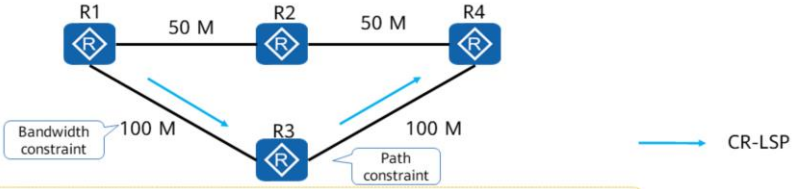
Path planning	Traffic optimization	Fault protection
		
• Different forwarding paths are planned for different service traffic.	• When traffic imbalance occurs due to major events or news, TE diverts some traffic to idle links so that network-wide traffic is properly distributed.	• TE supports fast protection switching in the case of a device or link fault.

Overview of MPLS TE

- RFC 2702 (Requirements for Traffic Engineering Over MPLS), published in September 1999, elaborates on the necessary conditions for implementing MPLS TE, laying a solid foundation for the development and application of MPLS TE.
- MPLS TE can effectively schedule, allocate, and utilize existing network resources and provide bandwidth and QoS guarantee for network traffic without needing to upgrade hardware, thereby minimizing costs. Because MPLS TE is implemented based on MPLS, it is therefore easy to deploy and maintain MPLS TE on existing networks.
- MPLS TE is a perfect combination of MPLS and TE and provides E2E service guarantee for core and backbone networks of large network service providers.

MPLS TE Tunnels

- MPLS TE often associates multiple LSPs with a virtual tunnel interface, and such a group of LSPs is called an MPLS TE tunnel.
- LSPs in an MPLS TE tunnel are called constraint-based routed LSPs (CR-LSPs).
- Constraints include bandwidth constraints and path constraints. A CR-LSP can be successfully established only when the links over which a tunnel traverses satisfies the constraints. Constraints allow services to be better planned based on the existing network conditions.



On the ingress (R1), create an MPLS TE tunnel that passes through R3, with the destination address being R4's address and bandwidth being 100 Mbit/s.

MPLS TE Implementation

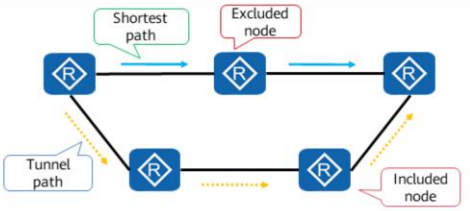
- Currently, mainstream MPLS TE implementation modes are as follows:
  - Resource Reservation Protocol-Traffic Engineering (RSVP-TE)
    - RSVP-TE is an extension of RSVP — Resource Reservation Protocol — designed to support TE.
    - RSVP-TE technology is mature and has been applied on a large scale. However, the technology is complex with poor extensibility.
  - Constraint-based Routing Label Distribution Protocol (CR-LDP)
    - CR-LDP is an extension of LDP.
    - CR-LDP technology is not mature and is seldom used. However, it is simple with good extensibility.
  - Currently, RSVP-TE is widely used in the industry.

- For details about RSVP-TE, see the section "MPLS TE Fundamentals".

Major MPLS TE Functions (1)

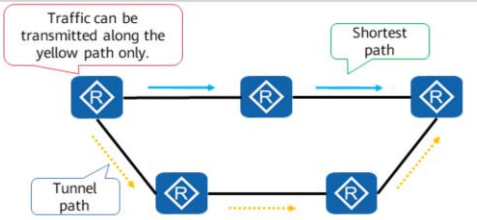
- MPLS TE tunnels support explicit paths, which can include or exclude specific nodes as required.
- Explicit paths include strict and loose explicit paths.

Loose explicit path



- In loose mode, no specific path is specified, and only the nodes to be included or excluded are specified on the tunnel path.

Strict explicit path



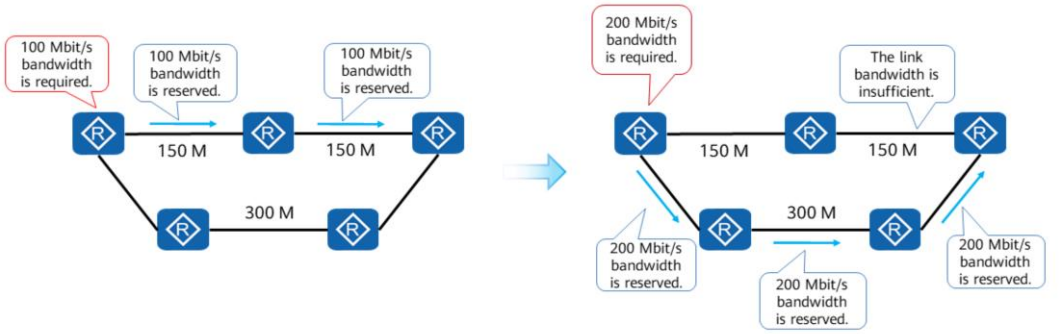
- In strict mode, tunnel paths are fixed.

HUAWEI

- When a CR-LSP is set up, you can manually specify the nodes that the CR-LSP must traverse or bypass. The path is called an explicit path in MPLS TE.
- For details about explicit paths, see the following sections.
- If no explicit path is configured as shown in the preceding two figures, traffic is forwarded along the shortest path. If an explicit path is configured, traffic is forwarded along the configured path.

Major MPLS TE Functions (2)

- MPLS TE supports resource reservation on LSPs to ensure E2E bandwidth.
- MPLS TE supports path re-optimization for LSPs upon network resource changes.
- MPLS TE supports automatic bandwidth adjustment for LSPs upon bandwidth changes.

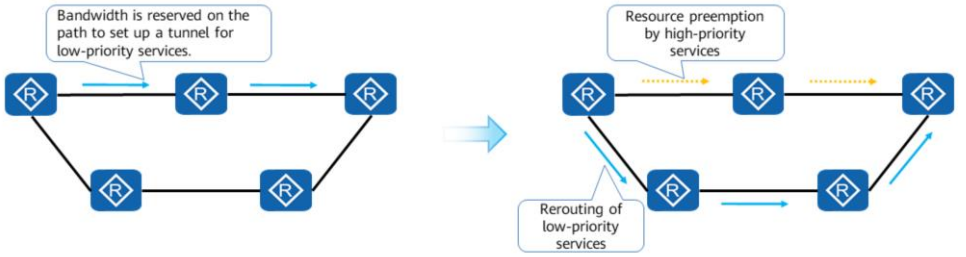


HUAWEI

- As shown in the figure, a tunnel with 100 Mbit/s bandwidth is established. If the bandwidth of the tunnel needs to be increased as services grow but the bandwidth of the links over which the tunnel traverses is insufficient, links are re-selected to establish a tunnel.

Major MPLS TE Functions (3)

- MPLS TE supports the configuration of tunnel priorities. In the scenario where a low-priority tunnel has been established over a path and a higher-priority tunnel is to be established over the same path, if the bandwidth resources of the path are insufficient, the higher-priority tunnel can preempt resources of the low-priority tunnel.

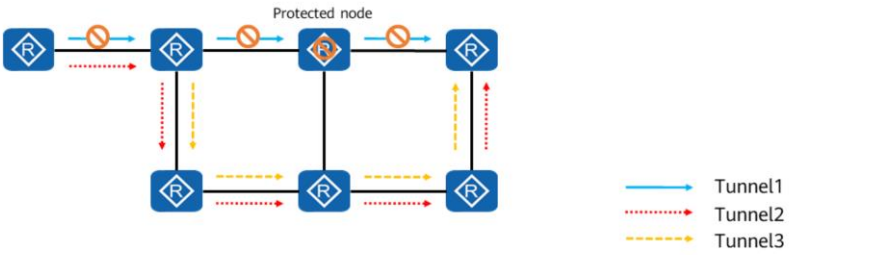


HUAWEI

- As shown in the preceding figure, when a tunnel needs to be set up for high-priority services but link resources are insufficient, high-priority services will preempt tunnel resources of lower-priority services. In this case, lower-priority services search for other links and re-establish a tunnel.

Major MPLS TE Functions (4)

- MPLS TE supports setup of E2E CR-LSPs.
- MPLS TE supports setup of a backup path in advance. The ingress and egress of the backup path are the same as those of the primary path. When the primary path fails, traffic can be switched to the backup path for forwarding, implementing E2E protection for the primary path.
- MPLS TE supports local protection. If a link or node on the primary path fails, fast local traffic switching can be triggered. Traffic is switched to the backup path after the tunnel ingress detects the failure.

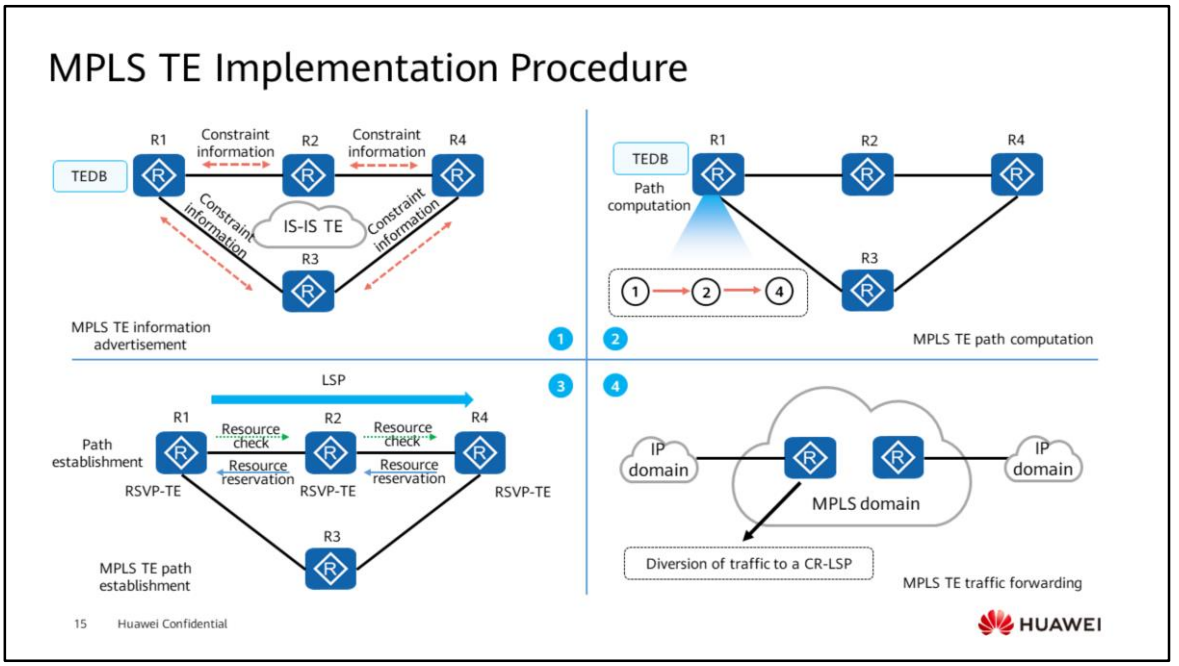


- Tunnel2 is the backup path of Tunnel1. When Tunnel1 fails, traffic is quickly switched to Tunnel2, implementing E2E protection.
- If a link or node fails, the tunnel ingress cannot immediately detect the failure. To ensure normal service transmission, fast reroute (FRR) can be used to rapidly switch traffic to Tunnel3. This allows traffic to be immediately switched away from the local failure point, ensuring traffic forwarding.
- Tunnel3 provides local protection for nodes. When an intermediate link or node fails, the ingress cannot immediately detect the failure. In this case, local protection can be used to ensure service continuity. After the ingress detects the failure, traffic is switched to Tunnel2 to implement E2E protection.

Four Major MPLS TE Components

- MPLS TE consists of the following four components:
  - Information advertisement component
  - Path calculation component
  - Signaling component (or path establishment component)
  - Packet forwarding component
- MPLS TE also supports the following advanced features to enrich the basic functions:
  - Fast reroute (FRR)
  - Tunnel backup
  - Auto bandwidth adjustment
  - Path re-optimization

- For details about the four components, see the following sections.



- Information advertisement component: advertises information such as constraints, bandwidth, and colors through OSPF-TE/IS-IS-TE, and generates a traffic engineering database (TEDB) on the ingress.
- Path computation component: The ingress uses CSPF to compute an optimal path based on the TEDB.
- Path establishment component: establishes CR-LSPs through RSVP-TE based on CSPF path computation results, and checks and reserves resources.
- Packet forwarding component: diverts traffic to an MPLS TE tunnel so that the tunnel forwards incoming traffic based on labels.

Contents

MPLS TE Information Advertisement

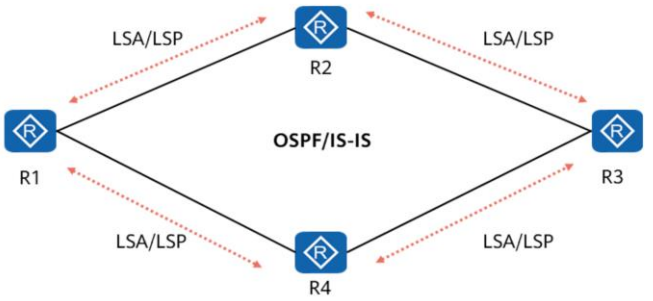
- MPLS TE allows routers to establish paths based on link information, instead of simply forwarding traffic through the shortest IP path.
- To enable network devices to compute paths more intelligently, each node on a network needs to know resource distribution on the network. Therefore, the following information needs to be advertised:
  - link-state information
  - TE metric
  - Bandwidth information
  - Administrative group
  - Shared risk link group (SRLG): A group of links that share a common physical resource (for example, a fiber). Links in an SRLG have the same level of risks. Specifically, if one of the links fails, other links in the SRLG also fail. Specifically, if one of the links fails, other links in the SRLG also fail. It is mainly used to enhance the reliability of TE tunnels.

- SRLG is mainly used in the networking where path protection and local protection are enabled. For details, see the section "MPLS TE Reliability."

Link-State Information | TE Metric | Bandwidth Information | Administrative Group

Link-State Information

- MPLS TE information advertisement is based on the extension of existing IGPs. As such, information (link-state information) collected by IGPs, such as interface IP addresses, link types, and link costs, is still advertised.



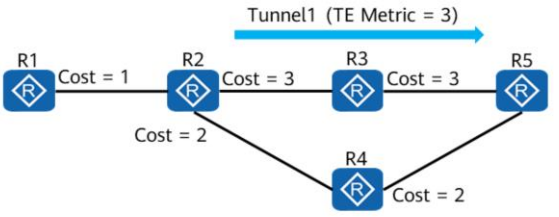
HUAWEI

- LSAs/LSPs are link-state information advertised by OSPF/IS-IS, including interface IP addresses, link types, and link costs.

Link-State Information | TE Metric | Bandwidth Information | Administrative Group

TE Metric

- To enhance the controllability of path computation for TE tunnels, MPLS TE provides the TE metric, allowing the computation process to be independent from IGP-based route computation.



A TE tunnel is established between R2 and R5 along the path R2 -> R3 -> R5. When R1 accesses R5:

If the TE metric of the link is used:
Path: R1 -> R2 -> R3 -> R5
Path cost: 4

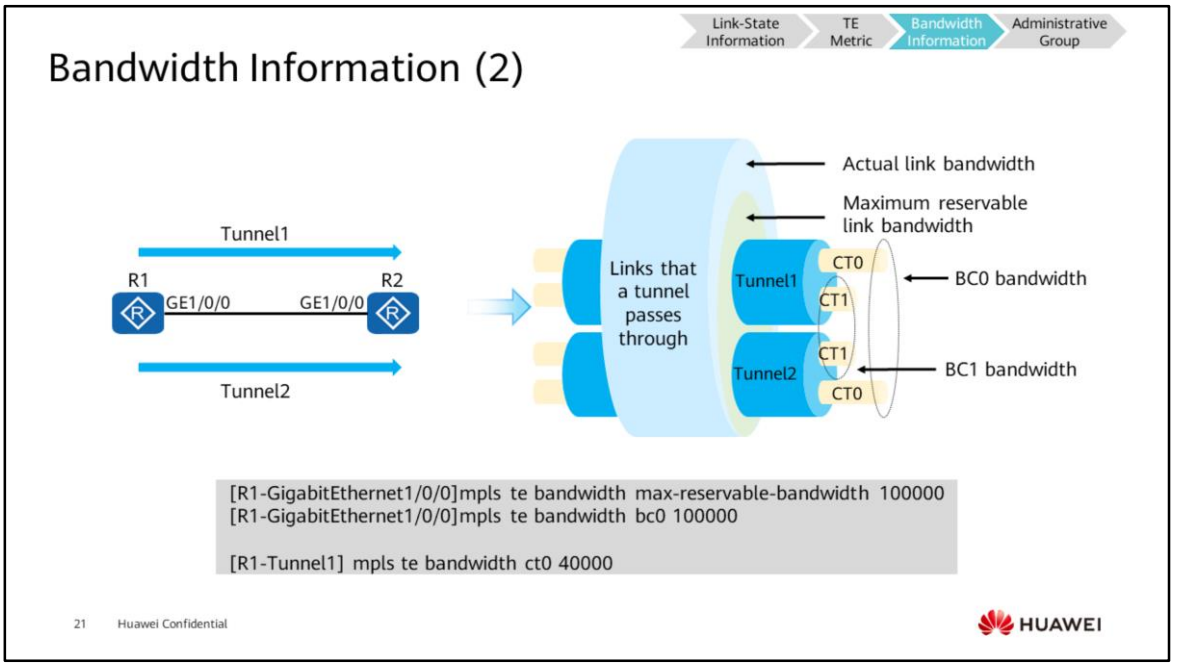
If the IGP metric of the link is used:
Path: R1 -> R2 -> R4 -> R5
Path cost: 5

HUAWEI

- In this scenario, traffic is forwarded in forwarding adjacency mode. If traffic is forwarded in forwarding shortcut mode, the path remains as R1 -> R2 -> R4 -> R5. For details, see TE traffic forwarding.
- By default, the IGP metric is used. Once the TE metric is enabled, the IGP metric becomes invalid.
- If the TE metric is configured on the ingress of a tunnel, only path selection policies of the local tunnel is affected.

Bandwidth Information (1)

- MPLS DiffServ-aware Traffic Engineering (DS-TE) combines MPLS TE and MPLS DiffServ to provide effective QoS guarantee.
  - MPLS DS-TE divides the LSP bandwidth into eight parts, each part corresponding to a service class. Such a collection of bandwidths of an LSP or a group of LSPs with the same service class are called a class type (CT). DS-TE maps traffic with the same per-hop behavior (PHB) to one CT and allocates resources to each CT. In this way, traffic with the same per-hop behavior (PHB) is mapped to one CT and resources are allocated to each CT.
  - DS-TE supports a maximum of eight CTs, which can be marked as CT0 to CT7.
- Bandwidth constraints (BCs): can be classified into BC0 to BC7 based on the CT bandwidth.
- Maximum reservable bandwidth: maximum bandwidth that a port can reserve for a tunnel.
- In a non-MPLS DS-TE scenario, BC0 is used. CT0 reserves bandwidth during tunnel establishment to provide bandwidth protection for services.

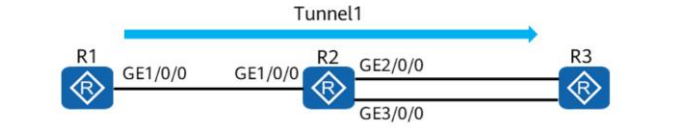


- The command output shows that CT bandwidth is the reserved bandwidth set for a tunnel, and the maximum reservable bandwidth and BC bandwidth are the maximum link bandwidth that a tunnel can use.

Link-State Information | TE Metric | Bandwidth Information | Administrative Group

Administrative Group Attribute

- MPLS TE can mark links with different colors so that tunnels can select paths based on link colors.
- Administrative group attribute:
  - Also called the color attribute of a link, is used to describe link attributes. It consists of 32 bits (each of which can indicate an attribute of a link) and is configured on a physical interface to manage links with user-defined attributes.
  - The link administrative group attribute is used together with affinities to control the paths for tunnels.




```
[R1-GigabitEthernet1/0/0]mpls te link administrative group 10001
[R2-GigabitEthernet2/0/0]mpls te link administrative group 10101
[R2-GigabitEthernet3/0/0]mpls te link administrative group 10011
```


HUAWEI

- Administrative group attribute: also called the color attribute of a link, is used to mark links with different colors. During route selection, paths can be selected based on colors. Administrative group information is advertised along with IGP-TE flooding information.
- Each bit of an administrative group can be set or not set. Network administrators can associate an administrative group bit with any required meaning, for example:
  - Link bandwidth or performance (such as delay).
  - Link function or attribute for management purposes. (For example, it can identify that an MPLS TE tunnel passes through a link or multicast services are transmitted over a link.)

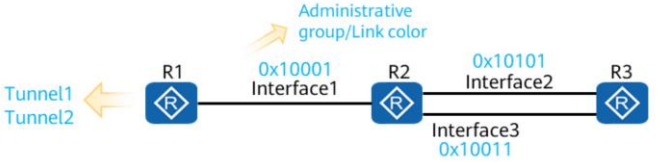
Affinity

- An affinity is a 32-bit vector used to describe the links required by a TE tunnel. Similar to the administrative group attribute, an affinity can be considered as a tunnel color and is configured on the ingress of a tunnel. (Tunnel information is not advertised through an IGP).
- The 32-bit mask determines which link attributes are to be checked by a device as well as which attributes a link must have for it to be selected for a tunnel. Data can be properly forwarded only when tunnel colors match link colors.
- A path can be selected for a tunnel only when the administrative group attribute of a link matches the affinity of the tunnel to be established over the link.


```
[R1-Tunnel1] mpls te affinity property 10101 mask 11011
[R1-Tunnel2] mpls te affinity property 10011 mask 11101
```

Example for Selecting Paths Based on Tunnel Attributes (1)

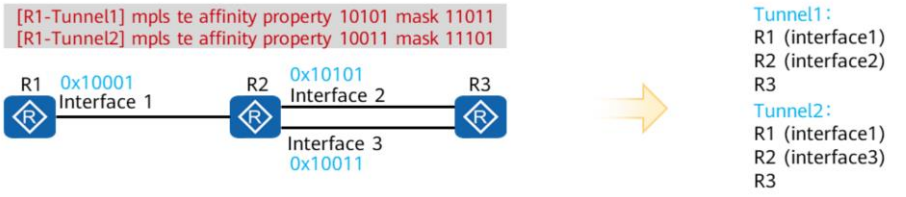
- Two tunnels to R3 named Tunnel1 and Tunnel2 are established on R1.
- It is required that the affinity and mask of a tunnel be used to match against the administrative group attribute of the link over which the tunnel is to be established so that Tunnel1 on R1 uses one physical link between R2 and R3 and Tunnel2 uses another physical link between R2 and R3.



Question: How can we configure affinities for tunnels so that paths can be selected for the tunnels as required?

Example for Selecting Paths Based on Tunnel Attributes (2)

- After a tunnel is assigned an affinity, a device compares the affinity with the administrative group attribute during path computation to determine whether a link with specified attributes is selected or not. Path selection for a tunnel complies with the following two rules:
  - The affinity bit whose mask value is 0 is not compared against the corresponding administrative group attribute bit.
  - For the affinity bits whose mask value is 1: If the affinity attribute bit value is 0, the administrative group attribute bit value must also be 0. If the affinity attribute bit value is 1, at least one bit of the administrative group attribute must be 1.



Question: When a tunnel from R3 to R1 is created, will the link administrative group attribute be reused?

HUAWEI

- For Tunnel1, the affinity attribute is 10101 and the mask is 11011. Therefore, the first two bits and last two bits of the affinity attribute are compared with those of the administrative group attribute. Therefore, the second and fourth bits of the administrative group attribute of the link selected by the tunnel must be 0. At least one of the first and fifth bits must be 1.
- For Tunnel2, the affinity attribute is 10011 and the mask is 11101. Therefore, the first three bits and last one bit of the affinity attribute are compared with those of the administrative group. The second and third bits of the administrative group attribute of the link selected by the tunnel must be 0. At least one of the first and fifth bits must be 1.
- For tunnels established in different directions, the system compares the affinity attribute of the tunnel with the link administrative group attribute configured on the outbound interface of the device along the tunnel path. Therefore, to establish a tunnel from R3 to R1, you need to reconfigure the administrative group attribute on the outbound interfaces of R3 and R2.

Who Advertises TE Information?

- TE information is mainly advertised using extensions of existing link-state routing protocols, including OSPF TE and IS-IS TE. OSPF TE and IS-IS TE automatically collect TE information and flood it to other nodes on the MPLS TE network.
- OSPF TE or IS-IS TE is used only in a single area by default.

OSPF TE	ISIS-TE
• The Type 10 Opaque LSA is added to collect and advertise TE information, including the maximum link bandwidth, maximum reservable bandwidth, current reserved bandwidth, and link color.	• The Extended IS reachability TLV (Type 22 TLV) is added to carry TE information configured on a physical interface. • The Traffic Engineering router ID TLV (Type 134 TLV) is added. It contains a 4-byte router ID, which is currently used as the MPLS LSR ID. • The Extended IP reachability TLV (Type 135 TLV) is added to carry routing information and extend the route metric (wide metric) range.

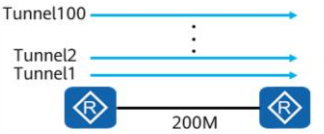
- IS-IS metrics are classified into the following types:
  - Narrow metric: 6 bits. The link metric ranges from 0 to 63.
  - Wide metric: extended from 6 bits to 24 bits. The link metric ranges from 0 to 1677214.
- The narrow metric supports only 64 vector values, difficult to meet TE requirements on large networks.
- To use IS-IS TE, the wide metric must be supported first. The wide metric is not necessarily related to MPLS TE, but it can enhance the scalability of MPLS TE. During the transition from the narrow metric to the wide metric of an IS-IS network, the compatibility between the metrics needs to be considered.
- The Type 135 TLV of the wide metric is sometimes called the IP-Extended TLV. You can run the cost-style { wide | compatible | wide-compatible } command to set the metric style. By default, the cost type of IS-IS routes is narrow.

When to Advertise Information?

The IGP status changes

1. IGP TE floods state information at intervals, which can be manually configured.
2. A link is activated or deactivated.
3. A link attribute (a parameter configured on an interface, such as the cost or administrative group attribute) changes.

The link bandwidth changes



- When a large number of tunnels that require bandwidth reservation are created on a node, the node frequently updates information in the TEDB and floods it. MPLS TE provides a bandwidth flooding suppression mechanism to reduce the frequency of TEDB updating and flooding.
  - The ratio of the bandwidth reserved for an MPLS TE tunnel to the available link bandwidth in the TEDB is greater than or equal to the configured threshold.
  - The ratio of the bandwidth released by an MPLS TE tunnel to the available link bandwidth in the TEDB is greater than or equal to the configured threshold.

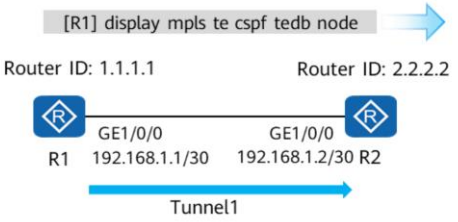
HUAWEI

- When the link bandwidth changes, bandwidth information flooding is performed as follows:
  - If either of the preceding conditions is met, an IGP floods link bandwidth information, and CSPF updates the TEDB.
  - For example, the remaining bandwidth of a link is 200 Mbit/s. If 100 TE tunnels with 2 Mbit/s bandwidth are established on the link and the flooding threshold is set to 10%, bandwidth information is not flooded when tunnels 1 to 9 are established. The bandwidth information (20 Mbit/s in total) of tunnels 1 to 10 is flooded only when tunnel 10 is established. The available bandwidth is 180 Mbit/s. According to the configured flooding threshold, the allowed bandwidth is 18 Mbit/s when the second batch of tunnels are established. Therefore, bandwidth information is not flooded when tunnels 11 to 18 are established. The bandwidth information is flooded only when tunnel 19 is established. The process repeats until tunnel 100 is established.

Flooding Result - Forming the Same TEDB (1)

- After an OSPF-TE or IS-IS TE flooding process is complete, all nodes in the local area generate the same TEDB.
- The TEDB contains information such as constraints and bandwidth usage of each link. A node calculates the optimal path to another node in the area based on the information in the TEDB. MPLS TE then establishes a CR-LSP over this optimal path.
- The TEDB and IGP LSDB are independent of each other.
  - Similarities:
    - Both the two databases are generated through IGP-based flooding.
  - Differences:
    - Content: A TEDB contains TE information in addition to all the information in an LSDB.
    - Function: An IGP uses information in an LSDB to calculate the shortest path, while MPLS TE uses information in a TEDB to calculate the optimal path for CR-LSPs.

Flooding Result - Forming the Same TEDB (2)

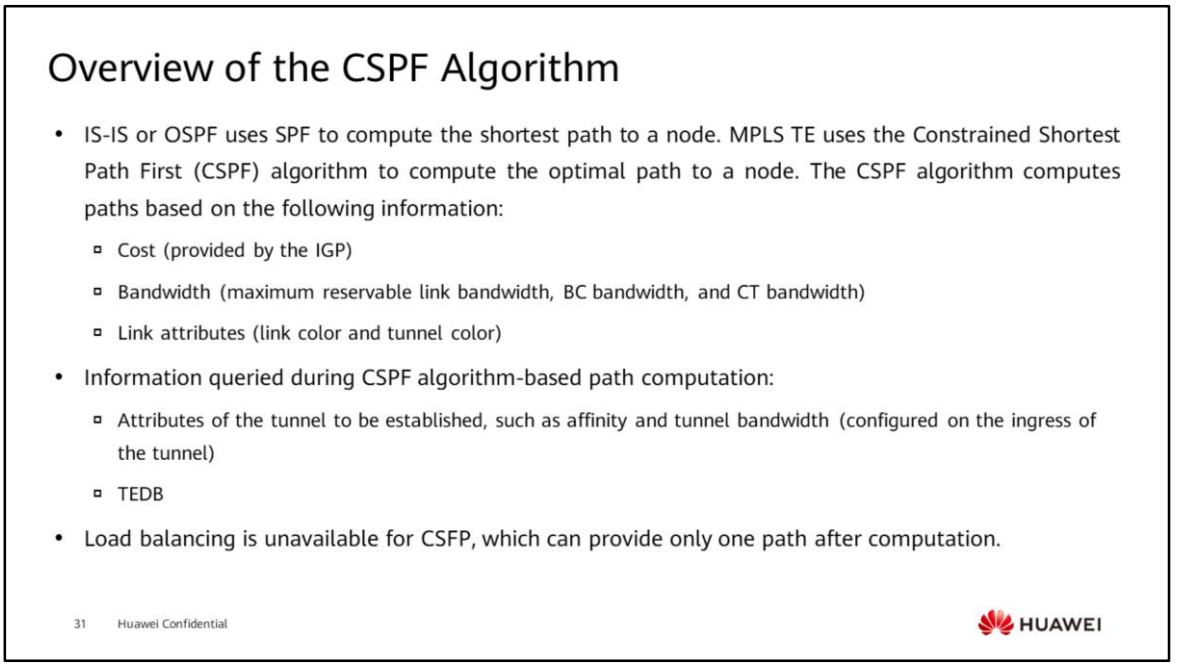



```
[R1] display mpls te cspf tedb node
 Router ID: 1.1.1.1
 IGP Type: OSPF      Process ID: 1      IGP Area: 0
  MPLS-TE Link Count: 1
  Link[1]:
    OSPF Router ID: 192.168.1.1      Opaque LSA ID: 1.0.0.1
    Interface IP Address: 192.168.1.1
    DR Address: 192.168.1.2
    IGP Area: 0
    Link  Type: Multi-access  Link Status: Active
    IGP Metric: 1              TE Metric: 1          Color: 0x10001
    Bandwidth Allocation Model : -
    Maximum Link-Bandwidth: 50000 (kbps)
    Maximum Reservable Bandwidth: 50000 (kbps)
    Operational Mode of Router: TE
    Bandwidth Constraints:          Local Overbooking Multiplier:
       BC[0]:      50000 (kbps)          LOM[0]:          1
    BW  Unreserved:
       Class  ID:
         [0]:      50000 (kbps),        [1]:      50000 (kbps)
         [2]:      50000 (kbps),        [3]:      50000 (kbps)
         [4]:      50000 (kbps),        [5]:      50000 (kbps)
         [6]:      50000 (kbps),        [7]:      50000 (kbps)
```


HUAWEI

- Color indicates the administrative group attribute of a link.

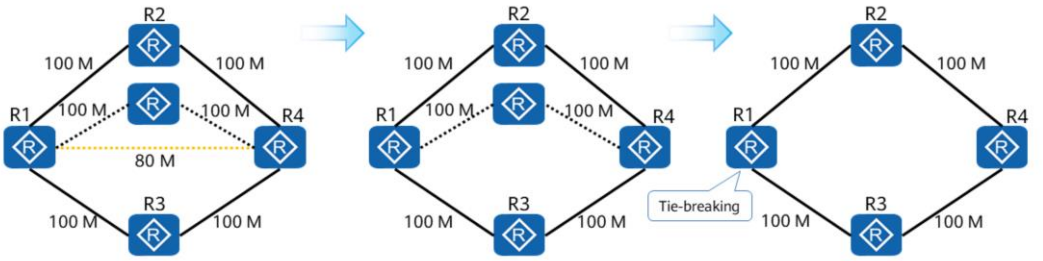
Contents



- CSPF is an improved shortest path first algorithm that takes certain constraints into consideration when computing the shortest path to a node. Based on the resource availability and whether the selected part violates user policy constraints, CSPF deletes the nodes and links that do not meet the constraints from the current topology, and then uses the SPF algorithm to compute the shortest path that meets the constraints, including a group of LSR addresses.
- The next hop calculated by the SPF algorithm of IS-IS or OSPF is the direct next hop. Each router needs to run the SPF algorithm. The result of CSPF computation is a constraint-compliant explicit route. CSPF computation is usually performed only on the ingress of a to-be-established CR-LSP (ingress of a TE tunnel). On the ingress of a TE tunnel, the route calculated by CSPF provides a logical outbound interface to the destination. Traffic enters a CR-LSP through this logical outbound interface, and this CR-LSP is called a TE tunnel.
- The MPLS signaling protocol transmits the explicit path calculated by CSPF to downstream nodes in signaling messages. Then, a TE tunnel is set up along the LSRs on the path. After a TE tunnel is successfully set up, the ingress of the TE tunnel adds an MPLS label to each IP packet that needs to enter the TE tunnel. The IP packets are then forwarded along the MPLS TE tunnel until they reach the egress of the TE tunnel.

How the CSPF Algorithm Works

- The links that do not meet tunnel attribute requirements in the TEDB are excluded.
- The SPF algorithm calculates the shortest path to a tunnel destination address.
- If multiple paths with the same weight still exist, tie-breaking is performed to select the optimal path.



On the ingress R1, create an MPLS TE tunnel with the destination address being R4's address, bandwidth being 100 Mbit/s, and affinity being black.

HUAWEI

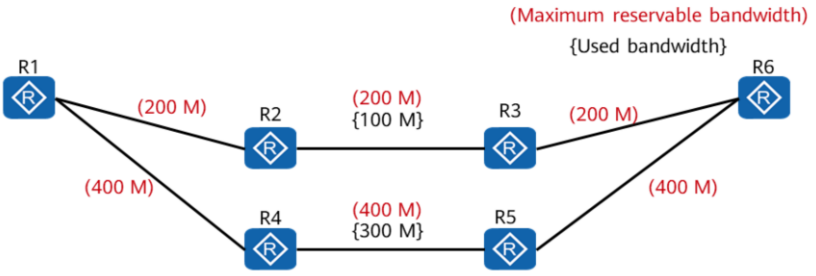
Tie-breaking of the CSPF Algorithm

- The CSPF algorithm provides only one path for a destination after computation. When there are multiple paths that meet the basic conditions, tie-breaking of the CSPF algorithm is required:
  - Most-fill: selects the link with the highest ratio of used bandwidth to the maximum reservable bandwidth. This policy enables efficient use of link bandwidth resources.
  - Least-fill: selects the link with the lowest ratio of used bandwidth to the maximum reservable bandwidth. This policy enables even use of link bandwidth resources.
  - Random: selects a link in a way that allows CR-LSPs to be as evenly distributed among links as possible, regardless of the bandwidth.
- The most-fill and least-fill modes take effect only when the bandwidth usage difference between two links exceeds 10%. For example, if the bandwidth usage of link A is 50% and the bandwidth usage of link B is 45% (a 5% difference), the most-fill and least-fill modes do not take effect and the random mode is used.

- A link is selected based on a bandwidth ratio using the tie-breaking mechanism. If the ratios are the same (for example, no reservable bandwidth is available or the bandwidth usage is the same), the first discovered link is used regardless of whether least-fill or most-fill is configured.

Tie-breaking

- If the most-fill mode is configured, the paths R1, R4, R5, and R6 are selected.
- If the least-fill mode is configured, paths R1, R2, R3, and R6 are selected.
- If the random mode is used, links are selected in a way that allows CR-LSPs to be as evenly distributed among links as possible, regardless of the bandwidth.
- By default, the random mode is used.



HUAWEI

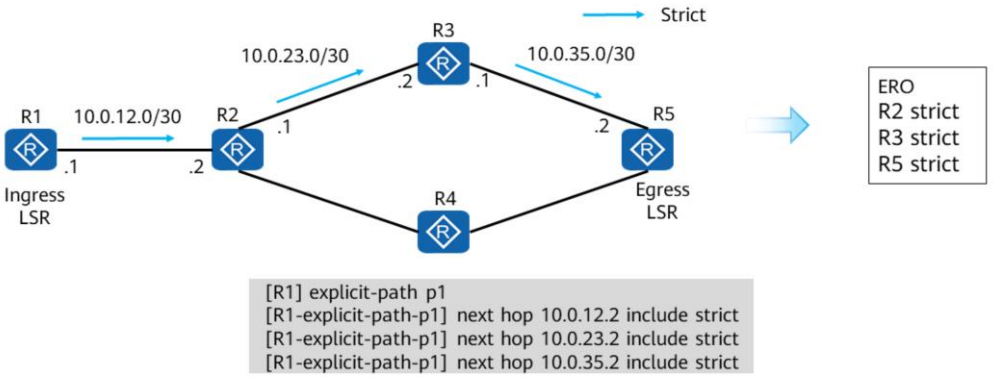
- For paths R1, R4, R5, and R6, the ratio of used bandwidth to the maximum reservable bandwidth is 3:4.
- For paths R1, R2, R3, and R6, the ratio of used bandwidth to the maximum reservable bandwidth is 1:2.
- Therefore, when the most-fill mode is configured, the paths R1, R4, R5, and R6 are selected. When the least-fill mode is configured, the paths R1, R2, R3, and R6 are selected.

Path Selection

- When a CR-LSP is set up, you can manually specify the nodes that the CR-LSP must traverse or bypass. These nodes constitute an explicit path in MPLS TE. Therefore, the explicit path can also be used to control path selection in addition to the CSPF algorithm.
- One of the greatest charm of MPLS TE lies in its support for explicit paths. You can define the path of a CR-LSP according to the actual requirements to improve operability and manageability.
- An explicit path consists of a series of nodes. Two adjacent nodes on an explicit path are connected in either of the following modes:
  - Strict: The two nodes are directly connected.
  - Loose: Other routers can exist between the two nodes.
- You can run the include or exclude command to configure a CR-LSP to traverse or bypass a node.

Path Selection - Strict Explicit Path

- A strict explicit path means that a hop is directly connected to its next hop.
- A strict explicit path precisely controls the path of a CR-LSP.



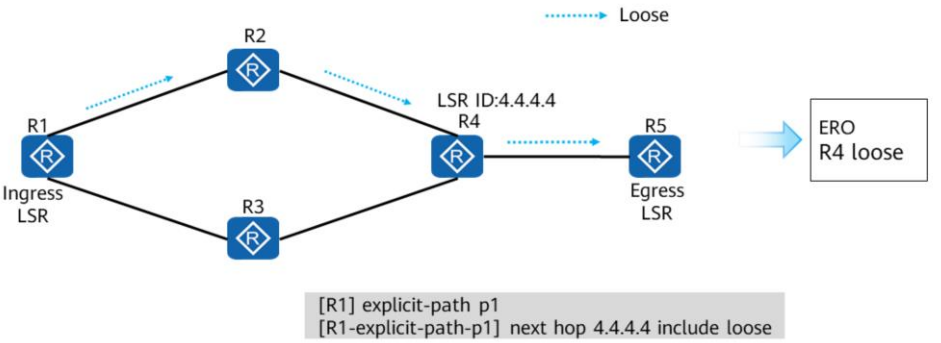

```
[R1] explicit-path p1
[R1-explicit-path-p1] next hop 10.0.12.2 include strict
[R1-explicit-path-p1] next hop 10.0.23.2 include strict
[R1-explicit-path-p1] next hop 10.0.35.2 include strict
```


HUAWEI

- ERO: indicates the explicit route object. It is mainly used to indicate explicit path information.
- In the example, "R2 strict" indicates that the CR-LSP must pass through R2 and the previous hop of R2 is Ingress R1. "R5 strict" indicates that the CR-LSP must pass through R5 and the previous hop of R5 is R3.

Path Selection - Loose Explicit Path

- A loose explicit path specifies the nodes through which a CR-LSP must pass; however, other routers can exist between a specified node and its next hop.



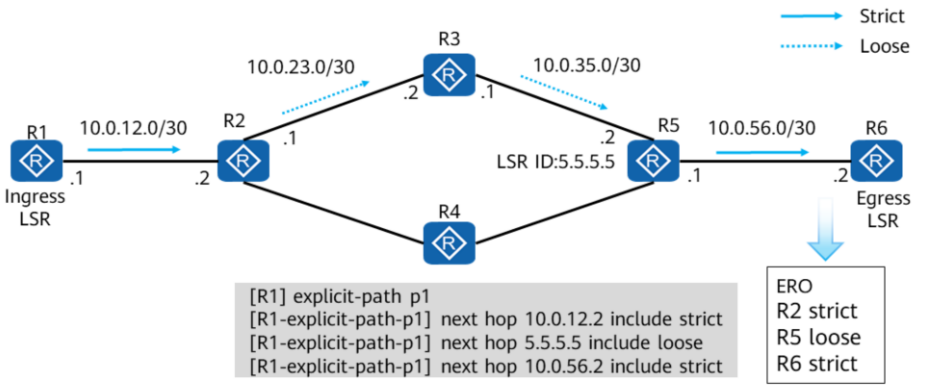

```
[R1] explicit-path p1
[R1-explicit-path-p1] next hop 4.4.4.4 include loose
```


HUAWEI

- In the example, "R4 loose" indicates that the CR-LSP must pass through R4, but other nodes can exist between R4 and Ingress R1.

Path Selection - Hybrid Explicit Path

- The strict mode and loose mode can be used together.



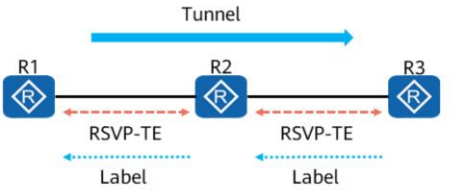
HUAWEI

- In the example, "R2 strict" indicates that the CR-LSP must pass through R2 and the previous hop of R2 is Ingress R1. "R6 strict" indicates that the CR-LSP must pass through R6 and the previous hop of R6 is R5.
- "R5 loose" indicates that the CR-LSP must pass through R5, but a router can be deployed between R5 and Ingress R1, which do not need to be directly connected.
- Therefore, the CR-LSP path can be R1 -> R2 -> R3 -> R5 -> R6. It can also be R1 -> R2 -> R4 -> R5 -> R6.

Contents

TE Signaling Protocols

- After path computation is complete, TE needs to reserve resources along the path to establish a CR-LSP.
- MPLS TE mainly uses RSVP-TE as the signaling protocol to establish CR-LSPs.
- RSVP is designed for the integrated service model in the QoS system. It is used to reserve resources on each node along a path. RSVP works at the transport layer and does not participate in the transmission of application data. It is a control protocol on the Internet and is similar to ICMP.
- RSVP-TE is an extension to RSVP and sets up or deletes CR-LSPs by using TE attributes carried in extended objects.



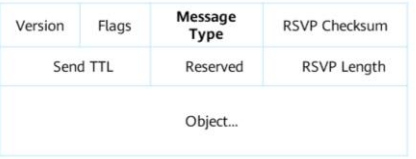
HUAWEI

- Currently, RSVP-TE is mainly used as the signaling protocol for MPLS TE, and CR-LDP is seldom used.
- As shown in the preceding figure, R1, R2, and R3 use RSVP-TE as the signaling protocol to distribute labels and set up CR-LSPs.

RSVP Packet Format and Message Type

- RSVP is a soft-state protocol. It needs to periodically and repeatedly advertise reservation information on the network.
- An RSVP message consists of a common header followed by one or more objects.

Message Type	Name	Function
1	Path	Used to request downstream nodes to distribute labels for the path.
2	Resv	Used to reserve resources on each node.
3	PathErr	Sent upstream by a node if an error occurs during the processing of a Path message.
4	ResvErr	Sent downstream by a node if an error occurs during the processing of an Resv message.
5	PathTear	Sent downstream by the ingress to delete information about the local state created on each node of the path.
6	ResvTear	Sent upstream by the egress to delete the reserved local resources assigned to a path.

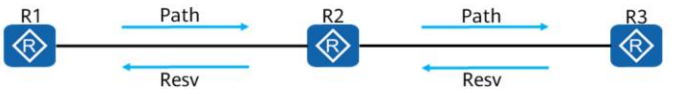


HUAWEI

- A soft state is a resource reservation state that a node maintains by periodically updating RSVP messages.
- Main RSVP-TE messages are described as follows:
  - Path message: used to request downstream nodes to distribute labels for a path. A Path message records the path information at each hop of a path and is used to establish a path state block (PSB) at a hop.
  - A Resv message carries the resource reservation information required by a sender and is sent in the reverse direction of data flows. The Resv message is used by each node on a path to establish a reservation state block (RSB) and record information about distributed labels.
  - PathErr message: sent upstream by an RSVP node if an error occurs during the processing of a Path message. A PathErr message is forwarded upstream by each transit node until it arrives at the ingress.
  - ResvErr message: sent downstream by an RSVP node if an error occurs during the processing of a Path message. A ResvErr message is forwarded downstream by each transit node until it arrives at the egress.
  - ResvTear message: sent upstream by the egress to delete the reserved local resources assigned to a path.
  - ResvTear message: sent upstream by the egress to delete the local reserved resources assigned to a path. After receiving the ResvTear message, the ingress sends a PathTear message to the egress.

How RSVP Works

- RSVP has three basic functions:
  - Path establishment and maintenance
  - Path teardown
  - Error advertisement
- The signaling process of RSVP is as follows:
  - The start point sends a Path message to the end point to apply for resource reservation, and the end point sends an Resv message to complete resource reservation.



HUAWEI

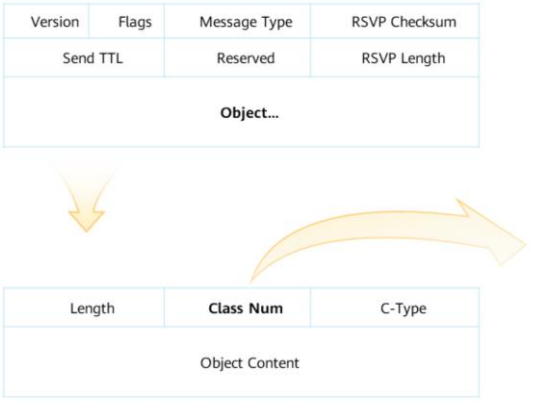
- RSVP is a soft-state signaling protocol used to reserve network resources periodically.
- As shown in the preceding figure, R1 sends a Path message downstream to apply for resource reservation. After receiving the message, R2 forwards the Path message downstream until the tunnel egress (R3) receives the message. R3 then generates an Resv message based on the Path message and sends the Resv message upstream until the ingress receives the Resv message while completing resource reservation on the link.

New Objects Extended by RSVP-TE (1)

- RSVP-TE extends TE CR-LSPs through extended objects:
  - Supports label distribution in downstream on demand (DoD) mode.
  - Allocates network resources to explicit CR-LSPs.
  - Supports teardown of a CR-LSP in make-before-break (MBB) mode when the established CR-LSP is preempted.
  - Records each node that a CR-LSP passes through to prevent loops.
  - Supports diagnosis of CR-LSPs.

- Detailed extended functions will be described in subsequent courses.
- The basic principle of MBB is as follows: A new CR-LSP is established first. After all traffic is switched to this CR-LSP, the original CR-LSP is torn down to ensure the forwarding of service traffic.

New Objects Extended by RSVP-TE (2)



- Label_Request
- Label
- Explicit_Route
- Record_Route
- Session_Attribute
- Sender_Template
- Filter_Spec
- Session
- Flowspec
- Sender_Tspec

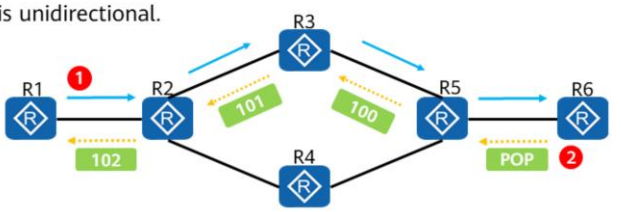
HUAWEI

New Objects Extended by RSVP-TE (3)

Object Type	RSVP Information	Description
Label_Request	Path	Identifies the LABEL_REQUEST object.
Label	Resv	Carries an assigned label.
Explicit_Route	Path	Explicit Route Object (ERO) describes information about the path through which a CR-LSP passes. The path can be a strict or loose one.
Record_Route	Path, Resv	Record Route Object (RRO) is a list of CR-LSRs through which a Path message passes.
Session_Attribute	Path	Specifies attributes, such as the setup priority, hold priority, reservation style, and affinity.
Sender_Template	Path	Carries the IP address and LSP ID of the node that sends the message.
Filter_Spec	Resv	Carries the IP address and LSP ID of the node that sends the message.
Session	Path, Resv	Carries RSVP session information, such as the destination address, tunnel ID, and extended tunnel ID.
Flowspec	Resv	Specifies QoS parameters for a data flow.
Sender_Tspec	Path	Specifies traffic characteristics of a data flow.

MPLS TE Tunnel

- An MPLS TE tunnel is a head-end CR-LSP initiated by the ingress of a tunnel.
- RSVP-TE uses CSPF to compute a path, and the receiver initiates a resource reservation request for the path. RSVP-TE uses Path and Resv messages to request CR-LSP establishment and maintain resource reservation information.
- An MPLS TE tunnel is unidirectional.



Path message: requests the egress to initiate resource reservation and set ERO.

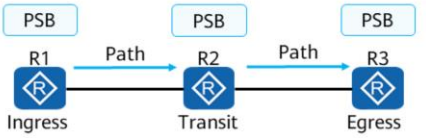
Resv message: reserves bandwidth resources and allocates labels.

HUAWEI

- Note that an MPLS TE tunnel is unidirectional. When bidirectional mutual access services exist, two CR-LSPs in opposite directions need to be established to forward traffic.

Process of Establishing an MPLS TE Tunnel (1)

- After RSVP-TE is configured on the ingress, the ingress creates a path state block (PSB) and sends a Path message to downstream nodes.
  - The Path message carries the information required for establishing a tunnel, such as the ingress, egress, ERO, and reserved bandwidth of the tunnel.
  - The PSB is used to record path information of each node.
- After receiving the Path message, the transit node processes and forwards this message, and creates a PSB based on the message.
- After receiving the Path message, the egress creates a PSB.



Process of Establishing an MPLS TE Tunnel (2)

```
<R1>display mpls rsvp-te psb-content
------------------------------------------------------------------------------
                The PSB Content
------------------------------------------------------------------------------
Tunnel Addr: 3.3.3.3              Exist time: 19h 29m 46s
Tunnel ExtID: 1.1.1.1             Session ID: 116
Ingress LSR ID: 1.1.1.1           Local LSP ID: 34926
Previous Hop: ----                Next Hop: 10.0.12.2
Incoming / Outgoing Interface: ---- / GigabitEthernet1/0/0
InLabel: NULL                     OutLabel: 32846
Send Message ID: 0                Recv Message ID: 0
Session Attribute:
  SetupPrio: 7                    HoldPrio: 7
  SessionAttrib: SE Style desired
CSPF Route Flag: False
LSP Type: -
FRR Flag: No protection           Local RRO Flag: 0x0
FRR Mode: -
```

```
ERO Information:
    L-Type          ERO-IPAddr        ERO-PrefixLen      Label
  ERHOP_STRICT      10.0.12.2         32
  ERHOP_STRICT      10.0.23.1         32
  ERHOP_STRICT      10.0.23.2         32
...
CT-BandWidth Information(Bytes/sec):
  CT0 Bandwidth: 0                  CT1 Bandwidth: 0
  CT2 Bandwidth: 0                  CT3 Bandwidth: 0
  CT4 Bandwidth: 0                  CT5 Bandwidth: 0
  CT6 Bandwidth: 0                  CT7 Bandwidth: 0
Path Message arrive on Unknown(0x0) from PHOP 0.0.0.0
Path Message sent to NHOP 10.0.12.2 on GigabitEthernet1/0/0
Resource Reservation OK
PADS Ability: Enable
```


HUAWEI

- Tunnel Addr: destination address of a tunnel.
- Tunnel ExtID: extended ID of a tunnel.
- Exist time: length of time since the PSB information was generated.
- Ingress LSR ID: ID of the ingress.
- Local LSP ID: ID of the local CR-LSP.
- Session Attribute: attributes of an RSVP session.
- SetupPrio: session setup priority.
- HoldPrio: session hold priority.
- SessionAttrib: RSVP session attribute.
    - SE Style desired: The expected resource reservation style is shared explicit (SE).
    - FF Style desired: The expected resource reservation style is fixed filter (FF).
    - Label Record: enables the label record function when exchanging Path and Resv messages.
    - Local Protection desired: provides CR-LSPs with FRR protection.
    - Node Protection desired: provides CR-LSPs with node protection.
    - Bandwidth Protection desired: provides LSPs with bandwidth protection.
    - Path Re-evaluation desired: Path re-evaluation is required for the CR-LSP.

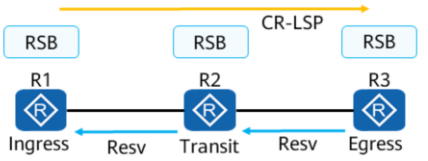
Tunnel Resource Reservation Style

- A reservation style defines how an RSVP node reserves resources after receiving a request sent by an upstream node. The NE40E supports the following reservation styles:
  - Fixed filter (FF): creates an exclusive reservation for each sender, which does not share its resource reservation with other senders and is assigned a unique label.
  - Shared explicit (SE): explicitly specifies the senders in a reservation for receivers. These senders share one reservation but send different labels to a receiver.
- The SE style allows old and new tunnels of the same session to share bandwidth resources. The old tunnel is torn down only after all traffic of the old tunnel is switched to the new tunnel.
- The SE style is an important condition for implementing make-before-break.

- How to determine whether the old and new tunnels belong to the same session
  - A unique identifier is obtained based on 5-tuple information {sender address, LSP ID, endpoint address, tunnel ID, and extended tunnel ID}.
  - The rules for MPLS TE SE reservation are as follows: If two reservations have the same 5-tuple information except for the LSP ID, the two reservations belong to the same session and can share bandwidth resources.
- The 5-tuple elements are contained in the Session and Sender_Template objects.
  - Session object
    - Endpoint address: RID of the tunnel endpoint
    - Tunnel ID: number of the configured TE tunnel
    - Extended tunnel ID: The default value is usually 0.
  - Sender_Template object
    - Sender address: RID of the tunnel ingress
    - LSP ID: functions as an instance counter. The LSP ID increases by 1 each time the bandwidth requirement of a tunnel changes or a new path is used.

Process of Establishing an MPLS TE Tunnel (3)

- After receiving a Path message, the egress generates an Resv message based on the Path message, creates a Reserve State Block (RSB), and sends the Resv message upstream.
  - The Resv message carries resource reservation information requested by the sender, such as labels and reserved bandwidth.
  - The RSB is mainly used to record information about allocated labels.
- The transit node processes and forwards the Resv message and creates state blocks such as the RSB.
- After receiving the Resv message, the ingress creates an RSB and confirms that the resources are successfully reserved.
- The label distribution on the path is complete, and the CR-LSP is set up.



Process of Establishing an MPLS TE Tunnel (4)

```
<R1> display mpls rsvp-te rsb-content
------------------------------------------------------------------------------
                    The RSB Content
------------------------------------------------------------------------------
 Tunnel Addr: 3.3.3.3                  Session Tunnel ID: 1
 Tunnel ExtID: 1.1.1.1
 Next Hop: 10.0.12.2                   Reservation Style: SE Style
 Reservation Incoming Interface: GigabitEthernet1/0/0
 Reservation Interface: GigabitEthernet1/0/0
 Message ID: 0
 Filter Spec Information-
   The filter number: 1
   Ingress LSR ID: 1.1.1.1       Local LSP ID: 24      OutLabel: 32846
Resv Message arrive on GigabitEthernet1/0/0 from NHOP 10.0.12.2
```

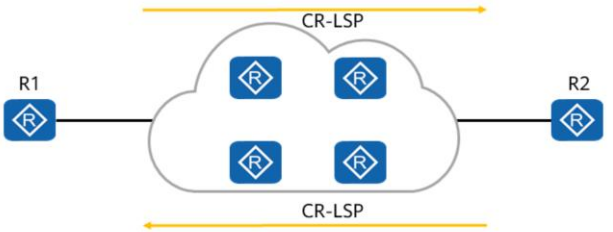
```
<R2> display mpls rsvp-te rsb-content
 Tunnel Addr: 3.3.3.3                  Session Tunnel ID: 1
 Tunnel ExtID: 1.1.1.1
 Next Hop: 10.0.23.2                   Reservation Style: SE Style
Filter Spec Information-
   The filter number: 1
   Ingress LSR ID: 1.1.1.1       Local LSP ID: 24      OutLabel: 3
Resv Message arrive on GigabitEthernet1/0/0 from NHOP 10.0.23.2
```

```
<R3> display mpls rsvp-te rsb-content
 Tunnel Addr: 3.3.3.3                  Session Tunnel ID: 1
 Tunnel ExtID: 1.1.1.1
 Next Hop: -----                       Reservation Style: SE Style
Filter Spec Information-
   The filter number: 1
   Ingress LSR ID: 1.1.1.1       Local LSP ID: 24      OutLabel: NULL
Resv Message arrive on Unknown from NHOP 0.0.0.0
```


- OutLabel: indicates the outgoing label.

Bidirectional CR-LSP

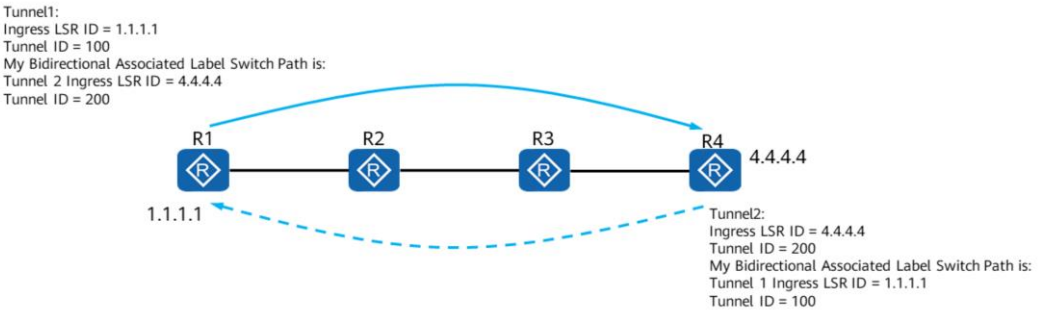
- When bidirectional services exist, two MPLS TE tunnels can be deployed on two devices that are the source and sink of each other to form a bidirectional CR-LSP to carry bidirectional traffic and ensure bandwidth.
- The two tunnels are independent of each other. When traffic switching is performed on one tunnel upon a fault, the other tunnel cannot be notified of the fault in time to also perform traffic switching. As a result, services are interrupted.



- If no reverse MPLS TE tunnel is configured, the traffic from R2 to R1 can be forwarded only through routes, and bandwidth protection cannot be provided. As a result, congestion cannot be avoided.

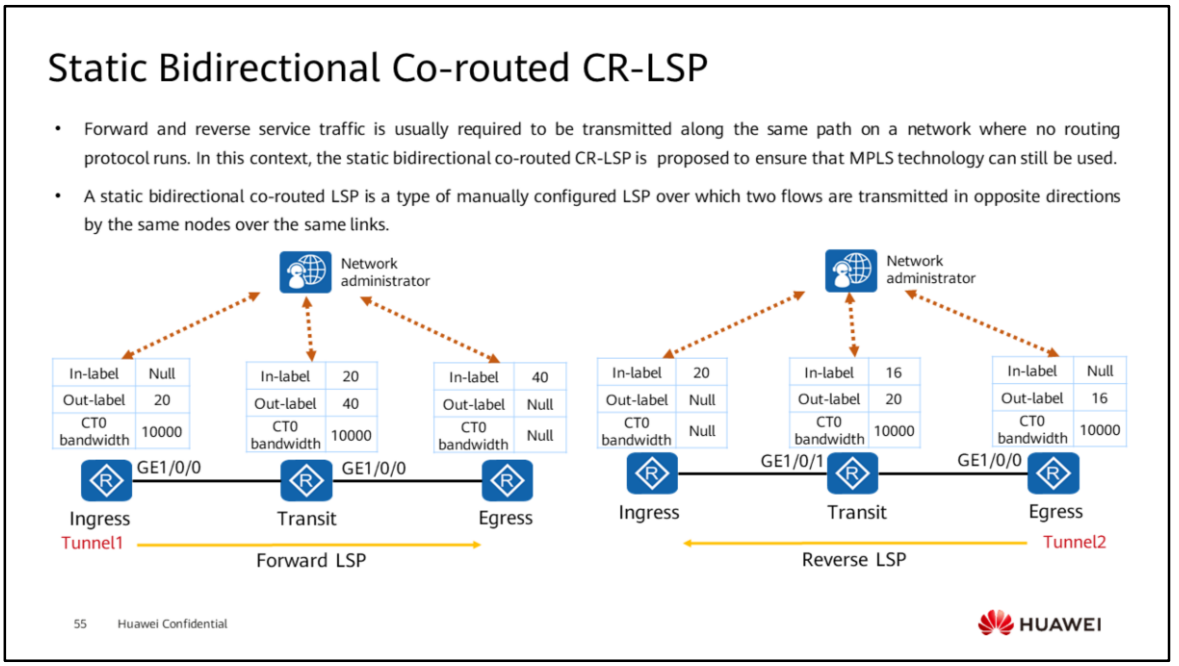
Bidirectional Associated CR-LSP

- If a pair of MPLS TE CR-LSPs in opposite directions is established between two nodes and each CR-LSP is bound to the ingress of its reverse LSP, the two LSPs form a bidirectional associated LSP. The associated bidirectional CR-LSP is configured to prevent traffic congestion. If a fault occurs on one end, the other end is notified of the fault so that both ends trigger traffic switching at the same time to ensure that bidirectional services are not interrupted.



HUAWEI

- Tunnel1 and Tunnel2 must both be MPLS-TE tunnels. They can be either dynamic CR-LSPs established using RSVP-TE or static CR-LSPs manually configured.
- The tunnel ID and ingress LSR ID of the reverse CR-LSP are specified on each tunnel interface so that the forward and reverse CR-LSPs are bound to each other. A reverse LSP must be specified for both the ingress and egress of a tunnel. In addition, the binding relationships must match each other. For example, in the preceding figure, set the reverse tunnel ID to 200 and ingress LSR ID to 4.4.4.4 on Tunnel1 so the reverse tunnel is bound to Tunnel1.
- R1 and R4 are ingress LSRs or egress LSRs for forward and reverse LSPs.



- As shown in the preceding figure, a network administrator manually configures the incoming and outgoing labels of each node on the forward and reverse LSPs and sets the bandwidth (CT0 bandwidth) required by the LSPs. The bandwidth must be less than or equal to the available bandwidth of the outbound interface of each LSP. Therefore, for the forward LSP, the CT0 bandwidth does not need to be configured on the egress. For the reverse LSP, the CT0 bandwidth does not need to be configured on the ingress.
- A static bidirectional co-routed CR-LSP is one LSP and corresponds to two forwarding entries. It can go up only when the conditions for forwarding traffic in both directions are met. If the conditions for forwarding traffic in one direction are not met, the LSP is in the Down state. In addition, the forwarding entries in both directions are associated with each other. When IP forwarding capabilities are unavailable, any transit node can send back a response packet along the original path. Compared with two independent LSPs in opposite directions, a static bidirectional co-routed LSP has the same delay and jitter in both directions, which guarantees QoS for bidirectional services.
- To establish a static bidirectional co-routed LSP, you need to manually specify labels and forwarding entries mapped to two forwarding equivalence class (FEC) for traffic transmitted in opposite directions.
- A static bidirectional co-routed LSP is meaningful only to the local node, and the local node cannot be aware of the entire LSP. You need to select the direction from the ingress to the egress as the forward direction, and the other direction becomes the reverse direction.

Contents

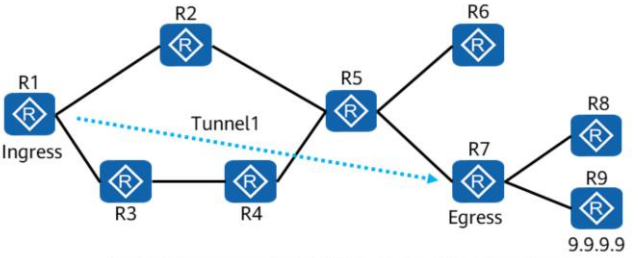
MPLS TE Traffic Forwarding

- An MPLS TE tunnel is established through information advertisement, path selection, and path establishment. Different from LDP, the LSP established by MPLS TE cannot automatically import traffic to the tunnel for forwarding. Traffic needs to be imported to the MPLS TE tunnel using a certain mode, so that the device can forward traffic based on labels.
- Traffic can be imported to an MPLS TE tunnel using:
  - Static route
  - Auto route
  - Policy-based routing (PBR)

- Unlike LDP, MPLS TE can forward traffic independently. TE requires manual configuration to forward traffic.

Forwarding Based on the Static Route (1)

- The simplest way to divert traffic to a TE tunnel is to configure a static route. Such a static route works in the same way as common static routes. You only need to configure the TE tunnel interface as the outbound interface of the static route.
- TE tunnel forwarding supports recursive static routes.
- As shown in the figure, a TE tunnel is established between R1 and R7. Tunnel1 is configured as the tunnel interface on R1. A static route is configured on R1 to divert traffic from R1 to R9 to the tunnel (the cost of each link is 10).




```
[R1] ip route-static 9.9.9.9 255.255.255.255 Tunnel1
```


HUAWEI

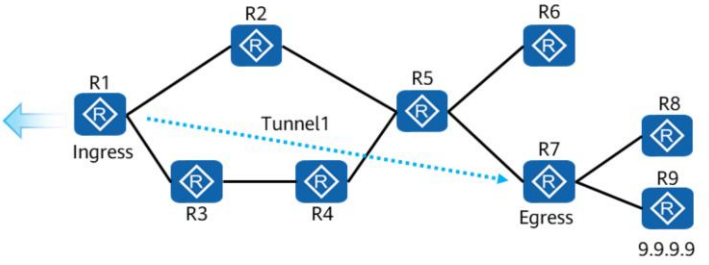
- Recursive static route: The next hop of the destination of the static route is not the TE tunnel interface. The TE tunnel interface is only an intermediate link to the destination. Simply put, route recursion needs to be performed on the TE tunnel to reach the next hop of the destination.

Forwarding Based on the Static Route (2)

- The routing table of R1 shows that the traffic from R1 to R9 is forwarded through the TE tunnel, but the traffic from R1 to R7 is still forwarded through an IGP, even if R7 is the endpoint of the tunnel.

[R1] display ip routing-table

Destination	Nexthop	Cost
R2	R2	10
R3	R3	10
R4	R3	20
R5	R2	20
R6	R2	30
R7	R2	30
R8	R2	40
R9	Tunnel1	40



Question: How can all service traffic from R1 to R7 be forwarded through the tunnel without the need to configure too many static routes?

HUAWEI

Forwarding based on the Auto Route

- Auto route: A TE tunnel is considered as a logical link participating in IGP route calculation. In this case, the tunnel interface is used as the outbound interface of a route, eliminating the need for a large number of manual configurations involved in static routes.
- A TE tunnel can be used as a P2P link during IGP route calculation. You can configure a TE metric for the TE tunnel, which allows for more controllability during path selection of nodes.
- Auto route forwarding modes:
  - IGP shortcut
  - Forwarding adjacency

- The configuration of static route forwarding is complex, and the workload is heavy. Like a routing protocol, auto route allows traffic to be automatically forwarded in a tunnel.

IGP Shortcut

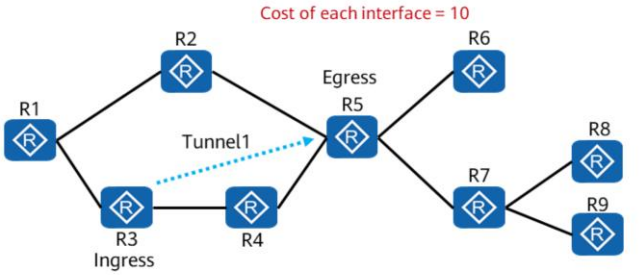
- In IGP shortcut mode, the TE LSP link is not advertised to neighboring nodes. Therefore, the tunnel can be used only by the ingress of the tunnel, not by other nodes.

[R1] display ip routing-table

Destination	Next hop	Cost
R5	R2	20
R6	R2	30
R7	R2	30

[R3] display ip routing-table

Destination	Next hop	Cost
R5	Tunnel1	20
R6	Tunnel1	30
R7	Tunnel1	30




```
[R3] interface tunnel 1
[R3-Tunnel1] mpls te igp shortcut isis
[R3-Tunnel1] mpls te igp metric absolute 10
[R3-Tunnel1] isis enable
```


HUAWEI

- IGP shortcut affects only the local routing policy and does not affect the routing of other routers.
- The IGP metric of the TE tunnel is configured using the mpls te igp metric { absolute | relative } value command.
  - If absolute is configured, the metric of the TE tunnel is the configured value.
  - If relative is configured, the metric of the TE tunnel is the sum of the metric of the corresponding IGP path and relative metric.

Forwarding Adjacency

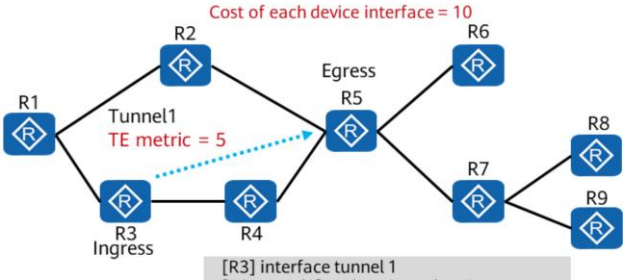
- On a live network, MPLS TE is usually deployed only on a few core nodes. In this case, edge nodes cannot detect the existence of TE tunnels and can only select paths based on the traditional IGP.
- In forwarding adjacency mode, the TE LSP is advertised to neighboring nodes. Therefore, all nodes can use this tunnel.
- If forwarding adjacency is used, the ingress and egress of a tunnel must be in the same area.

[R1] display ip routing-table

destination	nexthop	cost
R5	R3	15
R6	R3	25
R7	R3	25

[R3] display ip routing-table

destination	nexthop	cost
R5	Tunnel1	5
R6	Tunnel1	15
R7	Tunnel1	15

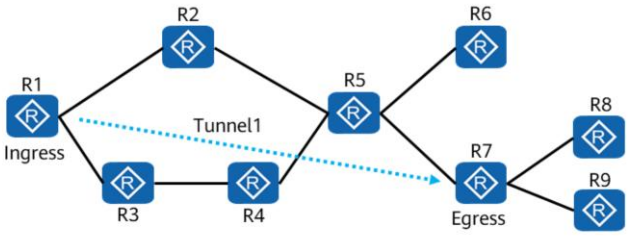



```
[R3] interface tunnel 1
[R3-tunnel1] mpls te igp advertise
[R3-tunnel1] mpls te igp metric absolute 5
[R3-tunnel1] isis enable
```


HUAWEI

Policy-Based Routing (PBR)

- TE policy-based routing is simple and does not change the routing table. Traffic is forwarded based on the configured policy. If packets do not match PBR rules, they are forwarded in IP forwarding mode; if they match PBR rules, they are forwarded over specific tunnels.



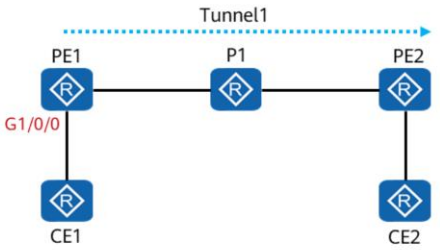

```
[R1] policy-based-route set_tunnel permit node 1 map-instance 1
[R1-policy-based-route-set_tunnel-1] if-match acl 3000
[R1-policy-based-route-set_tunnel-1] apply output-interface Tunnel1
```


HUAWEI

- ACL 3000 is defined according to the actual situation.

L2/L3VPN over TE

- In addition to the preceding MPLS TE traffic forwarding modes, tunnel recursion can be configured to divert service traffic to TE tunnels for forwarding in L2/L3VPN over TE scenarios.
- As shown in the figure, CE1 and CE2 belong to the same L3VPN and access the backbone network through PE1 and PE2, respectively. A TE tunnel is set up between PE1 and PE2. It is required that a tunnel policy be configured to divert traffic from CE1 to CE2 to the TE tunnel when the traffic reaches PE1.



Configure a tunnel policy named **policy1** to select CR-LSPs for traffic forwarding, and set the number of tunnels participating in load balancing to 1:

```
[PE1] tunnel-policy policy1
[PE1-tunnel-policy-policy1] tunnel select-seq cr-lsp load-balance-number 1
```

Configure L3VPN access on PE1 and apply the tunnel policy:

```
[PE1] ip vpn-instance vpn1
[PE1-vpn-instance-vpn1] ipv4-family
[PE1-vpn-instance-vpn1-af-ipv4] tnl-policy policy1
```

```
[PE1] interface gigabitethernet 2/0/0
[PE1-GigabitEthernet2/0/0] ip binding vpn-instance vpn1
```

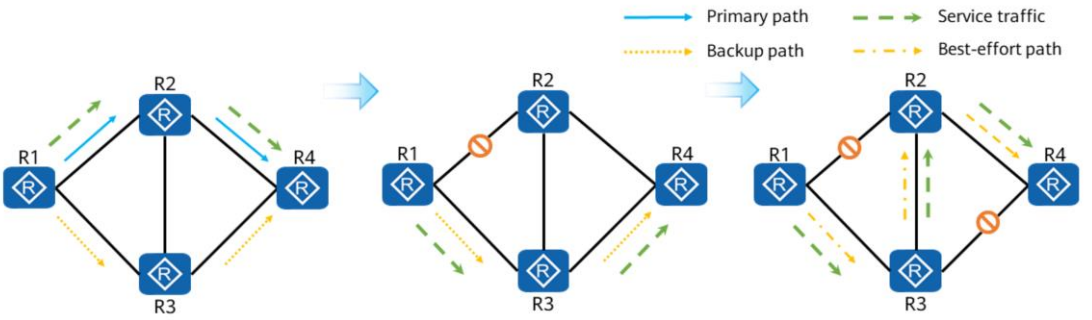
Contents

Path Protection Overview

- Path protection is also called E2E protection. A backup CR-LSP is set up to protect traffic on the primary CR-LSP.
- Path protection establishes multiple CR-LSPs between the ingress and egress of a tunnel, with each CR-LSP traversing a different path.
- If the ingress detects that the primary CR-LSP is unavailable, the ingress switches traffic to a backup CR-LSP. After the primary CR-LSP recovers, traffic is switched back.
- CR-LSP backup is performed in either of the following modes:
  - Hot standby
  - Ordinary backup
- Hot standby can be used together with fault detection technology (BFD) to switch traffic to the backup path within 50 ms. In ordinary backup mode, the switchover time depends on the number of affected tunnels when a fault occurs.

Hot Standby

- A backup CR-LSP is set up immediately after the primary CR-LSP is set up. When a message indicating that the primary CR-LSP fails is received by the ingress, traffic is switched to the backup CR-LSP. After the primary LSP recovers, traffic is switched back.
- Hot-standby CR-LSPs support best-effort paths. When both the primary and backup CR-LSPs fail, a temporary CR-LSP can be set up, and traffic is switched to the best-effort path.

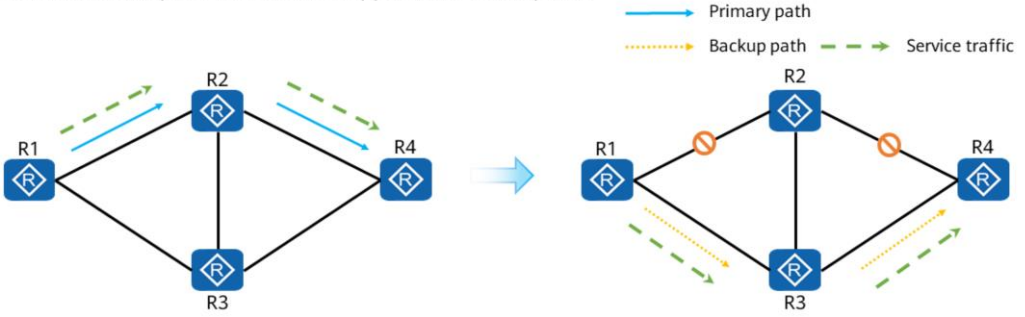


HUAWEI

- The primary and backup CR-LSPs can be specified using explicit paths. The best-effort path is automatically calculated by the system based on network faults and varies according to faulty nodes.

Ordinary Backup

- When the ingress receives a message indicating that the primary CR-LSP fails, the ingress initiates the setup of a backup CR-LSP. After the backup CR-LSP is set up, traffic is switched to the backup CR-LSP. After the primary CR-LSP recovers, traffic is switched back.
- Common backup CR-LSPs do not support best-effort paths.

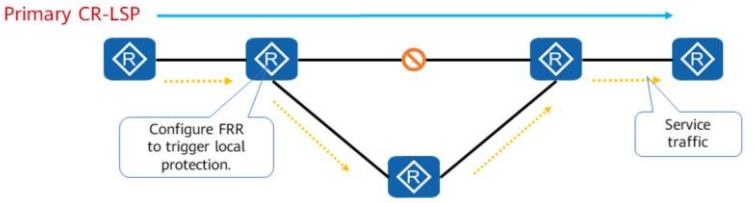


HUAWEI

Contents

Limitations of Path Protection

- Path protection provides E2E protection for TE tunnels. If a node or link on a tunnel fails, traffic is switched to a backup path. This process involves IGP route re-convergence on the backup path, CSPF path recalculation, and CR-LSP re-establishment. The slow process may cause packet loss.
- Fast reroute (FRR), also called local protection, is a temporary protection measure. When a transit node fails, local protection is triggered and a backup CR-LSP is set up locally for traffic switching. At the same time, the ingress of the tunnel is instructed to recalculate paths and switch traffic to the backup path in time.



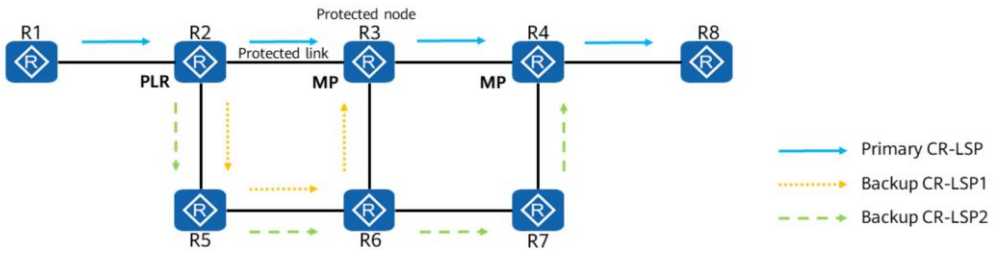
- FRR is a temporary protection measure. Compared with path protection, FRR allows traffic switching to be performed more quickly.

Related FRR Concepts (1)

- FRR can be classified into the following protection modes by tunnel establishment method:
  - One-to-one backup
  - Facility backup
- The roles of devices in FRR can be classified into the following types:
  - Point of local repair (PLR): ingress of the backup CR-LSP. The ingress must be on the path of the primary CR-LSP and cannot be the egress of the primary CR-LSP.
  - Merge point (MP): aggregation node of the primary and backup CR-LSPs. It cannot be the ingress of the primary CR-LSP.
- FRR is classified into the following types by protected object:
  - Link protection: A PLR and an MP are directly connected. A bypass CR-LSP only protects the direct link to the PLR.
  - Node protection: A PLR and an MP are indirectly connected. A bypass CR-LSP protects a direct link to the PLR and nodes on the primary CR-LSP's path between the PLR and MP.

- The bypass CR-LSP mentioned here is a local backup CR-LSP created by FRR, not a backup CR-LSP created in path protection.
- If node protection is enabled, only the link between the protected node and PLR is protected. The PLR cannot detect faults in the link between the protected node and the MP.

Related FRR Concepts (2)



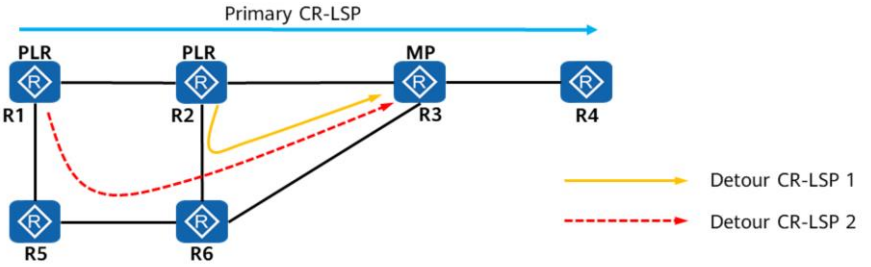
- R2 is the PLR of the primary tunnel.
- R3 is the MP of backup CR-LSP1, and R4 is the MP of backup CR-LSP2.
- Backup CR-LSP1 provides link protection for the primary CR-LSP.
- Backup CR-LSP2 provides node protection for the primary CR-LSP.
- R3 is the next hop (NHOP) router of the PLR.
- R4 is the next-next hop (NNHOP) router of the PLR

HUAWEI

- NHOP router: PLR's next hop router of the primary CR-LSP.
- NNHOP router: PLR's next-next hop router of the primary CR-LSP.
- The PLR of the primary CR-LSP already knows the NHOP and NNHOP. Link protection can be provided if the egress LSR ID of the bypass CR-LSP is the same as the NHOP LSR ID. Node protection can be provided if the egress LSR ID of the bypass CR-LSP is the same as the NNHOP LSR ID.
- When a link fault is detected, traffic is switched to the protection link within 50 ms. At the same time, the ingress of the primary tunnel is notified to re-optimize the primary tunnel.
- Link protection protects the link from PLR to MP (NHOP) instead of a specific LSP. Therefore, the bypass tunnel can protect multiple LSPs.
- Link protection works in one-to-many mode. Therefore, switching may cause partial congestion on the network.

Protection Modes Supported by FRR (1)

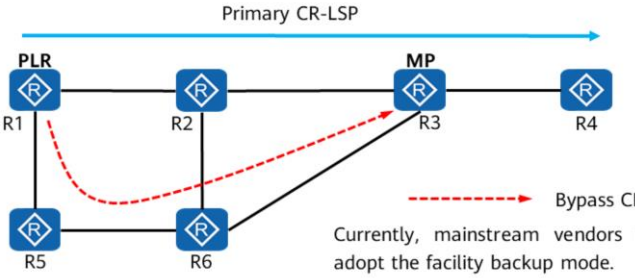
- One-to-one backup
  - A detour CR-LSP is automatically set up on each eligible node along each primary CR-LSP to protect downstream links or nodes.
  - CR-LSPs are used to protect CR-LSPs. The primary and backup CR-LSPs belong to the same tunnel.
  - In one-to-one backup mode, packets are forwarded with only one label.



- Detour CR-LSP: A detour LSP is automatically established on each node along the primary CR-LSP. Detour LSPs and the primary CR-LSP are in the same tunnel.
- This mode is easy to configure, eliminates manual network planning, and provides flexibility on a complex network. However, this mode has low extensibility, requires maintenance of the backup CR-LSP status on each node, and consumes more bandwidth than the facility backup mode.

Protection Modes Supported by FRR (2)

- Facility backup
  - A bypass tunnel is configured for each link or node that may fail on a primary tunnel. A bypass tunnel can protect traffic on multiple primary tunnels.
  - The primary and backup LSPs belong to different TE tunnels.
  - The facility backup mode requires two layers of labels during forwarding and supports label nesting.



Currently, mainstream vendors in the industry usually adopt the facility backup mode.

HUAWEI

- A bypass CR-LSP can protect multiple primary CR-LSPs. The bypass and primary CR-LSPs are established in different tunnels.
- This mode is extensible, resource efficient, and easy to implement. However, bypass tunnels must be manually planned and configured, which is time-consuming and labor-intensive on a complex network.

TE FRR Implementation

- Facility backup mode:
  - A bypass CR-LSP can be configured in manual or automatic mode. In automatic mode, a node enabled with auto FRR automatically establishes a bypass CR-LSP and binds it to the primary CR-LSP as long as the primary CR-LSP that passes through the node sends an FRR protection request and the topology meets the requirements of the FRR topology.
  - You can determine whether to configure bandwidth values for bypass CR-LSPs based on actual situations to implement bandwidth protection.
- One-to-one backup mode:
  - Detour CR-LSPs are automatically established on each node and do not need to be manually configured.
  - By default, a detour LSP has the same bandwidth as the primary CR-LSP and automatically provides bandwidth protection.

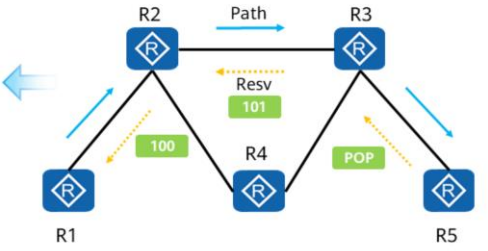
Implementation of TE FRR

- The following uses the facility backup mode as an example. The procedure is as follows:
  - Establish the primary and backup CR-LSPs.
  - Bind a bypass CR-LSP to the primary CR-LSP.
  - Perform fault detection.
  - Perform traffic switching.
  - Perform a switchback.

Setting up a Primary CR-LSP

- The process of setting up a primary CR-LSP is similar to that of setting up a common CR-LSP. The difference lies in that the ingress adds flags such as Local Protection desired and Label Record desired to the Session_Attribute object in a Path message during the setup of the primary CR-LSP. If bandwidth protection is required, a "bandwidth protection desired" flag is added to the Path message.

```
<R1> display mpls rsvp-te psb-content
------------------------------------------------------------------------------
                The PSB Content
------------------------------------------------------------------------------
Tunnel Addr: 5.5.5.5                Exist time: 16h 1m 13s
Tunnel ExtID: 1.1.1.1               Session ID: 100
Session Attribute:
  SetupPrio: 7                      HoldPrio: 7
  SessionAttrib:   SE Style desired
                   Label Record desired              //Indicates that
label recording is required.
                   Local Protection desired          //Indicates that FRR
protection is required.
                   Bandwidth Protection desired      //Indicates that
bandwidth protection is required.
```

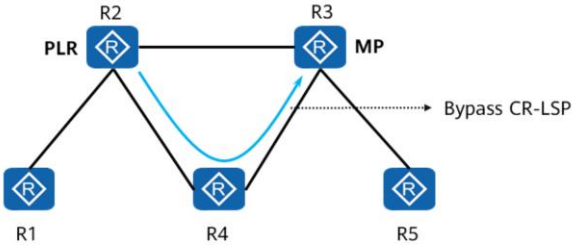


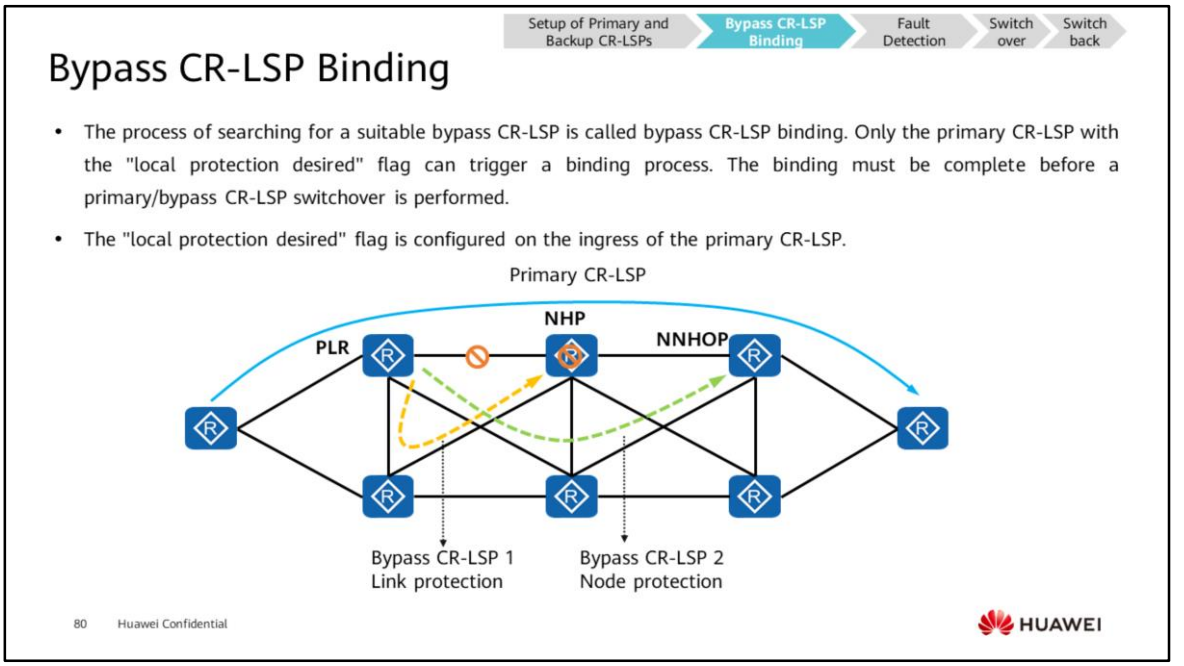
HUAWEI

- To obtain the label allocated by the NNHOP to the NHOP node, set the Label Recording Desired value in the Session_Attribute object of the Path message to 0x02.

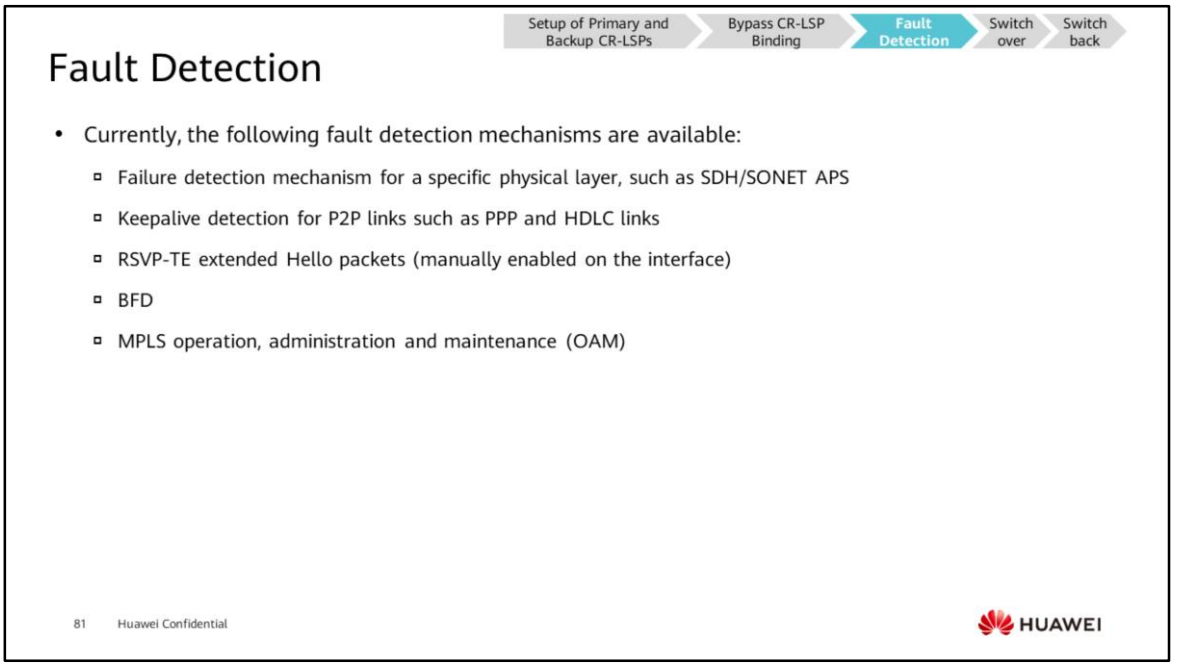
Backup CR-LSP Setup

- For the facility backup mode:
  - In manual mode, you need to manually configure a bypass CR-LSP on the PLR.
  - In automatic mode, you need to enable auto FRR on the PLR, which then automatically creates a bypass CR-LSP based on the network topology.





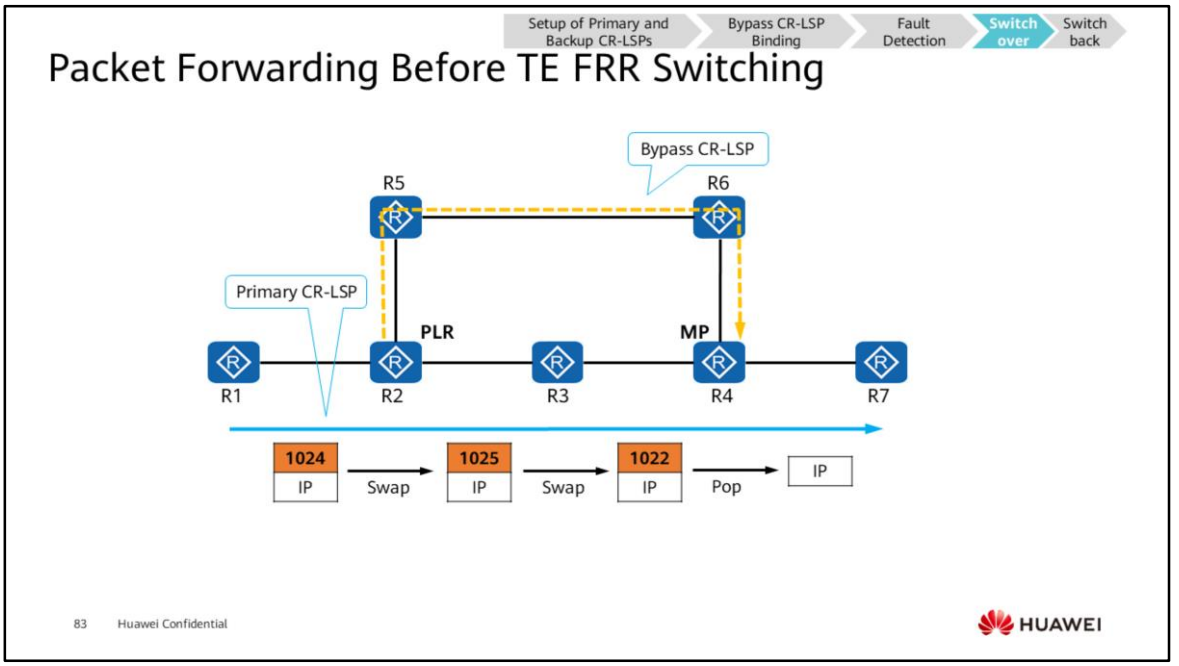
- The process of searching for a suitable bypass CR-LSP is called bypass CR-LSP binding. Only the primary CR-LSP with the "local protection desired" flag can trigger a binding process. The binding must be complete before a primary/bypass CR-LSP switchover is performed. During the binding, the PLR must obtain information about the outbound interface of the bypass CR-LSP, next hop label forwarding entry (NHLFE), LSR ID of the MP, label allocated by the MP, and protection type.
- The PLR of the primary CR-LSP already knows the NHOP and NNHOP. Link protection can be provided if the egress LSR ID of the bypass CR-LSP is the same as the NHOP LSR ID. Node protection can be provided if the egress LSR ID of the bypass CR-LSP is the same as the NNHOP LSR ID. For example, in the preceding figure, bypass CR-LSP 1 protects a link, and bypass CR-LSP 2 protects a node.
- After a bypass CR-LSP is successfully bound to the primary CR-LSP, the NHLFE of the primary CR-LSP is recorded. The NHLFE contains the NHLFE index of the bypass CR-LSP and the inner label assigned by the MP. The inner label is used to forward traffic during FRR switching.



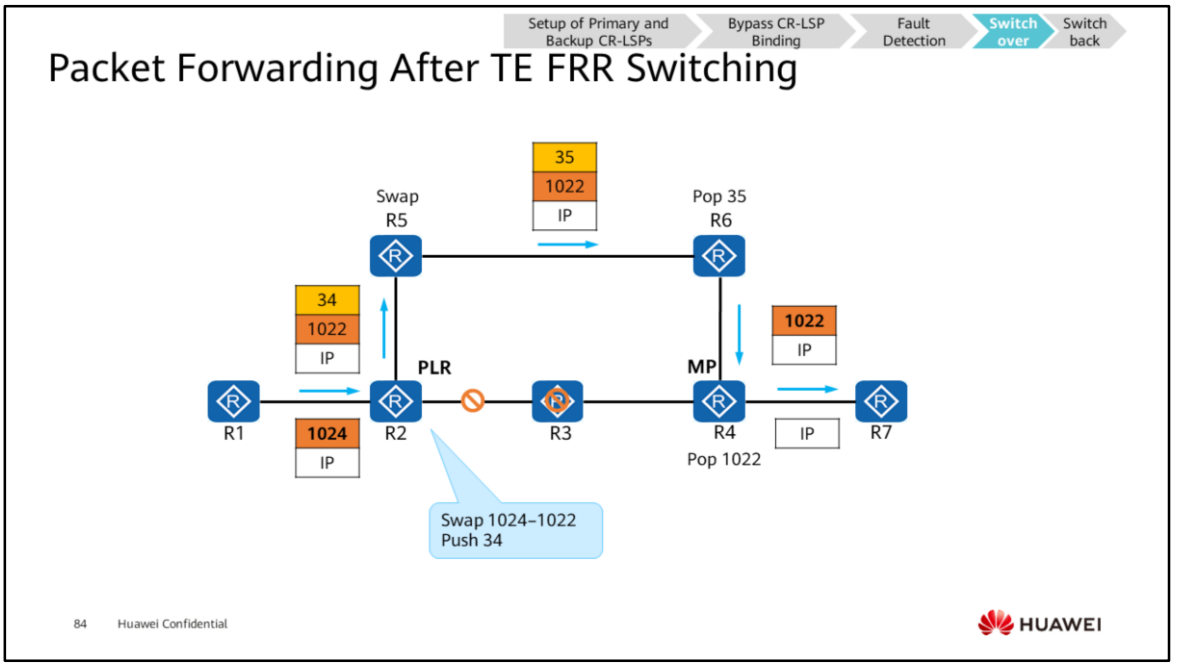
- BFD is a new fault detection mechanism that can be used to detect the reachability of vairous routing protocols (OSPF/IS-IS/BGP), MPLS LSP, VOIP, WG, and various upper-layer (for example, application layer) connections to speed up convergence.
- MPLS OAM can also quickly detect MPLS LSP faults and complete traffic switching within 50 ms.
- If node protection is enabled, only the link between the protected node and PLR is protected. The PLR cannot detect faults in the link between the protected node and the MP.

Traffic Switching

- The PLR switches traffic to the backup CR-LSP upon detection of a link failure. The internal processing of the PLR is as follows:
  - Ensure that the backup CR-LSP is ready and the label allocated to the backup tunnel is ready.
  - Traffic is switched to the backup CR-LSP after the preceding step is complete. During the switching, the corresponding parameters in the RSVP packets are modified or updated:
    - Clear the Local Protection Desired flag in the Session_Attribute object.
    - Change the IP address of the outbound interface on the PLR to the IP address of the outbound interface on the backup CR-LSP.
    - Generate a new ERO.
    - Update the RRO.
    - Change PLR labels — two-Layer label nesting



- Before TE FRR switching is performed, traffic is forwarded along the primary CR-LSP, and labels are pushed in and popped out along the primary CR-LSP.

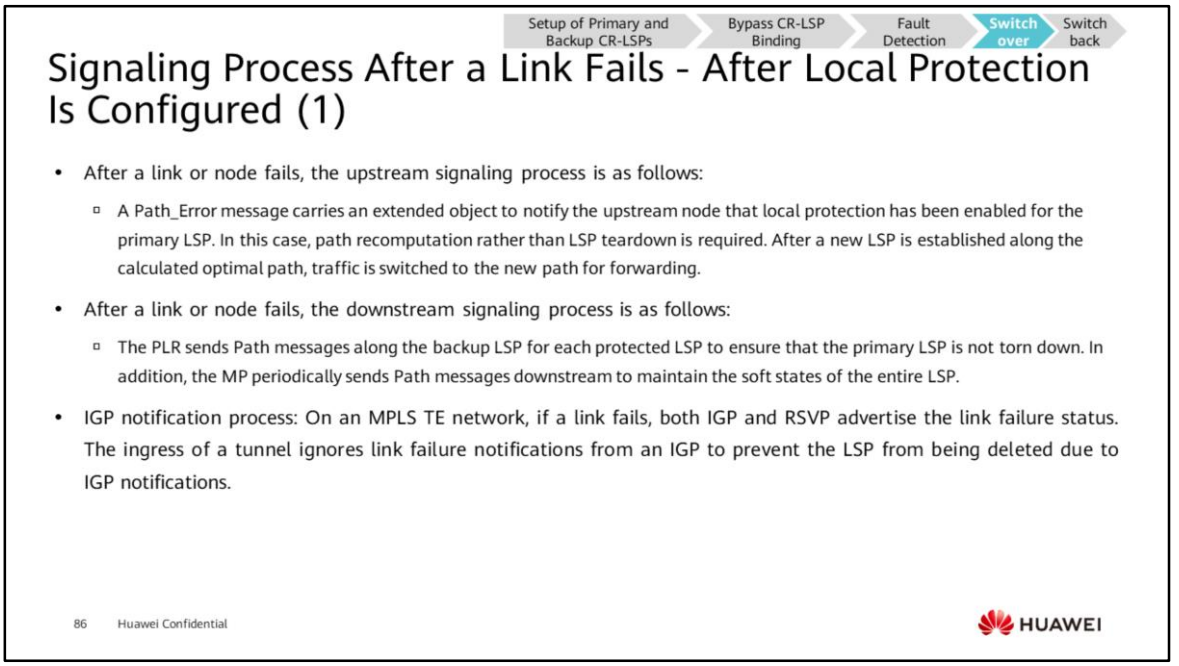


- If a primary CR-LSP fails, the PLR switches both service traffic and RSVP messages to a bypass CR-LSP and advertises the switchover event upstream. During the switchover, the MPLS label nesting mechanism is used. The PLR pushes the inner and outer labels that the MP assigns for the primary and bypass CR-LSPs, respectively. The penultimate hop along the bypass CR-LSP removes the outer label from the packet and forwards the packet only with the inner label to the MP. The MP forwards the packet to the next hop along the primary CR-LSP.
- The bypass CR-LSP provides node protection. If R3 or the link between R2 and R3 fails, traffic is switched to the bypass CR-LSP. During the switching, the PLR (R2) swaps an inner label 1024 for an inner label 1022, pushes an outer label 34 into the packet, and forwards the packet over the bypass CR-LSP. After the packet arrives at R4, the R4 forwards the packet to the next hop. The packet finally arrives at the tunnel egress.

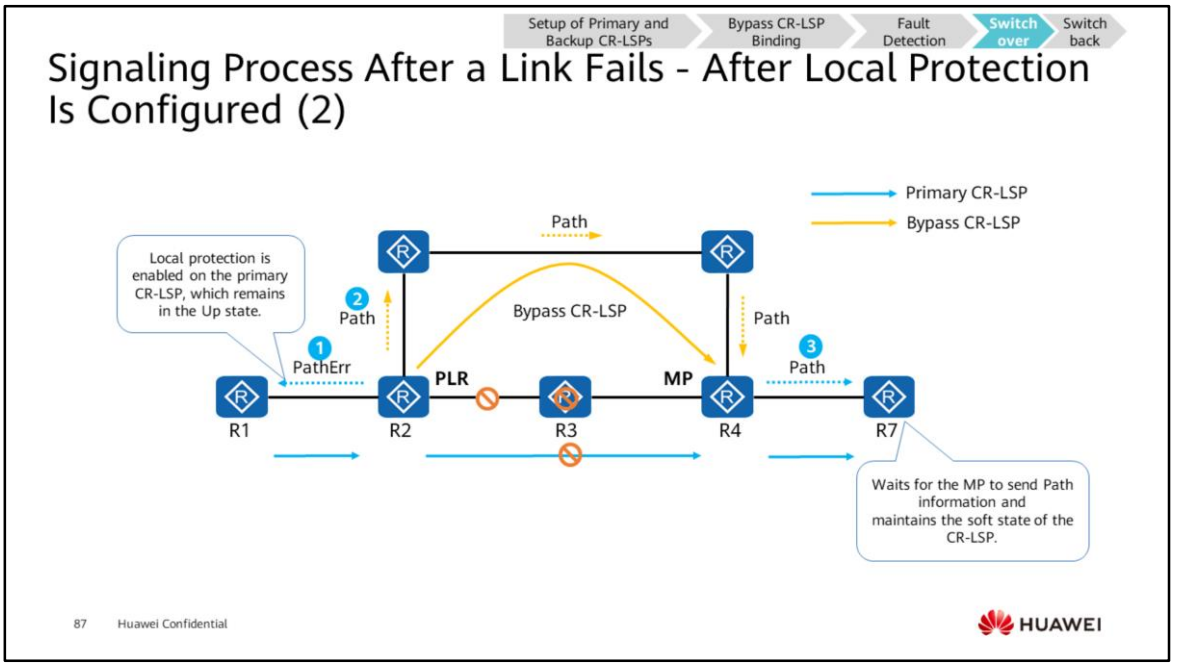
Signaling Process After a Fault Occurs - Local Protection Not Configured

- After a link or node fails, the upstream signaling process is as follows:
  - The PLR sends a Path_Error message to the ingress and disables the primary LSP.
- After a link or node fails, the downstream signaling process is as follows:
  - If local protection is not configured, the MP sends a Path_Tear message downstream to tear down the primary LSP if it does not receive Path information from upstream within a period of time.

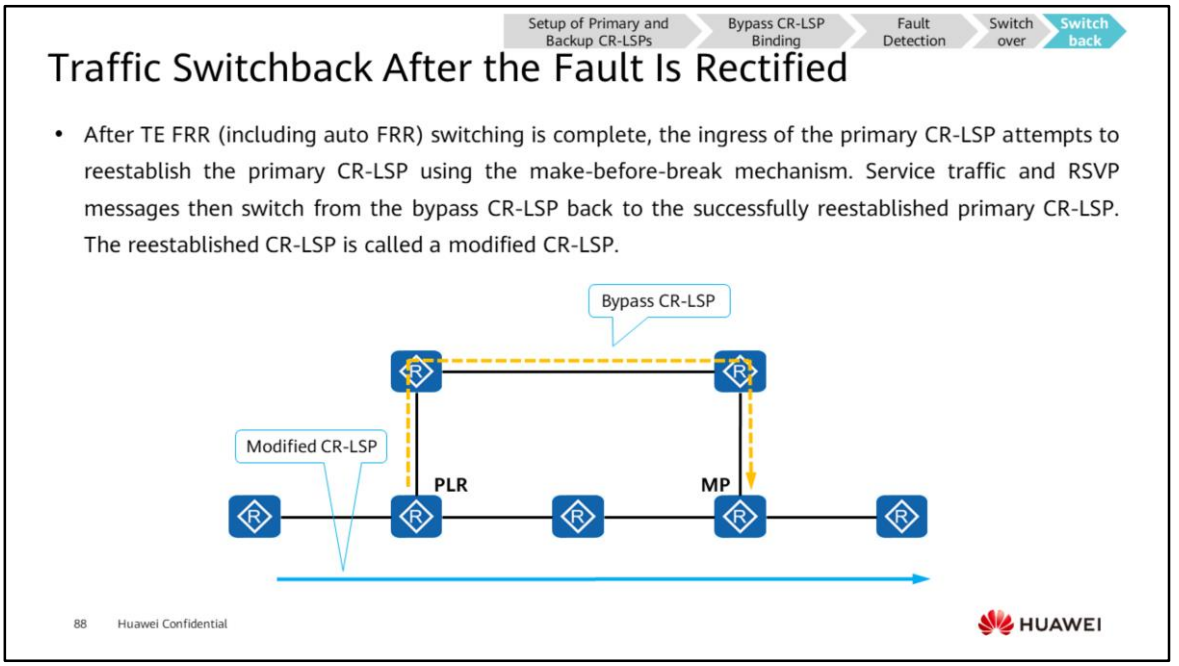
- If local protection is not configured, the PLR immediately sends a Path_Error message upstream and disables the primary LSP. If the MP does not receive a Path message upstream after a period of time, the MP sends a Path_Tear message downstream to delete the local state created by each node. Upon receipt, each node sends a ResvTear message upstream to delete the corresponding local resources.



- Upstream signaling process: The original Path_Error message is extended by adding Code=25 (Notification) and Subcode=3 (Tunnel locally repaired). This indicates that local protection has been enabled for the primary LSP and that path recalculation rather than LSP teardown is required. After a new LSP is set up along the optimal path, traffic is switched to the new path for forwarding.



- The upstream node is notified that local protection is enabled for the primary LSP and LSP teardown is not required. Step 1
- The PLR sends Path messages along the bypass LSP for each protected LSP to ensure that the primary LSP is not torn down. Step 2
- The MP periodically sends Path messages downstream to maintain the soft states of the entire LSP. Step 3



- For an established MPLS TE tunnel, topology and resource changes may cause the original LSP to fail to meet the requirements. This means that a new CR-LSP needs to be established. If traffic is switched to the new CR-LSP before it is established, as a result, traffic loss may occur.
- Make-before-break is a mechanism that allows a CR-LSP to be established using changed MPLS TE tunnel attributes (such as bandwidth and path) over a new path before the original CR-LSP is torn down. It helps minimize data loss and additional bandwidth consumption. The new CR-LSP is called a modified CR-LSP. Make-before-break is implemented using the shared explicit (SE) resource reservation style.
- The new CR-LSP may compete with the primary CR-LSP on some shared links for bandwidth. The new CR-LSP cannot be established if it fails the competition. The make-before-break mechanism allows the new CR-LSP to use the bandwidth of the original path, without needing to recalculate the bandwidth to be reserved for the new path. Additional bandwidth resources are consumed only when some links on the new LSP do not overlap with those on the original LSP.

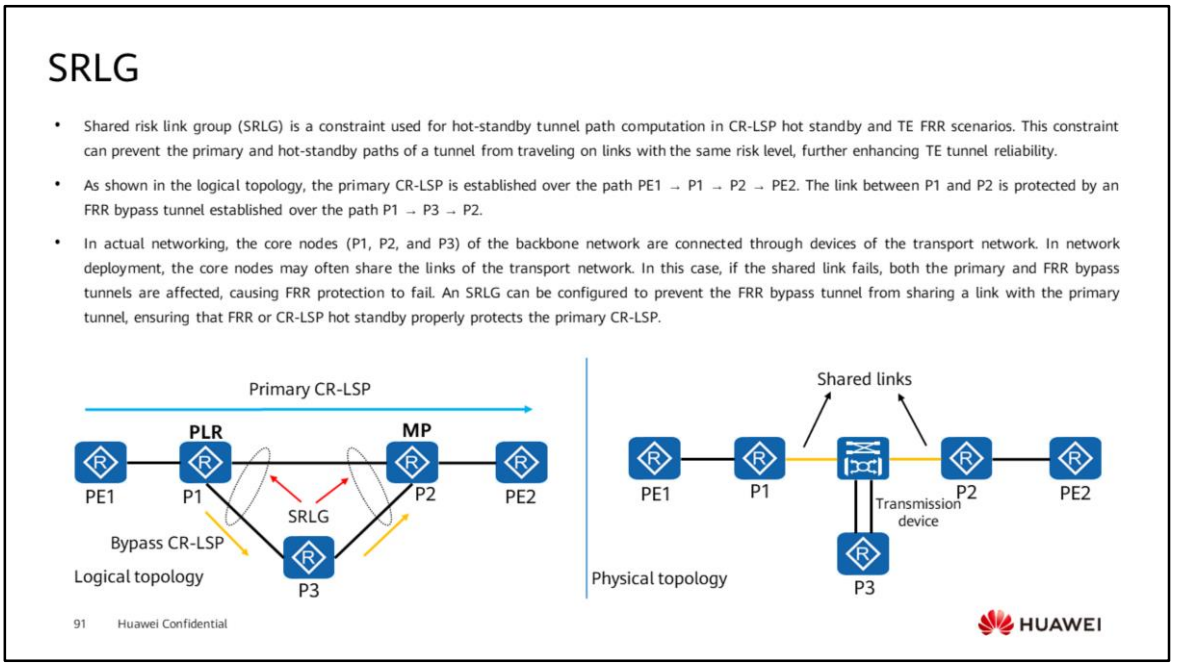
Joint Use of Path Protection and FRR (1)

- When hot standby and TE FRR are used together, the PLR location varies according to the fault location. Accordingly, the switching sequences of CR-LSP hot standby and TE FRR are different.
  - If the PLR is the ingress of the primary CR-LSP, it can fast detect the fault and immediately switch traffic to the hot-standby CR-LSP instead of entering the TE FRR process.
  - If the PLR is a transit node along the primary CR-LSP, the ingress of the primary CR-LSP may fail to detect the fault in time. To ensure uninterrupted traffic forwarding, the PLR switches traffic to the TE FRR bypass tunnel before transmitting the fault information to the ingress of the primary CR-LSP through RSVP signaling, triggering the ingress to switch traffic to the hot-standby CR-LSP.

- TE FRR is a temporary local protection mechanism used when the ingress does not detect a fault. It is a supplement to CR-LSP hot standby. Once the ingress of the primary CR-LSP detects the fault, it switches traffic to the hot-standby CR-LSP.
- If the PLR is a transit node along the primary CR-LSP, the ingress cannot quickly detect the fault due to slow RSVP signaling transmission. As a result, traffic is always forwarded over the TE FRR bypass tunnel. To speed up switching of traffic to the hot-standby CR-LSP by the ingress, you can enable CSPF fast switching. After this function is enabled, an IGP notifies the ingress of the primary CR-LSP when its topology changes. After detecting the fault, the ingress of the primary CR-LSP does not wait for RSVP signaling but switches traffic to the hot-standby CR-LSP in advance.
- If the hot-standby CR-LSP is down, the ingress keeps attempting to reestablish a hot-standby CR-LSP.

Joint Use of Path Protection and FRR (2)

- If ordinary backup and TE FRR are used together:
  - If ordinary backup is not associated with TE FRR: If a protected link or node fails, a PLR switches traffic to a bypass CR-LSP. Only after both the primary and bypass CR-LSPs fail, the ingress of the primary CR-LSP attempts to establish an ordinary backup CR-LSP and switches traffic to this CR-LSP.
  - If ordinary backup is associated with TE FRR: If a protected link or node fails, a PLR switches traffic to a bypass CR-LSP first. If the PLR is the ingress of the primary CR-LSP, the PLR attempts to set up an ordinary backup CR-LSP. After the ordinary backup CR-LSP is set up, the PLR switches traffic to this CR-LSP. If the PLR is a transit node on the primary CR-LSP, the PLR uses RSVP signaling to send fault information to the ingress of the primary CR-LSP, triggering the ingress to attempt to set up an ordinary backup CR-LSP. If the ordinary backup CR-LSP is set up successfully, the ingress switches traffic to this CR-LSP.

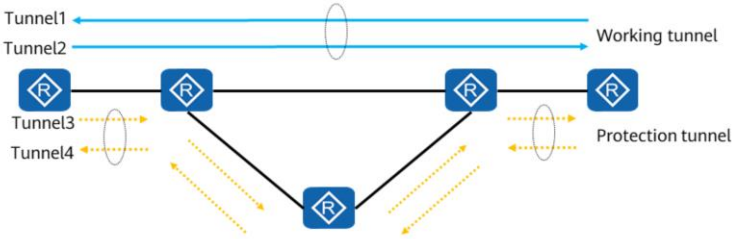


- An SRLG is a set of links at the same risk of faults. If a link in an SRLG fails, other links may also fail. If a link in this group is used by a hot-standby CR-LSP or FRR bypass CR-LSP of the failed link, the hot-standby CR-LSP or FRR bypass CR-LSP cannot provide protection.
- SRLG is a link attribute, in numerical notation. The links with the same SRLG value are in one SRLG.
- IGP TE advertises SRLG information to all nodes in a single MPLS TE domain through an IGP. The CSPF algorithm uses the SRLG attribute together with other constraints, such as bandwidth, to calculate a path.
- The MPLS TE SRLG works in either of the following modes:
    - Strict mode: The SRLG attribute is a necessary constraint used by CSPF to calculate a path for a hot-standby CR-LSP or an FRR bypass CR-LSP.
    - Preferred mode: The SRLG attribute is an optional constraint used by CSPF to calculate a path for a hot-standby CR-LSP or FRR bypass CR-LSP. For example, if CSPF fails to calculate a path for a hot-standby CR-LSP based on the SRLG attribute, CSPF recalculates the path without taking into account the SRLG attribute.

Contents

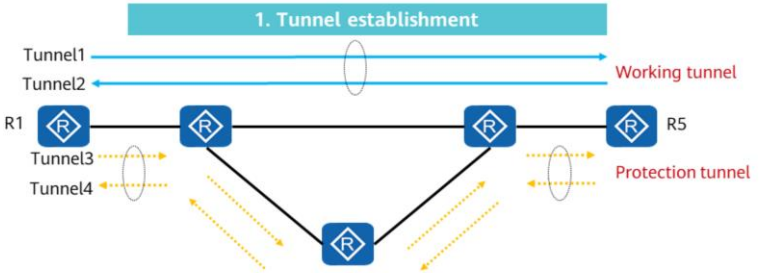
Tunnel Protection Group Overview

- A tunnel protection group provides E2E protection for MPLS TE tunnels. If a working tunnel in a protection group fails, traffic switches to a protection tunnel, minimizing traffic interruptions.
- The working and protection tunnels must be bidirectional. Therefore, an MPLS TE tunnel protection group supports associated bidirectional LSPs and static bidirectional co-routed LSPs.
- As shown in the figure, Tunnel1 and Tunnel2 are working tunnels, and Tunnel3 and Tunnel4 are protection tunnels. The four tunnels form an MPLS TE tunnel protection group. When a fault is detected on a working tunnel, traffic is switched to a protection tunnel for forwarding.



Tunnel Protection Group Implementation (1)

- Tunnel establishment: Bidirectional primary and protection tunnels are established. The tunnel establishment process is similar to the process of establishing a common TE tunnel. Note that the primary and protection tunnels must have the same ingress and destination IP address.
- A protection tunnel cannot be protected by another protection tunnel, and TE FRR cannot be enabled for the protection tunnel.
- Tunnel attribute inconsistency between the working and protection tunnels facilitates network planning.



Tunnel Protection Group Implementation (2)

2. Binding of working and protection tunnels

- After the tunnel protection group function is enabled for a working tunnel, the working tunnel and a protection tunnel are bound to form a tunnel protection group.
- As shown below, a tunnel ID is set for each tunnel, and each protection tunnel with a specified tunnel ID is bound to a working tunnel.

```
[R1-tunnel1] mpls te tunnel-id 100
[R5-tunnel2] mpls te tunnel-id 200
[R1-tunnel3] mpls te tunnel-id 101
[R5-tunnel4] mpls te tunnel-id 201
```

```
[R1] interface tunnel 1
[R1-tunnel1] mpls te protection tunnel 101
[R5] interface tunnel 2
[R5-tunnel2] mpls te protection tunnel 201
```

3. Fault detection

- MPLS OAM/MPLS-TP OAM is used to detect faults in a tunnel protection group to speed up protection switching.
- As shown below, MPLS-TP OAM is configured on the ingress of the forward and reverse primary tunnels to detect faults.

```
[R1] mpls-tp meg abc
[R1-mpls-tp-meg-abc] me te interface Tunnel 1 mep-id 1 remote-mep-id 2
[R5] mpls-tp meg abc
[R5-mpls-tp-meg-abc] me te interface Tunnel 2 mep-id 2 remote-mep-id 1
```

Tunnel Protection Group Implementation (3)

4. Protection switching	5. Switchback
• When the ingress detects a fault on the working tunnel, it triggers protection switching in the tunnel protection group. Two modes are available: ▫ Manual switchover: A network administrator runs commands to switch traffic. ▫ Automatic switching: Automatic traffic switching is performed after the working tunnel detects a fault. • Currently, an MPLS TE tunnel protection group supports only bidirectional switching. That is, if a switchover is performed for traffic in one direction, a switchover is also performed for traffic in the opposite direction.	• After a traffic switchover is complete, the ingress keeps trying to reestablish the working tunnel. If the working tunnel is reestablished, the ingress determines whether to switch traffic back to the working tunnel according to the configured switchback policy.

Configure the binding between the working and protection tunnels, specify whether to perform a switchback, and configured the WTR time:

```
[R1] interface tunnel 1
[R1-tunnel1] mpls te protection tunnel 101 mode revertive wtr 0
[R5] interface tunnel 2
[R5-tunnel2] mpls te protection tunnel 201 mode revertive wtr 0
```

Differences Between CR-LSP Backup and a Tunnel Protection Group

Comparison Item	CR-LSP Backup	Tunnel Protection Group
Object to be protected	Primary and backup CR-LSPs are established on the same tunnel interface. The backup CR-LSP protects the primary CR-LSP.	In a tunnel protection group, one tunnel protects another.
TE FRR	TE FRR protection is supported only by the primary CP-LSP, not the backup CR-LSP.	A tunnel protection group depends on the reverse LSP, which is mutually exclusive with TE FRR. Therefore, tunnels in a tunnel protection group do not support TE FRR.
LSP attributes	The primary and backup CR-LSPs have the same attributes (such as bandwidth, setup priority, and hold priority), except for the TE FRR attribute.	Tunnels in a tunnel protection group use independent attributes. This means that attributes of one tunnel are irrelevant to those of another tunnel. For example, a protection tunnel without bandwidth can protect a working tunnel requiring bandwidth protection.

Contents

RSVP Hello

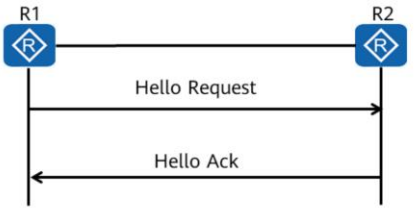
- After a TE tunnel is established, Path and Resv messages are sent periodically to check neighbor reachability and to maintain the resource reservation status (including PSB and RSB) of nodes. This method, however, takes a long time. If a fault occurs on the primary path, traffic cannot be switched to the backup path in a timely manner. Therefore, RSVP Hello is introduced to solve this problem.
- The RSVP Hello extension can rapidly monitor the reachability of RSVP nodes. If an RSVP node becomes unreachable, TE FRR protection is triggered.

Type = 20 Hello

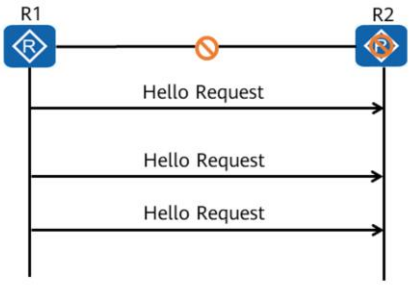
HUAWEI

- Hello messages are defined in RFC 3209 and used for fast detection (Keepalive) between MPLS TE neighbors.

How RSVP Hello Works



- After RSVP Hello is enabled on R1, R1 sends a Hello Request message to R2. After receiving the message, R2 replies with a Hello ACK message. Upon receipt, R1 determines that its neighbor R2 is reachable.



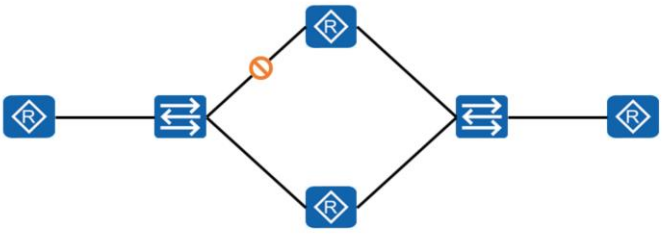
- If R2 does not respond Hello Ack messages to three consecutive Hello Request messages sent by R1, R1 triggers TE FRR switching, and re-initializes RSVP Hello.

HUAWEI

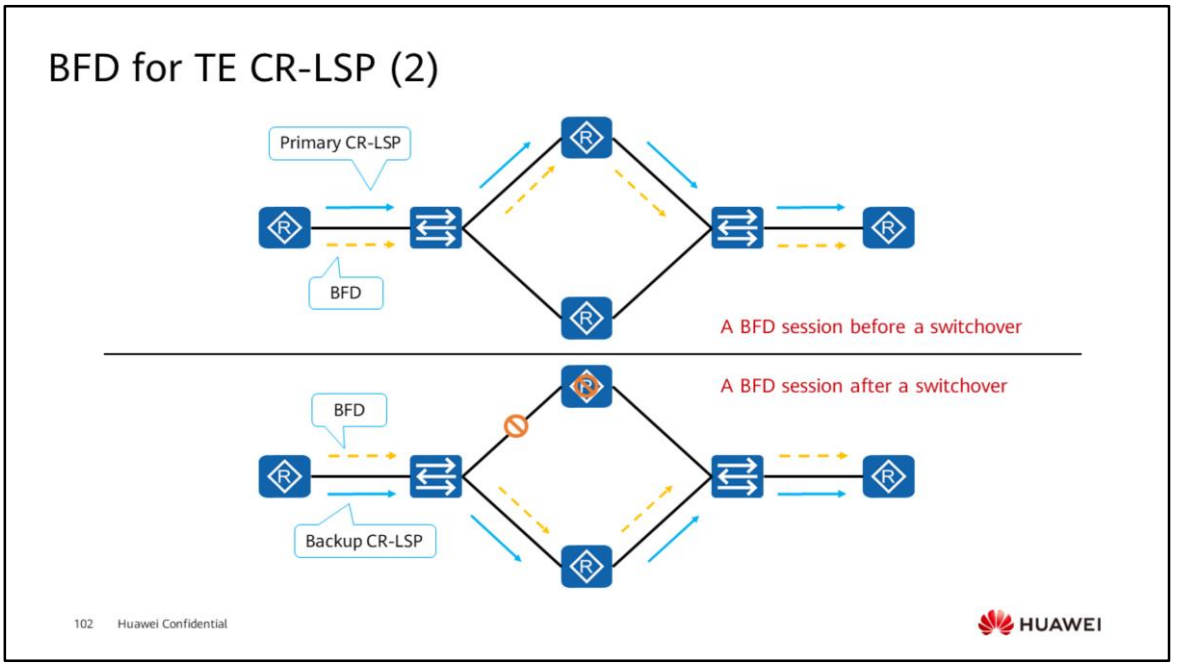
- R1 and R2 are connected through a direct link.

BFD for TE CR-LSP (1)

- As shown in the figure, the ingress of the tunnel cannot immediately detect a fault due to the Layer 2 switch. In this case, you can deploy the Hello protocol for fault detection, but the detection time is long. To resolve this problem, bind a BFD session to a CR-LSP. That is, set up a BFD session between the ingress and egress of a CR-LSP to solve the problem of long detection time. A BFD packet is sent by the ingress to the egress along the CR-LSP. Upon receipt, the egress responds to the BFD packet. The ingress can rapidly detect the status of links through which the CR-LSP passes.
- After detecting a link fault, the ingress searches for a backup CR-LSP and switches traffic to the backup CR-LSP.
- BFD for TE is usually used together with the hot-standby CR-LSP mechanism.



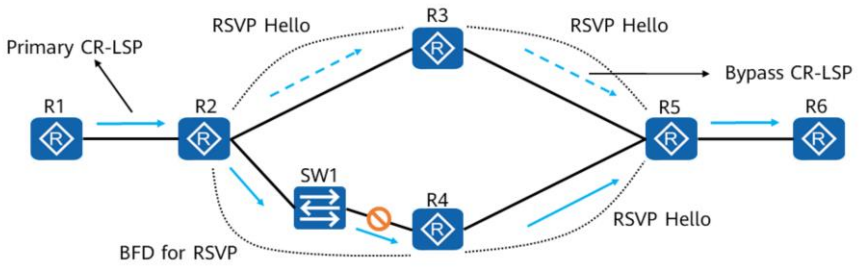
- Conventional detection mechanisms such as RSVP Hello and RSVP Srefresh (summary refresh) mechanisms detect faults slowly. In contrast, BFD rapidly sends and receives packets to detect faults in a tunnel. If a fault occurs, BFD triggers rapid service switching to protect service traffic.



- A BFD session is set up to detect the link through which the primary LSP passes. If a fault occurs on this link, the BFD session on the ingress immediately reports the failure. The ingress switches traffic to the backup CR-LSP and sets up a new BFD session to detect faults on the link of the backup CR-LSP.

BFD for RSVP

- When a Layer 2 device exists on a link between two RSVP neighbors, BFD for RSVP can be deployed to rapidly detect faults in this link. If a link fault occurs, BFD for RSVP rapidly detects the fault and trigger TE FRR switching.
- BFD for RSVP applies to the TE FRR networking where Layer 2 devices exist between the PLR and the RSVP neighbor along the primary CR-LSP.



HUAWEI

- When a Layer 2 device is deployed on the link between two neighboring RSVP nodes, the RSVP nodes can only use the Hello mechanism to detect link faults. As shown in the figure, a Layer 2 device (SW1) exists between R2 and R4. If the link between SW1 and R4 fails, R2 cannot quickly detect the fault through the link layer because SW1 isolates the fault. Instead, R2 can only be notified of the fault through the RSVP Hello mechanism. In this case, fault detection is performed in milliseconds, causing a large amount of data to be lost. BFD for RSVP can implement faster fault detection in this scenario, shorten the fault detection time, and trigger TE FRR to perform fast switchover, improving network reliability.

Contents

Tunnel Priority

- MPLS TE tunnel priority:
  - Eight priorities are provided. The value ranges from 0 to 7. A smaller value indicates a higher priority.
- Tunnel priorities are classified into the following types:
  - Setup priority
  - Hold priority
- Important services can preempt the resources of low-priority tunnels by setting high-priority tunnels to ensure service quality of high-end users.
- If the setup priority of a new tunnel is higher than the hold priority of the original tunnel, tunnel preemption occurs.

HUAWEI

- As shown in the figure, the hold priority of Tunnel1 is lower than the setup priority of Tunnel2. When Tunnel2 needs to set up a tunnel on the link between R1 and R2 but the resources are insufficient, Tunnel2 preempts the resources of Tunnel1 to set up a tunnel, ensuring the normal running of high-priority services.

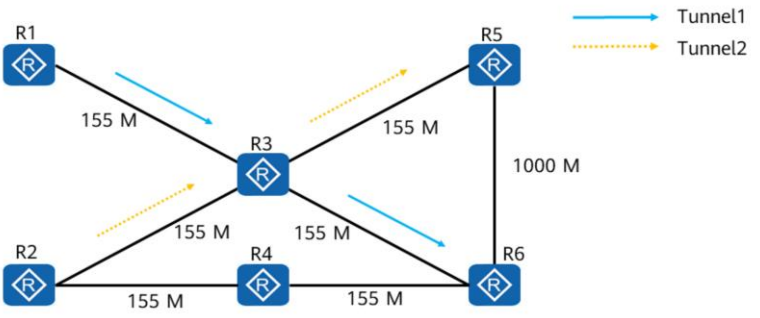
Tunnel Preemption

- The priority and preemption attributes are used in conjunction to determine resource preemption among tunnels. When multiple tunnels need to be set up, the tunnels with higher setup priorities preempt resources and are set up preferentially. If resources such as bandwidth are insufficient, a tunnel with a higher setup priority may preempt the bandwidth resources of an established tunnel with a lower hold priority.
- MPLS TE supports the following tunnel preemption modes:
  - Hard preemption: The original tunnel is directly torn down without being protected.
  - Soft preemption: The original tunnel is not torn down, and a new tunnel is established. After traffic is switched, the original tunnel is torn down.

- To prevent some tunnels from being preempted, you can set the setup priority to 7 and hold priority to 0.

Tunnel Preemption Instance (1)

- Tunnel1: R1 -> R3 -> R6. The required bandwidth is 155 Mbit/s, and the setup and hold priorities are both 0.
- Tunnel2: R2 -> R3 -> R5. The required bandwidth is 155 Mbit/s, and the setup and hold priorities are both 7.

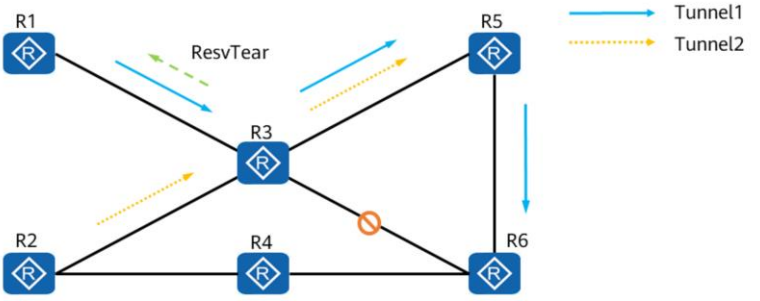


HUAWEI

- Two TE tunnels with different priorities are set up. The priority of Tunnel1 is 0 and that of Tunnel2 is 7.

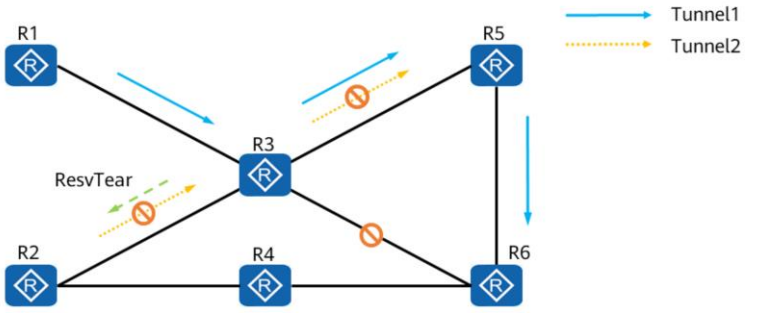
Tunnel Preemption Instance (2)

- When the link between R3 and R6 fails, R3 sends a ResvTear message to R1. R1 recalculates a new path R1 -> R3 -> R5 -> R6 for Tunnel1.
- As shown in the figure, the bandwidth (155 Mbit/s) of the link between R3 and R5 is insufficient for Tunnel1 and Tunnel2 (310 Mbit/s in total). In this case, preemption occurs.



Tunnel Preemption Mode - Hard Preemption

- In hard preemption mode, R3 directly sends an RSVP message to tear down Tunnel2 because Tunnel1 has a higher priority than Tunnel2.
- In hard preemption mode, if Tunnel2 has traffic, some traffic will be lost.

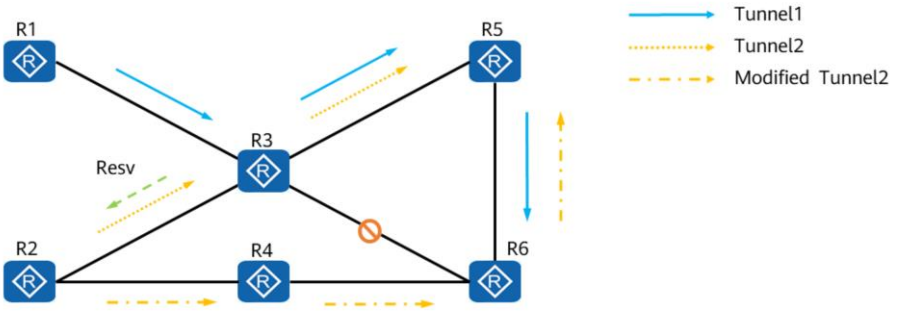


HUAWEI

- Hard preemption: When a high-priority tunnel competes for resources with a low-priority tunnel, the high-priority tunnel directly preempts resources of the low-priority tunnel without waiting. As a result, some traffic is lost.

Tunnel Preemption Mode - Soft Preemption

- In soft preemption mode, R3 sends a Resv message to R2 and sets Preemption pending in the message.
- In make-before-break mode, Tunnel2 is re-established along the path R2 -> R4 -> R6 -> R5 without tearing down Tunnel2 on R2. The original Tunnel 2 is torn down after traffic switching is complete.



HUAWEI

- Soft preemption: The make-before-break mechanism applies. A CR-LSP with a higher priority has to wait until traffic over a lower-priority CR-LSP switches to another CR-LSP before the higher-priority CR-LSP preempts bandwidth assigned to the lower-priority CR-LSP.

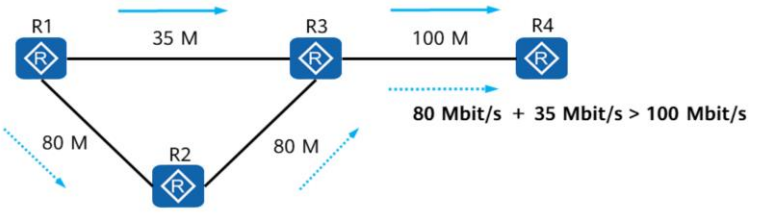
Tunnel Re-optimization

- After a CR-LSP is established in a tunnel, the path of the CR-LSP can be optimized based on changes in bandwidth, traffic requirements, management policies, and other factors of the network.
- A CR-LSP can be optimized when a better path is available. That is, the CR-LSP is re-set up only when a better path with a smaller metric and fewer hops exists.
- Tunnel optimization uses the make-before-break mechanism to ensure user service continuity.
- During tunnel re-optimization, the share explicit (SE) mode must be used to prevent overlapping of the old and new CR-LSPs.
- Re-optimization is classified into the following modes by triggering mode:
  - Automatic re-optimization: enables an ingress to run CSPF to calculate a path for a CR-LSP after a configured re-optimization triggering interval elapses. If CSPF calculates a path with a metric value smaller than that of the existing path, the CR-LSP is reestablished over the better path. If the CR-LSP is successfully established, the system notifies the forwarding plane to switch traffic and tear down the original CR-LSP. After the process, re-optimization is complete. If the CR-LSP fails to be reestablished, traffic still travels through the original CR-LSP.
  - Manual re-optimization: A network administrator runs a re-optimization command in the user view to trigger path re-optimization on the ingress.

- In some special circumstances, TE tunnel re-optimization is not desired. MPLS TE provides the tunnel locking function. If this function is enabled for a tunnel, the tunnel cannot be re-optimized after being established.
- If the tunnel uses the fixed filter (FF) reservation style, tunnel re-optimization cannot be configured.
- Although re-optimization can be successfully configured for TE tunnels established over explicit paths, the configuration does not take effect.

Re-optimization of Tunnels in SE Mode

- Originally, R1 notifies the network of the 35 Mbit/s bandwidth request, and the path selected by the tunnel is R1 -> R3 -> R4.
- The original bandwidth needs to be increased to 80 Mbit/s to meet new service requirements. If the original link bandwidth cannot meet the requirements, the ingress of the tunnel recalculates a tunnel path and selects the path R1 -> R2 -> R3 ->R4.
- The make-before-break mechanism is adopted. The total bandwidth required on the path between R3 and R4 will exceed 100 Mbit/s.
- The SE mode ensures that the old and new LSPs share bandwidth resources on the path between R3 and R4 until the old tunnel is torn down.

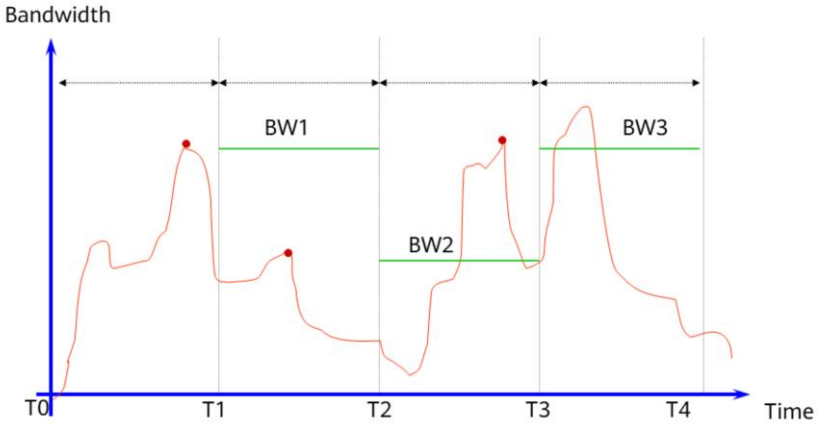


- The SE mode is an important condition for implementing the make-before-break mechanism.
- In SE mode, bandwidth resources can be shared only in the same session.

Automatic Bandwidth Adjustment

- The bandwidth change diagram within a sampling period can be obtained by periodically (sampling frequency) sampling traffic of each LSP.
- The highest sampling bandwidth can be obtained within a sampling period. Then, the highest sampling bandwidth is used as the tunnel bandwidth to initiate the establishment of a new LSP.
- After the LSP is established, traffic switches to this LSP and the original LSP is torn down.
- If the LSP fails to be established, traffic is still transmitted along the original LSP. Traffic is adjusted again after the net sampling period elapses.
- To avoid unnecessary adjustments, you can configure an adjustment threshold. The bandwidth is adjusted only when the ratio of the maximum sampling bandwidth of this time to the maximum sampling bandwidth of last time reaches a certain threshold.

Automatic Bandwidth Adjustment



HUAWEI

- The maximum sampling bandwidth BW1 is obtained in the time range from T0 to T1, and a new LSP uses BW1 as the reserved bandwidth in the time range from T1 to T2.
- The maximum sampling bandwidth BW2 is obtained in the time range from T1 to T2, and a new LSP uses BW2 as the reserved bandwidth in the time range from T2 to T3.
- The maximum sampling bandwidth BW3 is obtained in the time range from T2 to T3, and a new LSP uses BW3 as the reserved bandwidth in the time range from T3 to T4.

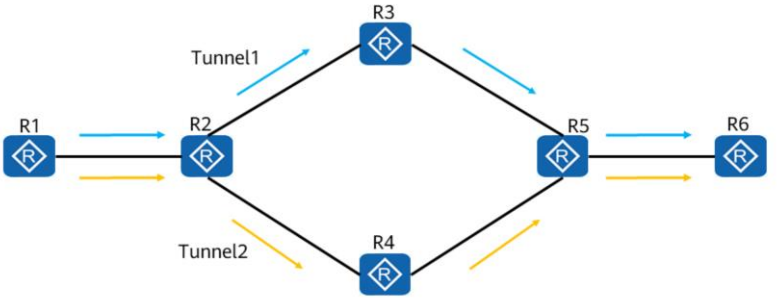
Load Balancing (1)

- Traditional IP routing protocols support equal-cost load balancing and unequal-cost load balancing. Load balancing is determined by the characteristics of routing protocols.
- MPLS TE tunnels also support equal-cost and unequal-cost load balancing.
  - Equal-cost load balancing mode: Similar to IP load balancing, per-packet load balancing or per-flow load balancing can be used. Generally, the per-flow mode is used.
  - Unequal-cost load balancing mode:
    - Traffic is load balanced based on the bandwidth ratio configured for each tunnel.
    - Traffic is load balanced based on the configured load balancing value.
- Note: Load balancing is not supported between TE tunnels and IGP routes. Traffic can be load balanced only by establishing multiple TE tunnels to the destination.

- Unequal-cost load balancing supports EIGRP.
- Equal-cost load balancing supports the following routing protocols: OSPF and IS-IS

Load Balancing (2)

- Traffic load balancing can be implemented by establishing two tunnels with the same destination address and reserved resources on the ingress of a tunnel.



- In most cases, the requirements differ from the actual situation. Therefore, after a TE tunnel is established, you need to monitor the traffic of the TE tunnel and adjust the reserved bandwidth of the tunnel in real time to ensure that network resources are properly used.

Quiz

1. (Multiple-answer question) In which of the following situations does MPLS TE advertise link information? ( )
   A. A link is activated or deactivated.
   B. The link bandwidth changes.
   C. The administrative group attribute of an interface changes.
   D. The affinity attribute of a tunnel changes.
2. (Multiple-answer question) After an MPLS TE tunnel is successfully established, which of the following methods can be used to import traffic to the tunnel? ( )
   A. Static route
   B. Auto route
   C. PBR
   D. Tunnel policy

- ABC
- ABCD

Summary

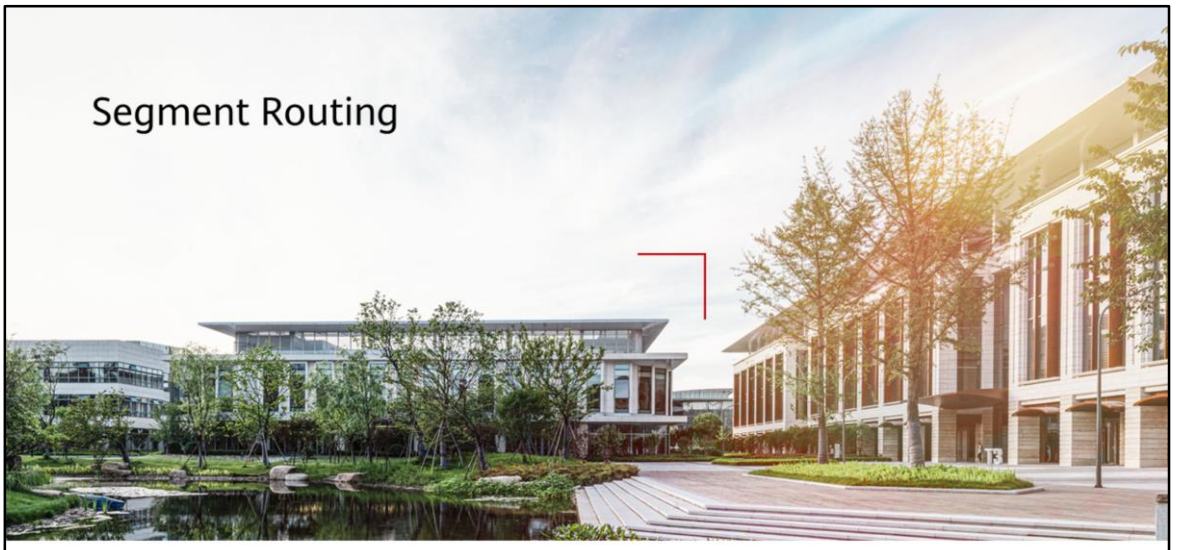
- MPLS TE establishes CR-LSPs and directs traffic to these LSPs for transparent transmission. In this manner, network traffic is transmitted along specified paths, achieving traffic engineering. MPLS TE uses the information advertisement component, path computation component, and path establishment component to establish an E2E MPLS TE tunnel. Then, the traffic forwarding component directs service traffic to the tunnel to complete forwarding of service traffic.
- MPLS TE supports multiple detection and protection mechanisms, such as TE FRR, CR-LSP backup, tunnel protection group, BFD for CR-LSP, and BFD for RSVP. If an MPLS TE tunnel fails or a node on the network is congested, the failure or congestion can be quickly detected and traffic can be switched to a backup path.
- This course describes the working principles, tunnel protection, fault detection, and some advanced features of MPLS TE, which lays a foundation for the subsequent study of segment routing.

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.

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.

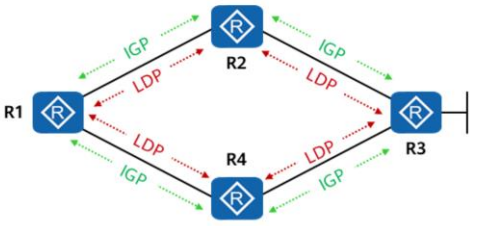
Objectives

- Upon 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.

Contents

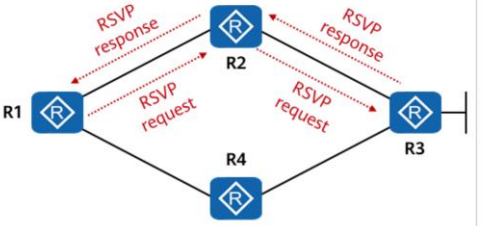
Problems in MPLS LDP and RSVP-TE

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



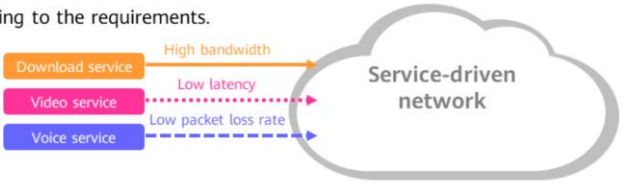
- RSVP-TE configuration is complex and load balancing is not supported.
- To implement TE, devices need to exchange a large number of RSVP packets to maintain neighbor relationships and path states, wasting link bandwidth and device resources.
- RSVP-TE uses a distributed architecture, so that each device 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

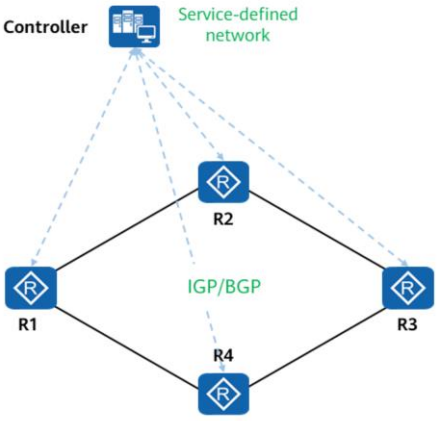
- The development of 5G and cloud services has changed the attributes and scope of network connections. More requirements are raised on connections, such as requiring better SLA guarantee, deterministic latency, or more information to be carried in packets.
- In this situation, the model that requires 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.

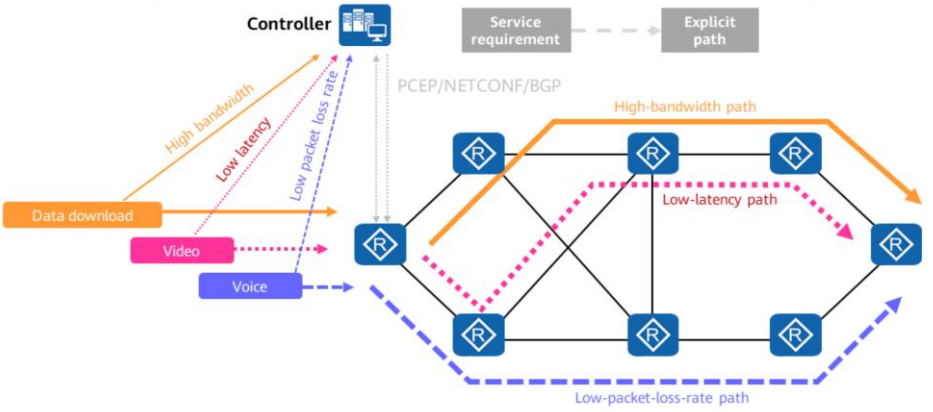
SR Roadmap



- 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.

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.



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.

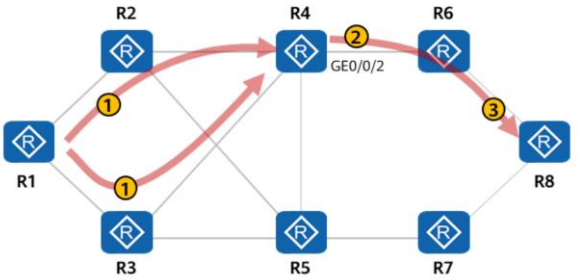
- https://datatracker.ietf.org/doc/rfc8402/

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.

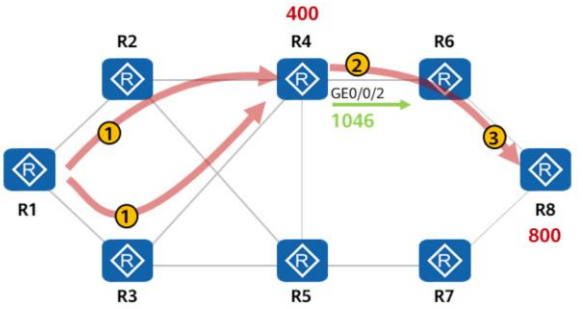
Contents

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.

Basic Concept: Segment ID

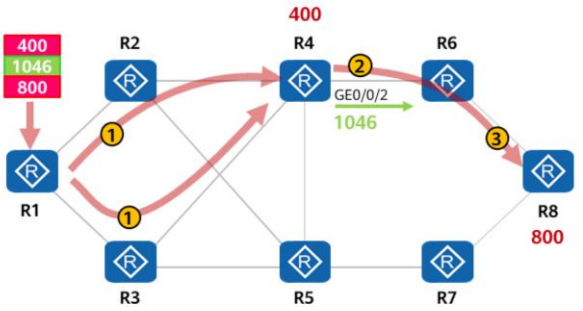


- Segment IDs (SIDs) identify segments. The SID format depends on the specific technical implementation. For example, SIDs can be MPLS labels, indexes in an MPLS label space, or IPv6 addresses.
- A segment list is an ordered list of one or more SIDs.
- For example:
  - Instruction 1 (400): Forward the 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

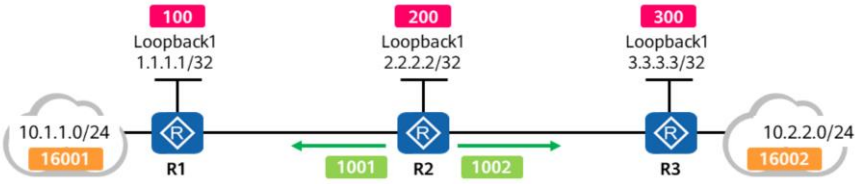
- The label values used in this course are only examples. For details about the label allocation scope, see the corresponding product documentation.

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.

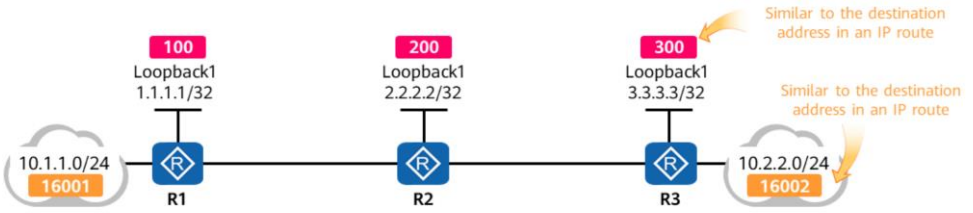
Basic Concept: Segment Classification



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.

Basic Concept: Prefix Segment



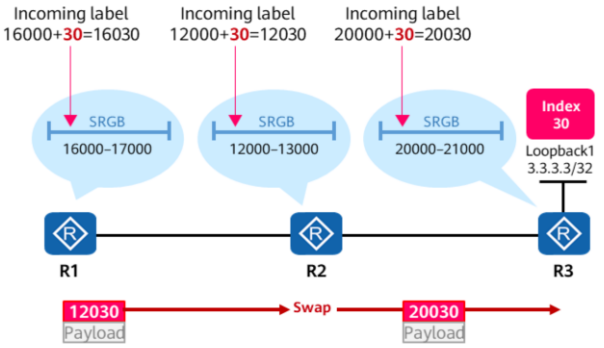
Prefix Segment

- 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.

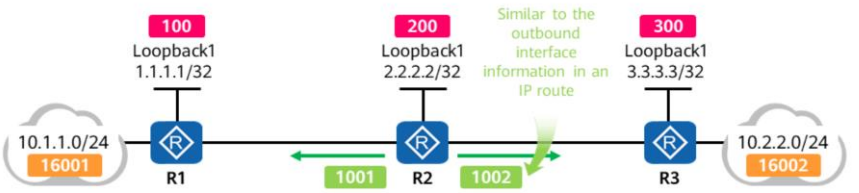
- Node segments are special prefix segments used to identify specific nodes.
- When an IP address is configured as a prefix for a node's loopback interface, the prefix SID of the node is the node SID.

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.

Basic Concept: Adjacency Segment

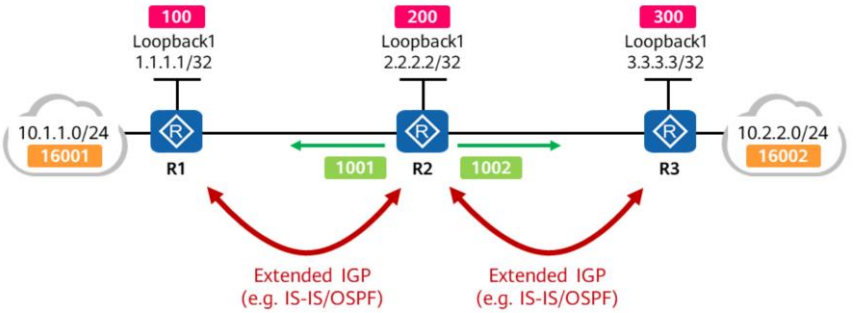


Adjacency Segment
Identifies an adjacency on a network. Adjacency segments are propagated to other devices through an IGP. They are visible to all the devices but effective only on the local device.

- Adjacency segments are identified using adjacency SIDs.
- Adjacency SIDs are local SIDs that are not in the SRGB range.

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.



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

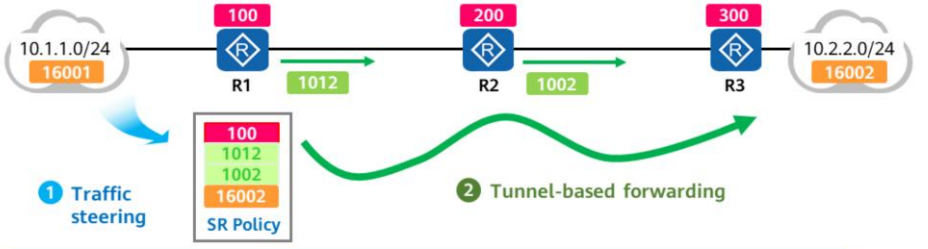
IS-IS for SR-MPLS

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

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.

SR Policy Example

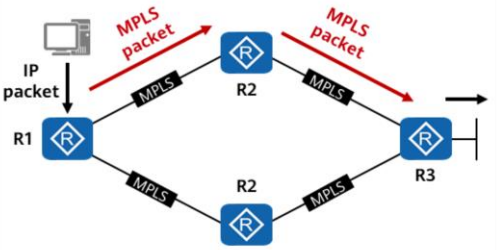


SR Policy:

- Can be generated using different modes, such as CLI, NETCONF, PCEP, and BGP SR Policy.
- Contains segment lists to guide traffic steering and forwarding.

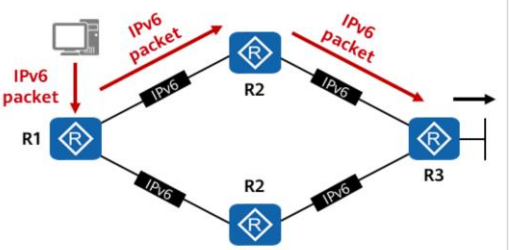
Basic Concept: SR-MPLS and SRv6

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.

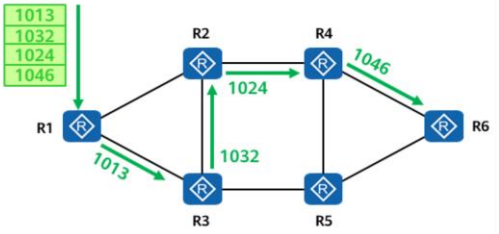
SRv6



- 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).

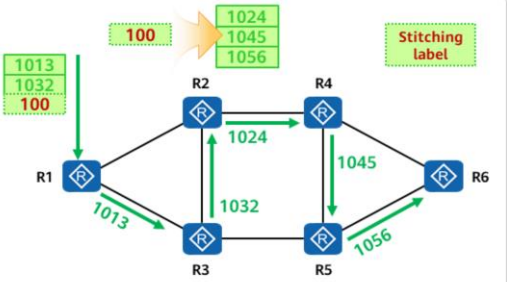
Label Stack, Stitching Label, and Stitching Node

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

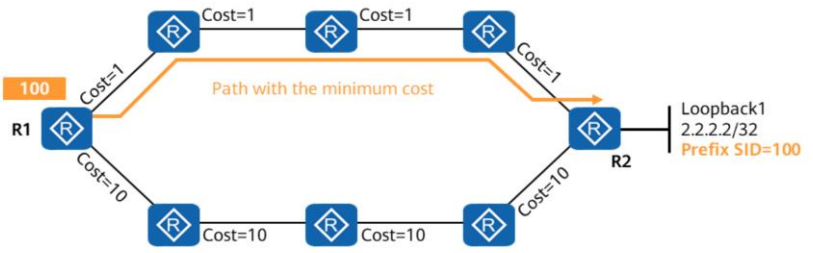


- If the label stack depth exceeds the maximum depth supported by forwarders, the controller needs to allocate multiple label stacks to the forwarders and a special label to an appropriate node to stitch these label stacks, thereby implementing segment-by-segment forwarding.
- This special label is called 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.

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.

Scenario 1: Prefix Segment-based Forwarding Path

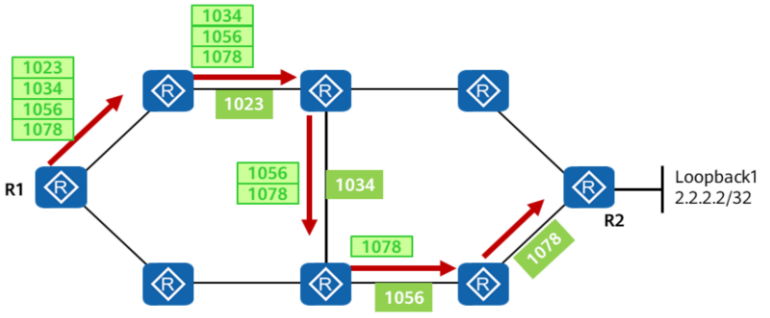


A prefix segment-based forwarding path is computed by an IGP using the SPF algorithm.

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.

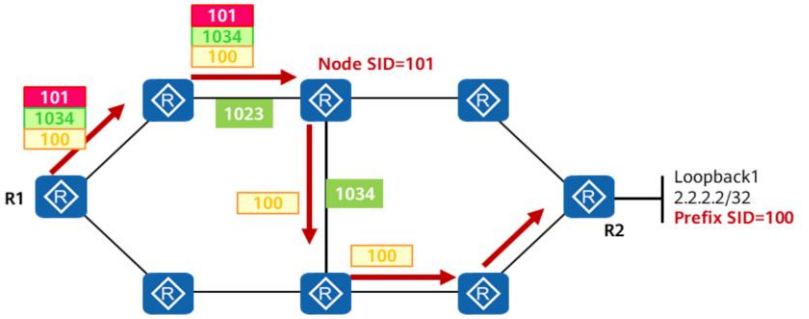
Scenario 2: Adjacency Segment-based Forwarding Path



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.

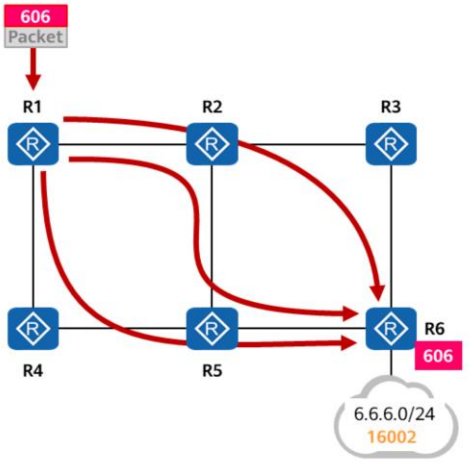
Scenario 3: Adjacency Segment+Node Segment-based Forwarding Path



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.

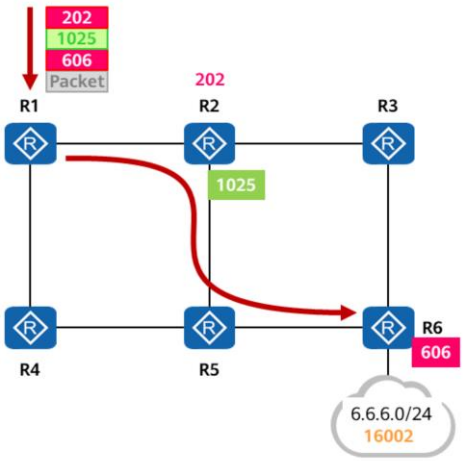
SR-MPLS BE



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.

SR-MPLS TE

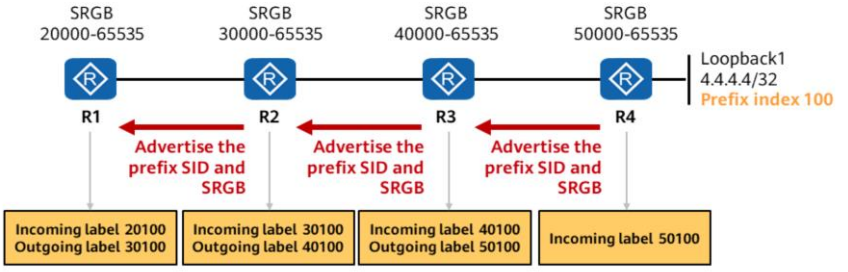


SR-MPLS TE

- 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.

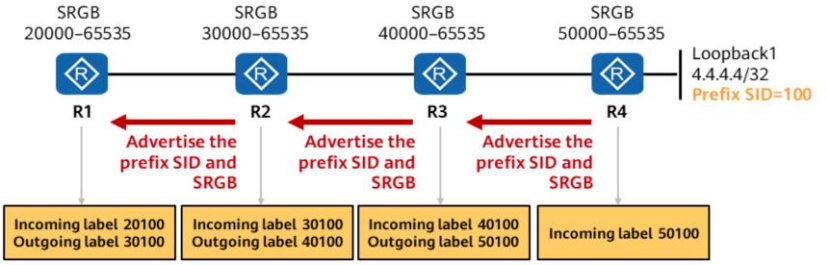
SR-MPLS BE LSP

- 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.



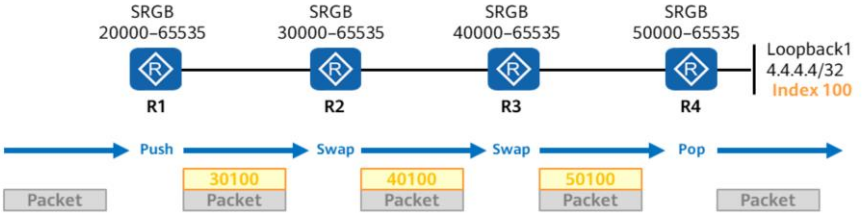
SR-MPLS BE LSP Creation

- 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.



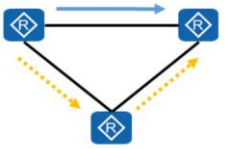
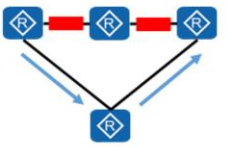
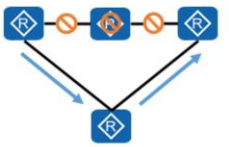
Data Forwarding Process

- 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.



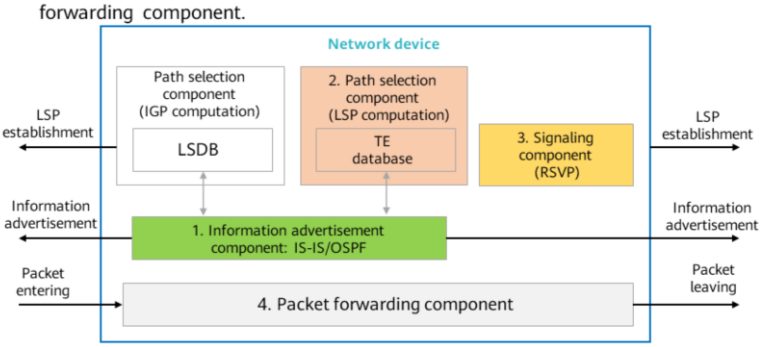
Traffic Engineering

- 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.

Traditional Distributed MPLS TE Architecture

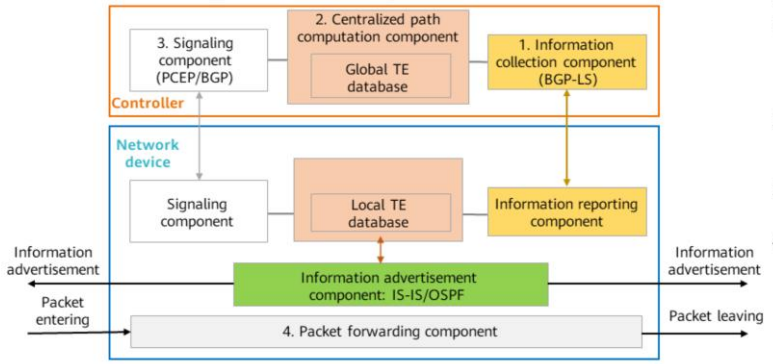
- MPLS TE uses the distributed architecture, in which the ingress computes paths according to constraints and uses RSVP-TE signaling to establish constraint-based LSPs.
- MPLS nodes are used to maintain a complete TE architecture through four components: information advertisement component, path computation 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.

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.

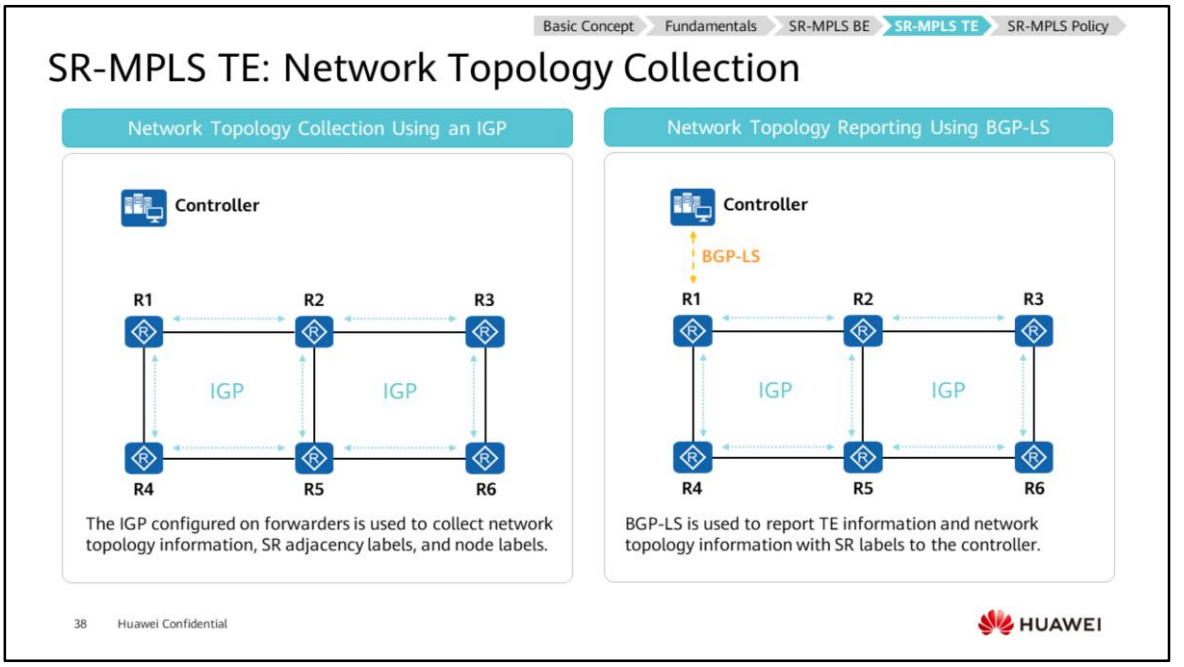


Centralized SR-MPLS TE:

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.

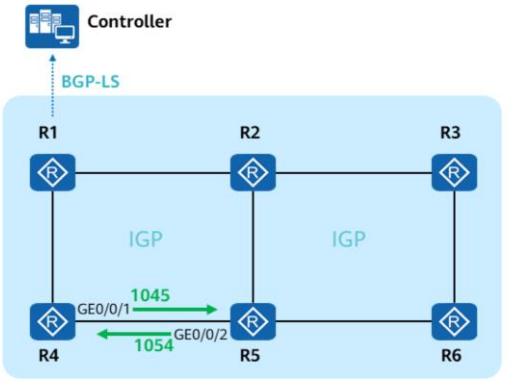
Comparison Between SR-MPLS TE and RSVP-TE

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.



- 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.

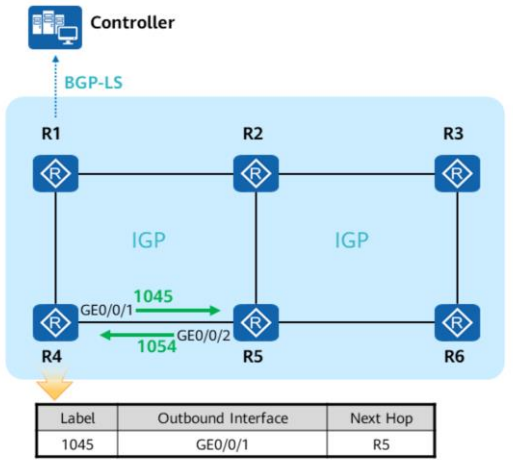
SR-MPLS TE: Label Allocation



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.

Label Allocation Example

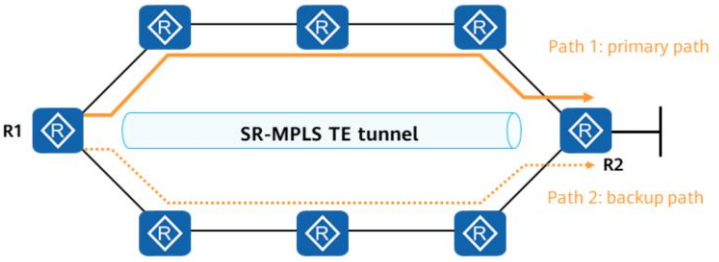


Label	Outbound Interface	Next Hop
1045	GE0/0/1	R5

- 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.

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

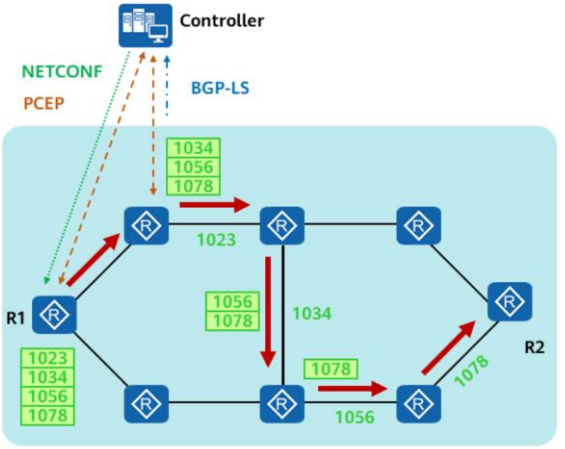
- SR-MPLS TE tunnel attributes must be configured before tunnel establishment. An SR-MPLS TE tunnel can be configured on a controller or forwarder.
- Tunnel configuration on a controller: After an SR-MPLS TE tunnel is configured on a controller, the controller uses NETCONF to deliver tunnel attributes 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.

- 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.

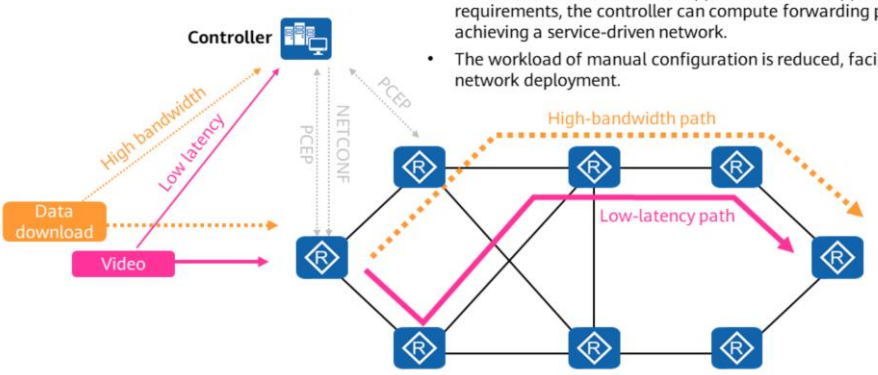
BGP-LS: used to report labels and network topology information by forwarders.

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.

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.

- 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.

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.

SR Policy Overview

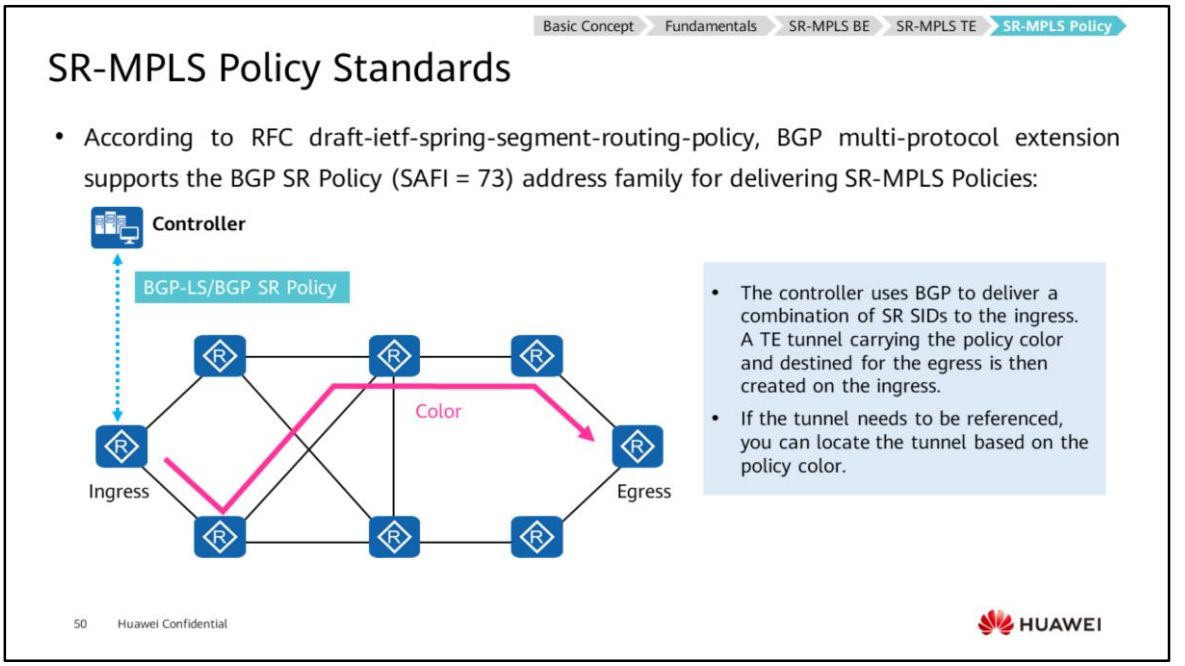
- An SR Policy uses a segment list to specify a forwarding path, without the need to use tunnel interfaces.
- SR Policies are classified into SR-MPLS Policies and SRv6 Policies based on segments. This document focuses on SR-MPLS Policies.
- The controller computes paths based on the color attribute that 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
```

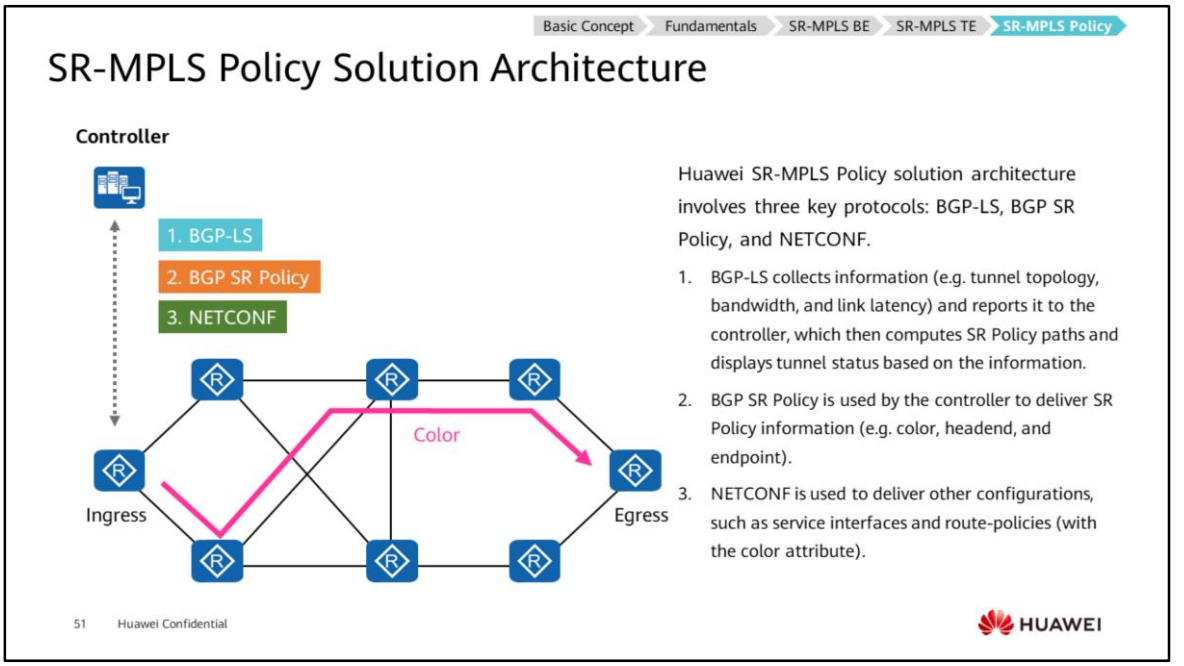

- https://datatracker.ietf.org/doc/draft-ietf-spring-segment-routing-policy/
- An SR Policy is a framework that enables instantiation of an ordered list of segments on a node for implementing a source routing policy with a specific intent for traffic steering from that node.

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.



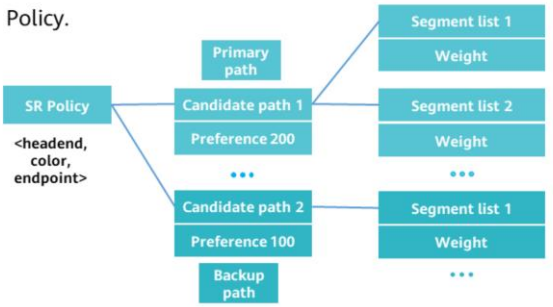
- There are three mainstream methods for SR Policy implementation.
  - BGP: BGP-LS is used to collect topology information, so that no new interface protocol needs to be introduced for customer-developed controllers. BGP SR Policy is used to deliver route information.
  - PCEP: PCEP is a mature southbound 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

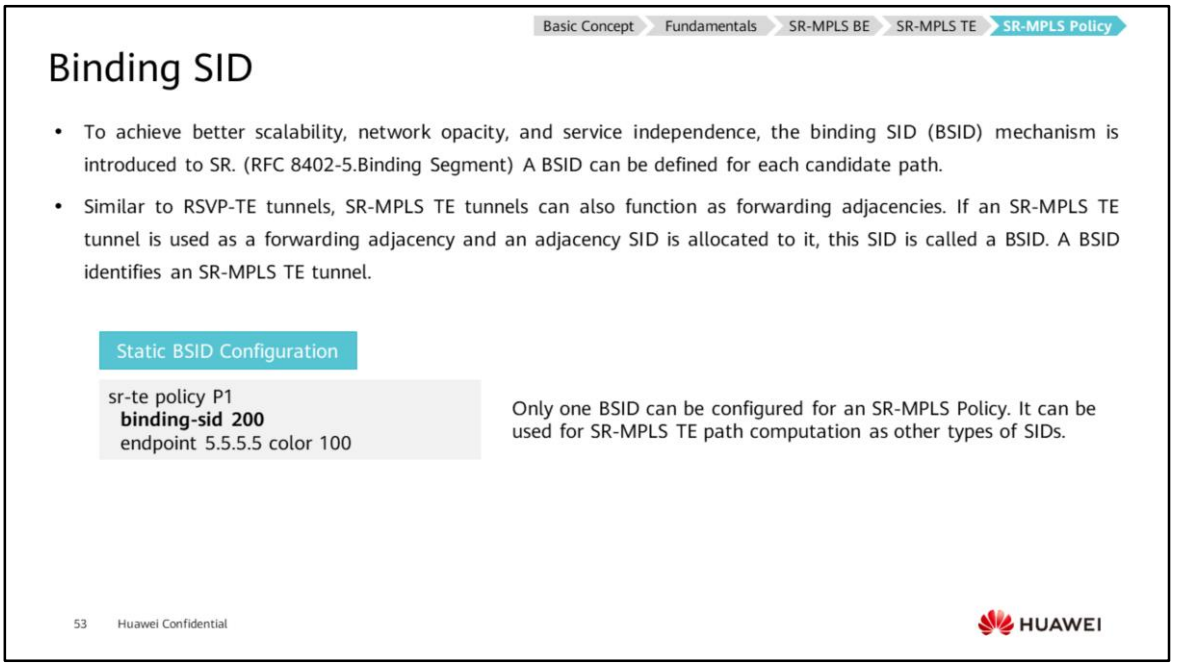
- An SR-MPLS Policy can contain multiple candidate paths with the preference attribute. The valid candidate path with the highest preference functions as the primary path of the SR-MPLS Policy, and the valid candidate path with the second highest preference functions as the backup path.
- A candidate path is an SR-MPLS Policy's segment list sent to the headend through PCEP or BGP SR Policy.



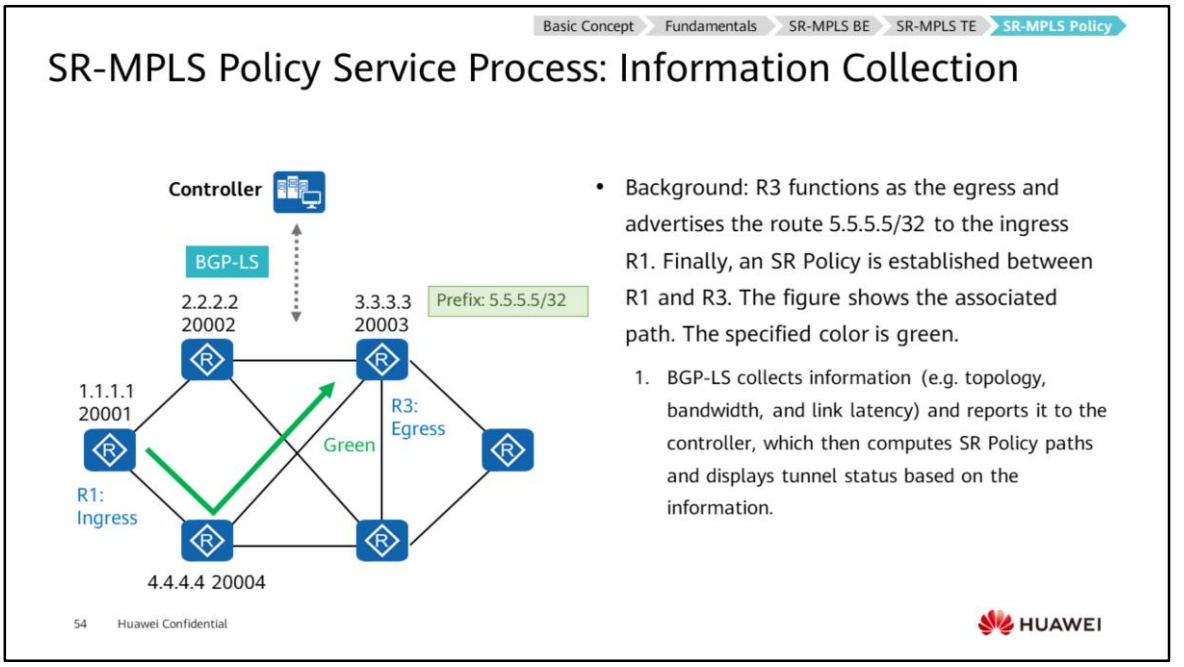
SR policy P1 <headend, color, endpoint>
Candidate-path CP1 <protocol, origin, discriminator>
Preference 200
Weight W1, SID-List1 <SID11...SID1i>
Weight W2, SID-List2 <SID21...SID2j>
Candidate-path CP2 <protocol, origin, discriminator>
Preference 100
Weight W3, SID-List3 <SID31...SID3i>
Weight W4, SID-List4 <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.

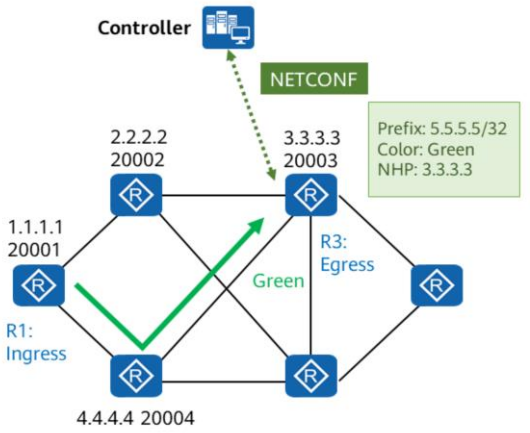


- Source of BSIDs: SRLB or SRGB
- Each candidate path of an SR Policy has a BSID. The BSIDs of different candidate paths of the same SR Policy are generally the same. The BSIDs of different SR Policies must be different. Generally, the BSID range needs to be planned and cannot be shared with other services.
- 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.



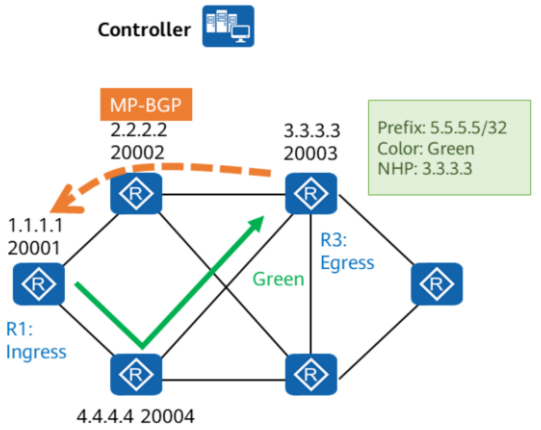
- Preparations:
    1. Controller planning: You can plan the color attribute and the mapping between the color attribute and SR tunnels' SLA requirements (path computation constraints) on the controller based on the SLA requirements of services.
    2. Enable SR on involved devices.
    3. Create a BGP 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



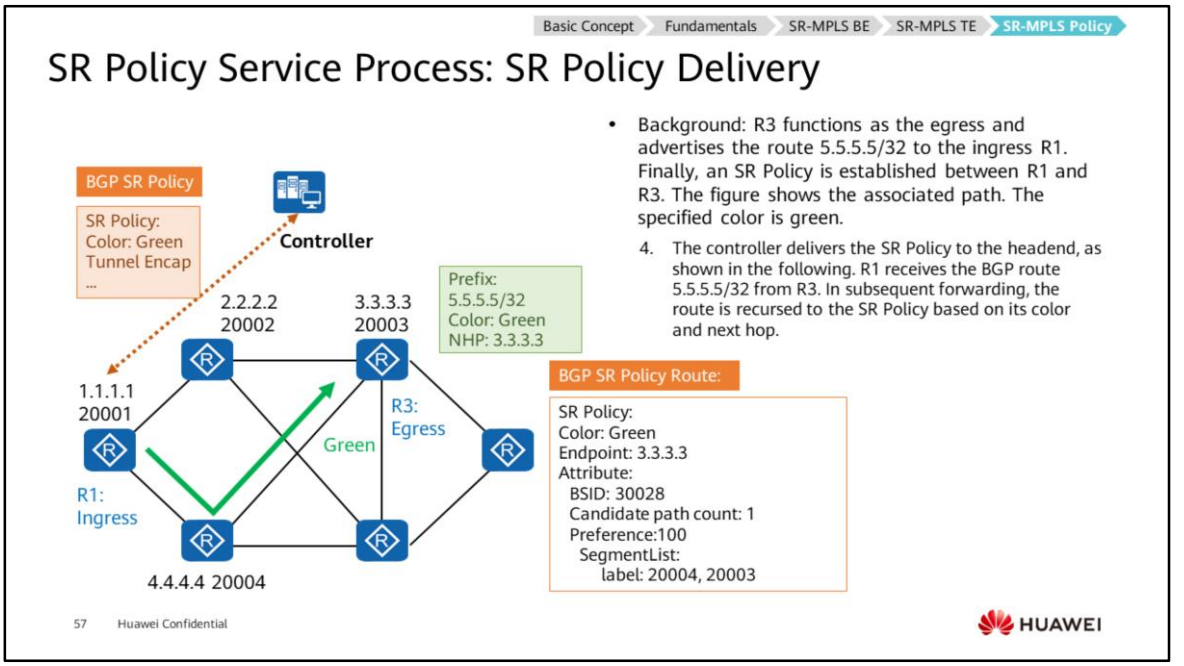
- Background: R3 functions as the egress and advertises the route 5.5.5.5/32 to the ingress R1. Finally, an SR Policy is established between R1 and R3. The figure shows the associated path. The specified color is green.
  2. The controller uses NETCONF to deliver a VPN or BGP export route-policy to the 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.

SR Policy Service Process: Route Advertisement

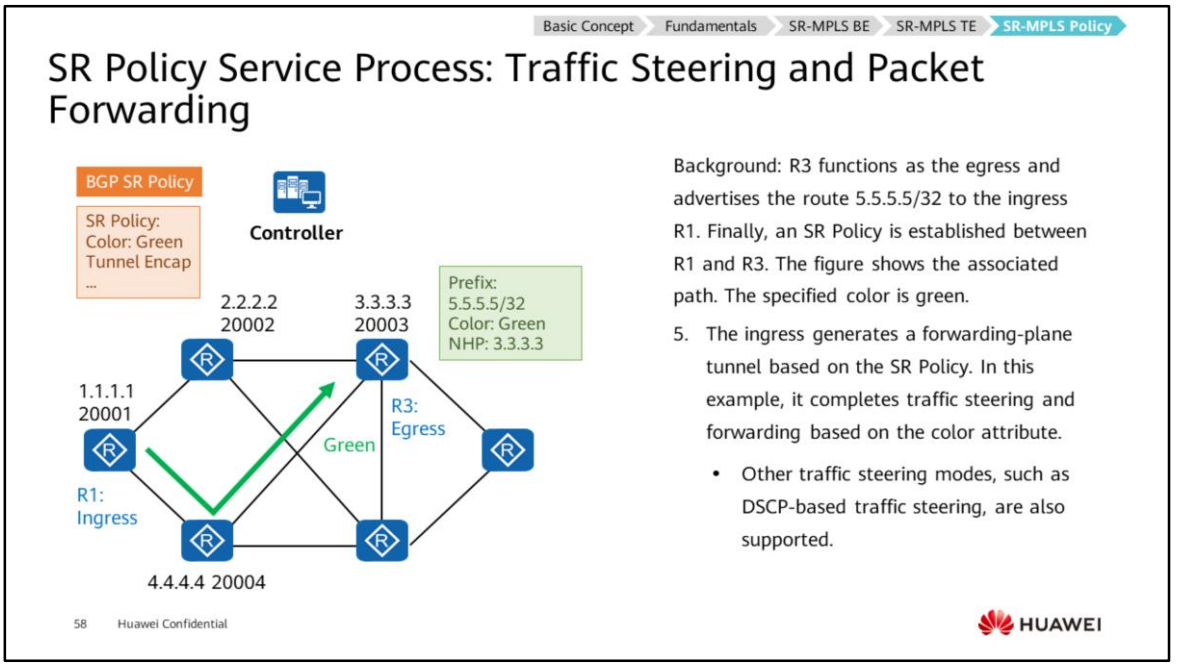


- Background: R3 functions as the egress and advertises the route 5.5.5.5/32 to the ingress R1. Finally, an SR Policy is established between R1 and R3. The figure shows the associated path. The specified color is green.

3. The egress advertises the colored route 5.5.5.5/32 to the ingress through MP-BGP.



- 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.

- Path Computation Element Communication Protocol (PCEP) Extensions for SR: https://datatracker.ietf.org/doc/rfc8664/?include_text=1

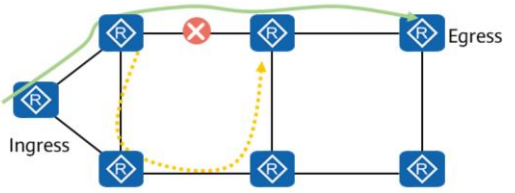
Contents

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

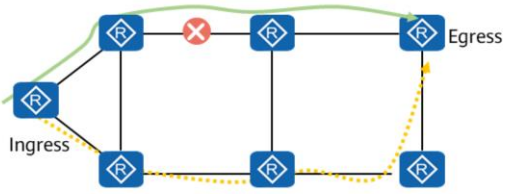


TI-LFA FRR

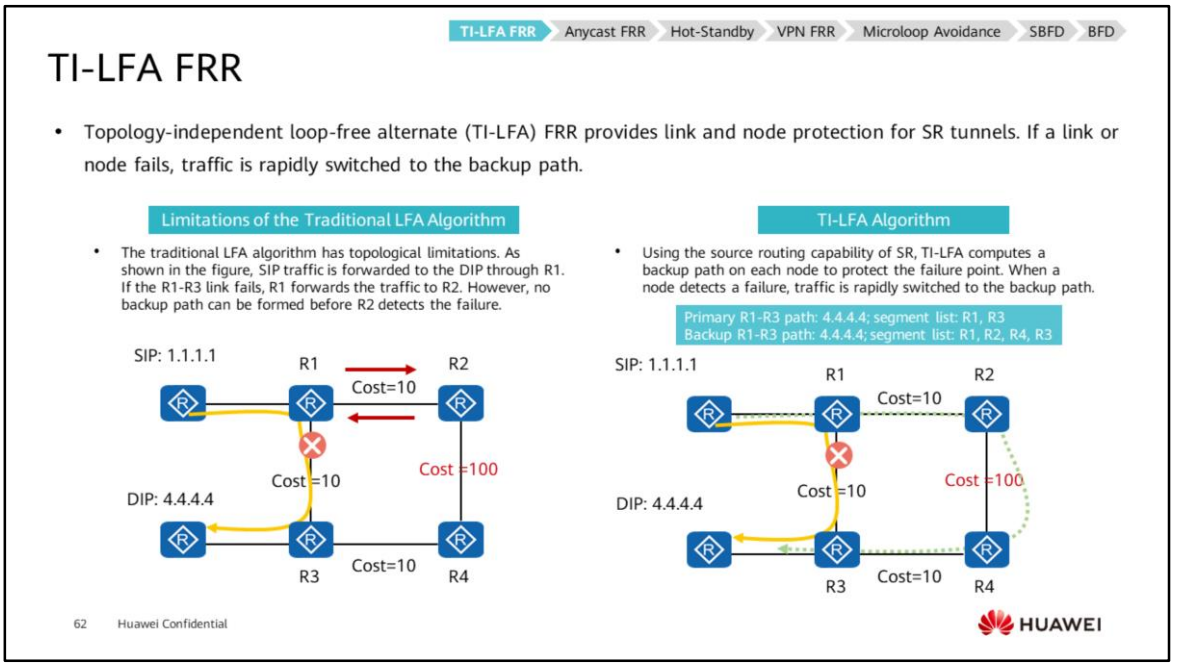
Anycast FRR

E2E protection

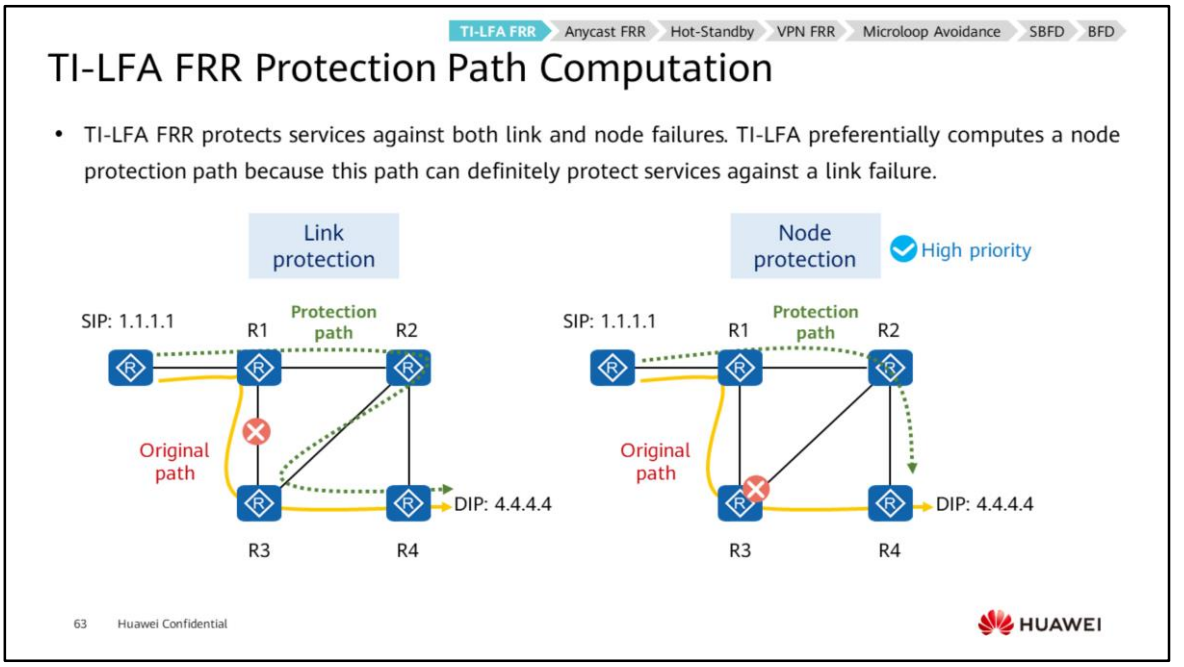
- Detection-dependent fast switching
- E2E paths protected



Hot Standby



- 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.

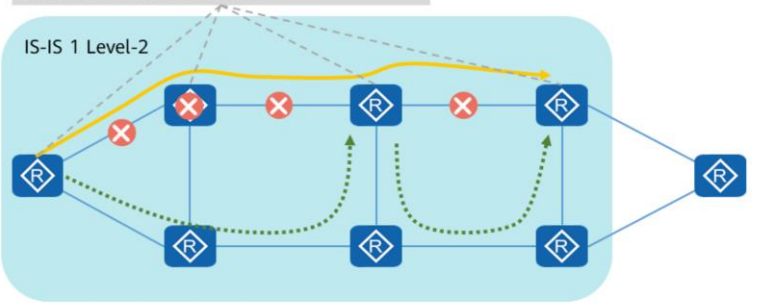


- For details about TI-LFA FRR, see the "TI-LFA FRR" section in NE series product documentation.

TI-LFA FRR Usage Scenarios and Configuration

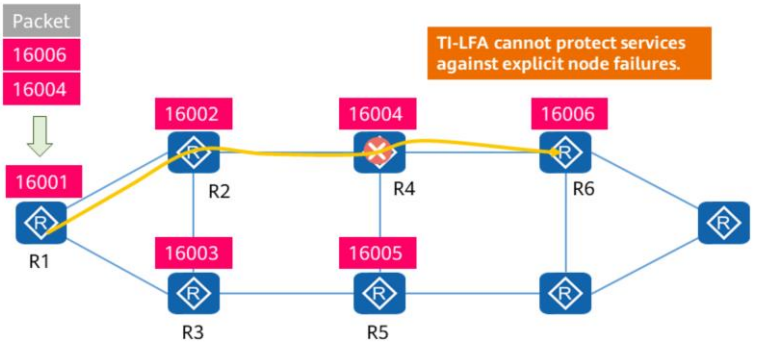
- To protect the entire path, you need to enable TI-LFA FRR local protection for the IGP processes of multiple nodes.

```
[Router] isis 1
[Router-isis-1] frr
[Router-isis-1-frr] loop-free-alternate level-2
[Router-isis-1-frr] ti-lfa level-2
```



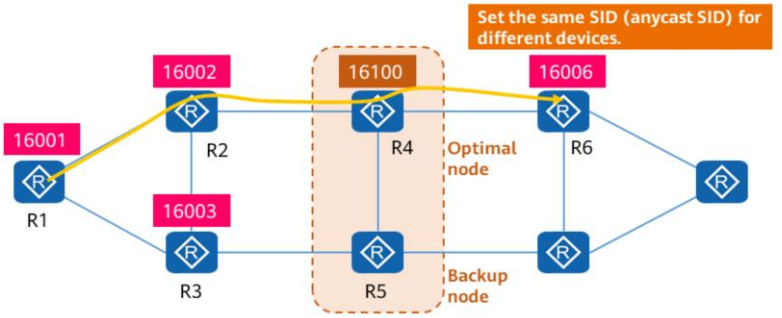
Limitations of TI-LFA FRR

- 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.



Anycast FRR

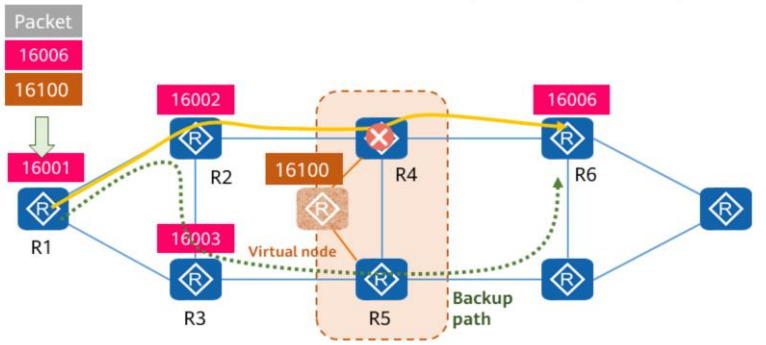
- Anycast FRR can protect services against failures of specified nodes.
- Assume that R4 and R5 advertise the same SID. This SID is called an anycast SID. The anycast SID is advertised in the IGP, with the next hop pointing to the nearest node on the path, such as R4. In this case, R4 is the optimal node of the anycast SID, and R5 is the backup node.



TI-LFA FRR | Anycast FRR | Hot-Standby | VPN FRR | Microloop Avoidance | SBFD | BFD

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.

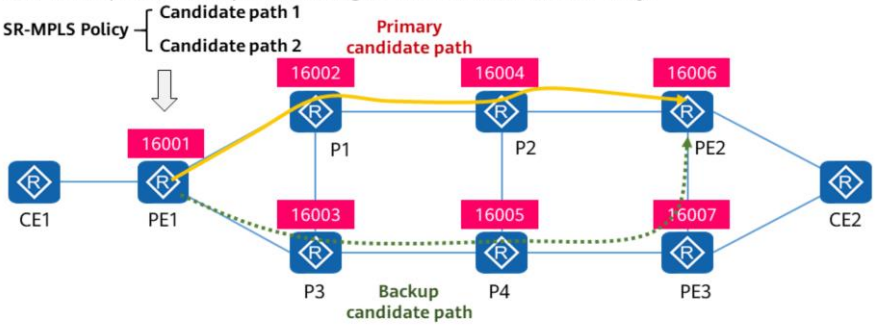


HUAWEI

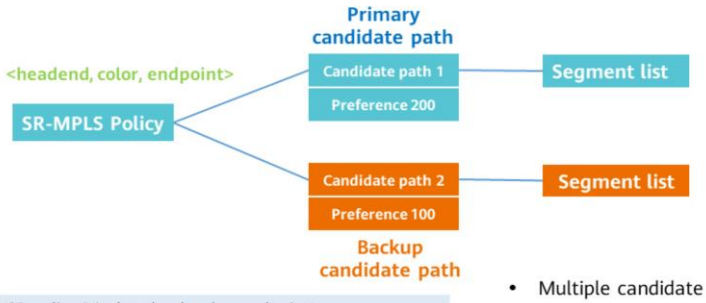
- The cost values of the links from R4 and R5 to the virtual node are both 0. However, the cost values of the links from the virtual node to R4 and R5 are infinite.

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.



Hot Standby Implementation for SR-MPLS Policy

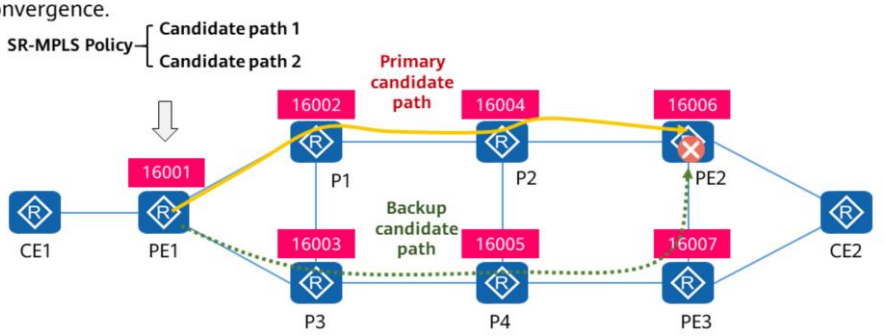


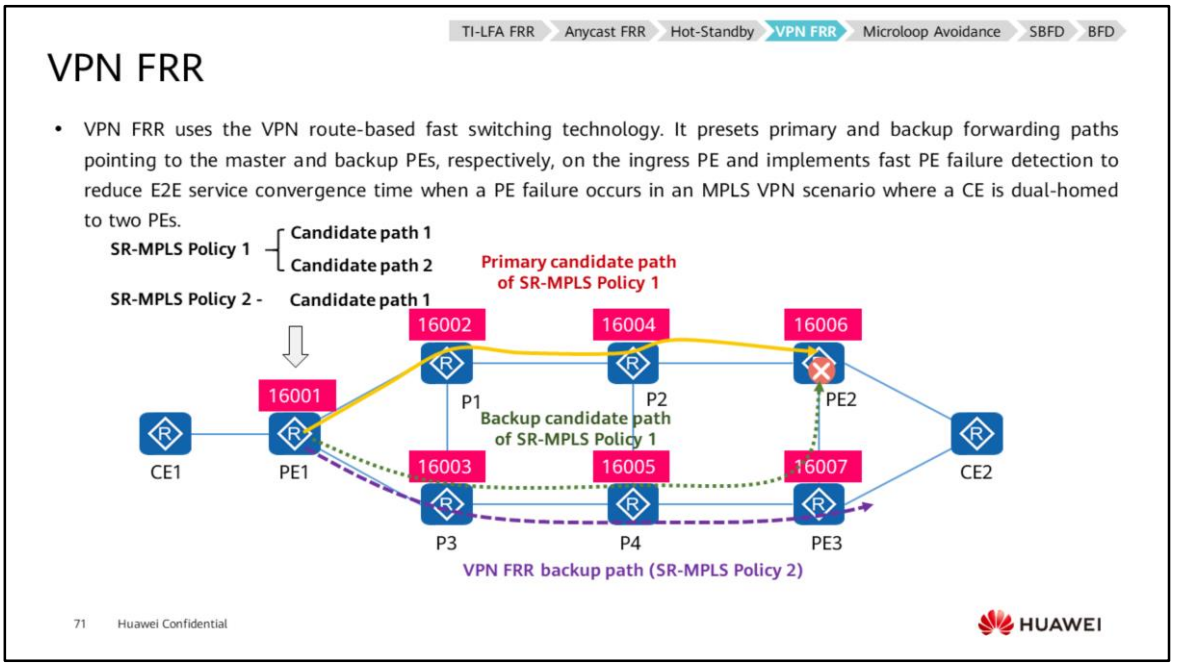

```
SR policy P1 <headend, color, endpoint>
  Candidate-path CP1 <protocol, origin, discriminator>
    Preference 200
      SID-List <SID11...SID1i>
  Candidate-path CP2 <protocol, origin, discriminator>
    Preference 100
      SID-List <SID21...SID2i>
```

- 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.

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.

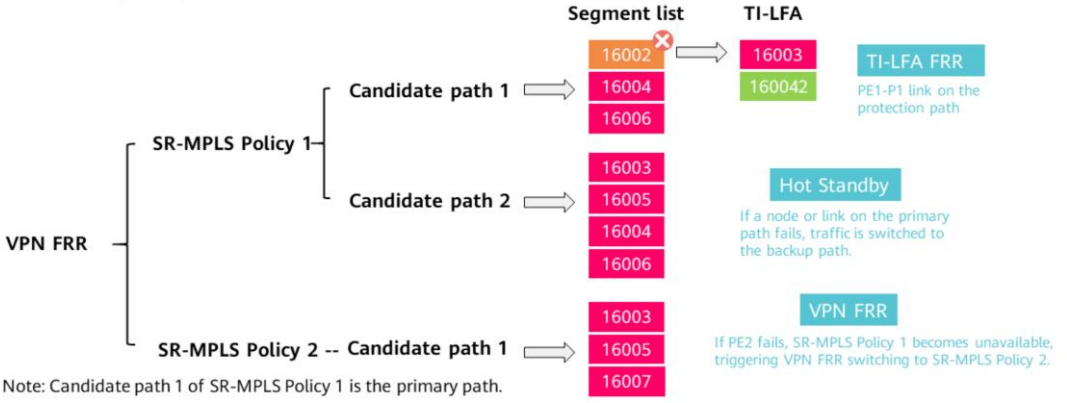




- In traditional TE tunnel protection technologies, if a PE fails, services can recover only through E2E route convergence and LSP convergence. The service convergence time is closely related to the number of internal MPLS VPN routes and the number of hops on the bearer network. The greater the number 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.

VPN FRR Failover Example

- In this example, when TI-LFA FRR, hot standby, and VPN FRR are used together, the protection switching is implemented as follows:



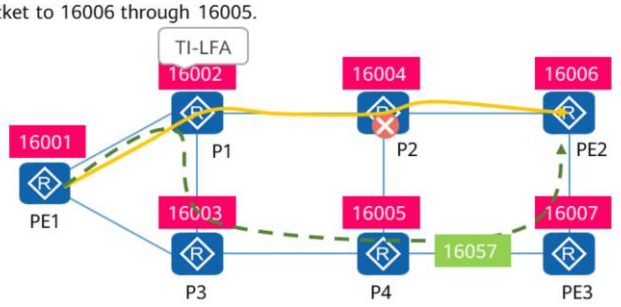
Note: Candidate path 1 of SR-MPLS Policy 1 is the primary path.

HUAWEI

- Fault detection in hot standby and VPN FRR scenarios depends on detection mechanisms such as BFD or SBFD.

SR Microloop Avoidance Overview

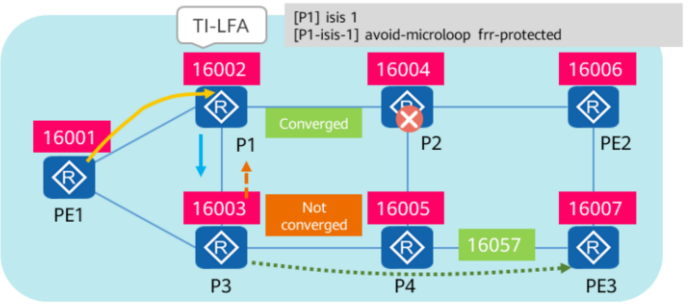
- Each node independently calculates the IGP LSDB, which may lead to a loop during unordered convergence. This may in turn result in microloops, a kind of transient loop that disappears after all the nodes on the forwarding path have converged.
- On the network shown in the following figure, TI-LFA FRR 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.

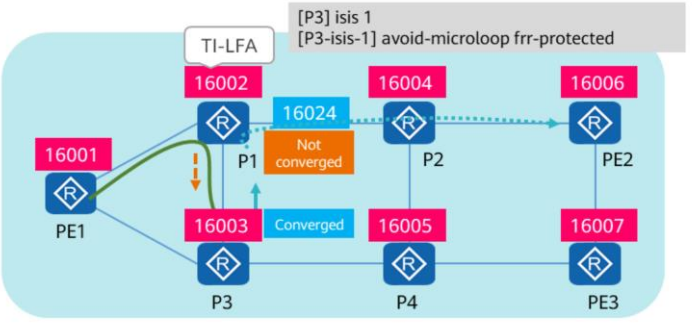


HUAWEI

TI-LFA FRR | Anycast FRR | Hot-Standby | VPN FRR | Microloop Avoidance | SBFD | BFD

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.



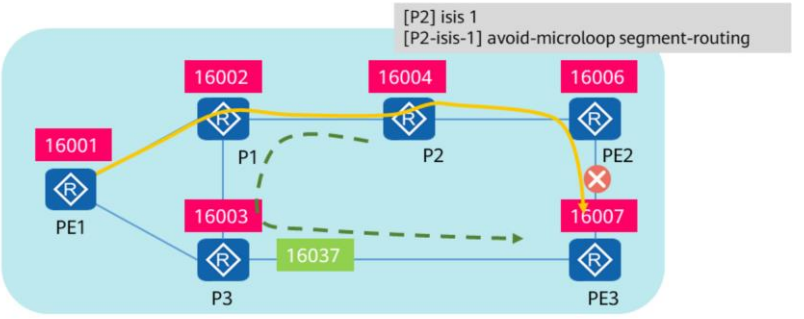
HUAWEI

- The microloop avoidance segment list generated by P3 takes effect within the timer T.

TI-LFA FRR | Anycast FRR | Hot-Standby | VPN FRR | Microloop Avoidance | SBFD | BFD

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

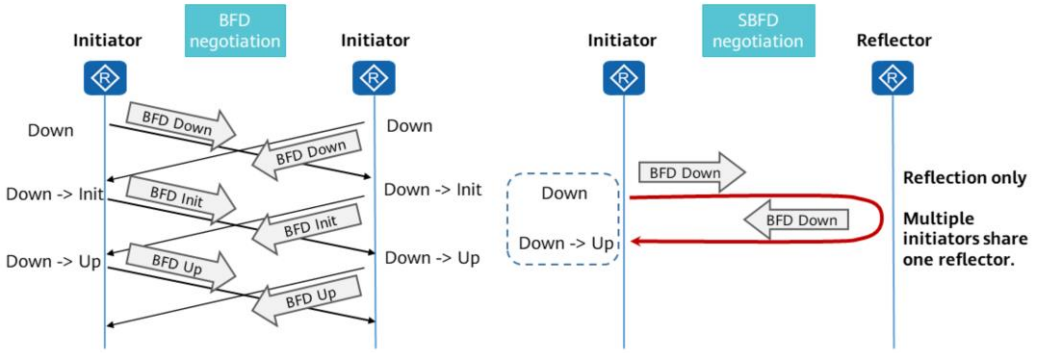
- The microloop avoidance segment list generated by P2 takes effect within the timer T.

Summary: Comparison Between TI-LFA and Microloop Avoidance

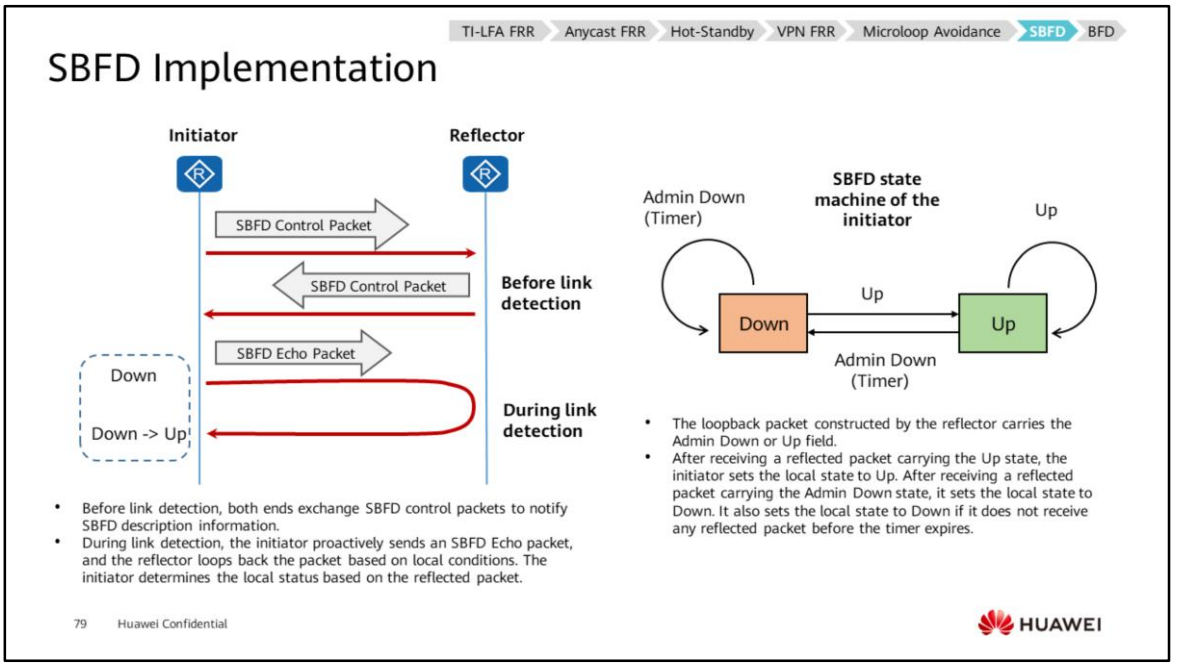
TI-LFA	Microloop Avoidance
• Purpose: to locally compute a backup path for the destination address • Trigger condition: link or node failure on the primary path	• Purpose: to prevent temporary loops during the update of the primary path • Trigger condition: primary path update

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



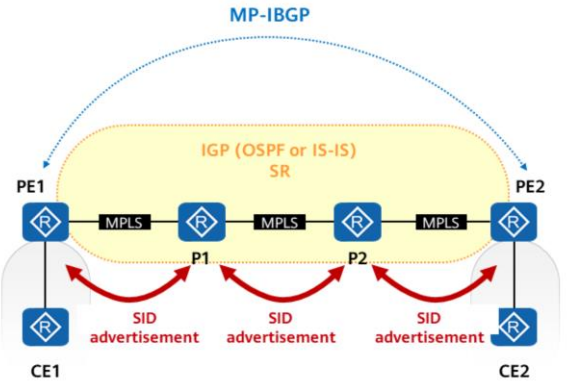
- Because the state machine has only Up and Down states, the initiator can send packets carrying only the Up or Down state and receive packets carrying only the Up or Admin Down state. The initiator starts by sending an SBFD packet carrying the Down state to the reflector. The destination and source 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.

Contents

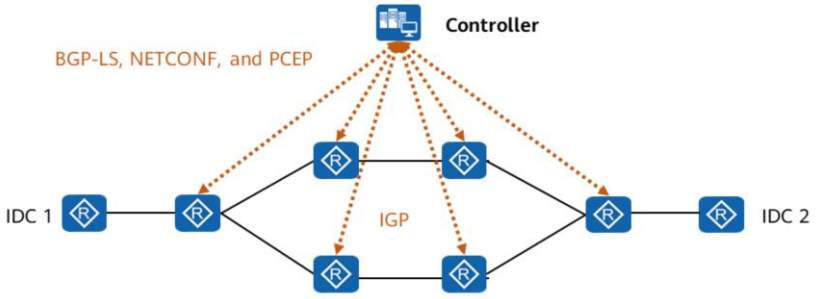
Intra-AS SR-MPLS BE



- 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.

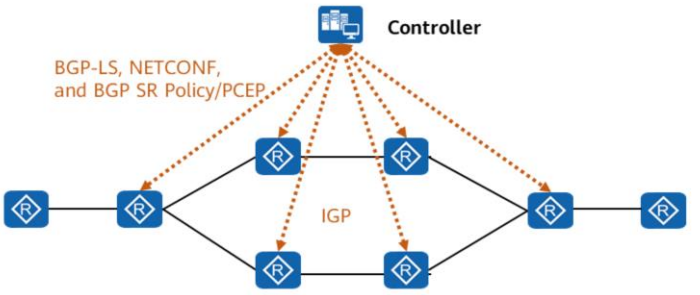
Intra-AS SR-MPLS TE

- 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.

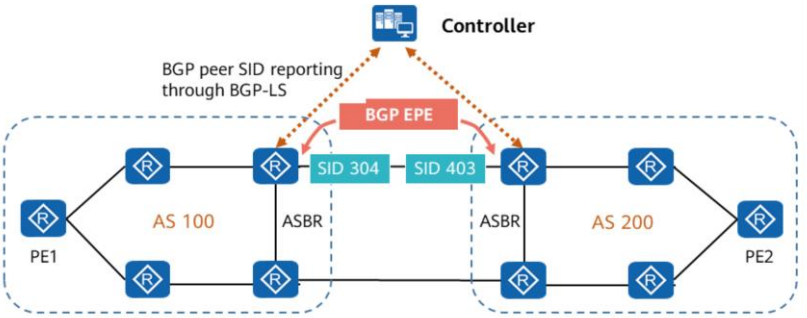


HUAWEI

- 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.

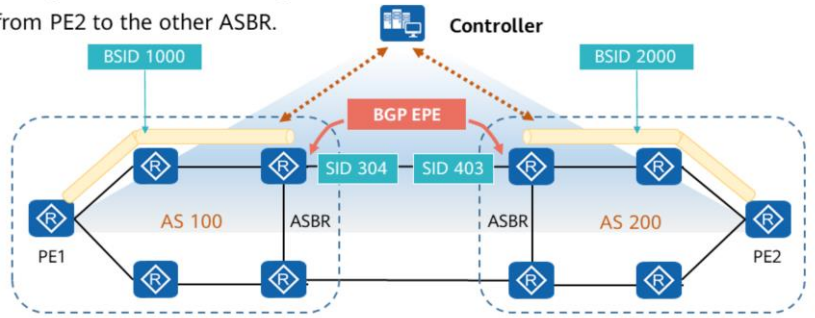
Inter-AS E2E SR-MPLS TE (1)

- 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.



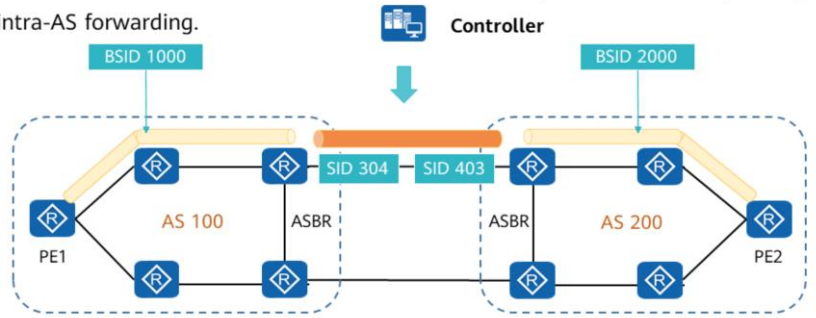
Inter-AS E2E SR-MPLS TE (2)

- 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.



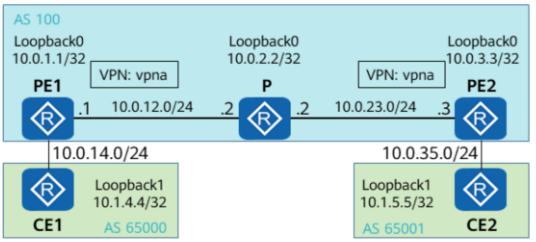
Inter-AS E2E SR-MPLS TE (3)

- 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.



Contents

L3VPN over SR-MPLS BE (1)



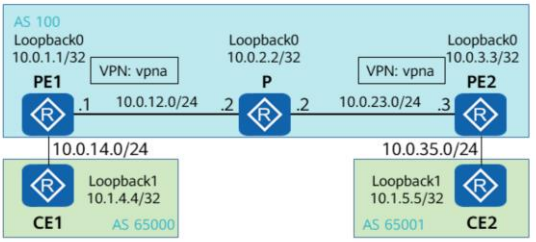
Networking requirements:

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.

L3VPN over SR-MPLS BE (2)



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: (P and PE2 configurations are not provided.)

```
[~PE1] ospf 1
[*PE1-ospf-1] opaque-capability enable
[*PE1-ospf-1] quit
[~PE1] mpls lsr-id 10.0.1.1
[*PE1] mpls
[~PE1-mpls] quit
[~PE1] segment-routing
[*PE1-segment-routing] quit
[*PE1] ospf 1
[*PE1-ospf-1] segment-routing mpls
[*PE1-ospf-1] segment-routing global-block 16000 23999
[*PE1-ospf-1] quit
[*PE1] interface loopback 0
[*PE1-LoopBack1] ospf prefix-sid index 1
[*PE1-LoopBack1] quit
[*PE1] commit
```

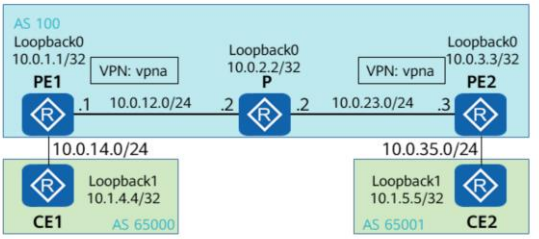
Configure the same SRGB.

P: index 2
PE2: index 3

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.

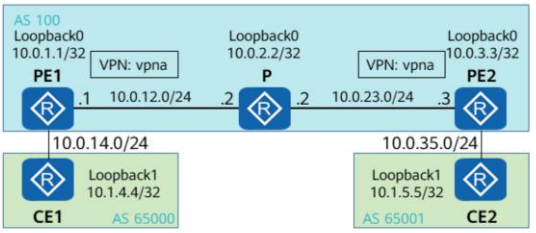
PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~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
```

L3VPN over SR-MPLS BE (4)



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] tunnel-policy p1
[*PE1-tunnel-policy-p1] tunnel select-seq sr-lsp load-balance-number 2
[*PE1-tunnel-policy-p1] quit
[*PE1] commit
[~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] quit
[*PE1-vpn-instance-vpna] quit
[*PE1] commit
```

Run the **display tunnel-info all** command on PE1 to check SR LSP information.

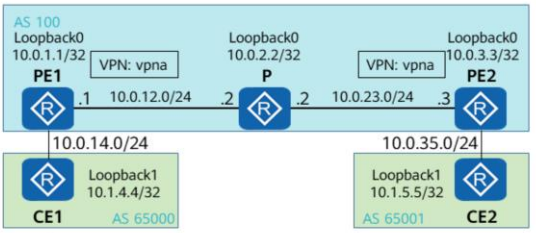
```
<PE1>display tunnel-info all
Tunnel ID                Type       Destination     Status
-------------------------------------------------------------
0x000000002900000042     srbe-lsp   10.0.3.3        UP
0x000000002900000043     srbe-lsp   10.0.2.2        UP
```

ID of the tunnel to PE2

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:

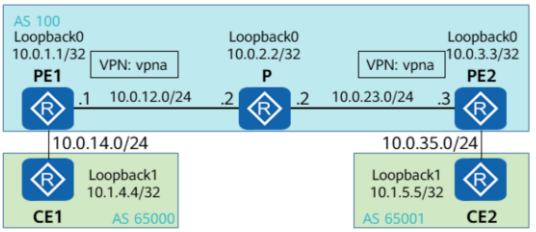
1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. **Configure 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

L3VPN over SR-MPLS BE (6)



Configuration roadmap:

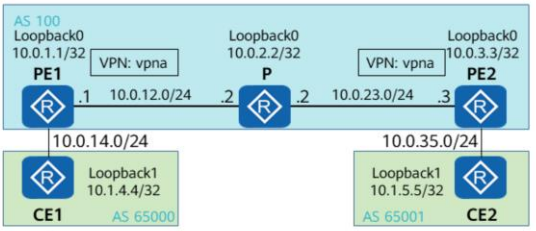
1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. **Configure 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.

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
```

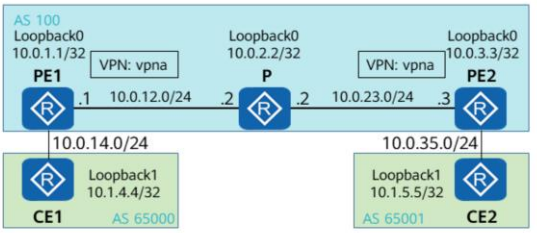
Question: How are the labels computed?

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
```

Contents

L3VPN over SR-MPLS TE (1)



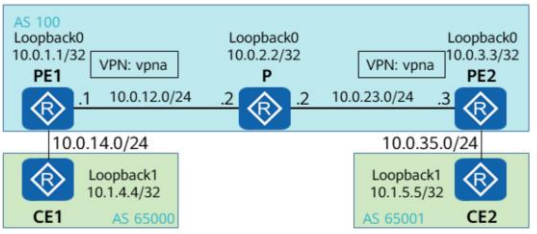
Networking requirements:

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.

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.

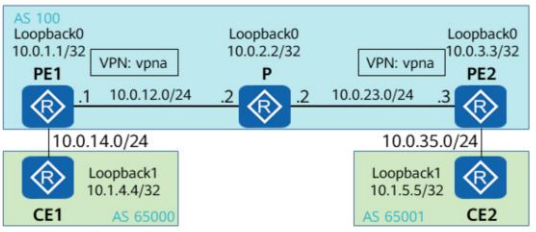
Configure basic SR-MPLS TE functions. PE1 configurations are as follows: (P and PE2 configurations are not provided.)

```
[~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
```

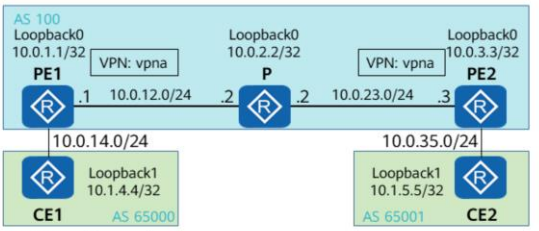
Configure an SR-MPLS TE tunnel interface. PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[*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.

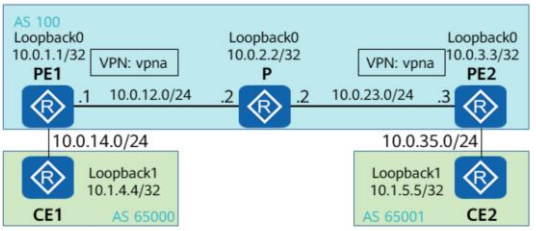
PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~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
```

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.)

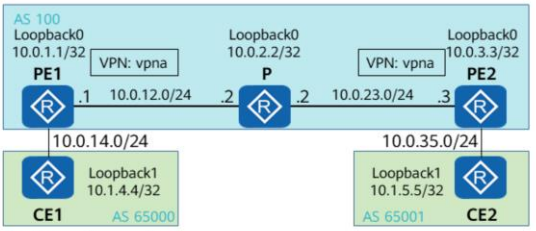
```
[~PE1] tunnel-policy p2
[*PE1-tunnel-policy-p2] tunnel select-seq sr-te load-balance-number 1
[*PE1-tunnel-policy-p2] quit
[*PE1] commit
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] tnl-policy p2
[*PE1-vpn-instance-vpna-af-ipv4] quit
[*PE1-vpn-instance-vpna] quit
[*PE1] commit
```

Run the **display tunnel-info all** command on PE1 to check SR LSP information.

Tunnel ID	Type	Destination	Status
0x0000000003000000001	sr-te	10.0.3.3	UP
0x0000000002900000042	srbe-lsp	10.0.3.3	UP
0x0000000002900000043	srbe-lsp	10.0.2.2	UP

ID of the SR-TE tunnel to PE2

L3VPN over SR-MPLS TE (6)



Configuration roadmap:

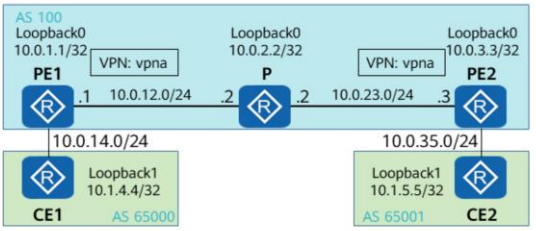
1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. Enable MPLS, configure SR, and establish SR-MPLS TE LSPs on the backbone network.
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish 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: 0x000000003000000001  Flags: RD
```

The VPNv4 label and SR TE LSP are combined to guide packet forwarding.

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
```

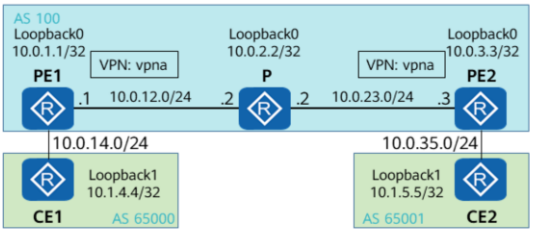
Question: How are the labels computed?

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
```

Contents

L3VPN over Static SR-MPLS Policy (1)



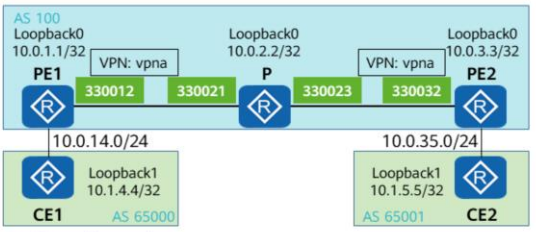
Networking requirements:

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.

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.)

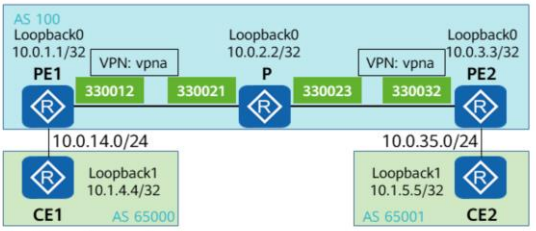
```
[~PE1] mpls lsr-id 10.0.1.1
[*PE1] mpls
[*PE1-mpls] mpls te
[*PE1-mpls] quit
[~PE1] segment-routing
[*PE1-segment-routing] ipv4 adjacency local-ip-addr 10.0.12.1 remote-
ip-addr 10.0.12.2 sid 330012
[*PE1-segment-routing] quit
[~PE1] ospf 1
[*PE1-ospf-1] opaque-capability enable
[*PE1-ospf-1] segment-routing 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:

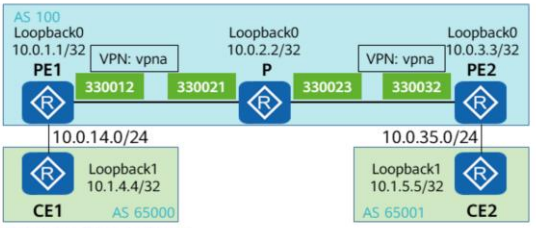
1. Configure interface IP addresses and OSPF. (Configuration details are not provided.)
2. **Enable MPLS and configure an SR-MPLS Policy on the backbone network.**
3. Establish an MP-BGP peer relationship between PE1 and PE2.
4. Enable the VPN instance IPv4 address family on each PE.
5. Configure the 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 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.

L3VPN over Static SR-MPLS Policy (4)



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.)

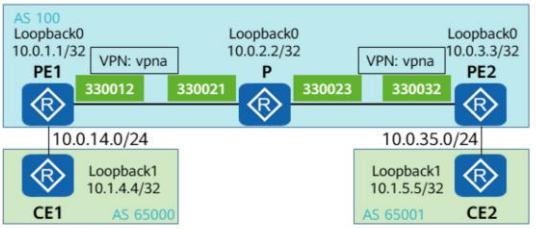
```
[~PE1] segment-routing
[~PE1-segment-routing] segment-list pe1
[*PE1-segment-routing-segment-list-pe1] index 10 sid label 330012
[*PE1-segment-routing-segment-list-pe1] index 20 sid label 330023
[*PE1-segment-routing-segment-list-pe1] quit
[*PE1-segment-routing] sr-te policy policy100 endpoint 10.0.3.3 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.

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.)

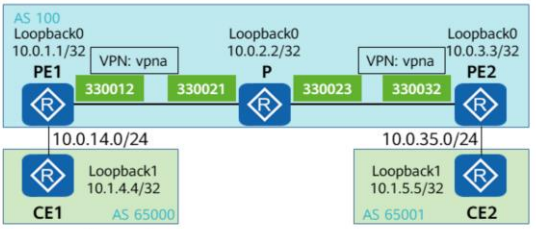
```
[~PE1] ip vpn-instance vpna
[*PE1-vpn-instance-vpna] ipv4-family
[*PE1-vpn-instance-vpna-af-ipv4] route-distinguisher 100:1
[*PE1-vpn-instance-vpna-af-ipv4] vpn-target 111:1 both
[*PE1] interface loopback1
[*PE1-LoopBack1] ip binding vpn-instance vpna
[*PE1-LoopBack1] ip address 10.1.4.4 24
[*PE1-LoopBack1] quit
[~PE1] route-policy color100 permit node 1
[*PE1-route-policy] apply extcommunity color 0:100
[~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] peer 10.0.3.3 route-policy color100 import
[~PE1-bgp-af-vpnv4] quit
[*PE1-bgp]ipv4-family vpn-instance vpna
[*PE1-bgp-vpna]import-route direct
[*PE1-bgp-vpna]commit
```

Add the color attribute to the received route.

HUAWEI

- 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 (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.**

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
```

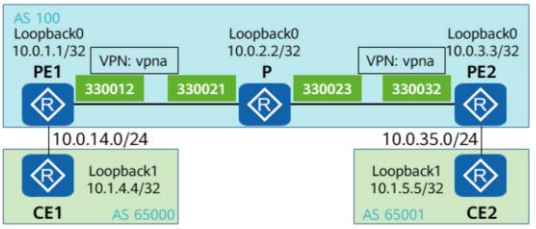
Run the **display tunnel-info all** command on PE1 to check SR LSP information.

```
<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 (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.**

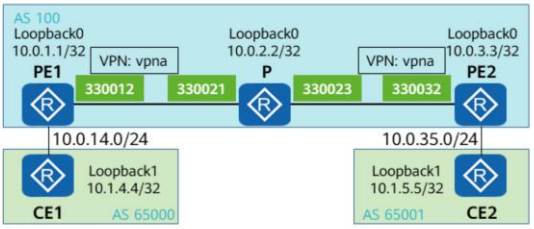
Check vpna's routing information on PE1.

```
[~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: 0x000000003200000001  Flags: RD
```

The VPNv4 label and SR-TE Policy LSP are combined to guide packet forwarding.

HUAWEI

L3VPN over Static SR-MPLS Policy (8)



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
```

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
2. (Multiple-answer question) Which of the following ports are used by SBFD packets by default? ( )
   A. 4784
   B. 3784
   C. 6784
   D. 7784

- D
- AD

Summary

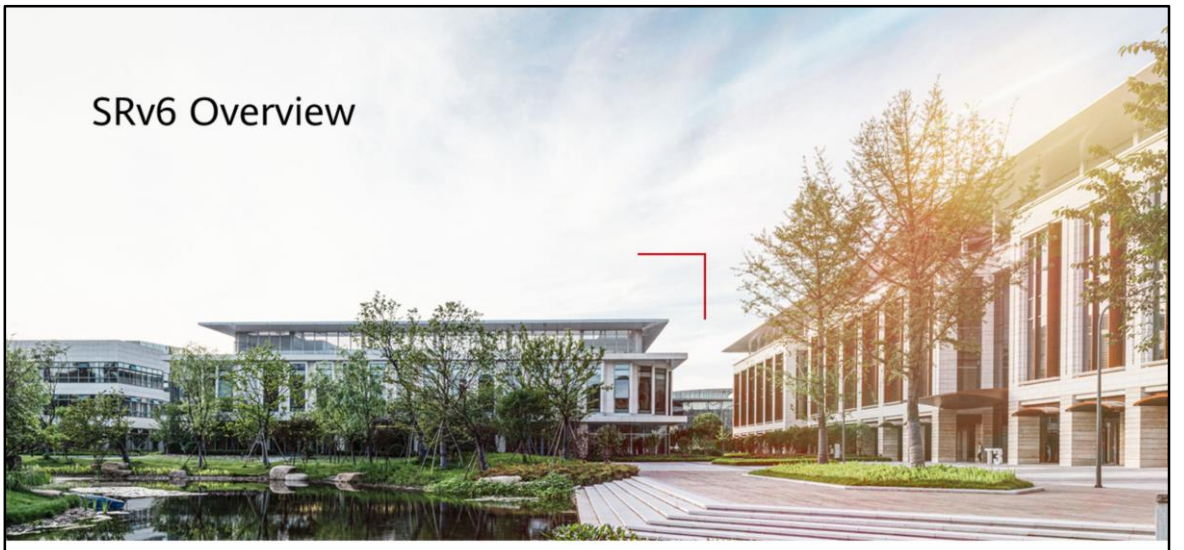
- SR is designed to forward data packets on a network using the source routing model. Compared with LDP and RSVP-TE, SR-MPLS simplifies the control plane of an MPLS network, enabling information such as labels to be carried only through IGP extensions. It provides higher scalability, freeing transit 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.

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.

SRv6 Overview



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.

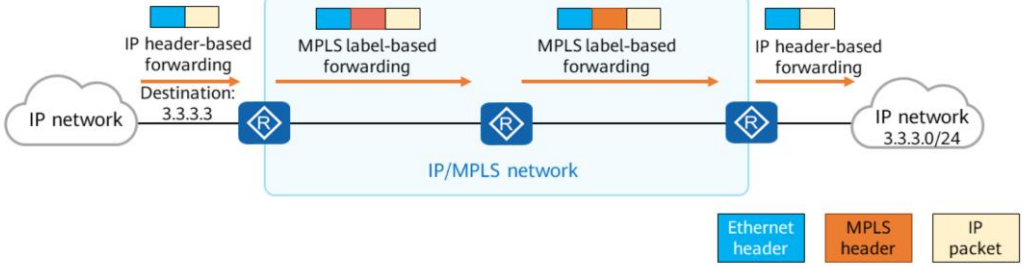
Objectives

- Upon completion of this course, you will be able to:
  - Describe the background of SRv6.
  - Describe the technical advantages of SRv6.
  - Describe the concepts and fundamentals of SRv6.
  - Understand how to configure SRv6 BE and static SRv6 Policies.

Contents

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.

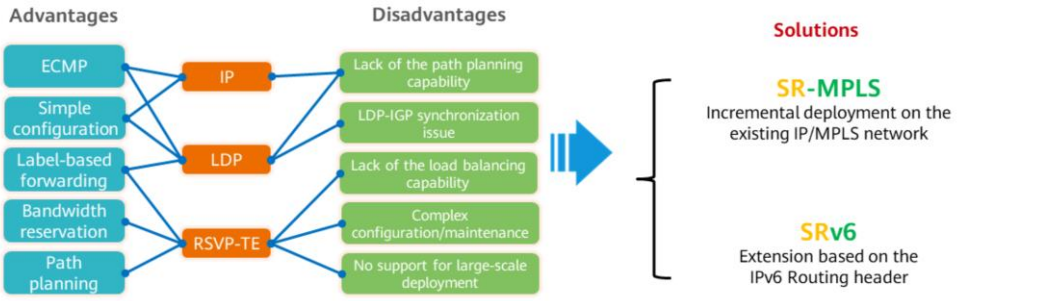


HUAWEI

- 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. Its design philosophy 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.

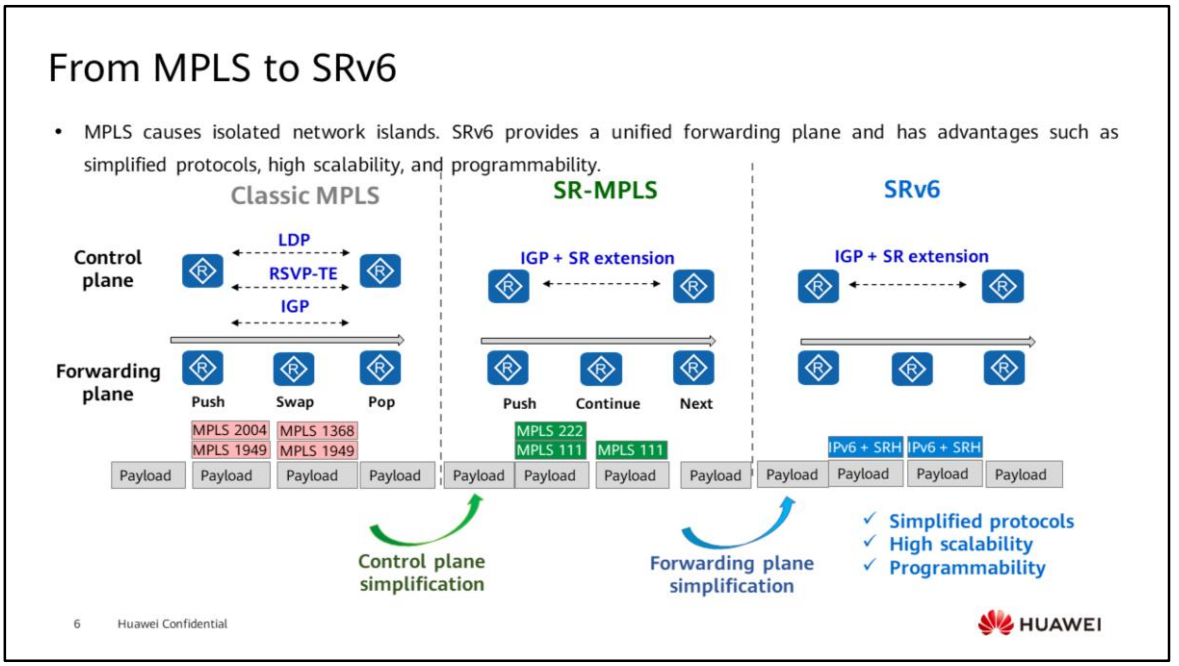
SR Origin and Solution

- 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.

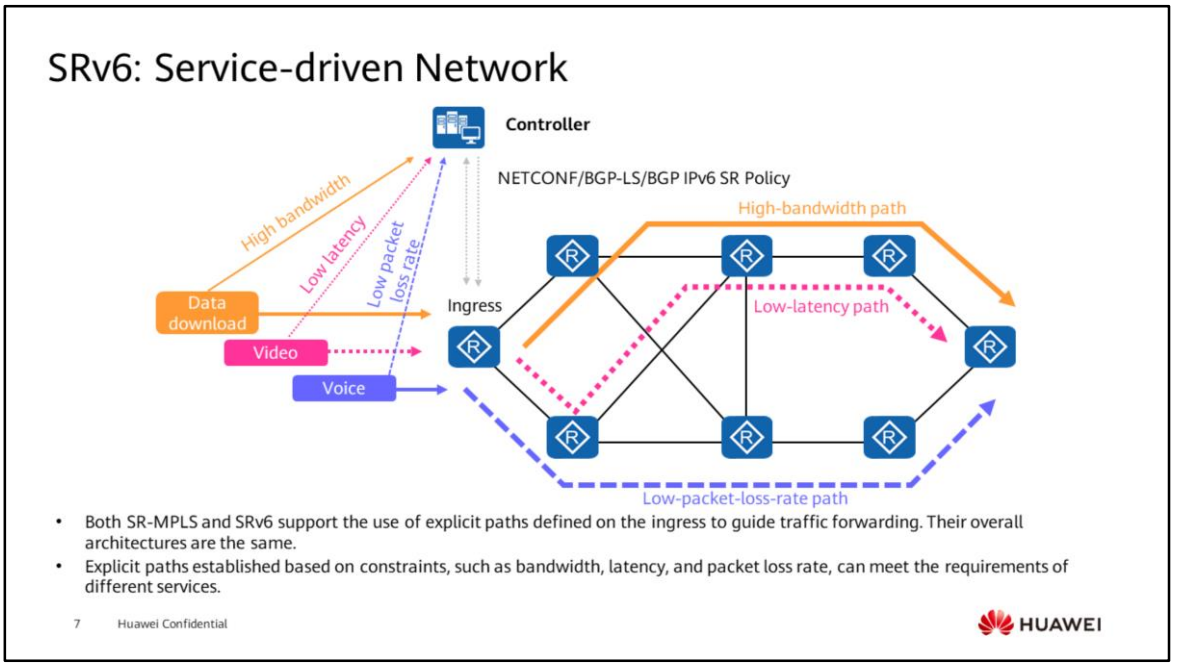


HUAWEI

- SR resolves many issues on IP/MPLS networks through two solutions: SR-MPLS (based on MPLS forwarding) and SRv6 (based on IPv6 forwarding).



- 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.



- The biggest advantage of SRv6 over SR-MPLS is that IPv6 offers stronger network programming capabilities. Although IPv4 packets also contain a programmable Option field, this field can be used only in scenarios such as fault locating. IPv6, however, takes the extensibility of packet headers into account from the very beginning. Multiple extension headers, including Hop-by-Hop Options, Destination Options, and Routing headers, were designed to support further extension.
- As of now, SRv6 network programming has grown rapidly, and SRv6 has been commercially deployed on multiple networks since the draft draft-ietf-spring-srv6-network-programming was proposed in 2019.

Value and Significance of SR

- MPLS is essentially an extension of IP functions and uses the shim mode to implement various flexible services. SRv6 integrates IP and MPLS functions, complying with the trend of technology development.
- During the development of SDN, there was too much emphasis on the building of SDN controller capabilities, but the impact of network infrastructure on SDN controller capabilities was ignored. SRv6 greatly reduces bearer network complexity.
- In the 5G and cloud era, more requirements are imposed on networks, such as stronger SLA assurance and deterministic latency. Network connection attributes are enhanced, and packets are required to carry more information. SRv6 extension can perfectly meet these requirements.

Contents

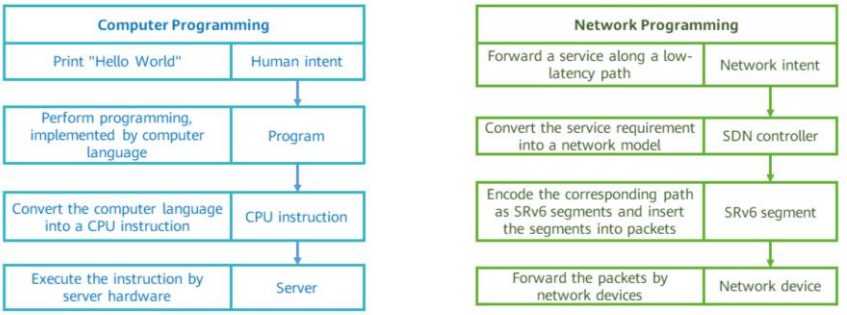
SRv6 Standards/Drafts

- RFC 8402 (Segment Routing Architecture) introduces the source routing model through which the ingress controls traffic steering and forwarding by encapsulating an ordered list of segments. This behavior can be considered as specifying an ordered list of instructions on the ingress, with each instruction representing a function to be executed at a specific network location. Each function is locally defined on a node, for example, instructing the node to perform simple forwarding based on the corresponding segment list or to implement complex user-defined behaviors.
- With greater focus on network programming, SRv6 (draft-ietf-spring-srv6-network-programming) combines simple and complex network functions to implement more network functions other than performing only forwarding. This draft defines the SRv6 network programming concept and main SR behaviors, such as SRv6 SIDs, SR endpoint behaviors, and SR Policy headend behaviors.

- The following terms in this document are defined in RFC 8402: Segment Routing, SR domain, segment ID (SID), SRv6, SRv6 SID, SR Policy, prefix-SID, and Adj-SID.
- The following terms in this document are defined in RFC 8754: SRH, SR source node, transit node, SR segment endpoint node, reduced SRH, Segments Left, and Last Entry.

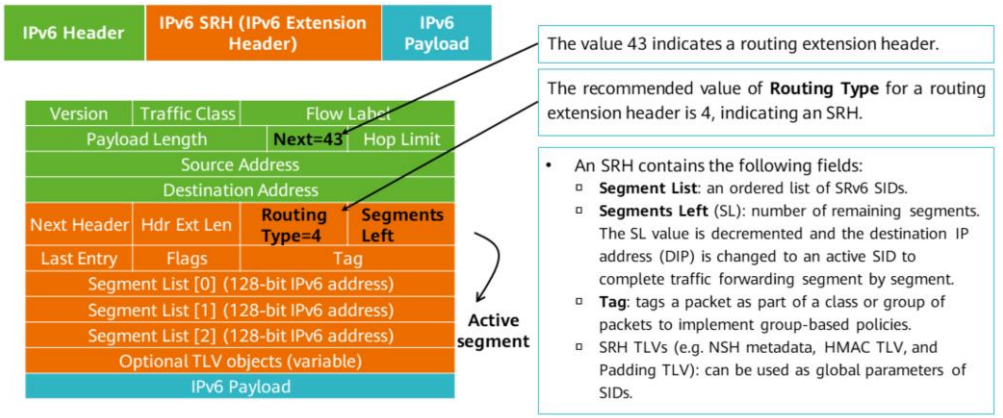
Introduction to Network Programming

- Network programming stems from computer programming. In computer programming, we can convert our intent into a series of instructions that computers can understand and execute to meet our requirements. Similarly, if a network can work as a computer to translate the network intent into a series of forwarding instructions to be executed by network devices, network programming can be achieved.
- Based on the preceding assumption, SRv6 was introduced to translate network functions into instructions and encapsulate the instructions into 128-bit IPv6 addresses, enabling network programming.



IPv6 Segment Routing Header (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 "2. Segment Routing Header" at https://datatracker.ietf.org/doc/rfc8754/.

SRv6 Network Programmability from the Perspective of SRHs

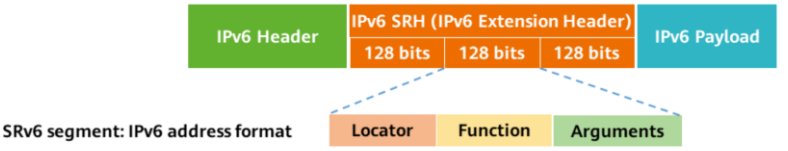
- Leveraging programmable SRHs, SRv6 offers more powerful network programming capabilities than SR-MPLS.
- Generally, SRHs provide a three-dimensional programming space:
  - First dimension: segment list. In SRv6, multiple segments are sequentially combined to represent a specific SRv6 path. This is similar to the application of an MPLS label stack.
  - Second dimension: flexible combination of fields in the 128-bit SRv6 SID. Each MPLS label contains four fixed-length fields: 20-bit label field, 8-bit Time to Live (TTL) field, 3-bit Traffic Class (TC) field, and 1-bit S field. In contrast, each SRv6 SID has 128 bits that can be flexibly divided into fields with variable lengths. This further demonstrates the programmability of SRv6.
  - Third dimension: flexible combination of optional TLVs following segment lists. During packet transmission on a network, irregular information can be encapsulated in the forwarding plane by flexibly combining TLVs in SRHs.

- Reference: SRv6 Network Programming: Ushering in a New Era of IP Networks

Contents

Network Instructions: SRv6 Segments

- A computer instruction typically consists of an opcode and an operand. The former determines the operation to be performed; the latter determines the data, memory address, or both to be used in the computation. Similarly, network instructions, which are called SRv6 segments, need to be defined for SRv6 network programming.



- SRv6 segments are expressed using IPv6 addresses and usually called SRv6 segment identifiers (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 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.

HUAWEI

- https://datatracker.ietf.org/doc/draft-ietf-spring-srv6-network-programming/

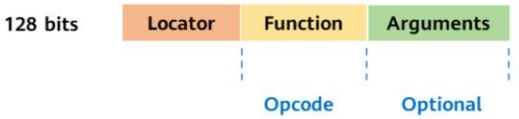
SRv6 Segment: Locator

- 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 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.

```
[Router] segment-routing ipv6
[Router-segment-routing-ipv6] locator srv6_locator1 ipv6-prefix 2001:DB8:ABCD:: 64
```


- 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 specifies the forwarding behavior to be performed, and is similar to the opcode in a computer instruction. In SRv6 network programming, forwarding behaviors are expressed using different functions. For example, RFC defines End, End.X, End.DX4, and End.DX6 behaviors.
- End.X is similar to an adjacency SID in SR-MPLS and is used to identify a link. A configuration example is as follows:

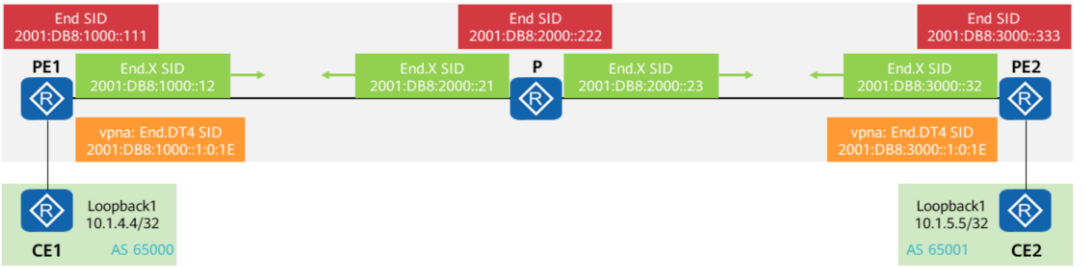
```
[Router-segment-routing-ipv6] locator srv6_locator1 ipv6-prefix 2001:DB8:ABCD:: 64
[Router-segment-routing-ipv6] opcode ::1 end-x interface G3/0/0 next-hop 2001:DB8:200::1
```

- 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 SID Examples



- An End SID is an endpoint SID that 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. An End.DT4 SID can be either manually configured or automatically allocated by BGP within the dynamic SID range of the specified locator.

SRv6 SID Configuration Commands (1)

1. Enable SRv6 and enter the SRv6 view.

   [Huawei] **segment-routing ipv6**

   After running the **segment-routing ipv6** command, you can configure an IPv6 SID in the SRv6 view so that an IPv6 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 Arguments field is determined by the **args** *args-length* parameter. This field is optional in SRv6 SIDs and depends on command configurations.

- **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 (2)

3. Configure a static End SID opcode.

```
[Huawei-segment-routing-ipv6] opcode func-opcode end [ no-psp ]
```

An End SID identifies an SRv6 node. The **no-psp** parameter is used to disable penultimate segment pop of the SRH (PSP).

4. Configure a static End.X SID opcode.

```
[Huawei-segment-routing-ipv6] opcode func-opcode end-x interface { interface-name | interface-type interface-number } nexthop nexthop-address [ no-psp ]
```

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.

SRv6 SID 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 run the opcode command 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--|

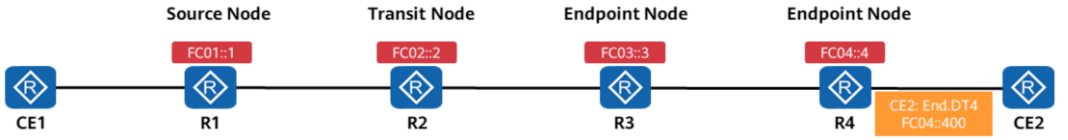
[Router-segment-routing-ipv6] **locator** srv6_locator1 **ipv6-prefix** 2001:DB8:ABCD:: 64 **static** 32

- 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.

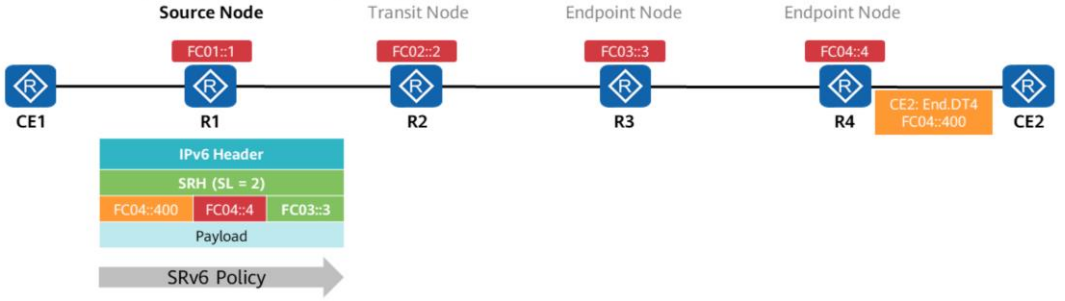
SRv6 Nodes

- RFC 8754 defines three types of SRv6 nodes:
  - SRv6 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 (endpoint node for short): destination node of SRv6 packets.



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.
- The following figure shows an L3VPNv4 over SRv6 Policy scenario where IPv4 traffic is transmitted over SRv6. The SRv6 Policy's explicit path specifies that the traffic must pass through R3 and R4. The source node R1 is responsible for steering the IPv4 traffic into the corresponding tunnel and encapsulating an SRH.

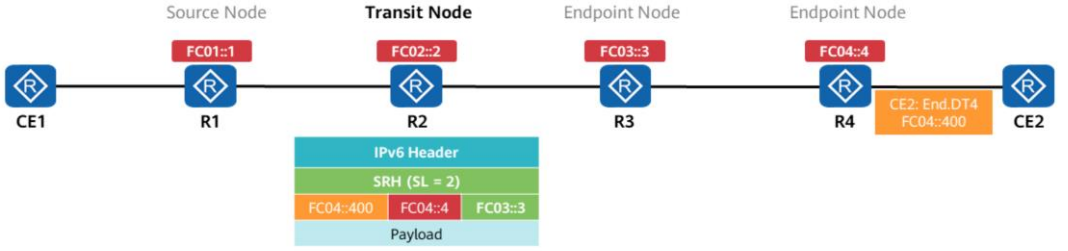


HUAWEI

- Examples for the SRv6 source node to encapsulate an SRH:
  - H.Insert: inserts an SRH into a received IP packet and searches the routing table for packet forwarding.
  - H.Encaps: encapsulates an outer IPv6 header and SRH for a received IP packet, and searches the routing table for packet forwarding.
  - H.Encaps.L2: encapsulates an outer IPv6 header and SRH for a received Layer 2 packet, and searches the routing table for packet forwarding.

Transit Node

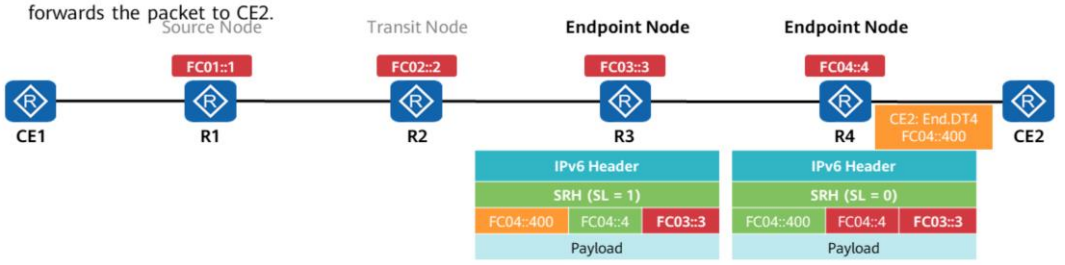
- A transit node is an IPv6 node that does not participate in SRv6 processing on the SRv6 packet forwarding path. That is, the transit node just performs ordinary IPv6 packet forwarding.
- After receiving an SRv6 packet, the node parses the IPv6 DA field in the packet. If the IPv6 DA is neither a locally 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

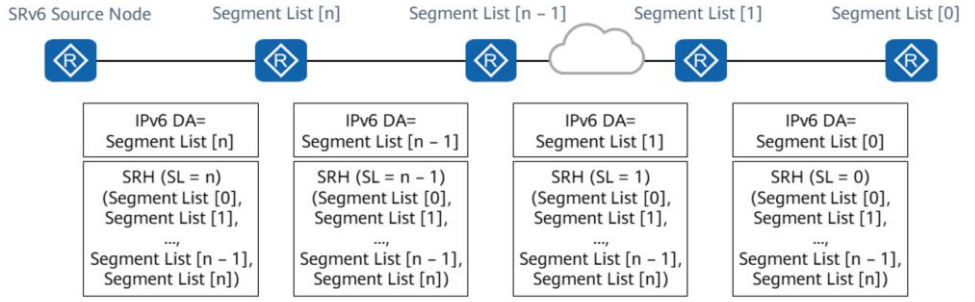
- An endpoint node is a node that receives an SRv6 packet destined for itself (a packet of which the IPv6 destination address is a local SID).
- For example, R4 searches its My Local SID table based on the IPv6 DA FC04::4 of the packet and finds a matching End SID. Then, R4 decrements the SL field by 1 and updates the IPv6 DA to the VPN SID FF04::400. Based on the VPN SID, R4 searches its My Local SID table, finds a matching End.DT4 SID, removes the SRH and IPv6 header, and forwards the packet to CE2.



HUAWEI

SRv6 Packet Forwarding Summary: SRH Processing

- In SRv6 forwarding, each time a packet passes through an SRv6 node, the SL field is decremented by 1 and the IPv6 DA changes. An IPv6 DA is determined by both the SL and Segment List fields.

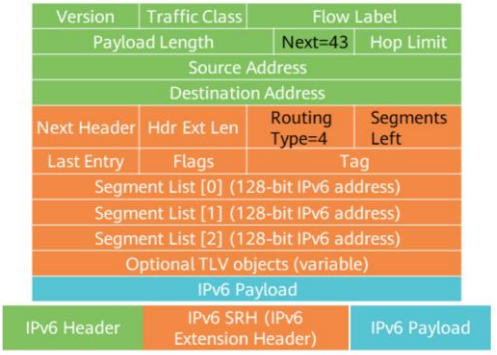


- Different from SR-MPLS label processing, SRv6 SRH processing is implemented from the bottom up, and segments in the SRv6 SRH are not popped after being processed by a node. Therefore, the SRv6 header retains path information, which can be used for path backtracking.

Contents

SRv6 Instruction Sets

- The SRH stores an ordered list of instructions for implementing network services, and it is equivalent to a computer program. Segment List [0] to Segment List [n] are equivalent to the instructions of the computer program. The first instruction to be executed is Segment List [n]. The SL field, which is equivalent to the program counter (PC) of the computer program, points to the instruction that is being executed
- ietf-spring-srv6-network-programming defines many behaviors, which are also called instructions.



SRv6 Instruction Sets: Endpoint Node Behaviors

- Behaviors defined by SRv6 instruction sets are classified as endpoint node behaviors and source node behaviors.
- Common endpoint node behaviors are as follows.

Instruction	Function Description	Application Scenario
End	Copies the next SID to the IPv6 DA and searches the forwarding table for packet forwarding.	Used for packet forwarding through a specified node. An End SID is similar to an SR-MPLS node segment.
End.X	Forwards a packet through a specified outbound interface.	Used for packet forwarding through a specified outbound interface. An End.X SID is similar to an SR-MPLS adjacency segment.
End.T	Searches a specified IPv6 routing table for packet forwarding.	Used in scenarios where multiple routing tables exist.
End.DX2	Decapsulates a packet and forwards it through a specified Layer 2 outbound interface.	Used in L2VPN scenarios, such as EVPN virtual private wire service (VPWS).
End.DX4	Decapsulates a packet and forwards it over a specified IPv4 Layer 3 adjacency.	Used in L3VPNv4 scenarios where packets are forwarded to a CE through a specified IPv4 adjacency.
End.DX6	Decapsulates a packet and forwards it over a specified IPv6 Layer 3 adjacency.	Used in L3VPNv6 scenarios where packets are forwarded to a CE through a specified IPv6 adjacency.
End.DT6	Decapsulates a packet and searches a specified IPv6 routing table for packet forwarding.	Used in L3VPNv6 scenarios.
End.DT4	Decapsulates a packet and searches a specified IPv4 routing table for packet forwarding.	Used in L3VPNv4 scenarios.
End.B6.Insert	Inserts an SRH and applies a specified SRv6 Policy.	Used in scenarios such as traffic steering into an SRv6 Policy in Insert mode, tunnel stitching, and software-defined networking in a wide area network (SD-WAN) route selection.
End.BM	Inserts an MPLS label stack and applies a specified SR-MPLS Policy.	Used in SRv6 and SR-MPLS interworking scenarios.

- For more endpoint behaviors, see ietf-spring-srv6-network-programming.

Naming Rules for SRv6 Instructions

- SRv6 instructions are named according to certain rules. You can quickly determine the 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.

SRv6 Instruction Sets: 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.

SRv6 Instruction Sets: Flavors

- Flavors are additional behaviors defined to enhance End series instructions. These behaviors are optional and used to meet diverse service requirements.
- *SRv6-Network-Programming* defines the following additional behaviors: 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, and End.T
USP	Removes the SRH on the ultimate endpoint node.	End, End.X, and End.T
USD	Decapsulates the outer IPv6 header on the ultimate endpoint node.	End, End.X, and End.T

- 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.

Contents

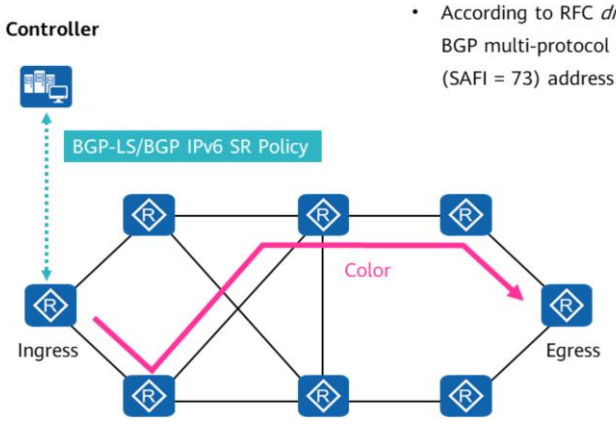
SRv6 Policy Overview

- SR-TE Policy is an application of SR Policy in the traffic engineering field. It leverages the source routing mechanism of SR to guide packet forwarding based on an ordered list of segments encapsulated by the headend. SR-TE Policy does not involve traditional tunnel interfaces.
- SR-TE Policy is classified as SR-MPLS Policy or SRv6 Policy by segment. This document uses SRv6 Policy as an example.
- SR Policy is the SR-TE implementation mode of mainstream vendors. In the following contents, SRv6 Policy will be used to refer to SRv6 TE Policy.

```
<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
0x00000000320000c001     srtepolicy    1.0.0.12         UP
0x000000003400002001     srv6tepolicy  FC01::12         UP
```


- 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 Policy Standards

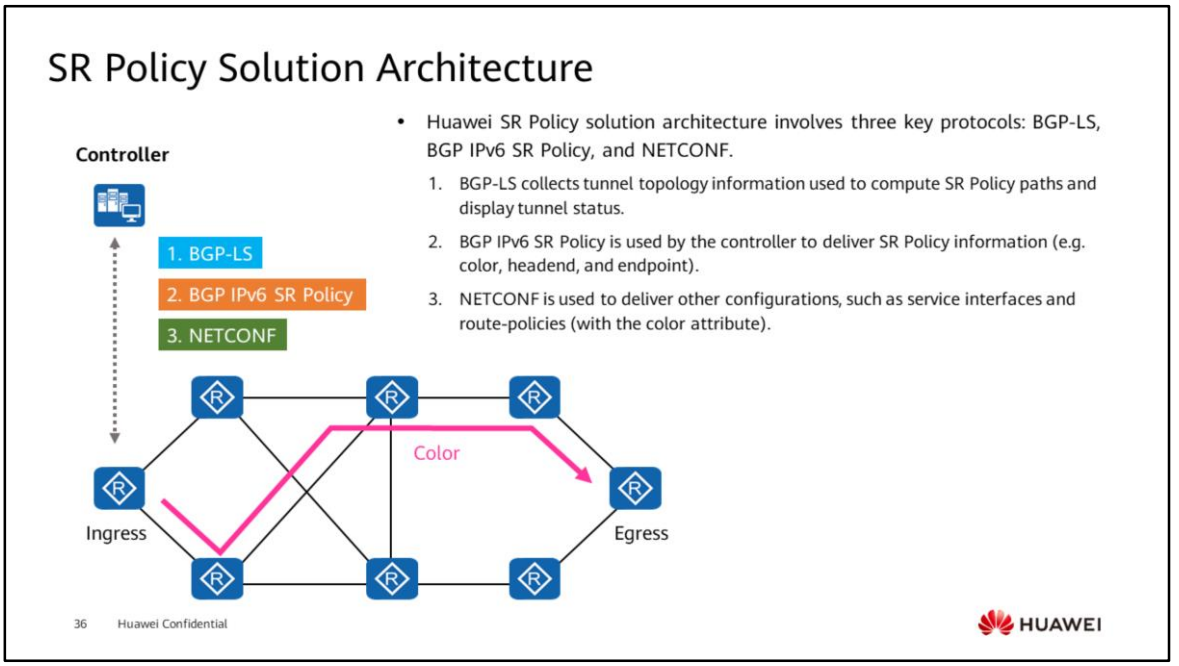


- According to RFC *draft-ietf-spring-segment-routing-policy*, BGP multi-protocol extension supports the BGP IPv6 SR Policy (SAFI = 73) address family for SR Policy delivery.

- 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, it can be located 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 IPv6 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/)



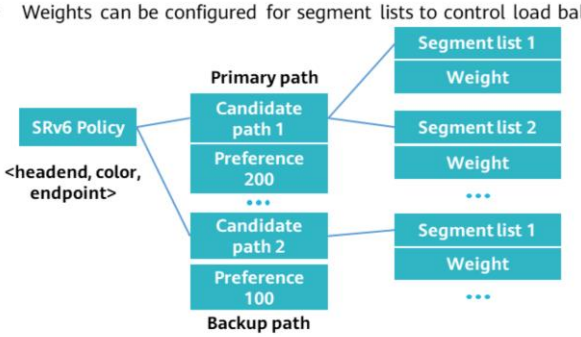
- BGP-LS connection:
  - Collects tunnel topology information for SR Policy path computation.
  - BGP-LS supports the collection of SR Policy status information, based on which the controller displays tunnel status. https://datatracker.ietf.org/doc/draft-ietf-idr-te-lsp-distribution/
  - BGP-LS supports Segment Routing Local Block (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 IPv6 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.

SRv6 Policy Tuple

- An SRv6 Policy is identified by <headend, color, endpoint>. For a specified node, the policy is identified by <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 latency).
  - Endpoint: destination address of an SRv6 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 SRv6 Policy.

SRv6 Policy 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 segment list sent to the headend through PCEP or BGP IPv6 SR Policy.
- Weights can be configured for segment lists to control load balancing among SR 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 SID-List <SID11...SID1i>, traffic is balanced according to W1/(W1+W2).

Extension: BGP IPv6 SR Policy Routing Table

- An SR Policy is represented as [distinguisher][policycolor][endpoint] in the BGP routing table.

```
<PE1>display bgp sr-policy ipv6 routing-table

 BGP Local router ID is 1.0.0.1
 Status codes: * - valid, > - best, d - damped, x - best external, a - add path,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
 RPKI validation codes: V - valid, I - invalid, N - not-found

 Total Number of Routes: 2
      Network                    Nexthop         MED        LocPrf    PrefVal Path/Ogn

 *>i  [10][5][FC01::12]          2000::102       4294967286 100       0       ?
 *>i  [14][3][FC01::12]          2000::102       4294967286 100       0       ?
```

Extension: BGP IPv6 SR Policy's Tunnel Encapsulation Attribute (1)

- In a BGP Update message, path attributes carry specific candidate path information. The following shows the Tunnel Encapsulation Attribute format.

```
SR Policy SAFI NLRI: <Distinguisher, Policy-Color, Endpoint>
    Attributes:
      Tunnel Encaps Attribute (23)
        Tunnel Type: SR Policy (15)
          Binding SID
          Preference
          Priority
          Policy Name
          Explicit NULL Label Policy (ENLP)
          Segment List
            Weight
            Segment
            Segment
          ...
```


HUAWEI

- RFC-Advertising Segment Routing Policies in BGP
- The BGP extensions for the advertisement of SR Policies include the following components:
  - New SAFI
  - New Tunnel Type ID and a series of TLVs to be inserted into the Tunnel Encapsulation Attribute
  - Route-target extended community used to indicate the intended headend of SR Policy advertisement
  - Color extended community for traffic steering (defined in [I-D.ietf-idr-tunnel-encaps])

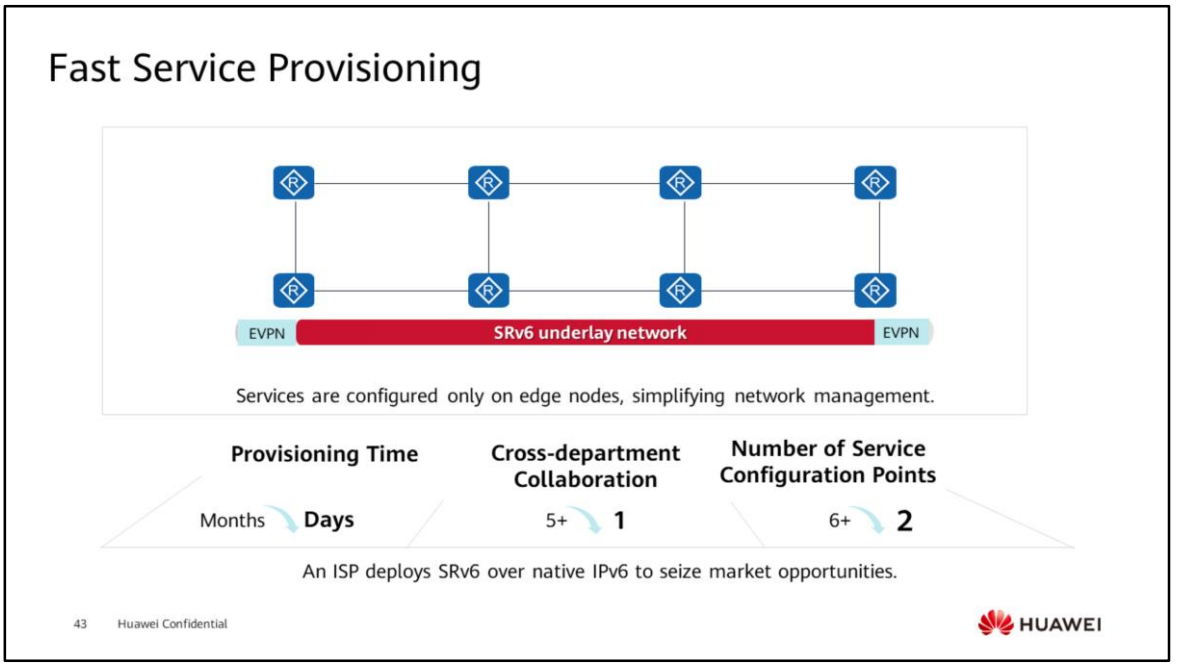
Extension: BGP IPv6 SR Policy's Tunnel Encapsulation Attribute (2)

- The BGP IPv6 SR Policy routing table contains complete tunnel encapsulation attribute information.

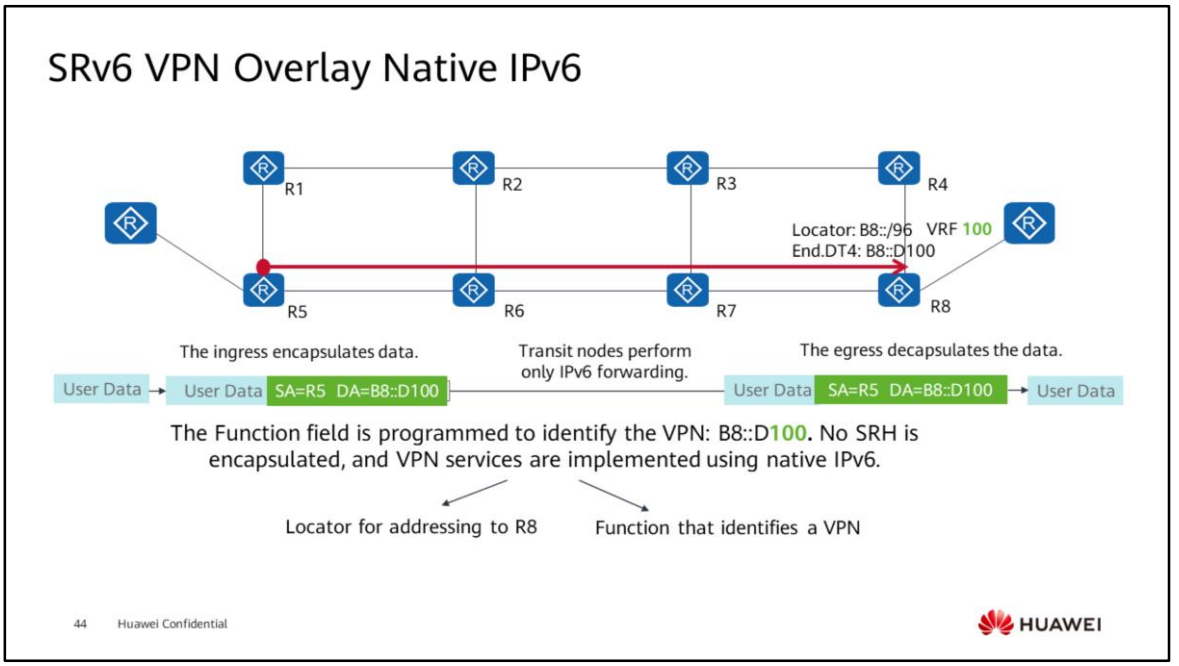
```
<PE1>display bgp sr-policy ipv6 routing-table [10][5][FC01::12]


 BGP local router ID : 1.0.0.1
 Local AS number : 100
 Paths:   1 available, 1 best, 1 select, 0 best-external, 0 add-path
 BGP routing table entry information of [10][5][FC01::12]:
 From: 2000::102 (172.21.17.102)
 Route Duration: 3d22h20m53s
 Relay IP Nexthop: ::
 Relay IP Out-Interface:GigabitEthernet0/0/0
 Original nexthop: 2000::102
 Qos information : 0x0
 Ext-Community: RT <1.0.0.1 : 0>, SoO <172.21.17.102 : 0>
 AS-path Nil, origin incomplete, MED 4294967286, localpref 100,
pref-val 0, valid, internal, best, select, pre 255
 ...
```

```
Tunnel Encaps Attribute (23):
 Tunnel Type: SR Policy (15)
  Preference: 200
  Binding SID: FC00:1::1:1, s-flag(0), i-flag(0)
  Segment List
    Weight: 1
    Path MTU: 9600
    Segment: type:2, SID: FC00:1::1:3D
    Segment: type:2, SID: FC00:2::1:22
    Segment: type:2, SID: FC00:6::1:24
    Segment: type:2, SID: FC00:4::1:26
  Segment List
    Weight: 1
    Path MTU: 9600
    Segment: type:2, SID: FC00:11::1:20
    Segment: type:2, SID: FC00:4::1:20
    Segment: type:2, SID: FC00:3::1:24
  Template ID: 4294967279
Not advertised to any peer yet
```

Contents

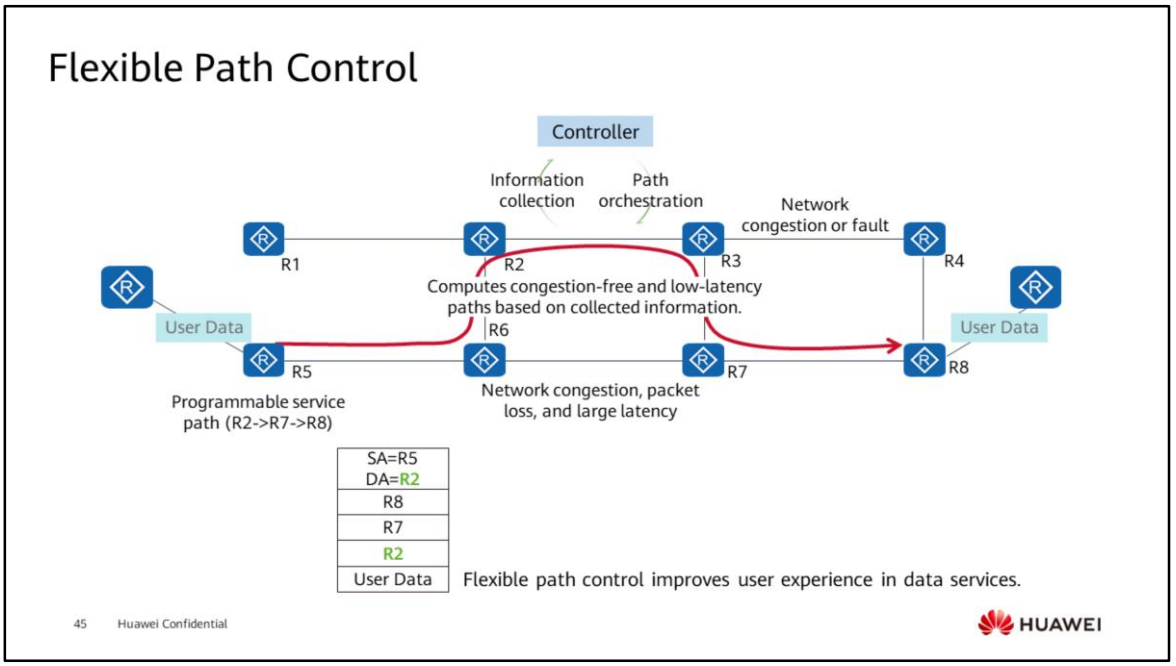


- A traditional network adopts the segmented networking and separate management mode. The bearer technologies used on the segmented network are not integrated, and different network segments need to be managed and maintained by different departments. As such, cross-domain service provisioning usually requires multiple departments (5+) to collaborate, slowing down the provisioning. SRv6 implements E2E collaboration on a cross-domain network. Cross-domain seamless connections can be automatically set up on the service network only by enabling SRv6 at the two ends of the network. Transit nodes are unaware of new services and do not require any changes. Through cross-domain E2E collaboration, the service network deployment time is reduced from months to days.

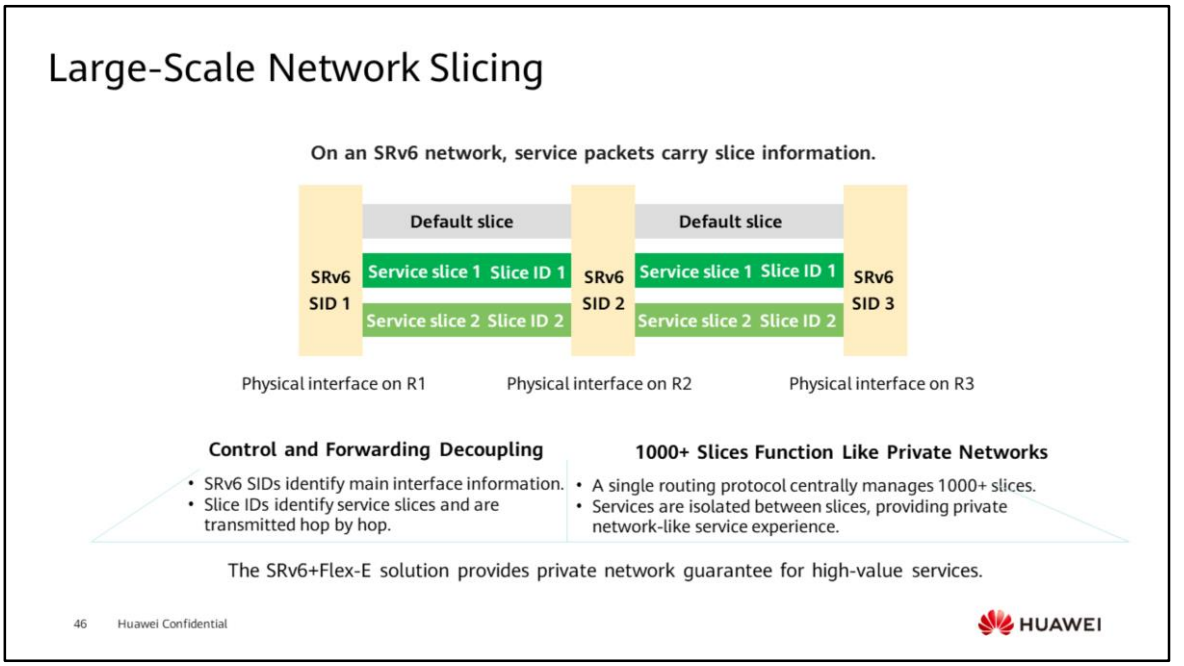


- SRv6 provides overlay VPN capabilities on native IPv6 networks to isolate different services.
- For a VPN service, a PE of the VPN service generates a VPN SID for the corresponding VPN instance, with the Locator field indicating the reachability information about the local node and the Function field identifying the VPN. As shown in the figure, R8 allocates VPN SID B8::D100 to VPN 100. B8 indicates the locator used to identify how the service can reach R8. D100 identifies the VPN and is used for service addressing and forwarding in VRF 100. The VPN SID is advertised together with VPN routes to all other PEs through BGP extension, and a local SID forwarding entry is delivered.
- When VPN service data reaches the ingress, the ingress searches for the corresponding VPN SID based on the VPN route and performs SRv6 encapsulation. After R5 receives the VPN packet destined for R8, R5 searches the local VPN forwarding table, finds the forwarding entry corresponding to VPN 100, and performs IPv6 encapsulation using the VPN SID B8::D100 allocated by R8 as the destination address. After packet encapsulation is completed, the ingress searches for the outbound interface and next hop based on the destination IPv6 address and forwards the packet accordingly.
- The packet finally reaches R8. R8 then searches the local SID forwarding table based on the destination address, finds the routing entry corresponding to the VPN, and forwards the VPN packet. This is the complete VPN packet forwarding process.
- SRv6-based VPN packet forwarding does not require additional encapsulation protocols. SRv6 tunnels natively support multi-point interworking, and therefore you do not need to manually configure tunnels one by one. IPv6 services need to be deployed only on edge nodes that are aware of services, and transit nodes only need to perform native IPv6 forwarding. This simplifies VPN network deployment and maintenance. In addition, the IPv6 route aggregation capability enables edge nodes to maintain only aggregated or default routes, reducing the

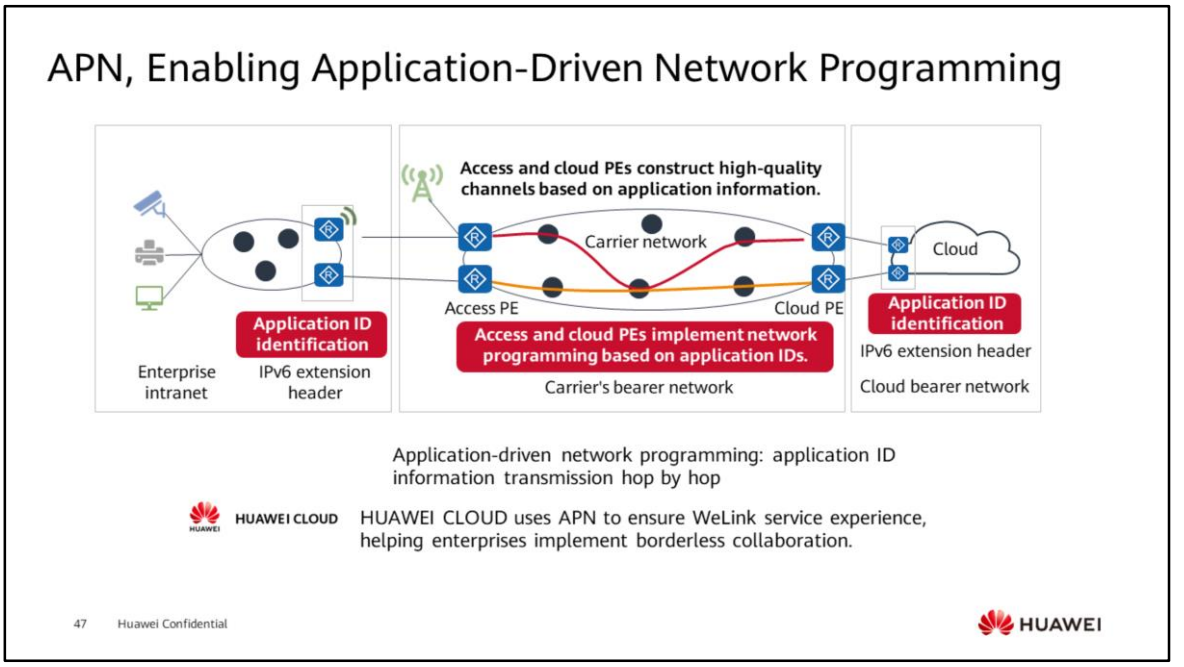
burden on edge nodes on a large-scale network.



- SRv6 implements service path programming through flexible segment list orchestration, thereby achieving service-driven network path selection. With the help of an SDN controller, SRv6 ensures that services are always transmitted over network paths meeting service requirements.
- The SDN controller collects network information in real time to obtain and maintain current network status and relevant information. To meet customers' SLA requirements, such as low latency, bandwidth guarantee, and primary and secondary paths, the controller computes and orchestrates the optimal network path based on the current network status. In addition, the controller delivers the orchestrated service path to the ingress to guide service forwarding on the network. For example, the service path R2->R7->R8 is delivered, as shown in the figure. After the user service packet reaches R5, SRv6 encapsulation is performed for it, and the destination address is set to the next SID, that is, R2 (the first SID to be encapsulated in the path).
- During service running, IFIT can be used to measure service quality in real time, and Telemetry can be used to collect measurement data in real time. Based on technologies such as big data analytics and machine learning, the SDN controller analyzes service quality on the network in real time. When service quality deteriorates, the SDN controller can re-trigger service path computation and switch service flows to a new path that meets SLA requirements.
- Through flexible service path control as well as network status awareness and collection, SRv6 implements real-time service quality measurement and assurance, ensuring user experience.



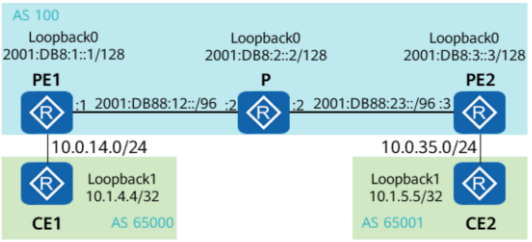
- In a traditional slicing solution, different IP addresses need to be configured for network slices and an IGP needs to be run for route distribution. This implementation mode requires a lot of configurations. In addition, because different slices must have independent IP addresses, the IGP route quantity increases exponentially with the number of slices. This increases the burden on network devices and limits the large-scale deployment of slices.
- In the SRv6+Slice ID solution, SRv6 SIDs identify network nodes, and slice IDs identify slices. All slices share the SID and IP address of the default slice, and slice IDs are used to differentiate slices. The SRv6 network programmability enables the data plane to be separated from the control plane, preventing the IGP load from affecting the slicing scale and allowing 1000+ slices to be created to meet the increasingly diversified slicing requirements in the future.



- SRv6 programmability enables networks to interact with applications, implementing application-based refined network traffic operation and management. For example, HUAWEI CLOUD can ensure WeLink service experience based on SRv6 APN. As shown in the figure, the network devices at the enterprise egress and cloud egress identify applications and transmit application information to the access and cloud PEs through IPv6 extension headers. According to specified service policies, the PEs perform differentiated service path orchestration based on application IDs so as to implement application-based differentiated service operation.

Contents

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.

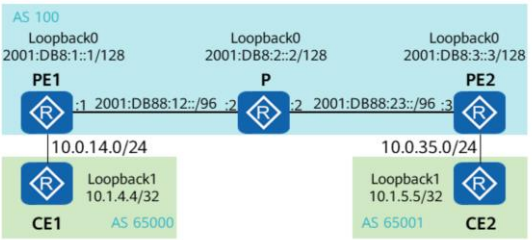
Configuration roadmap:

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.
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.
6. Verify the configuration.

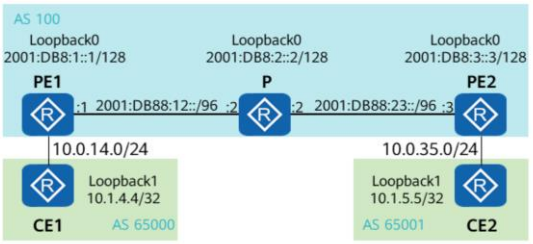
Establish an MP-IBGP peer relationship between the PEs.

```
[~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
```

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.
6. Verify the configuration.

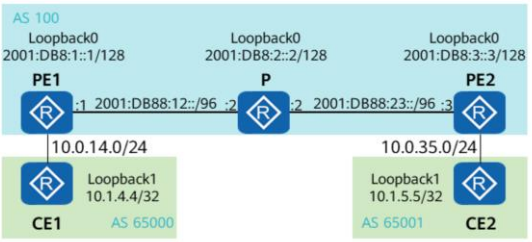
Establish an SRv6 BE path between the PEs. The configuration on PE1 is as follows: (The configuration on PE2 is not provided here, and the P does not require such configuration.)

```
[~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:
  - Run the segment-routing ipv6 command to enable SRv6 and enter the SRv6 view.
  - 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.
  - 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.

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.**
6. Verify the configuration.

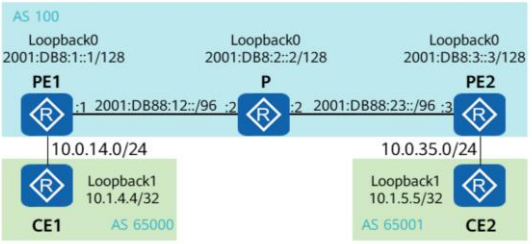
Enable the VPN instance IPv4 address family on each PE. PE1 configurations are as follows: (PE2 configurations are not provided.)

```
[~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.
  - Run the bgp {as-number-plain | as-number-dot} command to enter the BGP view.
  - Run the ipv4-family vpn-instance vpn-instance-name command to enter the BGP-VPN instance IPv4 address family view.
  - Run the segment-routing ipv6 best-effort command to enable VPN route recursion based on the SIDs carried by routes.
  - Run the segment-routing ipv6 locator locator-name [auto-sid-disable] command to enable the device to add SIDs to VPN routes.
  - 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.
  - Run the commit command to commit the configuration.

L3VPNv4 over SRv6 BE (5)



Configuration roadmap:

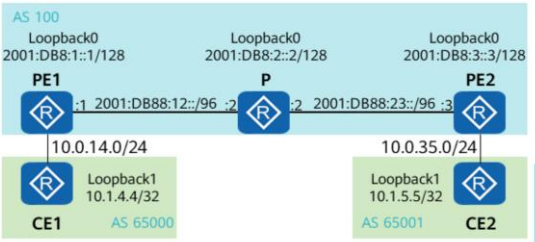
1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs.
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.

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.
6. **Verify the configuration.**

Check VPNv4 routing information on PE1.

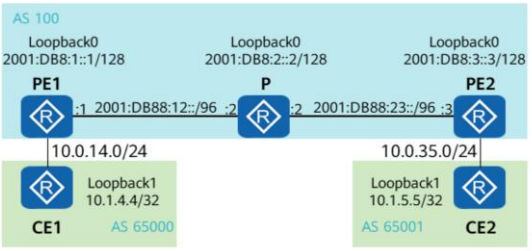
```
<PE1>display bgp vpnv4 al routing-table 10.1.5.5
 BGP local router ID : 10.0.1.1
 Local AS number : 100
 Total routes of Route Distinguisher(100:1): 1
 BGP routing table entry information of 10.1.5.5/32:
 Label information (Received/Applied): 3/NULL
 From: 2001:DB8:3::3 (10.0.3.3)
 Route Duration: 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.
6. **Verify the configuration.**

Check vpna's routing information on PE1.

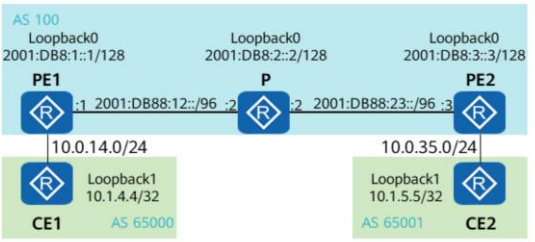
```
<PE1> display ip routing-table vpn-instance vpna 10.1.5.5 verbose
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B -
black hole route
------------------------------------------------------------------------------
Routing Table : vpna
Summary Count : 1

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:

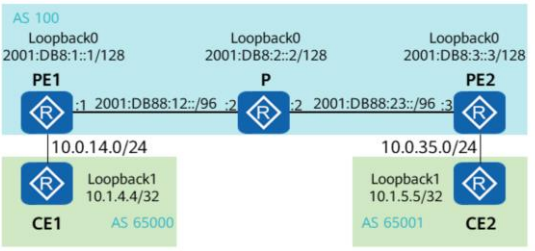
1. Configure interface IPv6 addresses and IS-IS. (Configuration details are not provided.)
2. Establish an MP-BGP peer relationship between PE1 and PE2.
3. Enable SR and establish an SRv6 BE path on the backbone network.
4. Enable the VPN instance IPv4 address family on each PE.
5. Establish an MP-IBGP peer relationship between the PEs.
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
```

Contents

L3VPNv4 over SRv6 Policy (1)



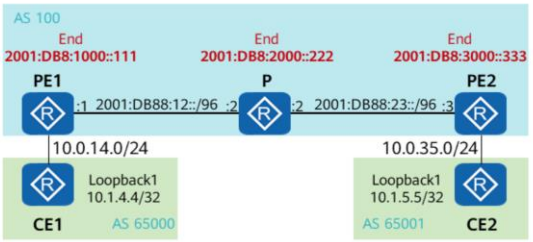
Networking requirements:

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.
5. Configure a tunnel policy and import VPN traffic.
6. Verify the configuration.

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.
5. Configure a tunnel policy and import VPN traffic.
6. Verify the configuration.

Establish an MP-IBGP peer relationship between the PEs.

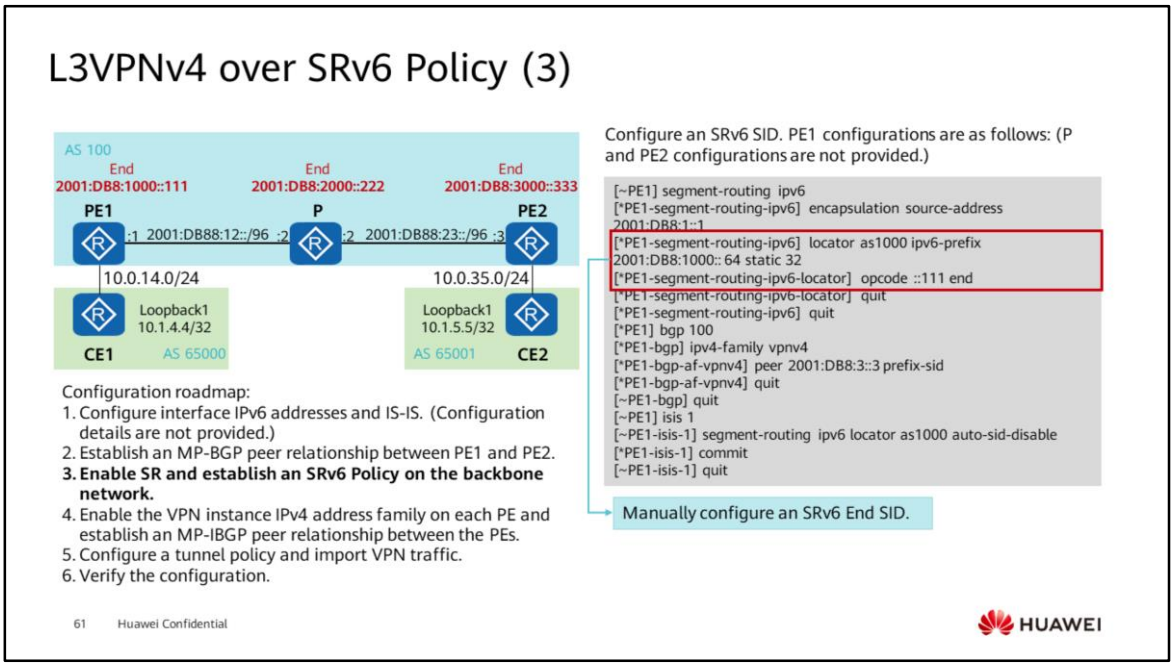
```
[~PE1] bgp 100
[~PE1-bgp] peer 2001:DB8:3::3 as-number 100
[*PE1-bgp] peer 2001:DB8:3::3 connect-interface loopback 0
[*PE1-bgp] ipv4-family vpnv4
[*PE1-bgp-af-vpnv4] peer 2001:DB8:3::3 enable
[*PE1-bgp-af-vpnv4] commit
[~PE1-bgp-af-vpnv4] quit
[~PE1-bgp] quit
```

Check the VPNv4 peer relationship on PE1.

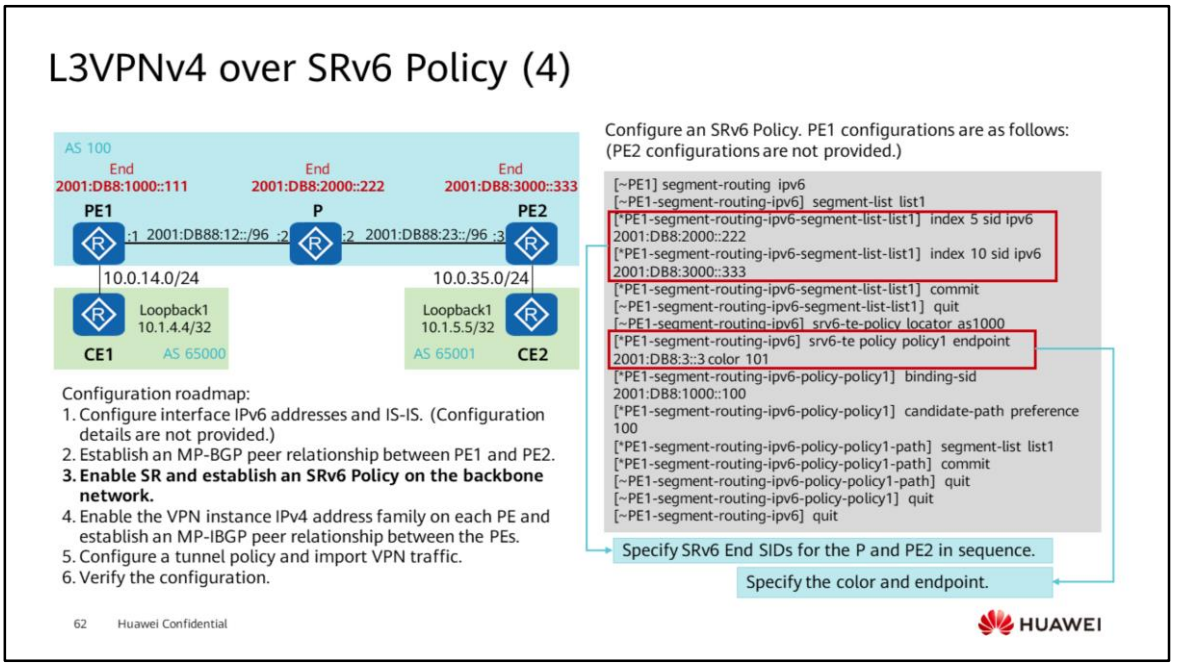
```
<PE1>display bgp vpnv4 all peer

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.

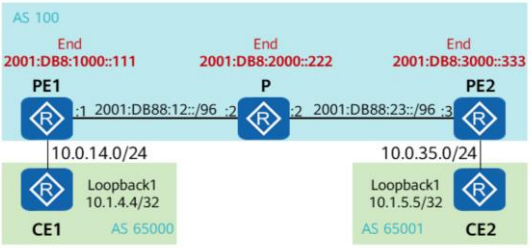


- SRv6 paths are established using SIDs. Static SRv6 SIDs are recommended. The configuration procedure is as follows:
  - Run the locator locator-name [ipv6-prefix ipv6-address prefix-length [static static-length | args args-length] *] command to configure an SRv6 locator.
  - Run the opcode func-opcode end command to configure a static End SID opcode.
  - Run the opcode func-opcode end-x interface interface-name nexthop nexthop-address [no-psp] command to configure a static End.X SID opcode.
  - Run the quit command to exit the SRv6 locator view.



- Configure a segment list.
  - Run the system-view command to enter the system view.
  - Run the segment-routing ipv6 command to enable SRv6 and enter the SRv6 view.
  - Run the segment-list list-name command to configure a segment list (an explicit path) for an SRv6 TE Policy candidate path and enter the segment list view.
  - 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 TE 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.
  - Run the commit command to commit the configuration.

L3VPNv4 over SRv6 Policy (5)



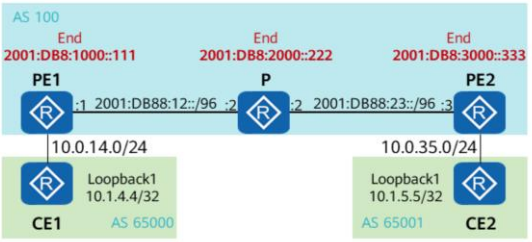
Configuration roadmap:

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.**
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
```

L3VPNv4 over SRv6 Policy (6)



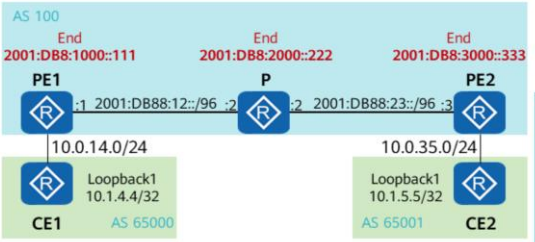
Configuration roadmap:

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.
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
```

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.
5. Configure a tunnel policy and import VPN traffic.
6. **Verify the configuration.**

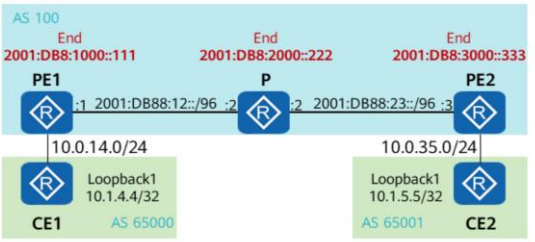
Check SRv6 TE Policy information on PE1.

```
<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.
5. Configure a tunnel policy and import VPN traffic.
6. **Verify the configuration.**

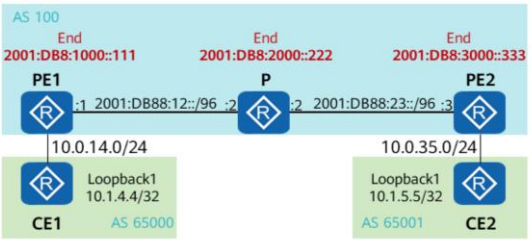
Check VPNv4 routing information on PE1.

```
<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.
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: 0x00000000003400000001   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
```

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?

- An SRv6 SID has 128 bits and consists of the Locator, Function, and Arguments fields.
- 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 device, such as a firewall, application acceleration gateway, or 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.

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.

Huawei CloudWAN Solution Architecture and Fundamentals

HUAWEI

Foreword

- With the rapid development of the Internet industry and the advent of the cloud era, new business models are emerging one after another, and enterprises are moving towards cloudification and digitalization in full swing. Service cloudification brings great flexibility and uncertainty to service applications. As the network scale and complexity continuously grows, O&M complexity increases accordingly. As such, it is imperative for WANs to enter the intelligent cloud-network era.
- Huawei's CloudWAN solution is based on advanced network devices and an efficient SDN architecture. This solution aims to build an intent-driven network that is first automated, then self-adaptive, and finally autonomous.
- This course describes the architecture of Huawei's CloudWAN solution and the fundamentals of centralized management, control, and analysis provided by the architecture.

Objectives

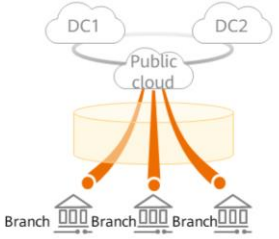
- Upon completion of this course, you will be able to:
  - Describe the main network management functions of the WAN controller.
  - Describe the protocols used for network management and their functions.
  - Describe the protocols used for network information collection and reporting.
  - Describe the meanings of path constraint parameters.
  - Describe the functions of network performance analysis.

Contents

3 Challenges Facing the WAN in the Cloud Era

Cloudification of millions of enterprises

How can networks be as agile as clouds?



Service deployment time

Dozens of days >>> < 1 hour

Production service bearer

Can IP provide deterministic experience?

Relay protection

10 ms One-way delay

200 µs Two-way delay variation

A failure to meet these requirements may mean a protection failure, incurring accidents.

Train control

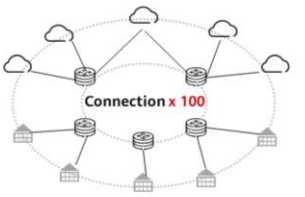
10 ms One-way delay

50 ms Failover duration

A failure to meet these requirements may mean train out-of-control.

Connection scale x 100 ↑

How can O&M be simpler and networks more reliable?

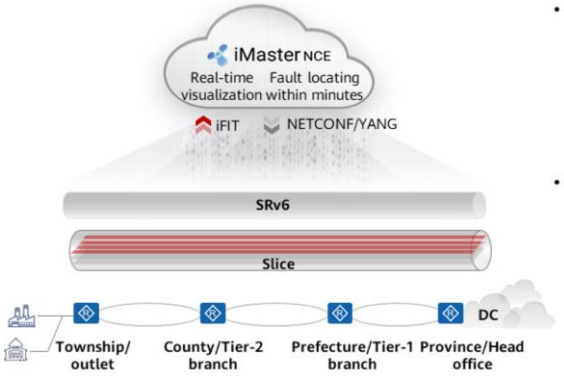


0 artifact, **0** frame freezing

24/7 online

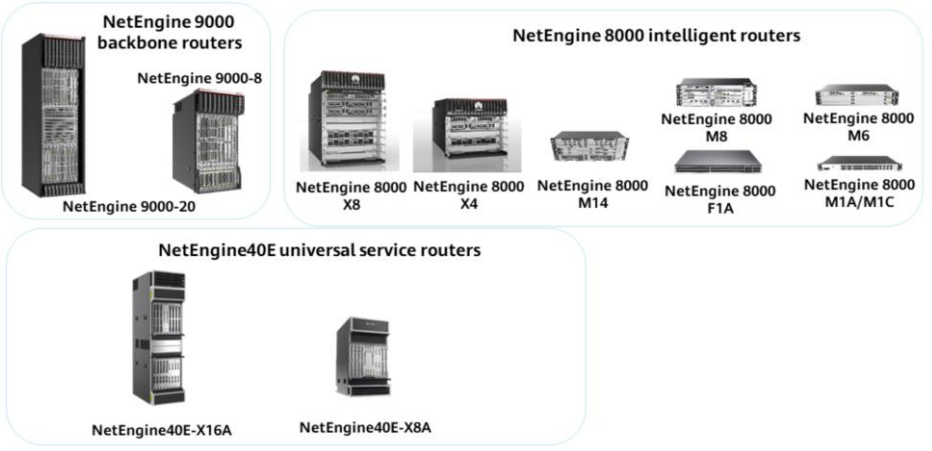
HUAWEI

CloudWAN: Accelerating the Digital Transformation of Various Industries



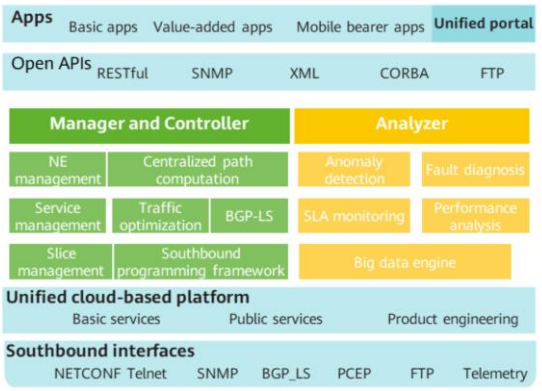
- CloudWAN solution advantages:
  - Service provisioning within minutes
  - Hierarchical slicing for deterministic delay
  - Hop-by-hop detection technology for real-time visualization of network-wide status
  - ...
- CloudWAN solution components
  - iMaster NCE-IP: implements centralized management, control, and analysis of the entire network. It enables resource cloudification, full-lifecycle automation, and data analysis-driven intelligent closed-loop management based on business and service intents. Moreover, it provides open network APIs for fast integration with IT systems.
  - NetEngine series intelligent routers: adopt the flexible, open, and efficient SDN architecture and provide large capacity and high reliability. They are easy to maintain and energy conserving.

Intelligent Universal Service Routers for the Cloud Era



HUAWEI

iMaster NCE-IP Architecture and Key Components

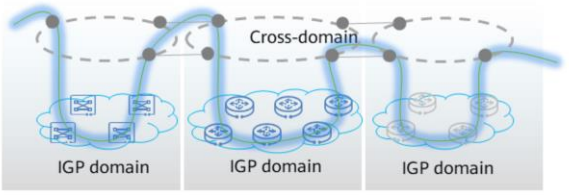


- NCE, which is based on the cloud architecture, supports distributed deployment on VMs. The overall architecture consists of modules for common services, management, control, and analysis, scenario-specific apps, and northbound and southbound openness.
  - Common service module: provides basic network services (such as alarm and log services) and product engineering capabilities (such as disaster recovery and backup).
  - Management, control, and analysis module: provides network topology, L2VPN/L3VPN, path computation and optimization, diagnosis, prediction, and other capabilities by module.
  - Scenario-specific app module: provides E2E service capabilities, such as agile private lines, for different business scenarios.
  - Northbound and southbound openness module: provides northbound APIs and southbound interfaces for quick interconnection or integration with third-party applications, other management and control systems, and devices.

CloudWAN: Fast Service Provisioning

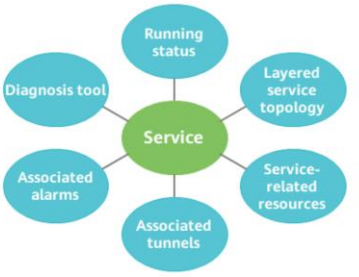
- Simplified, efficient E2E service provisioning within minutes, fast and differentiated service SLA assurance, and 360° graphical display.

Template-based rapid service deployment



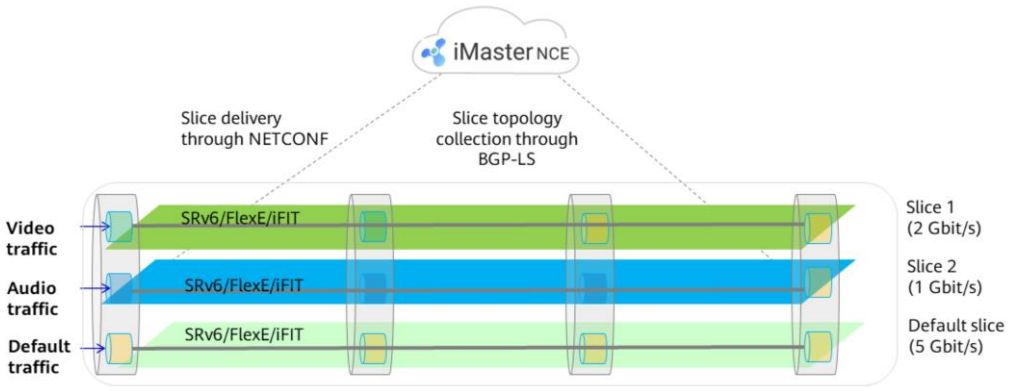
- **Simplified and efficient:** service-driven automatic tunnel creation
- **E2E optimal**: E2E path computation across IGP domains/ASs
- **Highly reliable**: path computation based on link availability/SRLG
- **SLA assurance:** path computation based on bandwidth, delay, etc.

Visualized service display (service 360°)

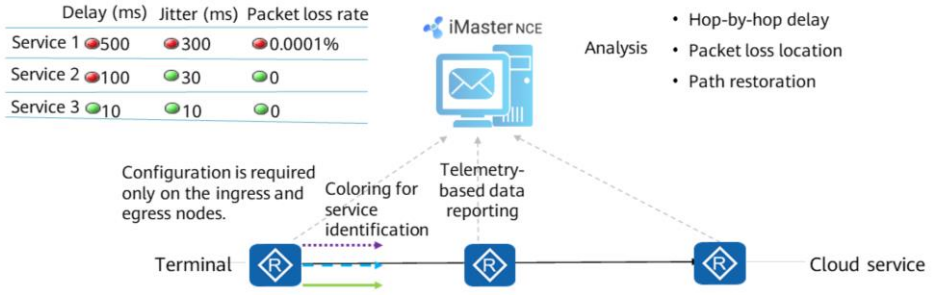


HUAWEI

CloudWAN: Key Service Assurance with Network Slicing

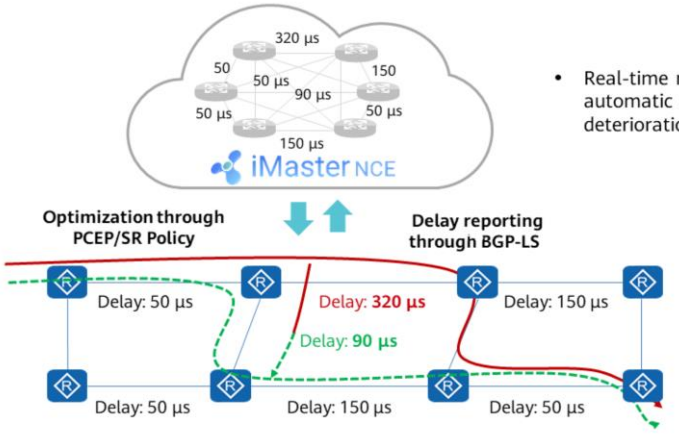


CloudWAN: Fast Fault Locating for VPN Services



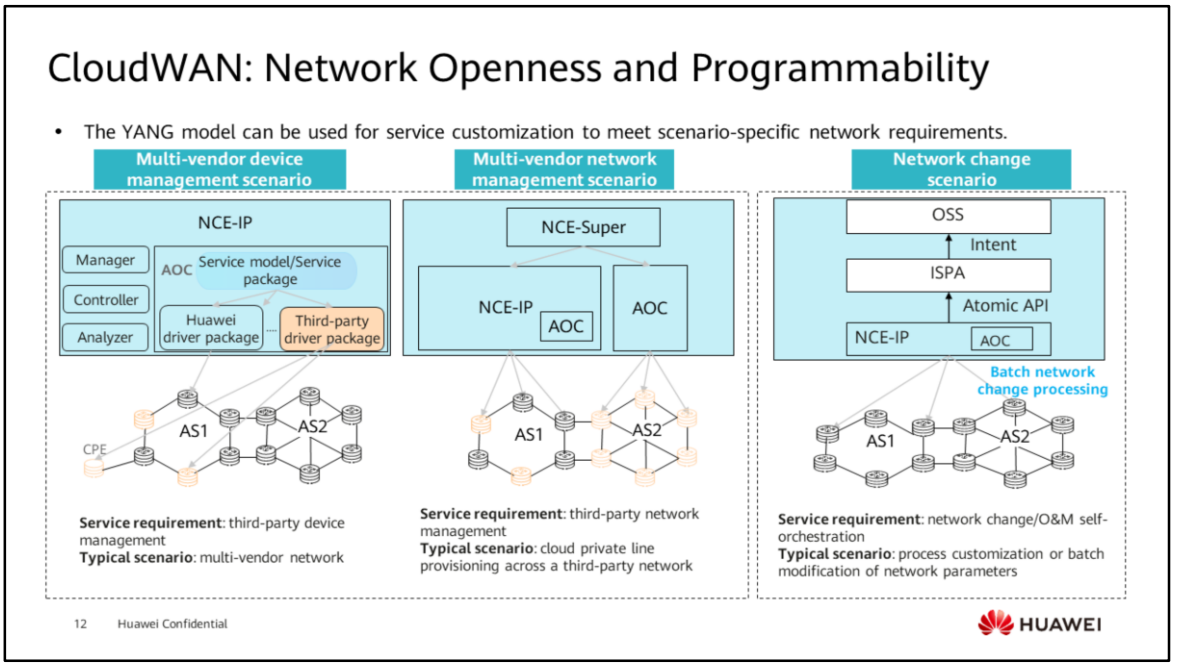
HUAWEI

CloudWAN: Automatic Network Optimization



- Real-time network delay monitoring and automatic service path adjustment upon delay deterioration for service experience guarantee.

HUAWEI

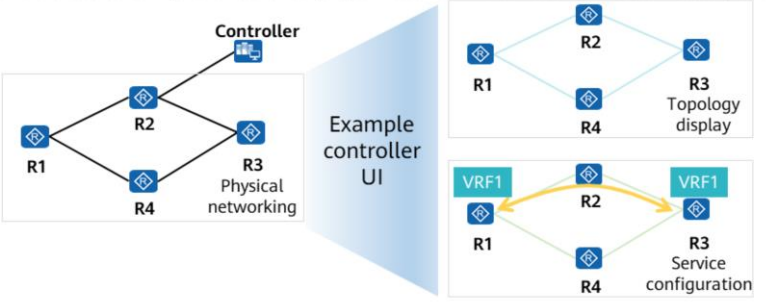


- In the multi-vendor device management scenario, Huawei provides driver packages. The customer develops or Huawei provides service packages.
- In multi-vendor network management scenarios, Huawei provides driver packages, NCE-Super, and Agile Open Container (AOC). AOC interconnects with third-party devices. NCE-Super implements E2E service provisioning through NCE-IP and AOC.
- In network change scenarios, AOC provides atomic APIs and network orchestration. The Intelligent Service Process Automation (ISPA) platform orchestrates service processes and provides intent APIs.

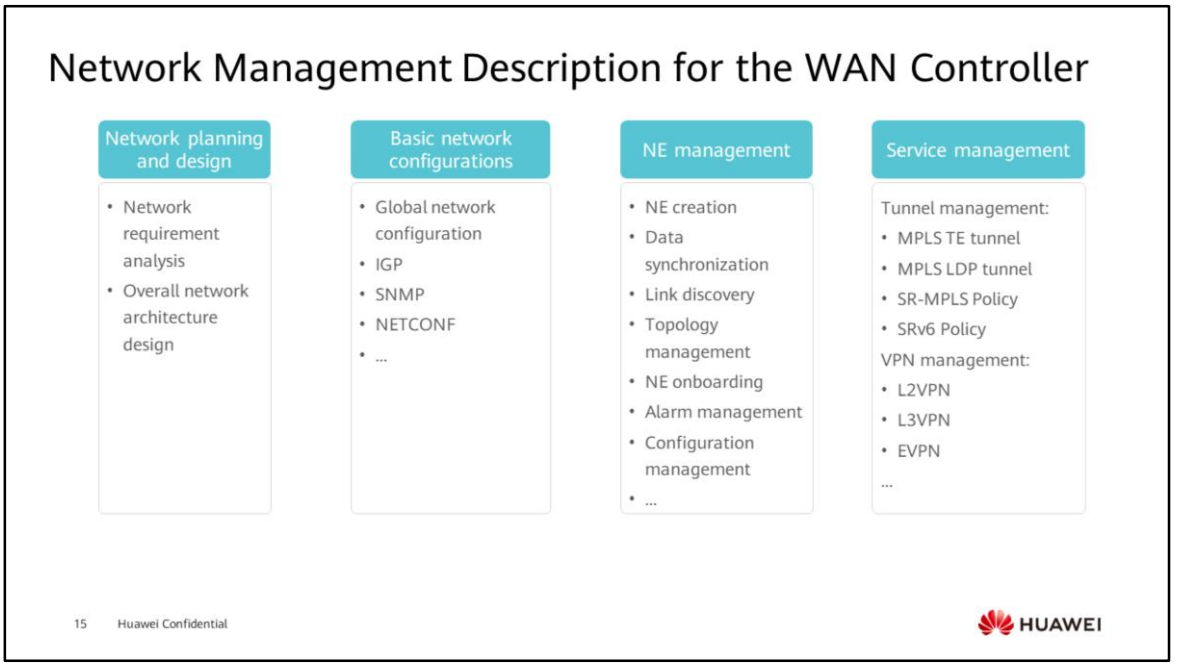
Contents

Network Management Overview for the WAN Controller

- Network management, a basic function of the WAN controller, consists of NE management and service management:
  - NE management is a basic function of the WAN controller. You can configure basic parameters (such as IP addresses, SNMP, NETCONF, and AAA) on the network devices and controller for them to communicate. In this way, the actual network topology can be displayed on the controller, and functions such as NE configuration management can be implemented in graphical mode.
  - Service management is used to configure various underlying tunnels and VPNs carried over these tunnels to provide services for customers.



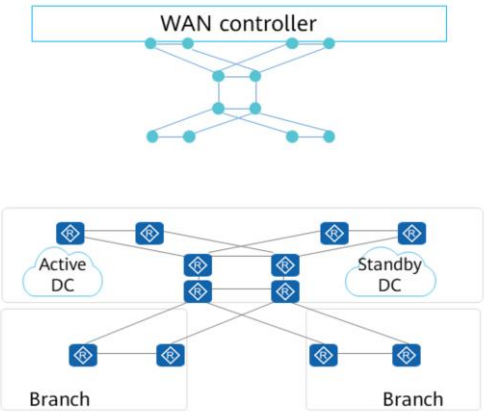
HUAWEI



- For details about basic network configuration, see the lab guide.

Contents

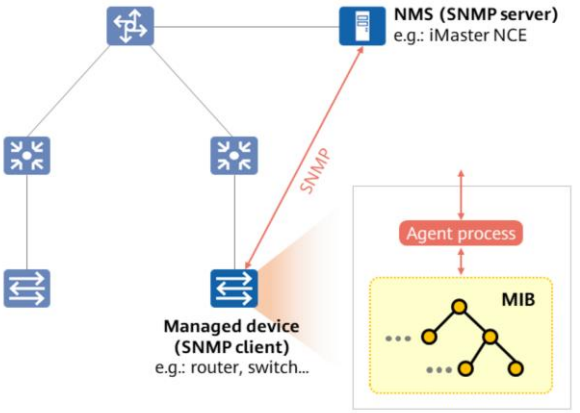
Network Management Overview



- Scenario description:
  - Network device, link, and alarm management for an enterprise's self-built WAN.
- Technical protocols:
  - SNMP
  - NETCONF
  - LLDP

SNMP

- SNMP is deployed for functions such as device alarm and topology management in WAN management.

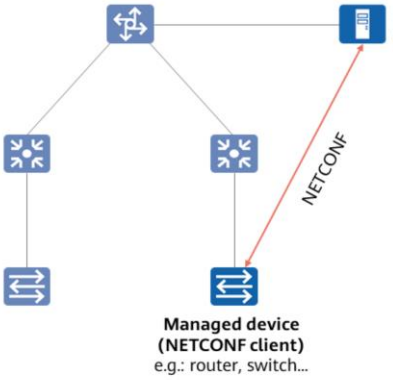


SNMP overview

- Simple Network Management Protocol (SNMP) is a network management standard widely used on TCP/IP networks.
- SNMP provides a method for managing devices through a network management station (NMS) — a central computer that runs network management software.
- 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.

NETCONF

- Network Configuration Protocol (NETCONF) is deployed for the controller to deliver configurations, such as VPN instance creation and routing protocol configurations, in WAN management.



SNMP's drawbacks

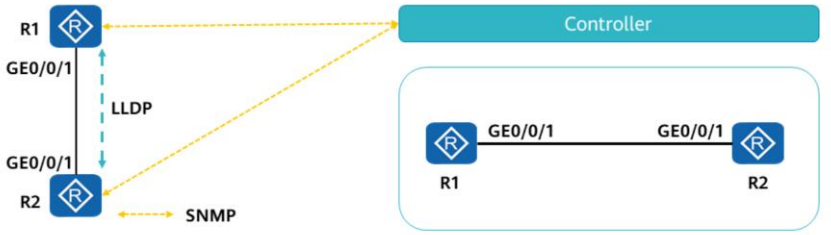
SNMP is not a configuration-oriented protocol. On a large-sized network with a complex topology, SNMP cannot meet network management requirements, especially the configuration management requirements.

NETCONF

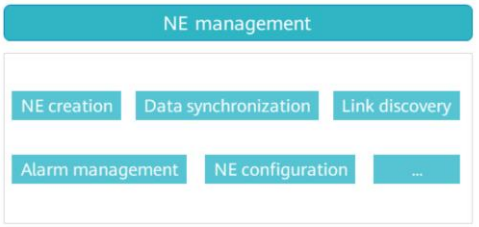
- NETCONF provides a mechanism for the NMS to communicate with network devices.
- To be specific, the network administrator can use this mechanism to add, modify, and delete the configurations of network devices as well as obtain the configurations and status of network devices.
- NETCONF is based on the Extensible Markup Language (XML).

LLDP

- Link Layer Discovery Protocol (LLDP) is defined in IEEE 802.1ab. It can identify the interfaces on devices and provide information about connections between devices. LLDP can also discover the paths between clients, switches, routers, application servers, and network servers.
- Devices use LLDP to collect their physical connection information and SNMP to report collected information to the controller, which then automatically discovers links.



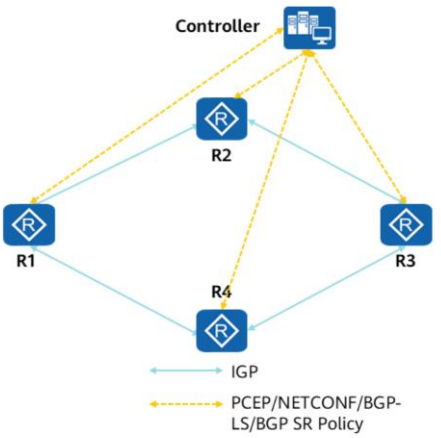
NE Management



- After synchronizing basic data from NEs, the controller can perform the following operations:
  - Creates NEs one by one or in batches.
  - Synchronizes NE data to the controller database.
  - Discovers links and displays physical connections. Manual link addition is also supported.
  - Configures NE information, such as interface IP addresses and router IDs.
  - Manages alarms and displays alarm information.

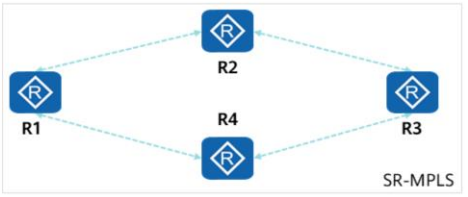
Contents

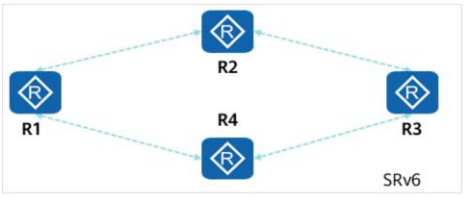
Tunnel Management Overview



- The controller uses the following protocols for tunnel management:
  - 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.
  - Path Computation Element Communication Protocol (PCEP): exchanges information between the controller and forwarders. For example, Huawei's CloudWAN solution uses PCEP to monitor tunnel status.
  - NETCONF: delivers tunnel configurations from the controller to forwarders.
  - BGP SR Policy: establishes BGP SR Policy peer relationships between the controller and PEs, so that the controller can deliver SR Policies through BGP peer relationships to instruct forwarders to forward packets. To reduce the number of peer relationships, you can deploy an RR and configure PEs and the controller to function as RR clients.

Topology Information Generation by Routers





- The key information that needs to be advertised for SR-MPLS tunnels is as follows:
  - SID: node SID, adjacency SID, and prefix SID
  - SRGB: a set of user-specified global labels reserved for SR-MPLS
  - Interface bandwidth information: physical bandwidth and reservable bandwidth
  - Interface delay information
- The key information that needs to be advertised for SRv6-based tunnels is as follows:
  - SID: END SID and END.X SID
  - Locator: 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. Other nodes on the network can locate this node through the locator route.
  - Interface bandwidth information: physical bandwidth and reservable bandwidth
  - Interface delay information

Topology Information Collection by BGP-LS

Drawbacks of direct topology information collection through an IGP by the controller

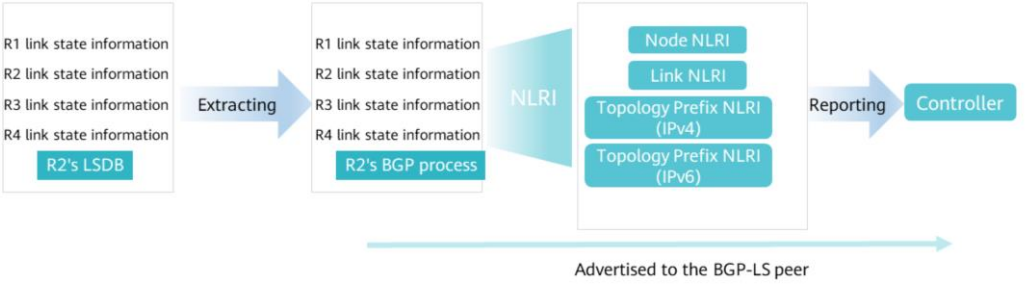
This has high requirements on the computing capability of the upper-layer controller. In addition, the controller must support the IGP and its algorithms. When cross-IGP domain topology information is collected, the upper-layer controller cannot obtain complete topology information and therefore cannot compute optimal E2E paths. When different routing protocols send topology information to the controller, the controller analyzes and processes the topology information in a complex manner.

BGP-LS

- The topology information discovered by IGPs is summarized and reported to an upper-layer controller through BGP. With powerful routing capabilities of BGP, BGP-LS has the following advantages:
  - Lowers the requirements on the controller's computing and IGP capabilities.
  - Facilitates path selection and computation on the controller by using BGP to summarize topology information in each process or AS and report the complete information directly to the controller.
  - Requires only one routing protocol (BGP) to report information about the entire network's topology to the controller.

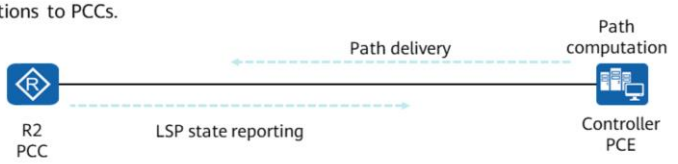
Topology Information Reporting to the Controller Through BGP-LS

- Each router maintains one or more LSDBs. The 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.



PCEP Overview

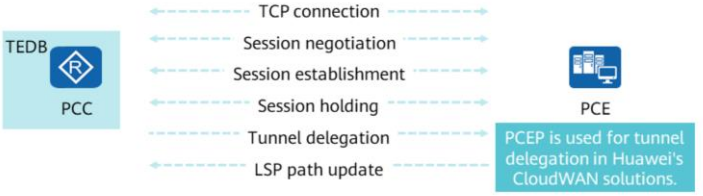
- Path Computation Element (PCE) can compute network paths or routing entities (components, applications, or network nodes) based on the network topology. A path computation client (PCC) is any client or application that requests path computation from a PCE. PCEP defines communication between the PCC and PCE and between two PCEs (RFC 5440).
- PCEP is a TCP-based protocol defined by the IETF PCE working group. It defines a group of messages and objects for PCEP session management, including TE LSP reporting and path delivery. PCEP provides a mechanism for the PCE to compute LSP paths for PCCs. In PCEP interaction, PCCs send LSP state reports to the PCE, and the PCE sends LSP path update instructions to PCCs.



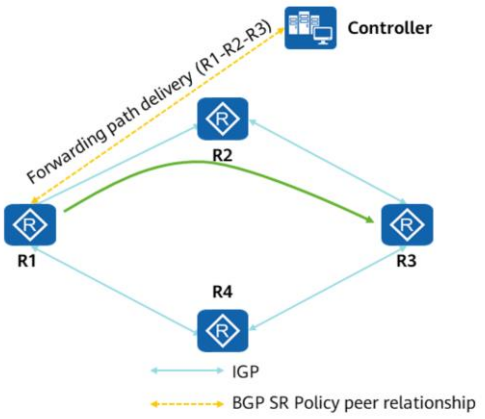
- For more information, visit https://datatracker.ietf.org/doc/rfc5440.

Tunnel Path Computation by PCEP

- A TCP-based PCEP session connects a PCC to a PCE. A PCC initiates a PCEP session and establishes a connection with a PCE. During session establishment, the PCC and PCE negotiate session capabilities. After a PCEP session is established, the PCC sends LSP state updates to the PCE. Upon receipt of LSP state updates, the PCE computes paths for LSPs that have path constraint changes as well as LSPs that do not have paths and instructs the PCC to update LSP paths. After a PCEP session is torn down, the TCP connection is immediately closed, and the PCC attempts to re-establish a PCEP session.
- PCEP provides the following functions:
  - Allows the PCC to delegate LSP control to the PCE. The PCE needs to synchronize LSP state information from the PCC.
  - Allows the PCE to compute paths based on delegated LSPs' attributes, including the bandwidth, explicit path, priority, and affinity attributes.
  - Transmits computed LSP attributes from the PCE to the PCC, which reports information about new LSPs to the PCE after updating paths.



BGP SR Policy Overview



- According to RFC 8402, an SR Policy is an ordered list of segments. In addition, it defines a framework for SR technologies used to provide functions such as computing/generating/maintaining the segment list and steering traffic.
- After an SR Policy peer relationship is established between the controller and PE, the controller delivers the computed path to the ingress of the SR Policy.
- For example, after R1 establishes an SR Policy peer relationship with the controller, if the path of the tunnel to be created is R1-R2-R3, the controller can deliver the path information together with the color attribute to R1 for subsequent data packet forwarding.

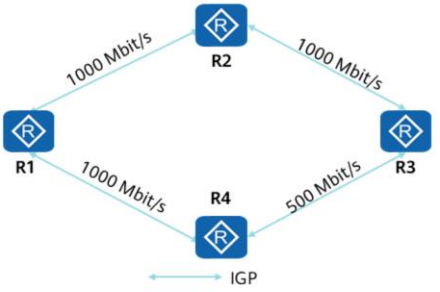
Tunnel Management Instance on the Controller

- To facilitate configuration, the controller provides the following functions:
  - Directly creates a bidirectional tunnel between the ingress and egress.
  - Allows you to configure tunnel templates and color templates to simplify the configuration of some parameters.
- In Huawei's CloudWAN solution:
  - The SR-MPLS TE tunnel configurations are delivered by NETCONF, and the tunnel status is reported by PCEP.
  - The SR-MPLS Policy configurations are delivered by BGP SR Policy, and the tunnel status is reported by BGP-LS.
  - The SRv6 Policy configurations are delivered by BGP SR Policy, and the tunnel status is reported by BGP-LS.

HUAWEI

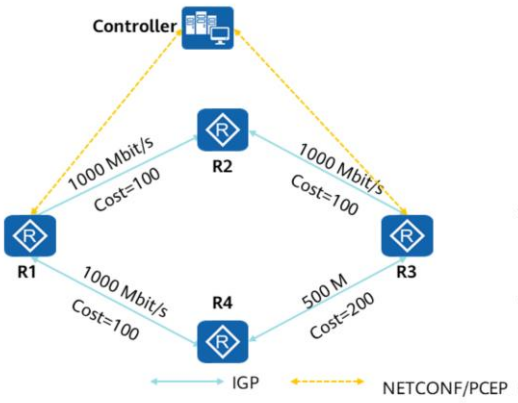
- The controller supports static configuration of various tunnels. Static tunnels are rarely used on live networks and are therefore not described in this course.

SR-MPLS TE Tunnel Creation (1)



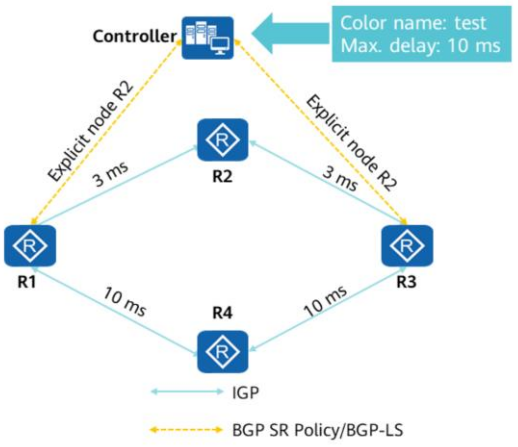
- Optional constraints:
  - Bandwidth constraint: ensures that the bandwidth configured for a service does not exceed the remaining bandwidth of the link that the service traverses.
  - PIR constraint: ensures that the peak bandwidth does not exceed the BC0 bandwidth of the link that the service traverses. 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.
  - Primary-backup path disjoint constraint: ensures that the PCE algorithm computes primary and backup paths that are as disjoint as possible in tunnel hot standby scenarios. The primary and backup paths are separated based on SRLG, node, and link information in sequence.

SR-MPLS TE Tunnel Creation (2)



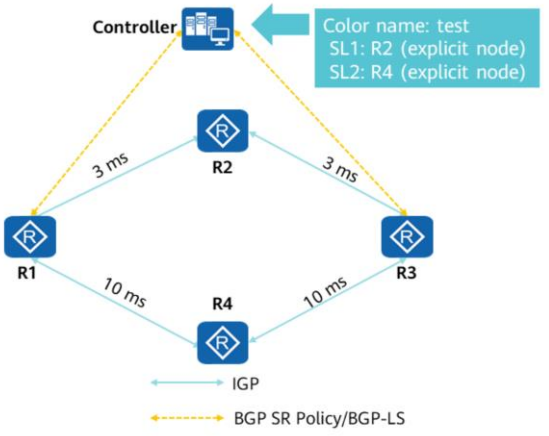
- With the path computation algorithm, the controller can provide the following path computation results if the constraints are met:
  - Least cost: path with the least cost among all paths that meet the constraint
  - Minimum delay: path with the minimum delay among all paths that meet the constraint
  - Bandwidth balancing: path with the most remaining bandwidth among all paths that meet the constraint and have the same cost
- In this example, the least cost constraint is selected, the ingress is R1, and the egress is R3. Given this, the forward path computed by the controller is R1-R2-R3, and the reverse path is R3-R2-R1.
- In Huawei's CloudWAN solution, the controller uses NETCONF to deliver tunnel configurations, and R1 and R3 report tunnel status through PCEP.
- To facilitate configuration and deployment, the controller provides templates. When creating a tunnel, you can skip the configuration of some parameters by selecting a template.

SR-MPLS Policy Creation



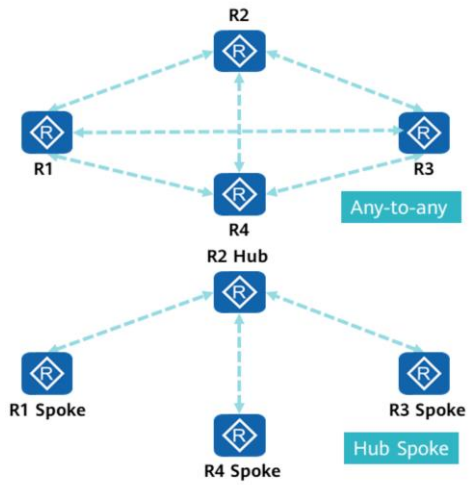
- The tunnel constraints and path computation algorithm of SR-MPLS Policies are similar to those of SR-MPLS TE tunnels.
- When creating an SR-MPLS Policy, you need to specify the color for each tunnel. In this example, the color name is test, the maximum delay is 10 ms, and the color ID is automatically generated by the controller.
- One SR-MPLS Policy can contain multiple candidate paths. Multiple segment lists can be configured for each candidate path to implement load balancing.
- In this example, the single-candidate-path single-segment-list mode is used. The forward and reverse traffic must pass through R2.
- In Huawei's CloudWAN solution, the SR-MPLS Policy configurations are delivered through BGP SR Policy, and R1 and R3 use BGP-LS to report tunnel status.

SRv6 Policy Tunnel Creation: Single Candidate Path and Multiple Segment Lists



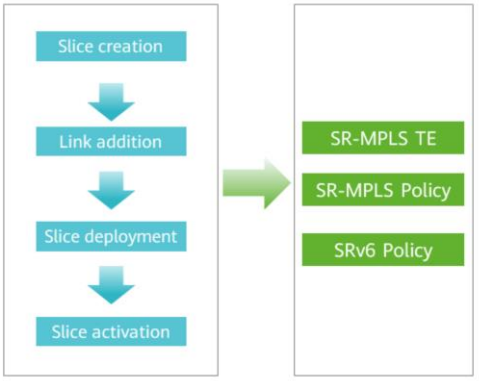
- The tunnel constraints and path computation algorithm of SRv6 Policies are similar to those of SR-MPLS TE tunnels.
- When creating an SRv6 Policy, you need to specify the color for each tunnel. In this example, the color name is test, and the color ID is automatically generated by the controller.
- An SRv6 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 SRv6 Policy, and the valid candidate path with the second highest preference functions as a backup path. Multiple segment lists can be configured for each candidate path to implement load balancing.
- In this example, the single-candidate-path multi-segment-list mode is used. R2 is the explicit node for path 1, and R4 is the explicit node for path 2.
- In Huawei's CloudWAN solution, the SR-MPLS Policy configurations are delivered through SRv6 Policy, and R1 and R3 use BGP-LS to report tunnel status.

SRv6 Policy Group



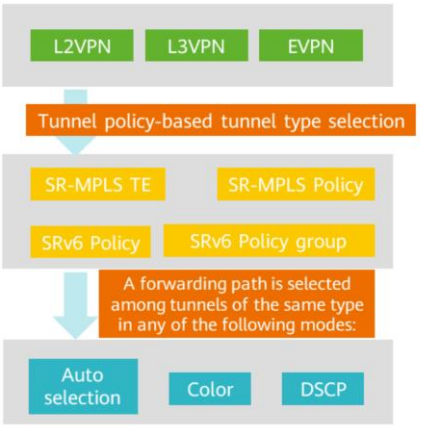
- SRv6 Policy group: allows a group of tunnels that share the same SLA requirements to be quickly established.
- Application scenario: Multiple SRv6 Policies need to be established in batches before service creation. The available tunnel types are as follows:
  - Any-to-any: indicates that a full-mesh network is established between NEs.
  - Hub-spoke: indicates that an NE is either a hub or spoke node. Communication between all spoke nodes must be implemented through the hub node.
- In Huawei's CloudWAN solution, the SRv6 Policy group is used in tunnel planning scenarios. Specifically, tunnel paths are computed in advance and delivered upon service request.

Network Slice-based Tunnel Creation



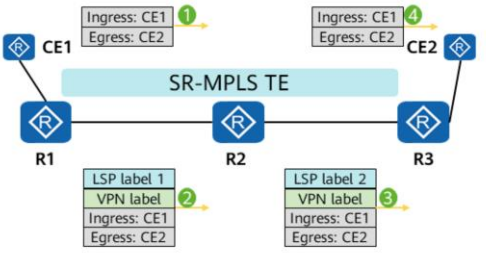
- Essentially, network slicing logically divides a physical network into multiple virtual E2E networks. As such, a specific network slice can then be selected when a tunnel is created.
- A network slice is deployed as follows:
  - Slice creation: Set the slice type.
  - Link addition: Select the links on which the slice is to be created.
  - Slice deployment: Deliver slice configurations, such as bandwidth, to devices for slice link generation.
  - Slice activation: Configure IP addresses for both ends of the slice link and configure data, such as IGP parameters, to achieve E2E connectivity at the network layer.
  - Tunnel creation: Select the specified slice during tunnel creation.

VPN and Tunnel Binding



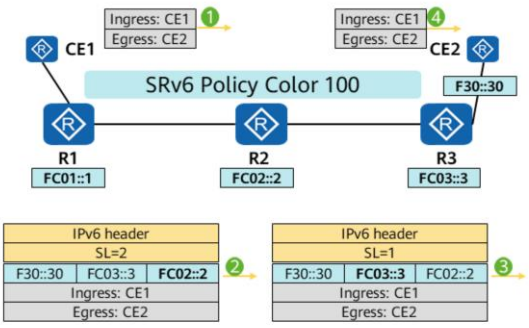
- The following types of VPNs are available in enterprise network scenarios:
  - L2VPN: The customer IP addresses are on the same network segment.
  - L3VPN: The customer IP addresses are on different network segments.
  - EVPN: The customer IP addresses are either on the same network segment or on different network segments.
- A tunnel policy is used by an application module to select tunnels for VPN services. There are two types of tunnel policies:
  - Tunnel type prioritizing policy: recurses a service to tunnels based on the tunnel type priority and the number of tunnels participating in load balancing (preferred mode).
  - Tunnel binding policy: binds a destination address to a tunnel, so that the traffic of VPN services referencing the policy will be transmitted over the bound tunnel.
- A VPN service first selects tunnels in the up state based on the tunnel policy, and then selects a forwarding path from tunnels that meet the requirements.

L3VPN over SR-MPLS TE: Automatic Tunnel Selection



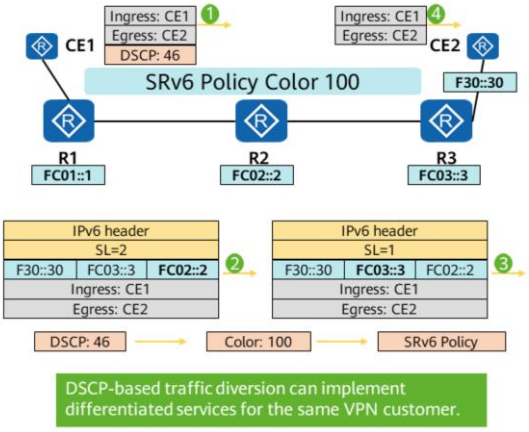
1. CE1 searches the routing table and forwards the packet to R1.
2. R1 searches the VPN routing table, adds the label allocated by R3 to the corresponding VPN instance to the packet, selects an SR-MPLS TE tunnel based on the tunnel policy, and adds a tunnel label to the packet.
3. After receiving the packet, R2 swaps the outer label and forwards the packet to R3.
4. After receiving the packet, R3 removes the outer label and forwards the packet to CE2 based on the inner label. The packet then reaches CE2.

L3VPN over SRv6 Policy: Color-based Traffic Diversion



1. CE1 searches the routing table and forwards the packet to R1.
2. R1 searches the VPN routing table and adds the END.DT4 SID (F30::30) allocated by R3 to the corresponding VPN instance to the packet. After finding that the corresponding route carries the color attribute of 100, R1 selects an SRv6 Policy tunnel with the color attribute being 100 based on the tunnel policy and adds a tunnel label to the packet. In this example, the forwarding path specified for the SRv6 Policy is R1-R2-R3.
3. After receiving the packet, R2 changes the destination IPv6 address to FC03::3, decrements the SL value by 1, and forwards the packet to R3.
4. After receiving the packet, R3 decrements the SL value by 1 and finds that FC30::30 is the SID locally assigned to CE2. R3 then removes the IPv6 header and sends the packet to CE2.

L3VPN over SRv6 Policy: DSCP-based Traffic Diversion



DSCP-based traffic diversion can implement differentiated services for the same VPN customer.

- DSCP-based traffic diversion is more flexible than color-based traffic diversion, as it allows traffic of the same VPN customer but with different priorities to enter different SRv6 Policies.
  1. CE1 searches the routing table and forwards the packet with DSCP 46 to R1.
  2. R1 searches the VPN routing table and adds an END.DT4 SID (F30::30) allocated by R3 to the corresponding VPN instance to the packet. R1 then maps DSCP 46 to color 100, selects an SRv6 Policy with color 100 based on the tunnel policy, and adds a tunnel label to the packet. In this example, the forwarding path specified for the SRv6 Policy is R1-R2-R3.
  3. After receiving the packet, R2 changes the destination IPv6 address to FC03::3, decrements the SL value by 1, and forwards the packet to R3.
  4. After receiving the packet, R3 decrements the SL value by 1 and finds that FC30::30 is the SID locally assigned to CE2. R3 then removes the IPv6 header and sends the packet to CE2.
- DSCP-based traffic diversion is more suitable for scenarios where there are multiple paths to the same destination and high-quality links need to be selected for services with high requirements.

HUAWEI

Quiz

1. (Multiple-answer question) Which of the following protocols are used for communication between the controller and NEs?( )
   A. PCEP
   B. SNMP
   C. NETCONF
   D. BGP-LS
2. (True or False) In Huawei's CloudWAN solution, PCEP is used to deliver path information computed by the controller to forwarders.( )

- ABCD
- False

Section Summary

- The prerequisite for the controller to perform network management is that the NEs and controller are reachable to each other at the IP layer. Basic interconnection parameters, such as IGP, SNMP, and NETCONF parameters, need to be configured on the controller and forwarders, so that the controller can manage these NEs.
- Compared with traditional CLI-based service management, controller-based service management has obvious advantages. The controller computes tunnel paths and uses NETCONF or BGP SR Policy to deliver information to forwarders. The controller uses NETCONF to deliver VPN information to forwarders, which select tunnels based on tunnel policies during data forwarding.
- In addition to NE management and service management, the controller also provides functions such as alarm management, which will be described in the following sections.

Contents

Analysis on Network Traffic Control Requirements for the WAN Controller

- As enterprises move towards digital transformation, the WAN calls for new traffic control technologies in the cloud era.

MPLS TE drawbacks	Traffic control requirements for the controller
• Inconsistent across domains ▫ IGP TE can flood only the topology information of the IGP area where it resides, and cannot compute or establish an E2E TE tunnel across IGP areas or ASs. ▫ The traditional path computation algorithm can only be determined by the ingress of a tunnel. As a result, collaborative management cannot be achieved, and network bandwidth resources cannot be optimized in a coordinated manner. • Resource preemption ▫ When resources are insufficient, high-priority tunnels preempt the resources of low-priority tunnels. • Difficult planning ▫ When a network is large, manual path planning is difficult.	• Topology processing for centralized path computation ▫ Collects topology information from multiple areas, combines and processes collected topology information, and computes paths based on the global topology across domains. • Multi-constraint-based path computation ▫ Provides the ability to optimize traffic based on multiple constraints, such as path locking, candidate path, and affinity attributes. • What-you-see-is-what-you-get ▫ Provides multiple path computation policies and intelligently plans service paths. • Computation of multiple types of tunnel paths ▫ Computes paths for multiple types of tunnels, such as MPLS TE tunnels, SR-MPLS TE tunnels, SR-MPLS Policies, and SRv6 Policies.

Traffic Control Process Overview of the WAN Controller

- The traffic control technology of the WAN controller provides E2E optimal path computation and optimization services. It allows you to perform centralized configuration and management of network topology information and tunnel constraints to optimize network bandwidth utilization and maximize the utilization of network resources.
- Traffic control 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

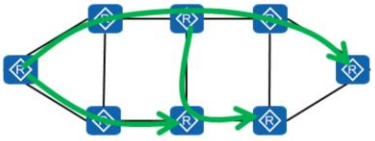
The controller computes paths based on the optimization target. Optimization targets include:

- Least path cost
- Shortest path delay
- ...

3. Optimization result delivery

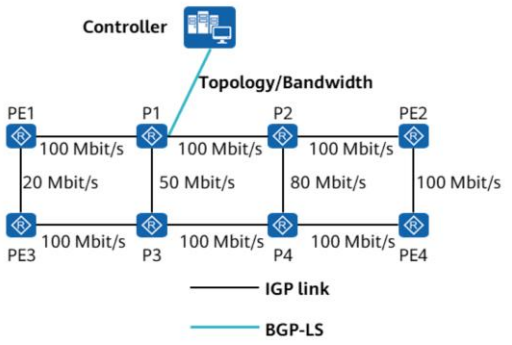
The controller delivers the computation result to network devices in any of the following modes:

- NETCONF
- BGP SR Policy
- ...



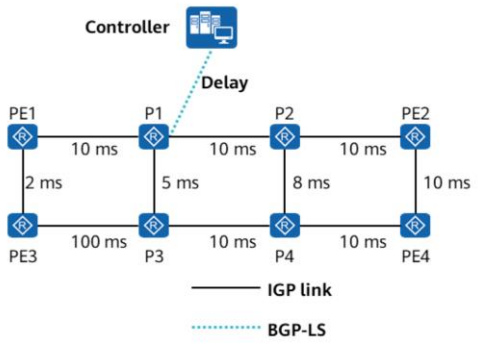
Contents

Topology Information and Interface Bandwidth Collection



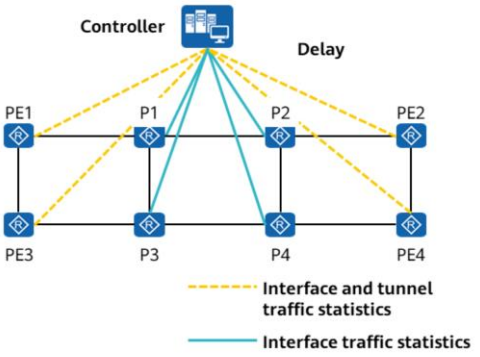
- The controller uses BGP-LS to collect topology and interface bandwidth information from routers.
  - Forwarders run an IGP (such as IS-IS) to advertise topology and bandwidth information.
  - One of the forwarders, for example, P1, establishes a BGP-LS peer relationship with the controller.
  - The controller processes topology information based on the node and link information reported by BGP-LS.

Link Delay Collection



- Link delay information can be collected in either static or dynamic mode.
  - In static mode, link delay is obtained through other means and then manually entered on the NCE-IP UI. Alternatively, link delay can be configured using commands on routers and then flooded in the IGP domain.
  - In dynamic mode, a router detects the link delay in real time through a detection protocol and reports the link delay to the controller.
- Compared with the static mode, the dynamic mode can reflect the link delay status in real time and is more suitable for scenarios where delay needs to be guaranteed. This section describes dynamic delay collection.
- The dynamic 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 optimization, the controller needs to collect interface and tunnel traffic statistics in real time:
  - SNMP: reports information such as the interface traffic rate and bandwidth usage within minutes.
  - Telemetry: reports interface traffic information and tunnel traffic statistics within seconds.
- Ps need to report interface traffic statistics, and PEs need to report both interface and tunnel traffic statistics.

Contents

Network Optimization Constraint Parameters and Optimization Policies

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.

- Optimization policies:
  - Least cost: selects a path with the least cost among all paths that meet the specified constraints.
  - Minimum delay: selects a path with the minimum delay among all paths that meet the specified constraints.
  - Availability+least cost: selects a path with the maximum link availability and least cost among all paths that meet the specified constraints.
  - Availability+minimum delay: selects a path with the maximum link availability and minimum delay among all paths that meet the specified constraints.

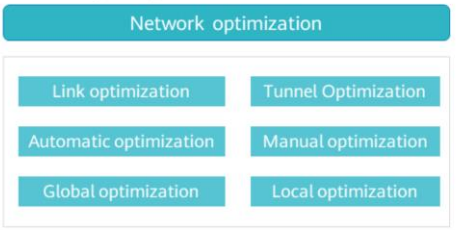
$$\text{Link availability} = \frac{\text{Total statistical time} - \text{Total link fault time}}{\text{Total statistical time}}$$

HUAWEI

- Bandwidth-balancing path: indicates the path with more remaining bandwidth among all paths that meet constraints and have the same cost.
- Maximum availability path: indicates the path with the maximum availability among all paths that meet the constraint.

Overview of Controller-based Optimization

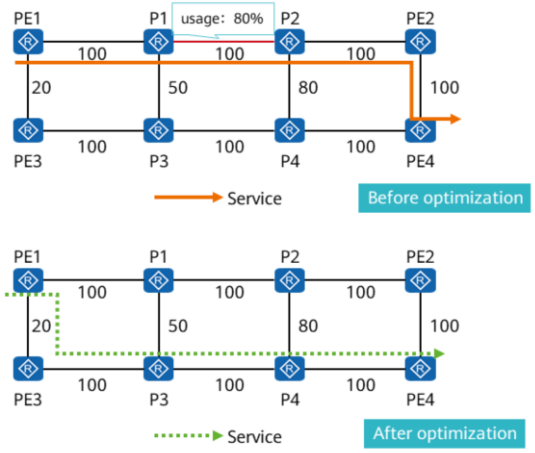
- Generally speaking, optimization is to optimize service paths (LSPs). LSPs are logical paths and can be considered as tunnels, whereas links are physical paths. One link can carry multiple tunnels, and each tunnel (LSP) corresponds to a service. Therefore, service paths can be changed based on either 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.
- The optimization solution supports both automatic and manual optimization:
  - 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.
- According to the optimization scope, optimization can be classified into the following two types:
  - Local optimization: optimizes specified tunnels or links.
  - Global optimization: optimizes all tunnels or links.



HUAWEI

- The link availability is calculated based on the ratio of the time when the link is normal to the total statistical time.

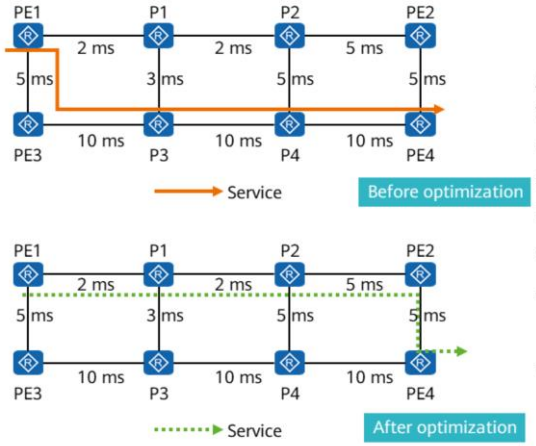
Example for Optimizing a Single Tunnel (Based on the Least Cost Policy)



An SRv6 Policy tunnel is established between PE1 and PE4. As services increase, the bandwidth usage of the link between P1 and P2 reaches 80%, which exceeds the preset optimization threshold 60%. In this case, some service traffic needs to be diverted to other links.

- Before optimization, the path is PE1-P1-P2-PE2-PE4.
- Path optimization is performed based on the least cost policy.
- After optimization, the path is PE1-PE3-P3-P4-PE4.

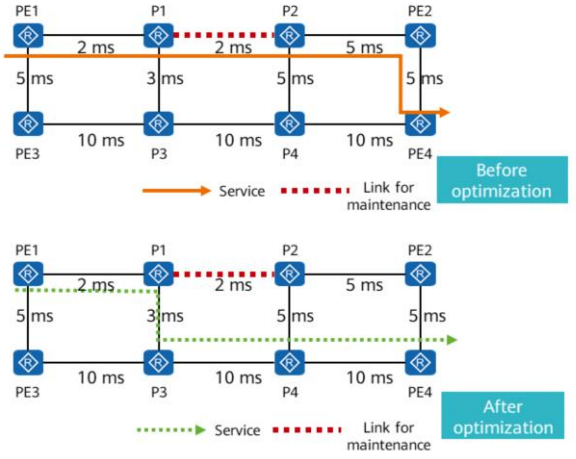
Example for Optimizing a Single Tunnel (Based on the Minimum Delay Policy)



An SRv6 Policy tunnel is established between PE1 and PE4. The service monitoring result shows that the delay of the service carried by the tunnel is too long. Therefore, the service path needs to be adjusted to reduce the network delay.

- Before optimization, the path is PE1-PE3-P3-P4-PE4.
- Path optimization is performed based on the minimum delay policy.
- After optimization, the path is PE1-P1-P2-PE2-PE4.

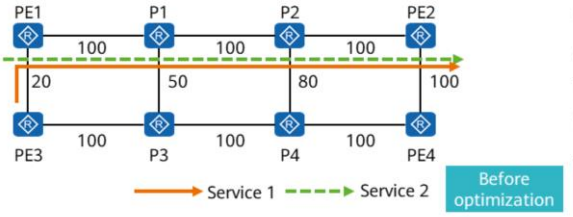
Example for Optimizing a Single Link (Maintenance State Setting)

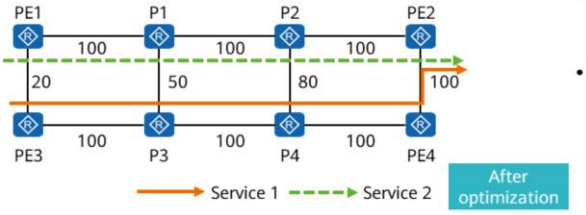


An SRv6 Policy tunnel is established between PE1 and PE4. The optical module connecting P1 to P2 is unstable and needs to be replaced.

- Before optimization, the path is PE1-P1-P2-PE2-PE4.
- Set the link between P1 and P2 to the maintenance state to prevent the replacement from affecting services.
- After optimization, the path is PE1-PE3-P3-P4-PE4 (minimum delay).

Controller-based Multi-tunnel Optimization from the Global Perspective



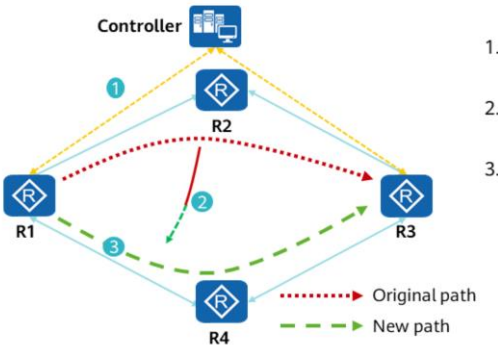


There are two tunnels between PE1 and PE2 that carry different services. As the service traffic increases, the bandwidth usage of the tunnel exceeds the threshold.

- Before traffic optimization, two service flows pass through the same path PE1-PE1-P2-PE2
- Service 1 has a higher priority and is optimized from the global perspective by the controller.
- After optimization, the path of service 1 is PE1-P1-P2-PE2, and that of service 2 is PE3-P3-P4-PE4-PE2.

MBB-enabled Zero Packet Loss During Tunnel Path Switching

- The make-before-break (MBB) mechanism ensures that the original path is deleted after a new forwarding path is established. During this period, traffic is still forwarded along the original path. The system deletes the original path only after the switch-delay timer expires. This prevents traffic interruption caused by path switching.



1. The controller delivers the new path (R1-R4-R3) to the forwarder R1.
2. R1 starts to establish a new path and starts the switch-delay timer.
3. After the switch-delay timer expires, R1 forwards traffic along the new path.

Quiz

1. (Multiple-answer question) Which of the following traffic optimization policies are supported by the controller? ( )
   A. Least cost
   B. Minimum delay
   C. Availability+least cost
   D. Availability+minimum delay

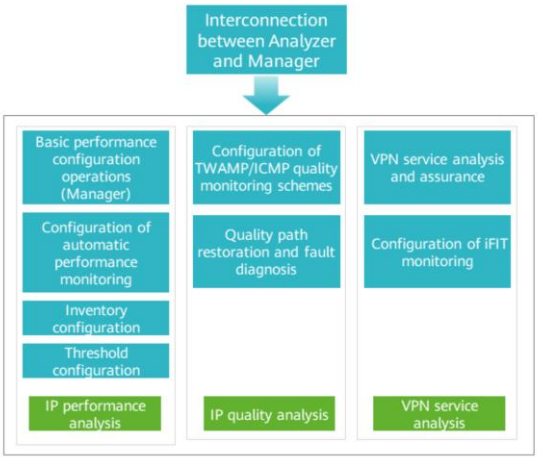
- ABCD

Section Summary

- A tunnel is established based on multiple constraints, such as the explicit path and affinity attributes. However, as the number of services increases, situations such as tunnel congestion, excessive delay, and high bandwidth usage may arise. In this case, you can re-optimize the entire traffic path through the controller from a global perspective.
- Tunnels still need to meet basic tunnel constraints during path optimization. When the network is large, manual tunnel optimization is difficult, highlighting the advantage of automatic optimization by the controller.

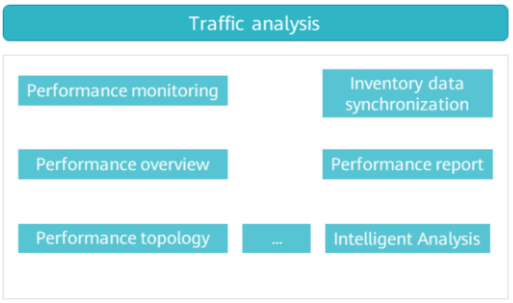
Contents

Classification and Process of Network Performance Analysis



- Network performance analysis is classified into the following types:
  - IP performance analysis
    - Network traffic analysis: processes and analyzes the collected network traffic data and displays traffic data in multiple modes, such as reports, dashboards, topologies, and maps, for data visualization.
    - Network quality analysis: collects performance data such as link bandwidth, delay, jitter, and packet loss, and displays link quality for quick locating of links with deteriorated quality.
  - VPN service analysis: detects service quality through iFIT in-band measurement in real time and quickly identifies private line service faults, so that the faults can be located and rectified in a timely manner.

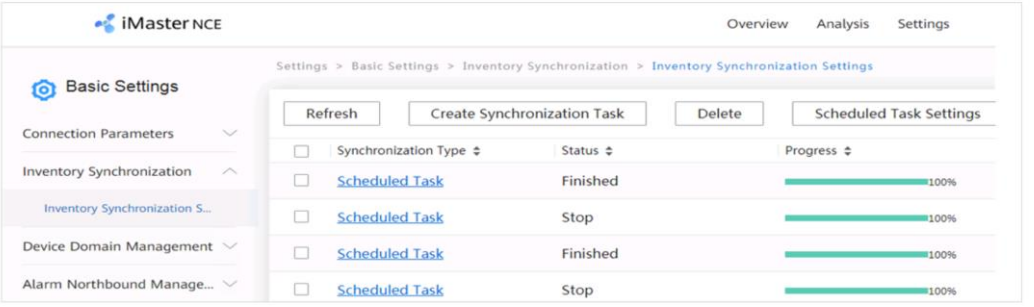
Overview of Controller Traffic Analysis Functions



- Performance monitoring: Performance monitoring instances are the basis of traffic analysis and provide data sources for analysis.
- Inventory synchronization: This function synchronizes resource information, such as NE, board, port, link, and ring information, from Manager to Analyzer. You can create a synchronization task and set a period to periodically synchronize inventory information.
- Performance overview: The controller displays performance data from different dimensions on different home pages for management.
- Performance report: The controller provides various performance report-related functions, such as displaying performance reports by time range, sorting report data in ascending or descending order, filtering report data, customizing report headings, and customizing the number of data records to be displayed on each page of a report.
- Performance topology: The IP performance topology reflects the networking and running status of devices. You can monitor the running status of the entire network in real time by viewing the color and status of device icons in the topology view.
- Intelligent analysis: The controller predicts the future resource status based on the current resource status.

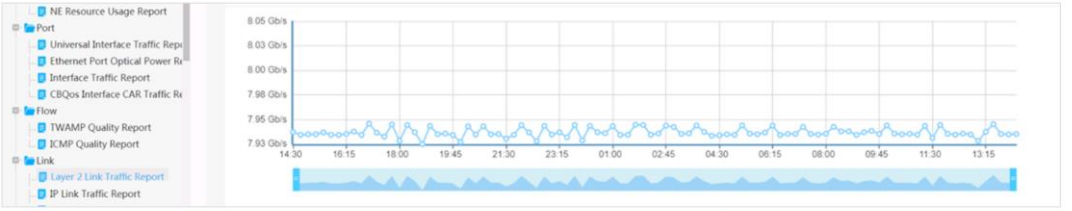
Analyzer obtains inventory data from Manager.

- If you want to use the performance topology and traffic monitoring functions, you need to synchronize the inventory data of related resources.
- After obtaining all inventory data, Analyzer performs background processing on the data and displays task details on the foreground.



Performance Reports

- The NCE solution provides various types of reports to help users learn performance data, such as network-wide resource status and port, link, and ring traffic. Users can view reports to evaluate the current network status. This effectively supports O&M and capacity expansion.

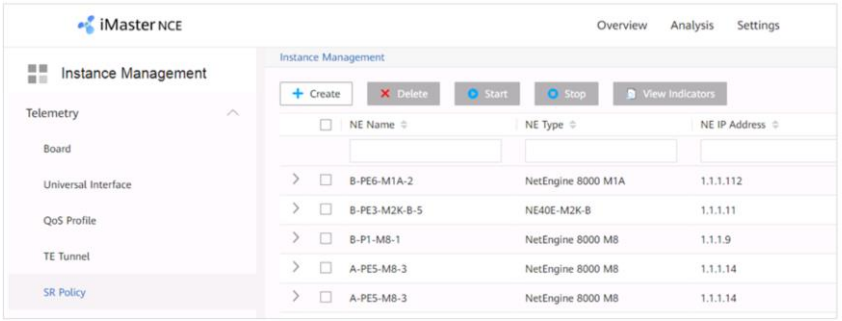


Layer 2 link traffic report

Traffic Analysis | Quality Analysis | VPN Service Analysis

Telemetry Deployment

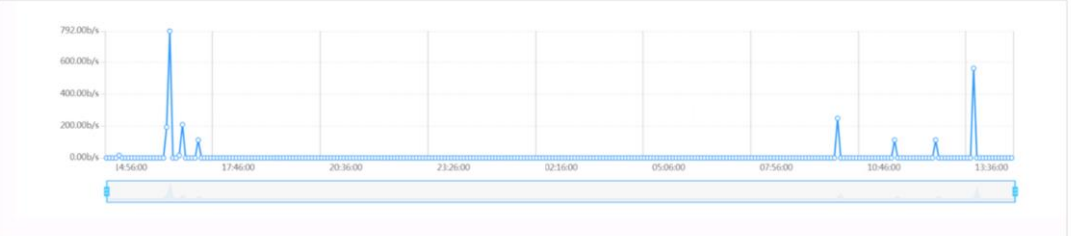
- Telemetry, which has advantages such as large application scope and short collection period, is suitable for large-scale network O&M.
- Static subscription: Devices function as gRPC/UDP clients and periodically report data based on the collection subscription configuration. Analyzer functions as the gRPC/UDP server, receives and processes data reported by devices, and provides an entry for users to query historical performance data.
- Dynamic subscription: Devices function as gRPC servers, and Analyzer functions as a gRPC client to proactively query the real-time performance data of devices. The devices return real-time performance data based on the query configuration. Analyzer provides a unified entry for users to query real-time performance data.



HUAWEI

- SNMP supports real-time performance data collection at an interval of 5s, 10s, or 15s, but not 1s.

Telemetry Traffic Statistics Reporting



Historical traffic data display

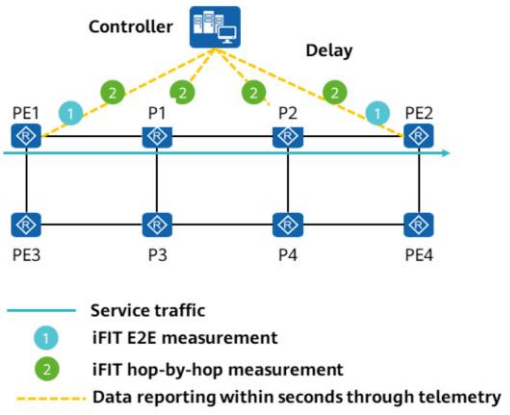
TWAMP Deployment

Description of TWAMP indicators

Indicator	Description
Two-way packet loss rate	Rate of lost packets to transmitted packets on a link within a period.
Two-way delay	Duration from when the source end sends a packet to the destination end to when the source end receives an acknowledgment packet from the destination end.
Two-way jitter	Difference between the two-way delays of two adjacent packets with the same destination.
One-way packet loss rate	Rate of lost packets to transmitted packets on a link within a period in the source-to-destination or destination-to-source direction.
One-way delay	Duration from when the source end sends a packet to when the destination end receives the packet or from when the destination end sends an acknowledgment packet to when the source end receives the acknowledgment packet.
One-way jitter	Delay variation from the source end to the destination end or from the destination end to the source end.
Number of error packets	Number of error packets during data transmission.

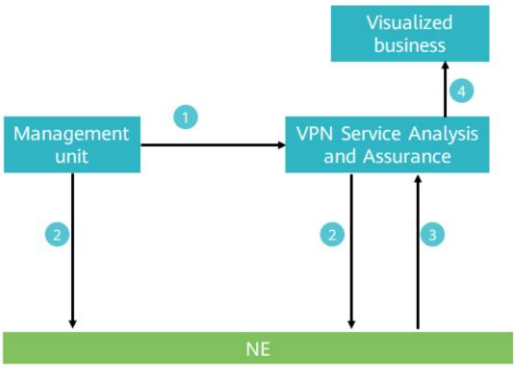
- TWAMP threshold configuration: During a TWAMP test, you need to set different levels of alarm thresholds for different indicators of different resources. If the indicator value generated by the network or service exceeds the set threshold, an alarm is triggered for quick fault locating.
- TWAMP test task deployment: TWAMP test tasks are deployed to monitor E2E network service quality and help O&M personnel quickly demarcate and locate faults.
- TWAMP performance reports: By viewing TWAMP performance reports, you can intuitively learn the information and quality of each service and discover services that require special attention.

VPN Service Analysis and Assurance



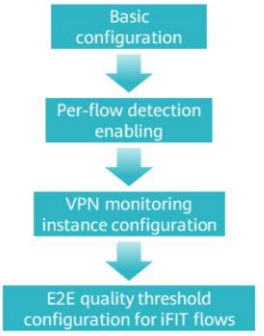
- The controller uses iFIT to implement proactive service SLA visualization and awareness and can proactively and quickly demarcate and locate faults, ensuring user experience.
  - E2E mode: After the ingress (for example, PE1) of a VPN service is specified for iFIT E2E measurement initiation, PE1 inserts the iFIT E2E measurement header into the VPN service flow destined for PE2. PE1 and PE2 report the detection results respectively, and NCE computes and displays the E2E SLA.
  - Hop-by-hop mode: For a VPN service with deteriorated quality, the system automatically triggers iFIT hop-by-hop measurement on the service ingress (for example, PE1). PE1 inserts the iFIT hop-by-hop measurement header into the VPN service flow destined for PE2, devices report the measurement results hop by hop, and NCE displays the nodes/links that cause poor service quality.

Principles of VPN Service Analysis and Assurance



- Analyzer obtains inventory data, including NEs, ports, links, and tunnels, from Manager.
- Analyzer delivers the iFIT monitoring configuration and telemetry subscription configuration to NEs.
- The unified southbound collection module collects NE iFIT performance data (including flow-related data) through telemetry and uploads the data to Analyzer.
- Analyzer performs aggregation and calculation on performance data and displays the data to users from multiple dimensions, including topologies and reports.

VPN Service Analysis and Assurance Deployment



- Basic configuration: Configure 1588 clock synchronization or NTP time synchronization (only packet loss detection is supported), interconnect Manager and Analyzer, synchronize inventory data, and configure port performance indicator thresholds.
- Per-flow detection enabling: Select the device to be monitored and enable the detection function.
- VPN monitoring instance configuration: Select the VPN to be monitored.
- E2E quality threshold configuration for iFIT flows: Configure a threshold for iFIT flows. When the indicator of an iFIT flow exceeds the threshold, hop-by-hop flow measurement will be triggered to quickly identify the abnormal indicator and faulty point.

Display of iFIT-measured Actual VPN Service Quality

- After iFIT is deployed, the iMaster NCE-IP visualized analysis platform can clearly display the E2E delay and packet loss rate.

HUAWEI

Quiz

1. (True or False) TWAMP supports both one-way and two-way packet loss detection.( )
2. (Single-answer question) When VPN service analysis and assurance are deployed, which of the following modes is used by devices to report detection data?( )
   A. Telemetry
   B. SNMP
   C. IFIG
   D. BGP-LS

- False
- A

Summary

- Huawei's CloudWAN solution leads WANs into the intelligent cloud-network era.
- NetEngine series high-end routers are based on Huawei-developed chips, an advanced SDN architecture, and a mature VRP software platform. They support smooth evolution of WANs.
- iMaster NCE effectively bridges the gap between physical networks and business intents, and implements centralized management, control, and analysis of entire networks. It enables resource cloudification, full lifecycle automation, and data analytics-driven intelligent closed-loop management according to business and service intents. Additionally, it has open network APIs that support rapid integration with IT systems.
- This course describes the fundamentals and procedures of network management, network traffic control, and network performance analysis based on iMaster NCE-IP. For more controller-related operations, see the related lab guide.

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.

Huawei CloudWAN Solution O&M and Troubleshooting

HUAWEI

Foreword

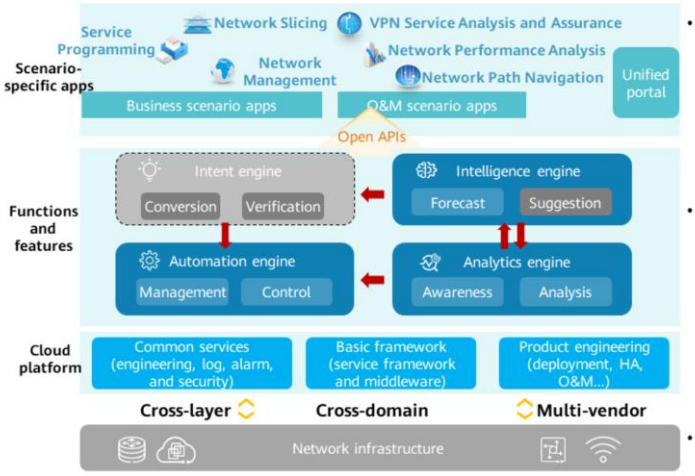
- In the intelligence era, especially with the rise of 5G and cloud services, the bearer WAN faces structural challenges. With the rapid development of new services such as 4K/8K, ultra-HD video, and VR, data traffic on the entire bearer WAN increases 10-fold every 5 years. In the 5G era, connectivity of everything will bring even faster traffic growth. Increasing network complexity leads to low O&M efficiency and OPEX increase year after year. In addition, passive customer experience management is in urgent need of improvement in the experience-centric era.
- These problems can be systematically solved only when a digital twin driven by the business logic and service intents of users is built over the physical network.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the iMaster NCE-IP architecture.
  - Describe the basic troubleshooting process.
  - Use the controller for routine O&M.

Contents

Huawei CloudWAN Solution Controller



- iMaster NCE-IP is an innovative network cloudification engine of Huawei. Positioned as the brain of the future cloud-based network, it integrates functions such as network management, service control, and network analysis. It is the core enabling system that implements network resource pooling, network connection automation and self-optimization, and O&M automation.
- iMaster NCE-IP is located at the management and control layer of the cloud-based network. On the one hand, it manages and controls the lower-layer network composed of IP, transmission, and access devices. On the other hand, it opens capabilities and enables services for the upper layer, supporting interconnection or integration with the OSS, BSS, and service orchestrators as well as quick, customized development of the application layer.
- iMaster NCE-IP aims to build an intent-driven network (IDN) that is first automated, then self-adaptive, and finally autonomous.

iMaster NCE-IP Features (1)

Converged management and control, supporting smooth network evolution

- Integrates traditional NMS and SDN controller functions.
- Achieves unified management and control of SDN and non-SDN networks, leverages SDN network automation, maximizes legacy network values, and reduces technical difficulties and risks of network evolution.

Cloud-based deployment

- Provides a unified cloud platform, user portal, and API gateway as well as unified installation, deployment, and upgrade, greatly simplifying O&M.
- Adopts a microservice-based architecture, enabling scenario-specific on-demand deployment.

iMaster NCE-IP is a full-lifecycle network automation platform that integrates management, control, and analysis. It focuses on service automation, O&M automation, and network autonomy to support bearer WAN cloudification and digitalization.

iMaster NCE-IP Features (2)

Open programmability

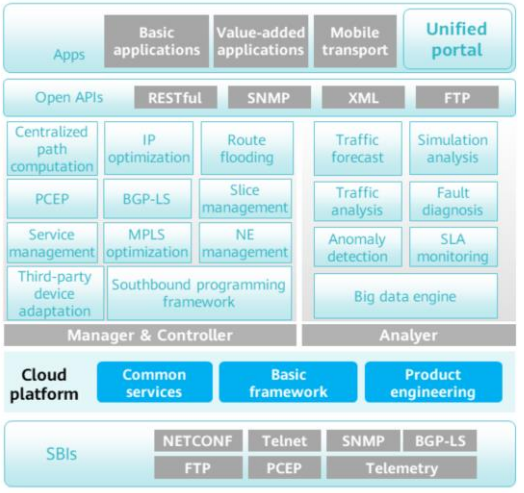
- Provides standard RESTful APIs to connect to external systems.
- Provides a DevOps-based tool platform to quickly customize innovative services based on flexible business scenarios and network technologies.

Intelligent O&M

Uses the big data algorithm for intelligent analysis and proactive prediction based on a massive amount of data.

- Collects network-wide data, and uses the big data platform to perform panoramic and in-depth analysis on the quality and traffic data of the entire network, achieving intelligent analysis and predictive maintenance.
- Continuously verifies network design and configuration to ensure that the network accurately implements user intents and runs stably.

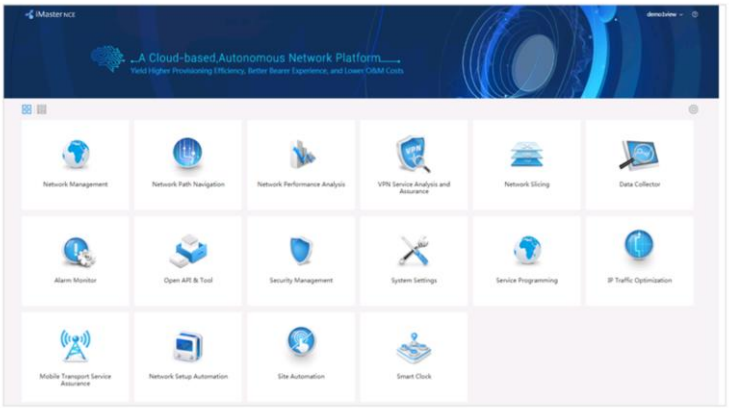
iMaster NCE-IP Software Architecture



- iMaster NCE-IP is a cloud-based system that uses a service-oriented software architecture. It is deployed on a virtualization platform and can be scaled flexibly. iMaster NCE-IP provides three logical modules (Manager, Controller, and Analyzer) and various scenario-specific apps based on a cloud-based platform to achieve flexible and modular deployment based on customer requirements.

Unified Portal

- iMaster NCE-IP provides a unified user portal based on the unified cloud platform. This portal provides access to various apps, and these apps call the interfaces of underlying components to implement various functions.



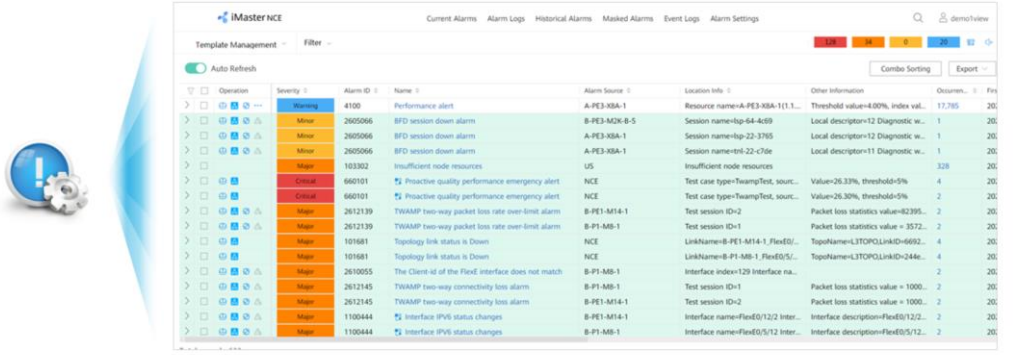
HUAWEI

- Network Management: manages all Huawei transmission, IP, and access devices in a unified manner, implements visualized service management and E2E provisioning of cross-domain services, detects network faults in a timely manner, and effectively locates faults across domains to meet the requirements of network informatization development.
- Alarm Monitor: allows network maintenance personnel to learn the operating status of the network in a timely manner through centralized alarm monitoring. Alarm management includes monitoring alarms, handling alarms, setting alarm monitoring/processing rules, and remotely sending alarm notifications.
- Data Collection: provides data collection monitoring and management capabilities, covering dimensions such as collection indicators, collection instances, and collection tasks. This helps users effectively monitor the running status of the collector.
- Network Path Navigation: manages network paths, computes and optimizes E2E forwarding paths, and centrally configures and manages network topology information to optimize network resource utilization.
- Network Performance Analysis: serves as a center for network experience awareness, decision making, and big data analytics. It provides visualized management to help users monitor network traffic and quality in real time, detect the network change trend, and optimize the network, improving O&M efficiency and reducing OPEX.

Contents

Alarm Monitoring Overview

- Alarm Management allows you to monitor and manage alarms and events reported by the system or managed objects. Alarm Management also provides a variety of monitoring and processing rules to meet requirements in different monitoring and processing scenarios. In this way, network faults can be efficiently monitored, located, and rectified.



Alarm Categories

Category	Description
Current alarms	Current alarms include unacknowledged and uncleared alarms, acknowledged and uncleared alarms, and unacknowledged and cleared alarms.
Historical alarms	Historical alarms refer to acknowledged and cleared alarms.
Masked alarms	You can mask alarms that you do not need to pay attention to. Masked alarms are moved to the Masked Alarms list and will not be displayed in the Current Alarms list even if they are generated later.
Event alarms	Event alarms are of the lowest severity. They are used to inform users of event occurrence. Unlike fault alarms, event alarms do not need to be handled.

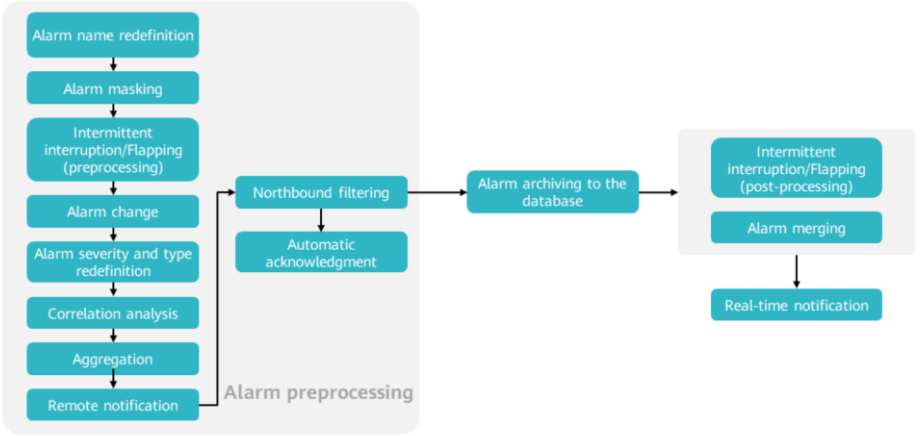
Alarm Severities and Status

Alarm Severity	Default Color	Description	Handling Policy
Critical		A service-affecting fault has occurred, and measures must be taken immediately.	Handle critical alarms immediately. Otherwise, services may be interrupted or the system may break down.
Major		A service-affecting fault has occurred. If the fault is not rectified, it will lead to a severe result.	Handle major alarms in a timely manner. Otherwise, important services will be affected.
Minor		Trivial impact has been caused on services, but corrective measures need to be taken to prevent more severe faults.	Identify causes and eliminate risks.
Suggestion		A potential or imminent fault is detected, but services are not affected currently.	Handle suggestion alarms based on network and NE running status.

Status Type	Alarm Status	Description
Acknowledgment status	Acknowledged or unacknowledged	The initial acknowledgment status is unacknowledged. After users acknowledge an alarm because they plan to handle it, the alarm status changes to acknowledged. After users unacknowledge an alarm, the alarm status changes to unacknowledged. Users can configure rules to automatically acknowledge alarms.
Clearance status	Cleared or uncleared	The initial clearance status is uncleared. When the fault that causes an alarm is rectified, a clearance notification is automatically reported to Alarm Management, and the alarm status changes to cleared. For some alarms, clearance notifications cannot be automatically reported to the alarm management system. They must be manually cleared after the corresponding faults are rectified. The background color of cleared alarms is green.

Internal Alarm Processing Flowchart

- Internal alarm processing includes alarm masking, correlation analysis, and severity redefinition.



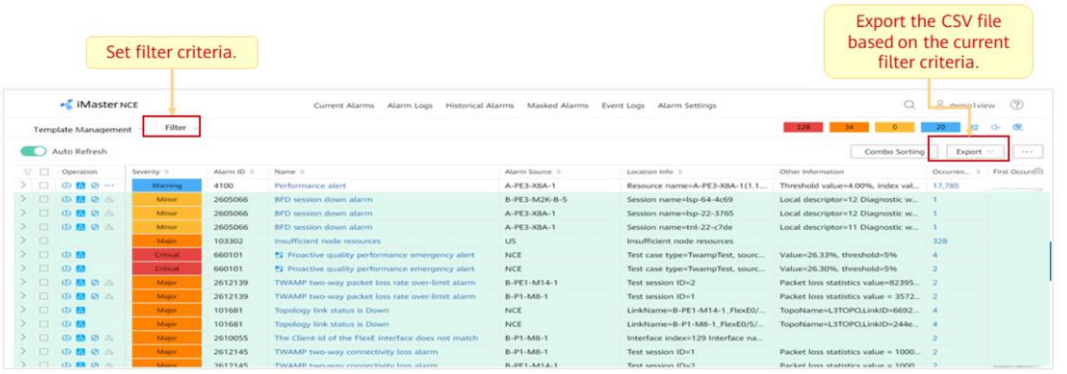
Internal Alarm Processing Description (1)

Operation	Description
Alarm name redefinition	After receiving alarms, Alarm Management matches these alarms against name redefinition rules and modifies the names of alarms that meet these rules.
Alarm masking	Alarm Management discards alarms that meet the masking rules (such alarms are not saved to the database) or records these alarms in the masked alarm list. Alarm Management does not preprocess such alarms.
Intermittent interruption/Flapping (preprocessing)	Alarm Management records alarms that meet the intermittent interruption/flapping rules in the intermittent interruption/flapping alarm list. Alarm Management does not preprocess such alarms.
Alarm change	Alarm Management updates the current alarm information based on reported alarm changes, such as alarm clearing or severity modification.
Alarm severity and type redefinition	Alarm Management redefines alarms that meet severity and type redefinition rules.
Correlation analysis	Alarm Management marks alarms that meet the correlation rules as root/correlative alarms and processes the root/correlative alarms based on actions in the rules.
Aggregation	Alarm Management aggregates alarms that meet aggregation rules based on the aggregation action.
Remote notification	When an alarm that meets the remote notification rules is reported, Alarm Management sends an email or SMS message to notify the O&M personnel.
Northbound filtering	Alarm Management sends alarms that meet the reporting conditions to the upper-layer NMS.
Automatic acknowledgment	Alarm Management automatically acknowledges the alarms that meet automatic acknowledgment rules. Automatically acknowledged alarms are then recorded in the historical alarm list.

Internal Alarm Processing Description (2)

Operation	Description
Alarm archiving to the database	Alarm Management records alarms processed through the preceding steps in the database. Alarms masked or moved to the historical alarm list during alarm preprocessing do not undergo post-processing.
Intermittent interruption/Flapping (post-processing)	Alarm Management analyzes alarms in the intermittent interruption/flapping alarm list and processes alarms that meet the intermittent interruption/flapping policy.
Alarm merging	Alarm Management merges alarms that meet the merging conditions.
Real-time notification	Alarm Management updates alarm information in real time on the alarm page.

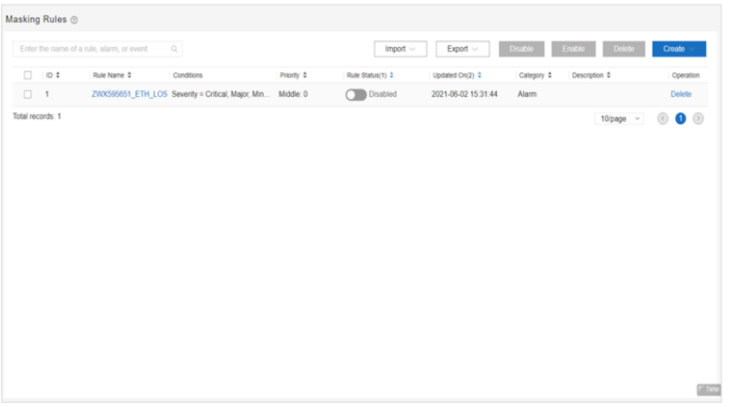
Alarm Page Overview



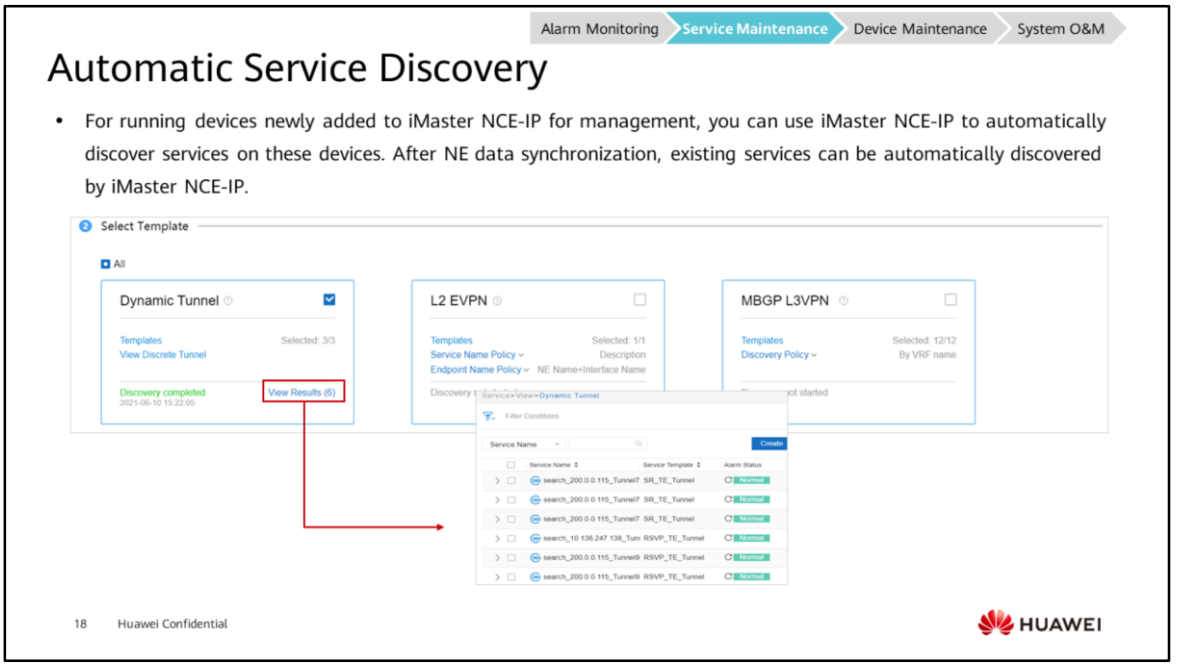
- Used to clear the current alarm
- Used to acknowledge the current alarm
- Used to set masking rules

Setting Masking Rules

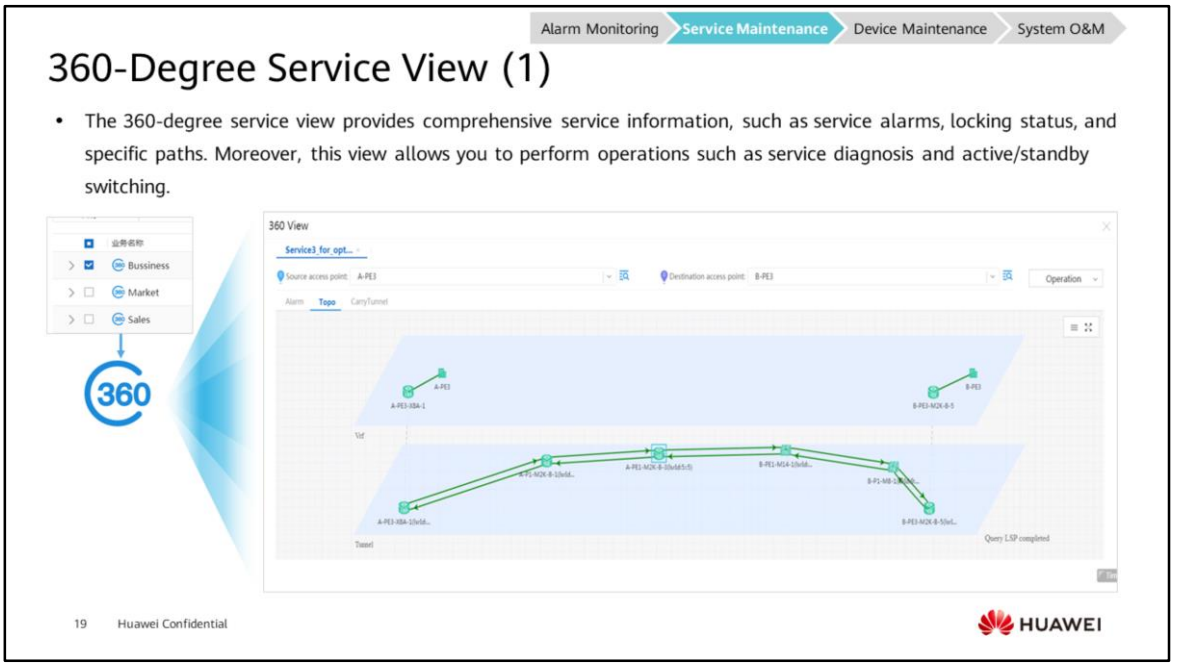
- You can create masking rules to prevent events and alarms that do not require attention from appearing in event logs and current/historical alarms, respectively.



HUAWEI



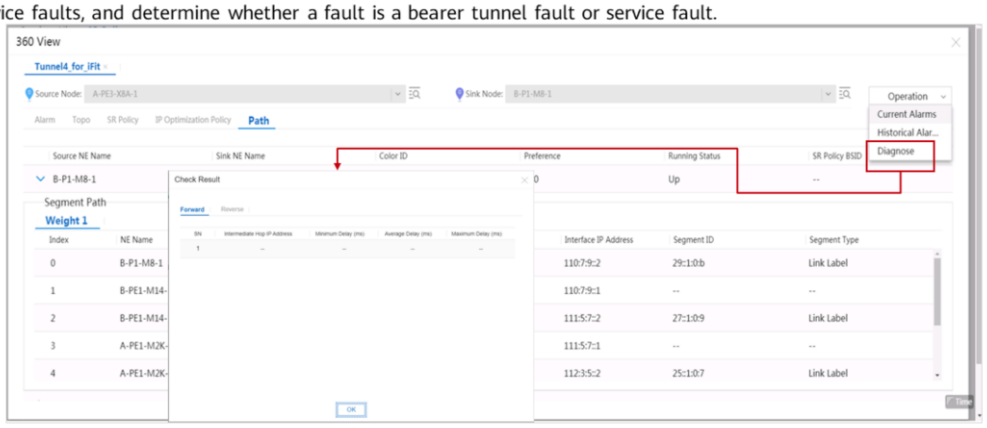
- Open the Network Management app and choose Service > Auto Service Discovery from the main menu. On the page that is displayed, expand Search Service by Policy and click Search Service by Policy.



- The 360-degree service view can be accessed from each service view:
  - Open the Network Management app and choose Service > View > Dynamic Tunnel from the main menu. Click the 360 icon next to a service.
  - Open the Network Management app and choose Service > View > SR Policy from the main menu. Click the 360 icon next to a service.
  - Open the Network Management app and choose Service > View > L2 EVPN Service from the main menu. Click the 360 icon next to a service.
  - Open the Network Management app and choose Service > View > MBGP L3VPN from the main menu. Click the 360 icon next to a service.

360-Degree Service View (2)

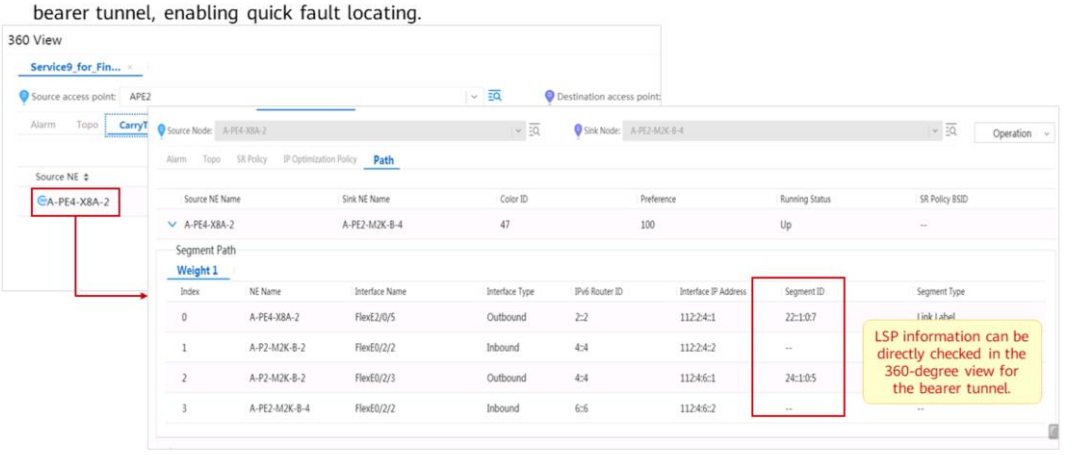
- The diagnosis function provided by the 360-degree service view allows you to check service connectivity, demarcate service faults, and determine whether a fault is a bearer tunnel fault or service fault.



HUAWEI

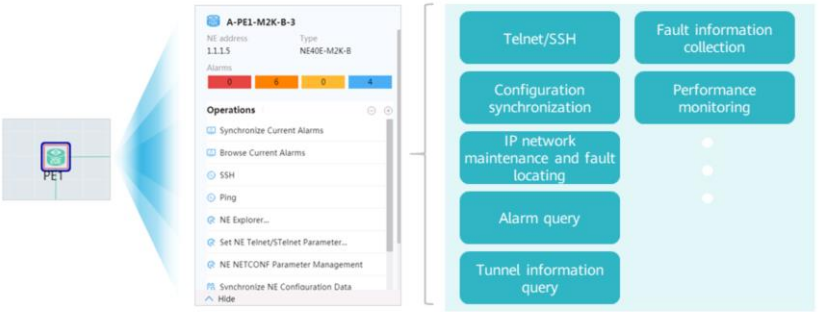
360-Degree Service View (3)

- The 360-degree view for a VPN service allows you to directly check the 360-degree view for the corresponding bearer tunnel, enabling quick fault locating.



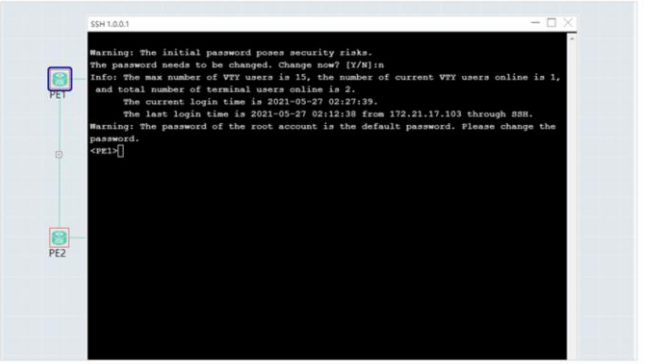
Device Maintenance

- The controller delivers convenient device management, facilitating routine maintenance and fault locating for O&M personnel.



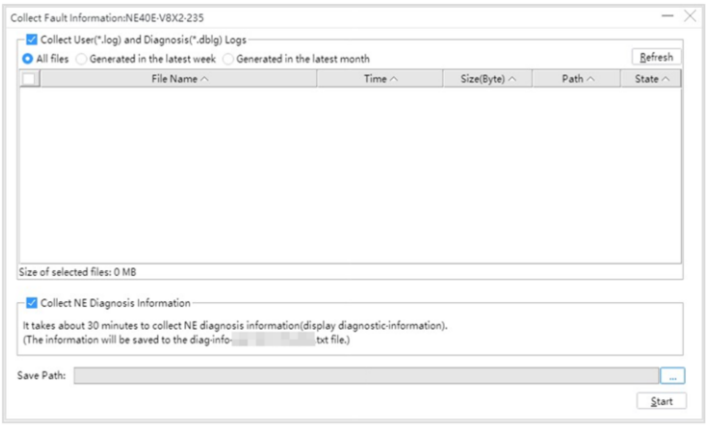
Remote Management

- The Telnet/SSH login function allows administrators to run commands on the controller web UI to operate and maintain devices. In this case, the clients used by administrators do not need to be routable to devices.



Fault Information Collection

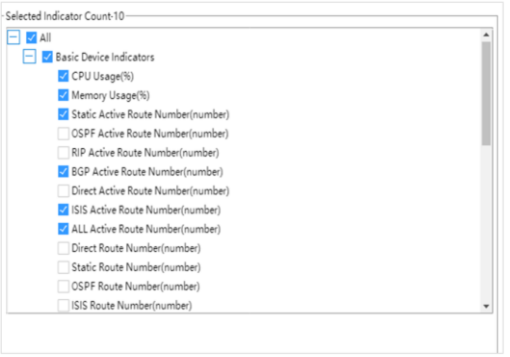
- Fault information collection collects device diagnosis information, including a large amount of device running status information. Such information can be used for device inspection and fault diagnosis.

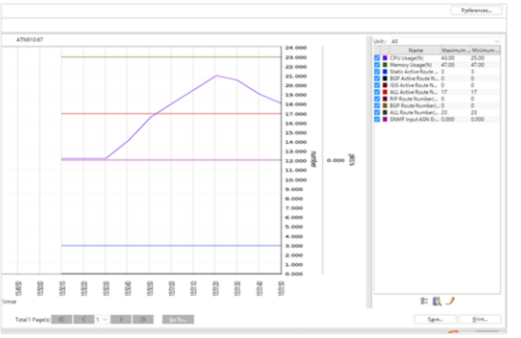


HUAWEI

Performance Monitoring

- Real-time performance monitoring allows you to view the performance data of devices in real time and customize the content to be displayed.

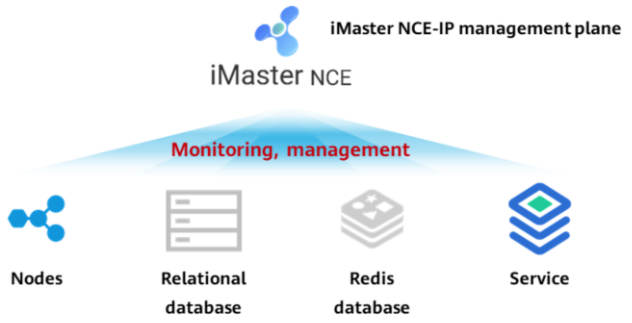




HUAWEI

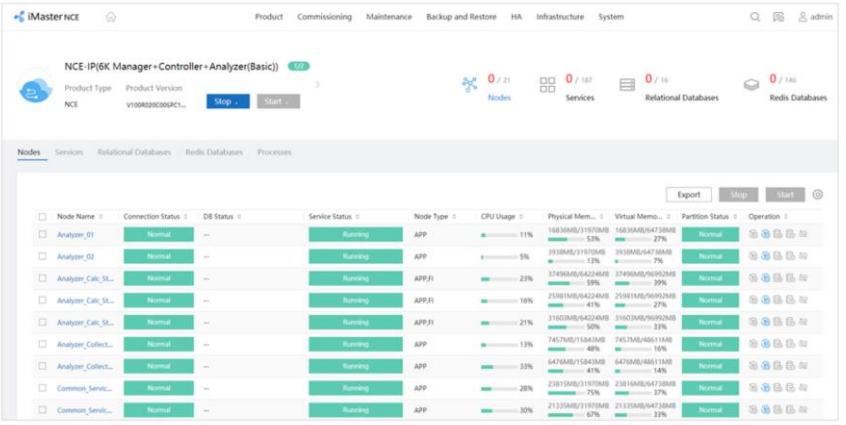
System Maintenance (1)

- On the management plane, you can maintain and manage the iMaster NCE-IP platform, view the running status of underlying nodes, servers, and databases, stop and restart related nodes and servers, and monitor the performance of the controller in real time.



System Maintenance (2)

- On the management plane, you can view the status of each node, service, database, and process, and perform operations such as stop, start, and restart on these objects.



HUAWEI

O&M Dashboard

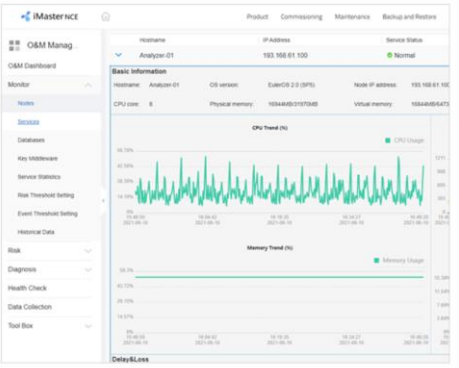
- The dashboard displays a summary of system evaluation, risk, and service health information as well as a summary of monitored system resource indicator information for the current system, helping O&M personnel learn the system health status during system running and reduce running risks.

- In the preceding figure, 1, 2, 3, and 4 provide the following capabilities:
  - 1: provides an overview of system status evaluation based on the number of current risks and risk levels.
  - 2: displays risks when system resource or service status anomalies are monitored.
  - 3: displays service health status, specifically, service operation statistics within the latest 1 minute.
  - 4: Displays system resource indicator status, specifically, the average value of each VM resource (CPU, memory, I/O, and network traffic) within the latest hour and the top 3 information.
- The dashboard also provides the redirection function. In other words, you can click the risk, service health status, or system resource monitoring area to go to the corresponding details page.

Unified Monitoring

- The O&M dashboard module monitors the indicators of NCE service, process, node, database, and key middleware resources in a centralized manner. By analyzing resource indicators, this module can detect and resolve potential risks in a timely manner. For key resources, you can set thresholds to trigger alarms and handle exceptions promptly.

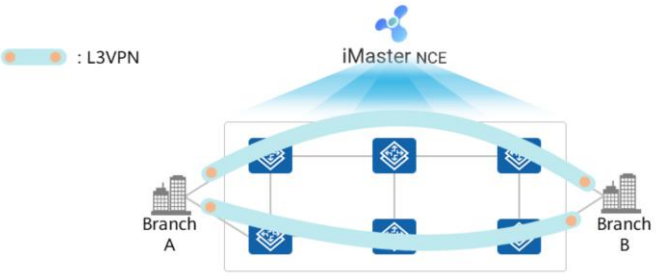
Function UI	Indicator
Node monitoring	Node CPU, memory, disk space, network indicators, and process status
Service monitoring	Service CPU, memory, number of threads, etc.
Database monitoring	GaussDB 100 V3 status, number of connections, etc.
Key middleware monitoring	Kafka and ETCD indicators
Service statistics	Number of operation times, operation duration, etc. The data is obtained based on the call chain statistics.
Risk and event threshold settings	Threshold settings for nodes: CPU, memory, I/O, packet loss rate, delay, etc. Threshold settings for databases: number of sessions, password expiration time, etc.



HUAWEI

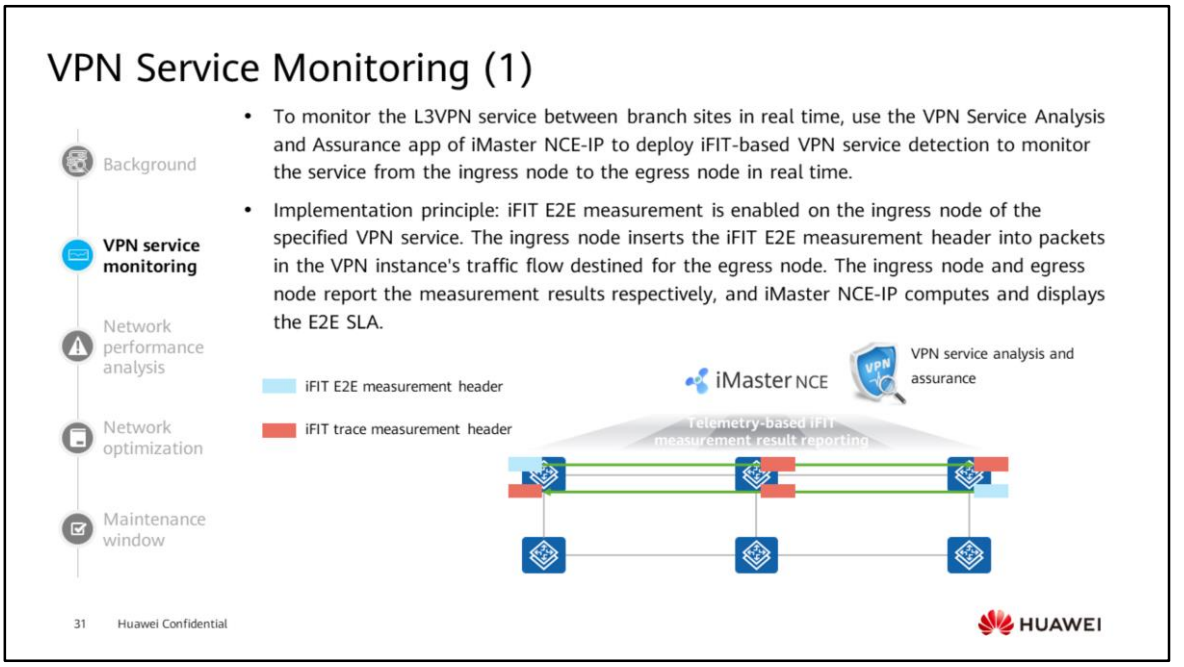
Typical Routine O&M Cases

- **Background**
- VPN service monitoring
- Network performance analysis
- Network optimization
- Maintenance window



- A company's branches communicate with each other over the company's own bearer WAN. iMaster NCE-IP is deployed to manage the WAN and perform routine maintenance on the WAN.

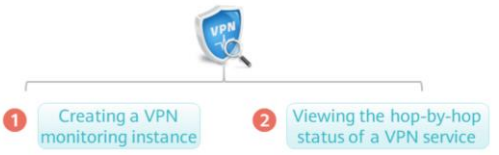
HUAWEI



- Here, technologies such as in-situ Flow Information Telemetry (iFIT) are used to meet the SLA assurance and O&M requirements of VPN services, implement proactive and visualized service SLA awareness, and proactively and quickly demarcate and locate faults. This helps ensure user experience.

VPN Service Monitoring (2)

- Background
- **VPN service monitoring**
- Network performance analysis
- Network optimization
- Maintenance window



Creating a VPN monitoring instance

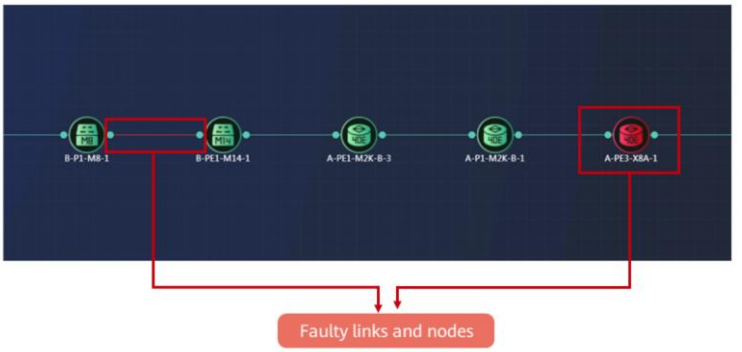
Service7_for_Finan...	A-PE5-M8-3	1.1.1.14	25GE0/7/2	A-PE4-X8A-2	Locator-A-PE4	
Service7_for_Finan...	A-PE1-M2K-B-3	1.1.1.5	GigabitEthernet0/3...	A-PE5-M8-3	A-PE5	

HUAWEI

VPN Service Monitoring (3)

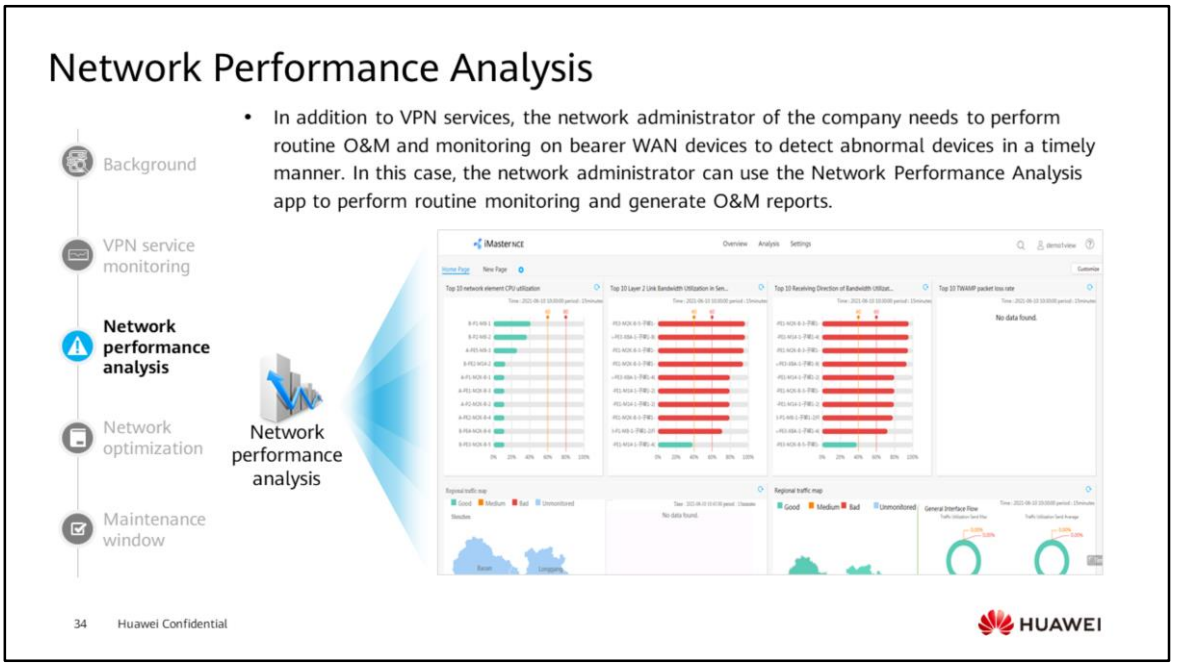
- Background
- **VPN service monitoring**
- Network performance analysis
- Network optimization
- Maintenance window

In the VPN list, drill down to the VPN service details.



- The VPN Service Analysis and Assurance app allows network administrators to quickly locate faulty links and nodes when L3VPN services are faulty.

HUAWEI

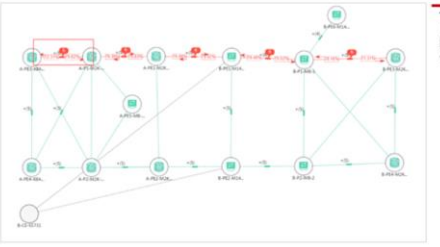


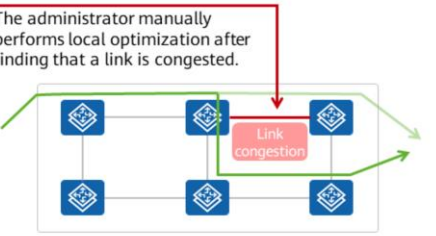
- As a center for network experience awareness, decision making, and big data analytics, the Network Performance Analysis app provides visualized management to help users monitor network traffic and quality in real time, detect the network change trend, and optimize the network, improving O&M efficiency and reducing OPEX.

Performance Topology

- Background
- VPN service monitoring
- **Network performance analysis**
- Network optimization
- Maintenance window

- The performance topology view provided by the Network Performance Analysis app allows you to build and manage the entire network topology and learn the networking and running status of devices. The color and status displayed for each device icon in the physical view help you learn the running status of the entire network in real time.
- The network administrator can view device and link loads in the performance topology in real time. If a link or device is overloaded, the administrator can manually adjust tunnel paths to prevent services between sites from being affected.





HUAWEI

Network Optimization

- Background
- VPN service monitoring
- Network performance analysis
- **Network optimization**
- Maintenance window

- If a link or node is faulty, you can use the Network Path Navigator app to ensure L3VPN service continuity through network optimization. After selecting a faulty link, you can view tunnels carried over the link and schedule the bearer tunnel of the L3VPN service to another link.

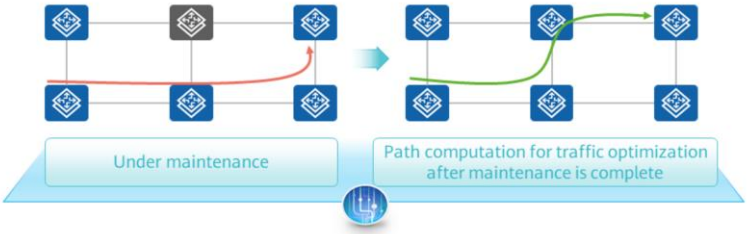
Manually performing local optimization

HUAWEI

Maintenance Window (1)

- Background
- VPN service monitoring
- Network performance analysis
- Network optimization
- **Maintenance window**

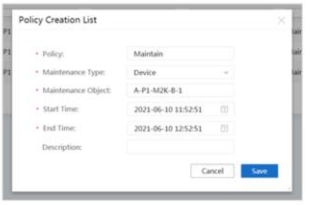
- The faulty node needs troubleshooting. If a device needs to be replaced, you can use the maintenance window function to ensure that services automatically bypass the target node during device replacement and service paths are automatically recomputed after device replacement. This allows the new device to take over service forwarding.
- 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 performs path computation for traffic optimization.



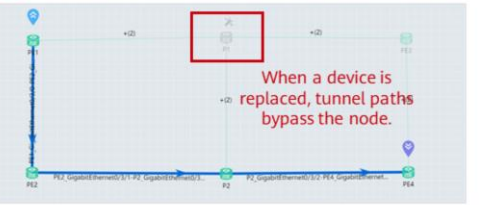
Maintenance Window (2)

- Background
- VPN service monitoring
- Network performance analysis
- Network optimization
- **Maintenance window**

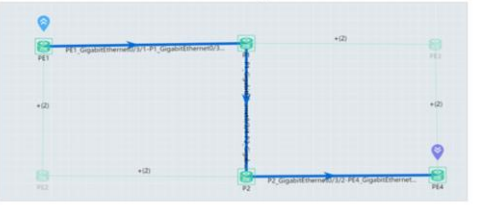
1 Configuring a maintenance window



2 Under maintenance



3 Tunnels switch back to original paths after the maintenance is complete.



HUAWEI

Contents

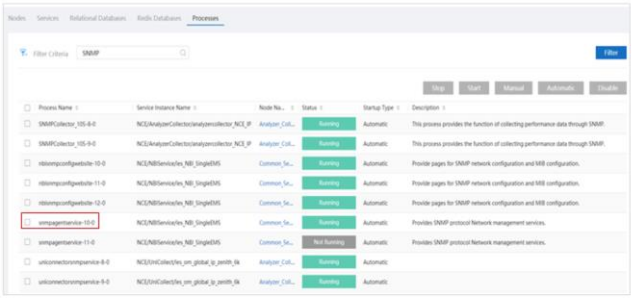
Topology Management Fault (1)

- Symptom: When an NE is added, a message indicating that the NE already exists is displayed.
- Cause: The NE has been added to NCE.

- Procedure:
  - Open the Network Management app and choose **Topology** > **View** > **Physical Topology** from the main menu. Then click the search button next to the toolbar.
  - Enter a keyword, such as an NE name or device type, and click the search button again.
  - If the NE can be found in the search result, the NE has been added to NCE.

Topology Management Fault (2)

- Symptom: During automatic NE discovery, an SNMP NE is created as a third-party NE.
- Cause: The agent corresponding to the SNMP NE is not enabled. As a result, the SNMP NE is created as a third-party ICMP NE.



- Procedure:
  - Log in to the management plane of iMaster NCE-IP.
  - Choose **Product** > **System Monitoring** from the main menu. On the **System Monitoring** page, click the **Processes** tab and enter **agent** in the search box to filter processes.
  - Check the status of the **snmpagentservice** process. If the process is not running, click **Start** to start the process.

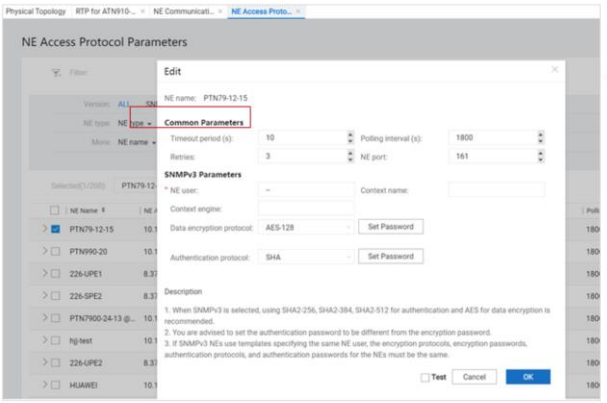
Topology Management Fault (3)

- Symptom: After a user selects an NE on the web UI and clicks Ping on the information panel that is displayed, the ping operation fails, and a timeout message is displayed. However, the NE can be pinged through the CLI.
- Cause: The built-in ICMP tool of iMaster NCE-IP is abnormal.

- Procedure:
  - Log in to the management plane of iMaster NCE-IP.
  - Choose **Product** > **System Monitoring** from the main menu. On the **System Monitoring** page, click the **Services** tab and enter **topology** in the search box to filter services.
  - Check the status of the topology service. If the service is not in the running state, click **Start** to start the service.

Topology Management Fault (4)

- Symptom: When a link is added, the corresponding connection fails to be created.
- Cause: The NE is busy during data synchronization. As a result, some data fails to be synchronized.

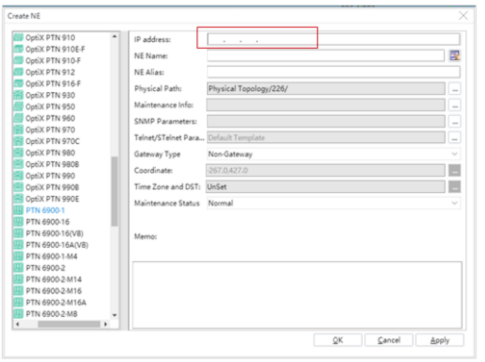


- Procedure:
  - Change the **Timeout Interval** setting of SNMP.
  - Open the Network Management app and choose **System** > **NE Communication Parameters** from the main menu. Then expand **NE Communication Parameter** and click **NE Access Protocol Parameters**.
  - Double-click the SNMP record of the NE. In the detailed information area, increase the value of **Timeout Interval (s)**.

HUAWEI

Failure to Obtain NE Alarm Information

- Symptom: SNMP parameters have been set on NCE and an NE. NCE, however, fails to receive trap packets from the NE or does not process received trap packets. As a result, NCE cannot obtain NE alarm information.
- Cause: The NE management address configured on NCE is different from the source address used by the NE to report trap packets.

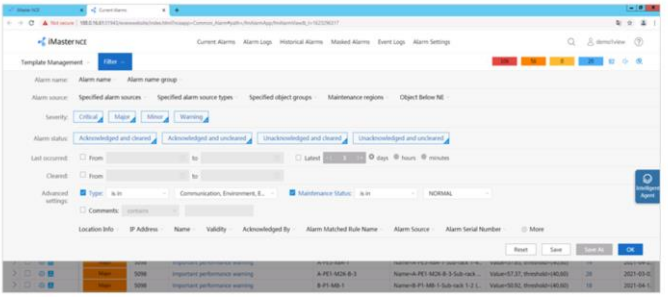


- Procedure:
  - Delete the NE from iMaster NCE-IP. Specifically, open the Network Management app and choose **Topology** > **View** > **Physical Topology** from the main menu. In the topology view, right-click the NE and choose **Delete** from the shortcut menu.
  - Add the NE again and set the NE address as the source address for sending trap packets.

HUAWEI

Failure to View Certain Alarms

- Symptom: Alarms received by NCE cannot be found on the Current Alarms or Historical Alarms page.
- Cause: Some alarms are filtered out because they meet the filter criteria or alarm template rules.

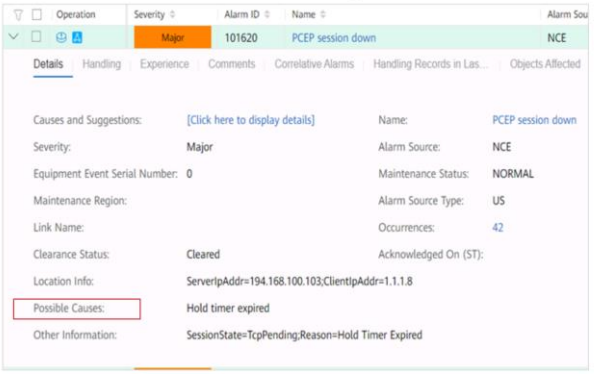


- Procedure:
  - On the page for browsing alarms, click **Filter**. In the **Filter** dialog box, view and modify the filter criteria.
  - Then click **OK**.

HUAWEI

PCEP Session Interruption

- Symptom: The PCEP session between the controller and a device is interrupted.
- Cause: The source address configuration changes, the number of delegated LSPs exceeds the PCE limit, or the controller receives incorrect PCEP messages.



- Procedure:
  - Open the Alarm Monitor app and choose Current Alarms from the main menu.
  - Click Filter and then click Alarm Name to display all alarms similar to **PCEP_Session_Down.**
  - Locate the target alarm and click the alarm name to view alarm details.
  - Check the possible causes one by one based on information displayed in the Possible Causes field on the alarm details page.

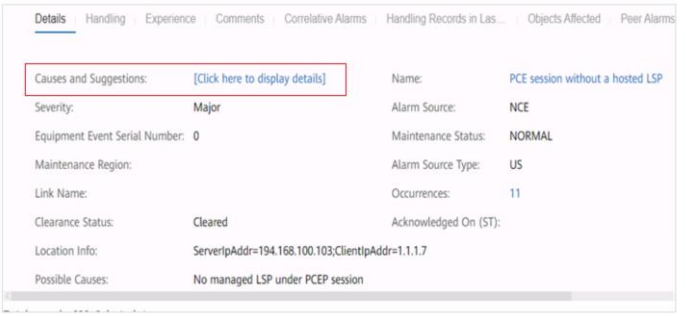
HUAWEI

- Possible causes:
  - The network control node received incorrect PCEP messages. In this case, the peer end automatically re-establishes the connection, and you are advised to wait for some time. If the PCEP session is still down, contact technical support for the forwarder.
  - The hold timer expired. In this case, you need to check the network connection between the controller and forwarder and restore the communication between them as required.
  - The controller source address was changed. In this case, check whether the actual controller IP address is the same as the controller IP address configured on the forwarder. If they are the same, you are advised to wait for some time. If the PCEP session is still down, contact technical support for the forwarder.
  - The number of delegated LSPs exceeded the PCE limit. Check the PCE delegation limit. Set the limit to a larger value or reduce the number of LSPs delegated by PCCs.
  - Enable/disable TLS authentication. In this case, the value of Possible Causes is Enable TLS authentication or Disable TLS authentication. After PCEP TLS authentication is enabled or disabled on the forwarder as required, wait for the forwarder to re-establish the connection. If the PCEP session is still down, contact technical support for the forwarder.
  - PCEP configurations were deleted. In this case, check the PCEP

configurations on the controller and reconfigure PCEP as required.

No Tunnel Delegated to the Controller

- Symptom: The controller detects that no tunnel is delegated by PCCs.
- Cause: Tunnel delegation is not configured on the forwarder.



- Procedure:
  - Open the Alarm Monitor app and choose Current Alarms from the main menu.
  - Click Filter and then click Alarm Name to display all alarms about **PCE sessions without delegated LSPs**.
  - Locate the target alarm and click the alarm name to view alarm details.
  - Rectify the fault based on information displayed in the Cause and Suggestion field on the alarm details page.

Quiz

1. (Single-answer question) Which of the following device maintenance modes is not supported by iMaster NCE-IP? ( )
   A. Remote management (Telnet and SSH)
   B. Fault information collection
   C. Remote file management
   D. Performance monitoring

- C

Summary

- This chapter describes the logical architecture of iMaster NCE-IP, which provides various apps (such as Network Management, Network Path Navigation, Network Performance Analysis, and VPN Service Analysis and Assurance) based on a universal basic platform to meet the function requirements of different scenarios. Moreover, iMaster NCE-IP has good scalability.
- This chapter also describes common O&M methods used by the controller, helping you understand how to maintain services through the controller and how to perform controller O&M through the management plane.
- Finally, this chapter describes the common faults of the controller and their troubleshooting methods.

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.

Huawei CloudWAN Solution Design Practice (Financial Scenario)

HUAWEI

Foreword

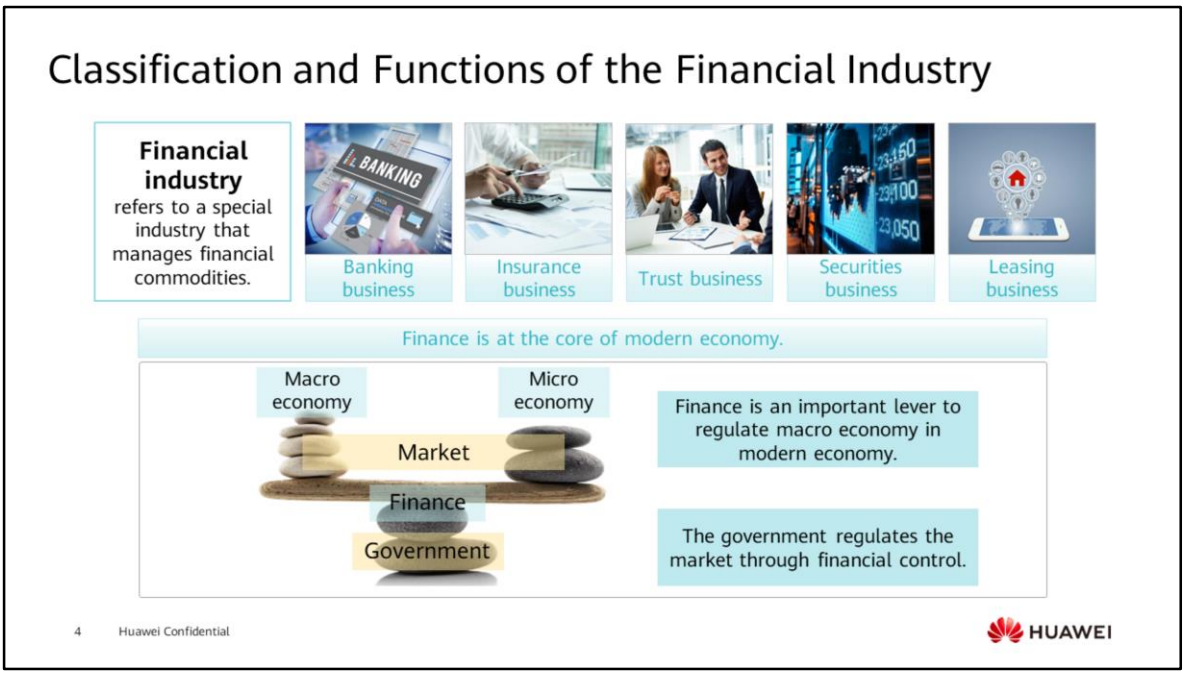
- As the financial industry enters the cloud era, financial backbone networks face the challenge of providing varying network bearer quality based on service characteristics in a fine-grained manner. In addition, problems such as increasing network costs and difficult troubleshooting make it difficult for the networks to meet the requirements of cloud services.
- To help the financial industry address the challenges brought by cloudification, Huawei launches the CloudWAN solution.
- This course describes the design roadmap of the financial cloud WAN in terms of physical networking, IP address planning, VPN design, routing design, SLA design, and so on based on the Huawei CloudWAN solution.

- This course is based on Huawei's CloudWAN solution.

Objectives

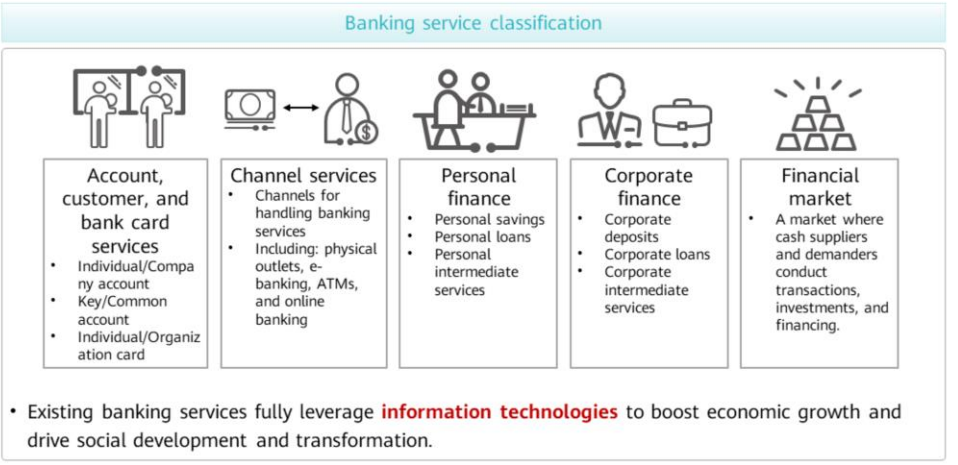
- Upon completion of this course, you will be able to:
  - Describe the current status and trend of financial networks.
  - Describe the basic design roadmap of the financial cloud backbone network.
  - Describe the tunnel and VPN design roadmap of the financial cloud backbone network.
  - Describe the SLA and reliability design roadmap of the financial cloud backbone network.
  - Describe the optimization and O&M design roadmap of the financial cloud backbone network.

Contents

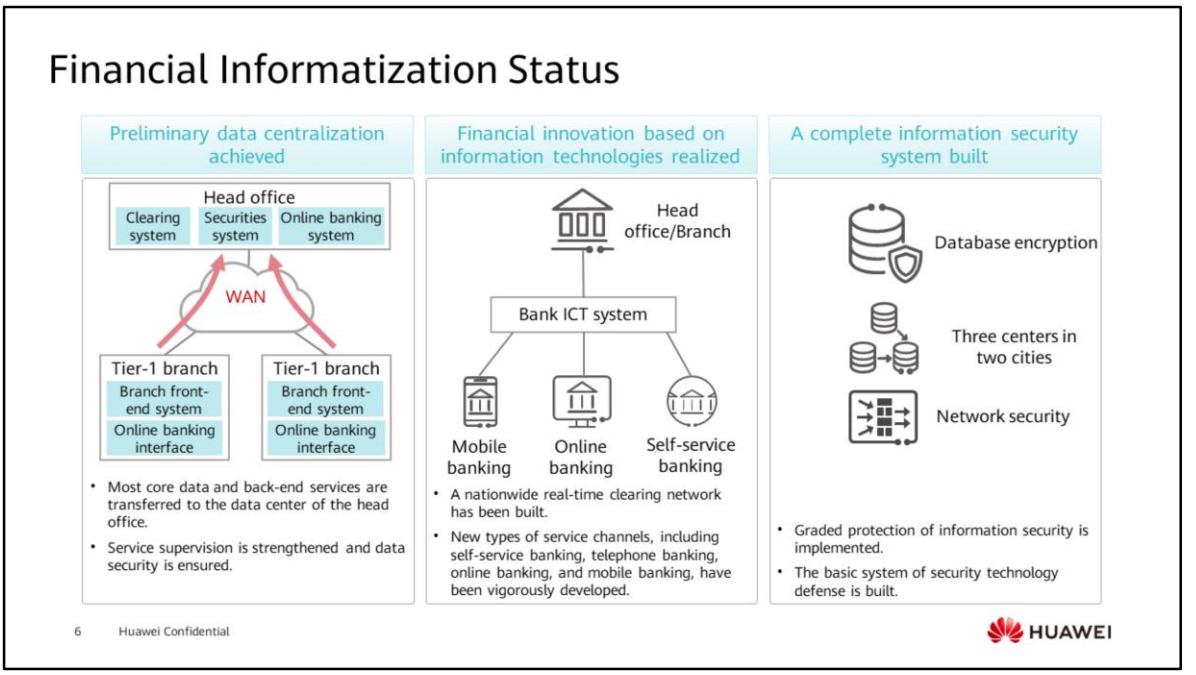


- The financial industry bridges all aspects of the national economy. Meanwhile, financial means such as interest rate, exchange rate, credit, and settlement exert direct influence on microeconomic entities.
- Finance is related to economic sovereignty and wealth control of a country, and plays an important role in safeguarding economic growth and national interests as well as serving the real economy and citizens.
- Finance is at the core of not only modern economy, but also modern politics and modern society.
- Banking: Financial service institutions that undertake credit intermediary through deposits, loans, remittances, and other services.
- Insurance: An industry in which funds are pooled in the form of contracts to compensate the insured for their economic interests.
- Trusts: A trustor entrusts properties to a trustee. A trust allows a trustee to manage assets on behalf of a beneficiary or beneficiaries.
- Securities: A special industry engaged in securities issuance and trading services. It mainly consists of stock exchanges, securities companies, securities associations, and financial institutions.
- Leasing: A business that provides credit services in a way that combines financial credit and materials.

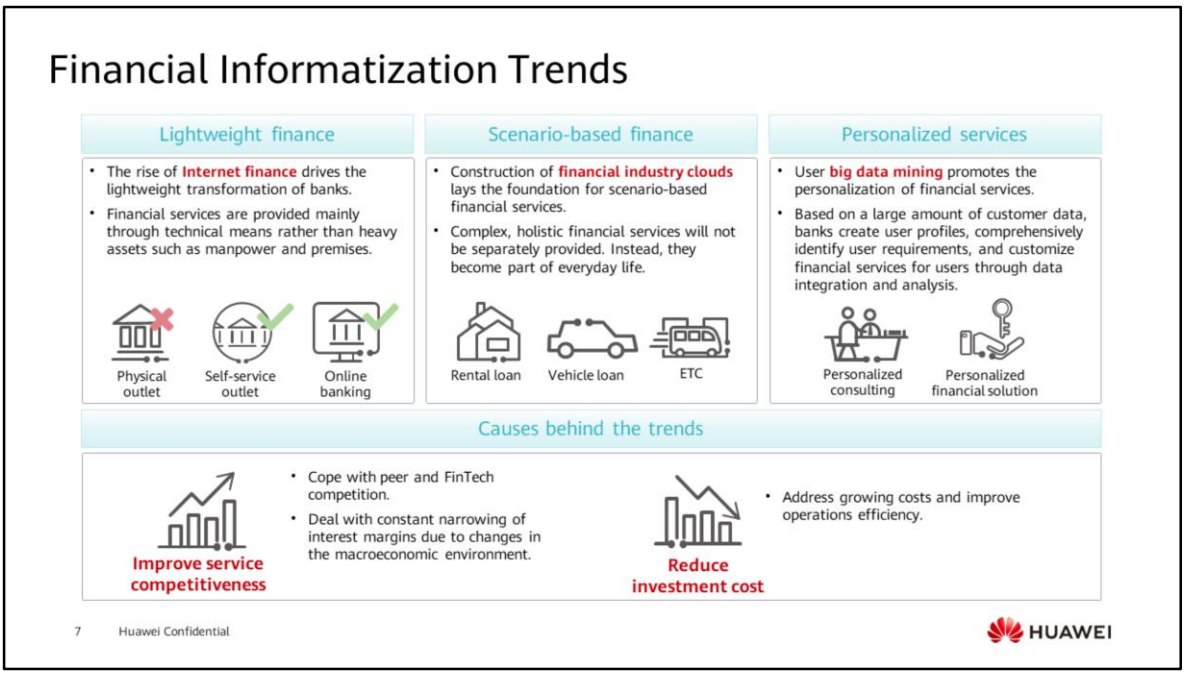
Overview of Banking Services



- This course uses the banking industry as an example to describe the informatization requirements of the financial industry.
- Banking services can be classified into front-end, middle-end, and back-end services based on the functions and architecture of banks:
  - Responsible for business development, the front end is a customer-oriented department that provides one-stop and comprehensive services for customers. Bank tellers, account managers, and lobby managers are all front-end positions.
  - Major middle-end responsibilities include formulating business development policies and strategies, providing professional management and guidance for the front end, and controlling risks by analyzing the macro market environment and internal resources.
  - Major back-end responsibilities include supporting and processing services and transactions, including accounting processing, IT support, and call center. The centers that centrally process loan approval can also be included in the back-end system.

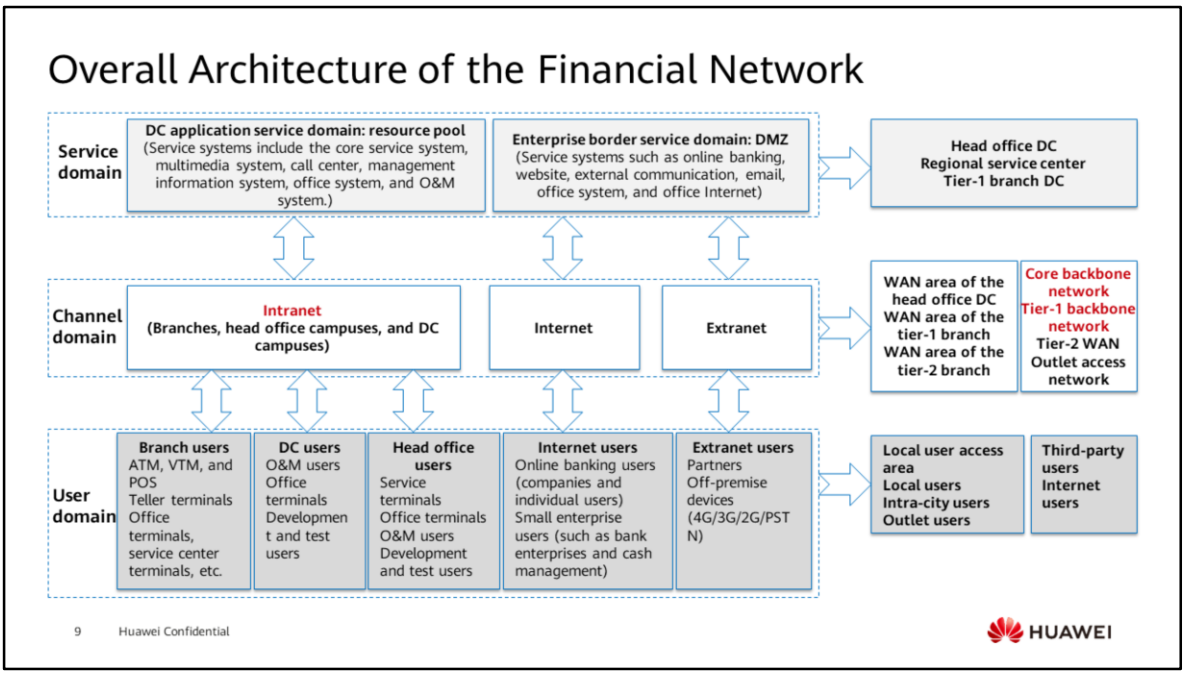


- Informatization of banks is not simply computerizing manual services. Instead, banks need to integrate technical transformation and institutional transformation, establish and improve the financial risk mechanism, and reconstruct business models and business processes while introducing information technologies.

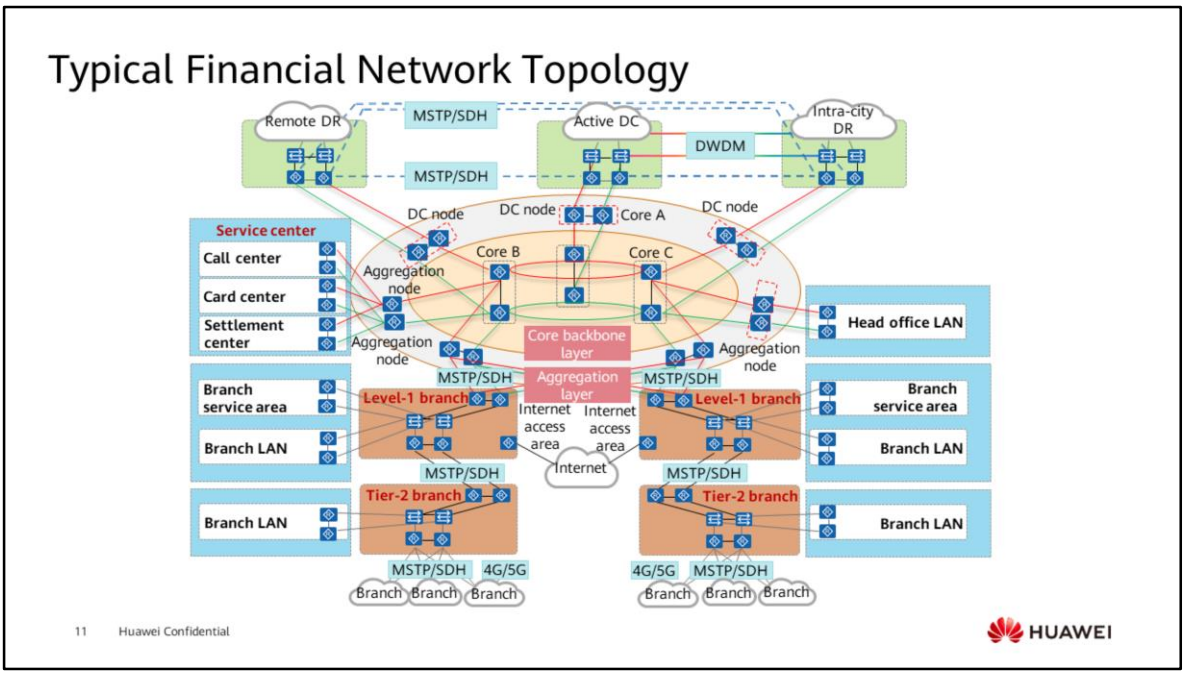


- FinTech: Short for financial technology, FinTech is an economic industry formed by a group of enterprises that use technological means to make financial services more efficient.
- Leveraging FinTech revolution, digital banking and mobile finance that win customers with services and experience gradually change the future service mode of banks, breed new sources of growth for digital finance, and play an increasingly important role in banking services.
- Big data, AI, IoT, and cloud computing also provide a new technical engine for full-process evolution of customer management, process reengineering, risk prevention and control, open ecosystem, and channel convergence.
- The informatization trends of the financial industry and the causes behind the trends pose great challenges to financial WANs.

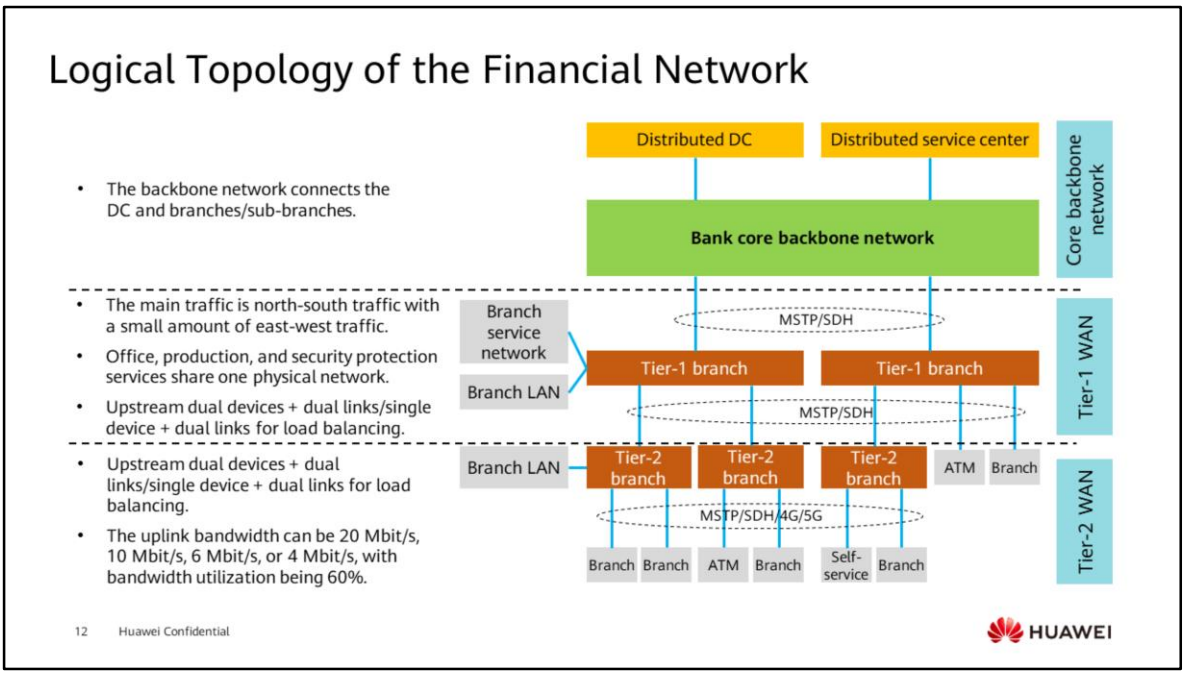
Contents



- The overall architecture of the financial enterprise network consists of the service domain, channel domain, and user domain:
  - The user domain contains internal users and external users. Internal users include branch users, DC campus users, and head office users. External users include Internet users and extranet third-party users.
    - Branch users: access the data center network (DCN) through the Intranet WAN in the channel domain.
    - DC campus users: locally access the DC through the Intranet DCN in the channel domain.
    - Head office users: access the intranet WAN in the channel domain through the metro network and then access the WAN access area of the DCN.
    - Internet users: access the Internet access area of the DCN through the Internet channel domain.
    - Extranet users: access the extranet access area of the DCN through the extranet channel domain (mainly through private lines).

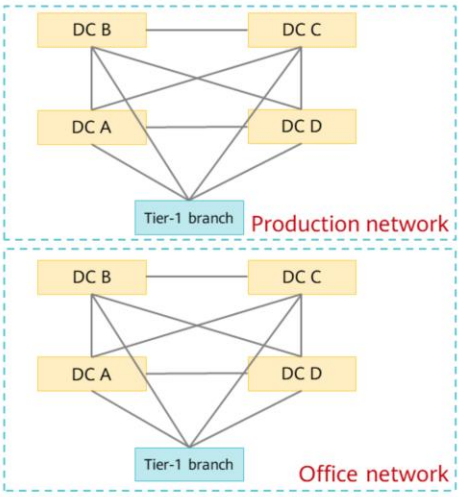


- Multi-DC DR
  - Hierarchical design of the physical network and stable core backbone network.
  - Multiple DCs are built in multiple cities, implementing cloud-based interconnection between DCs through the core backbone network.
- Hierarchical WAN networking
  - The tree-shaped network structure requires hierarchical network construction and level-by-level aggregation.
  - Prevents horizontal traffic detour and improves link utilization.
  - Clarifies responsibilities of network O&M, preventing cross-area maintenance.
- Branch flattening trend
  - Densely-distributed outlets and abundant line resources enable outlets to be directly connected to branches, thereby forming a flat network.
  - Reduces the network architecture by one layer and maintenance costs.
  - Comprehensively evaluate the impact of increasing line leasing costs on the overall costs.



- The three-center, two-city architecture is widely adopted on the financial (bank) backbone network. The active and intra-city DCs are interconnected through WDM devices, and connected to the remote DC through carriers' private lines to implement remote DR. The bandwidth can be 622 Mbit/s, 2.5 Gbit/s, or 10 Gbit/s based on service requirements. Some banks are evolving towards the multi-center, multi-city architecture.

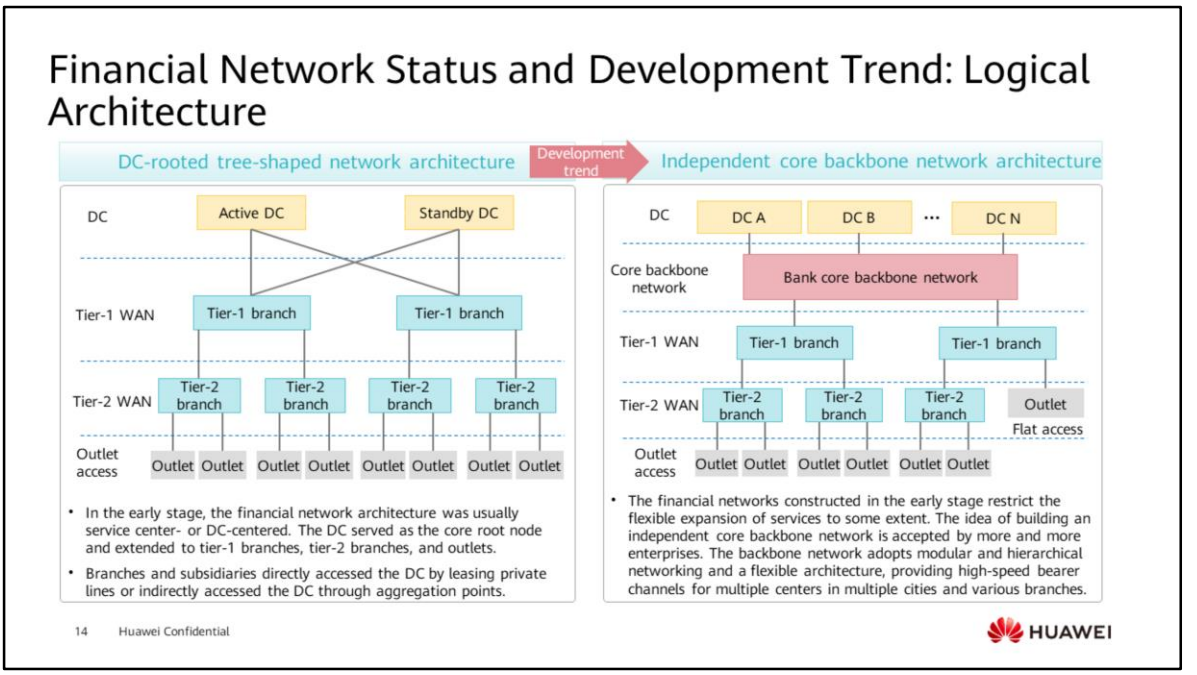
Pain Points of the Financial Network



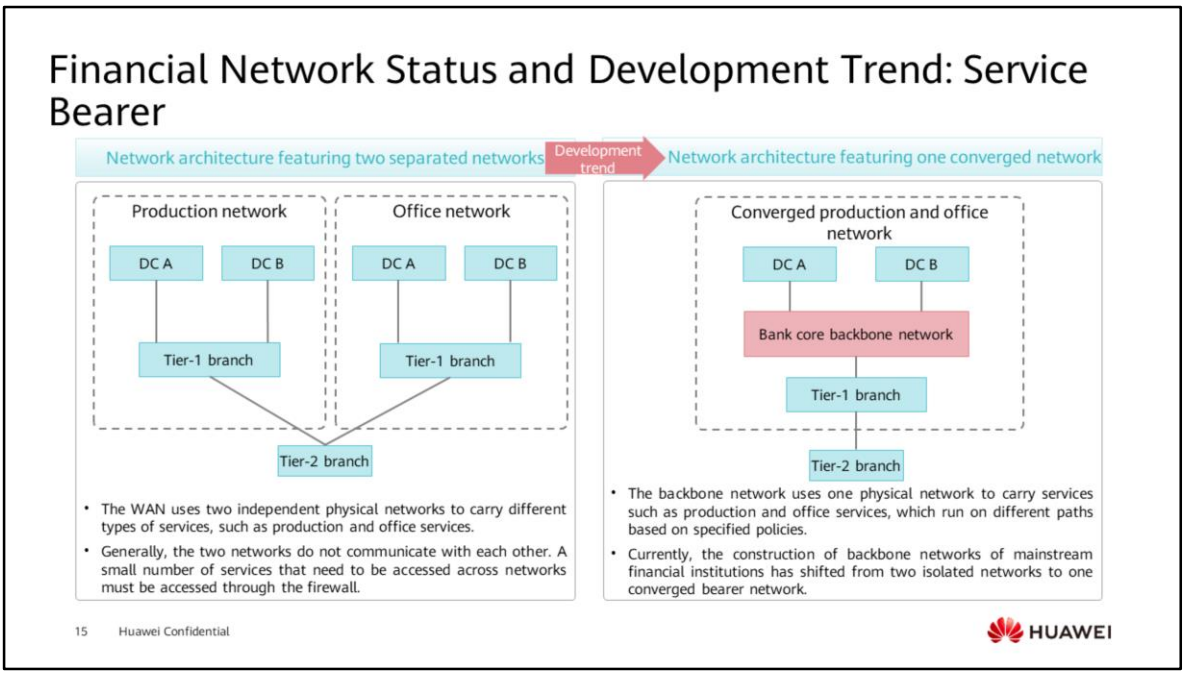
- To improve service reliability and user experience, banks start to build multiple DCs in multiple cities. However, this may bring the following problems:
  - The horizontal traffic between multiple DCs will increase sharply, which may cause partial link congestion and affect critical services.
  - Existing bank networks are DC-rooted. Therefore, the flexibility is poor, and the cost of building a new DC is high.
  - Banks have multiple physical private networks, and service streamlining is a bottleneck.

HUAWEI

- Multiple DCs in multiple cities:
  - Separation of the front-end, middle-end, and back-end services and centralized service processing by the back end are inevitable requirements for constructing process-based banks. Centralized back-end operations trigger the emergence of multiple centralized service processing centers, such as the centralized operation and post-supervision center, bank card center, call center, audit center, and financing settlement center. These centers will be separated from DC areas. As DC technologies develop, distributed cloud DCs gradually become the mainstream for construction. The WAN horizontal service traffic between multiple DCs in multiple cities is increasing, and the tree-shaped network structure overloads DC core nodes.



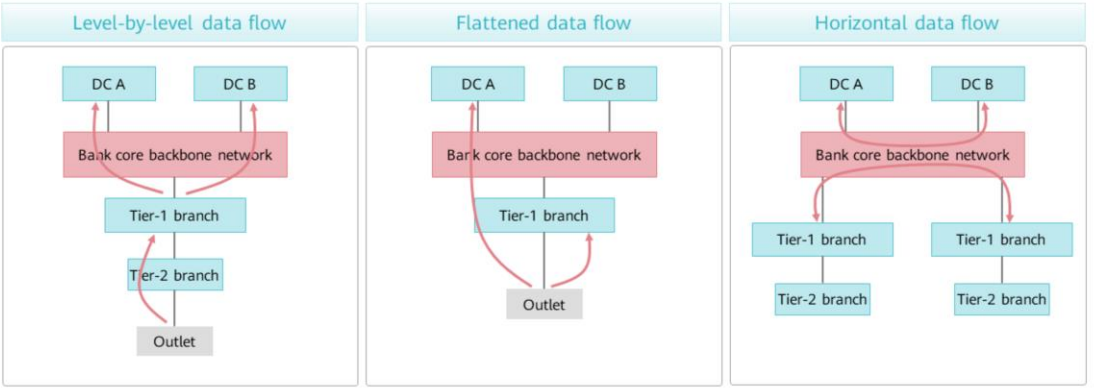
- The three-center, two-city architecture is widely adopted on the financial backbone network. The active and intra-city DCs are interconnected through WDM devices, and connected to the remote DC through carriers' private lines to implement remote DR. The bandwidth can be 622 Mbit/s, 2.5 Gbit/s, or 10 Gbit/s based on service requirements. Some banks are evolving towards the multi-center, multi-city architecture.
- In terms of network architecture, two commonly used architecture models are available for banks' backbone networks.
    - One is the DC-rooted, tree-shaped network architecture.
    - The other is an independent core bearer network built to implement interconnection between three centers in two cities or multiple centers in multiple cities. Network services are mainly processed in head office and tier-1 branch DCs, and users are distributed both inside and outside the banks. Branches access head office DCs through the backbone network. The backbone network mainly uses BGP and OSPF. Some financial institutions also use Enhanced Interior Gateway Routing Protocol (EIGRP).



- One type of the financial backbone network consists of two independent physical networks, for example, No. 1 network (production network) and No. 2 network (office network), which carry different types of services.
- Another type uses one physical network to carry all services, and services are logically isolated using VPNs.
- Financial customers who have two independent networks generally plan to integrate their two networks for unified service bearing.

Common Financial Service Traffic Models

- From the perspective of data traffic, the traffic models of common banking services can be classified into the following types:



- Level-by-level data flow:
  - Currently, service applications are generally deployed in a centralized manner. The integrated application front-end and intermediate service platforms are deployed in tier-1 branches. All front-end users/channels access the front-end or intermediate service platform, which then communicates with applications in DCs. For example, credit management and email services are deployed level by level. Users at the current level can only access the server at the current level, and application communication and forwarding are performed between servers.
- Flattened data flow:
  - Some financial enterprise applications are deployed in a flattened manner. For example, new outlets of banks and insurance companies do not need to be aggregated at the provincial or municipal level. Instead, they directly access the applications of the tier-1 branch or head office DC.
- Horizontal data flow:
  - Most commonly seen financial applications are hierarchical, such as communication between tier-2/tier-1 branches and head office DCs, and horizontal communication between branches. With the construction of multiple DCs in multiple cities, the horizontal mutual access traffic between DCs will increase.

Requirements on Networks Posed by Financial Services

- Common bank services usually place the following requirements on networks:

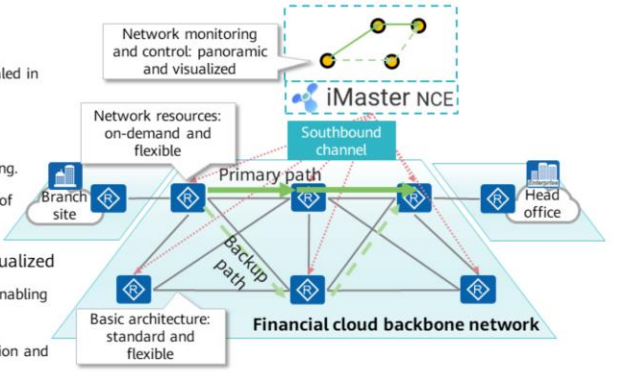
Service Type	Service name	Requirements on Network Services	User or Terminal
Data	Online transaction	Delay-sensitive	Teller
	Online batch	Bandwidth-sensitive	Branch front-end system
	End-of-day batch	Bandwidth-sensitive	Head office
	O&M management	Delay-sensitive	O&M terminal
	Operation management	High requirements for rapid response, bandwidth-sensitive	Office terminal
	OA	Delay-insensitive	Office terminal
Voice	VoIP service	Sensitive to delay and jitter	Conference and customer service
Video	Video	Sensitive to delay and jitter	Videoconferencing and surveillance

- Online transactions: refer to single transactions at the counter, initiated by ATMs, or accessed through various channels that require timely response, such as over-the-counter deposits and withdrawals, loans, ATM withdrawals, large and small transactions, etc.
- Batch processing: Batch processing can be divided into two modes: 1. Daily (online) batch processing, for example, debiting of intermediate services (collection and payment on behalf) and receipts. 2. Post-online batch processing services. After the online status is switched, various accounting processing (such as interest settlement), registers, and daily statements are performed. The output is various reports. Online batch processing is usually performed by a back-end host invoking a group of programs at night.
- End-of-day batch processing: Banks perform batch processing for all services on the current day at about 12:00 every day to record the services into accounts and modify account information in batches.

Contents

Financial Cloud Backbone Network Design Objectives

- Based on the future-oriented design concept, the design objectives of the financial cloud backbone network are as follows:
  - Basic architecture: standard and flexible
    - Backbone network architectures and access models are standardized to support quick service rollout.
    - Backbone network nodes and links can be elastically scaled in or out.
  - Network resources: on-demand and flexible
    - WAN link resource pooling enables on-demand scheduling.
    - Backbone network virtualization allows flexible bearing of different types of services.
  - Network monitoring and control: panoramic and visualized
    - Comprehensively display application flow information, enabling application flow visualization and prediction.
    - Network quality visualization allows fast fault demarcation and recovery.



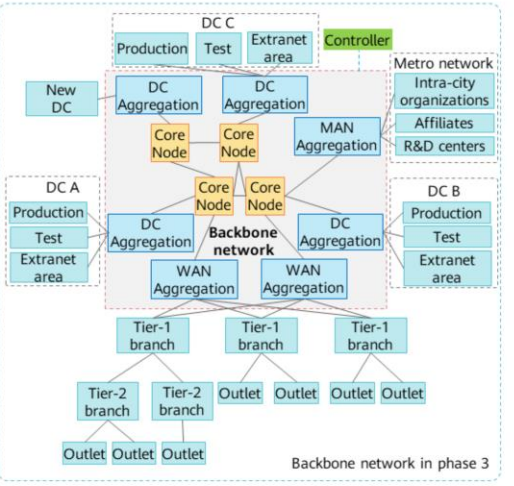
HUAWEI

Financial Cloud Backbone Network Design Objectives

Compliance	• The design solution complies with related network specifications, including backbone network construction specifications, IP address specifications, and DC construction specifications. • Ensure that backbone network construction is consistent with that of other systems.
Standard	• Technical standardization: Use open and standard mainstream technologies and protocols to ensure network openness, interconnection, upgrade, and expansion.
High reliability	• Carry out multi-dimensional high reliability design based on networks, links, devices, and controllers to ensure 24/7 running of networks. • Decouple the control plane from the forwarding plane, so that faults on the control plane do not affect online services.
High scalability	• Adopt the standard hierarchical architecture for the backbone network so that it is decoupled from the access network, making it easy to expand core nodes in the future. • Standardize the access model so that subsidiaries and third parties can directly access the system based on the standard model.
High performance	• On-demand bandwidth resource allocation enables on-demand service access at any time and ensures service interaction quality. • The delay and convergence of the entire network are controllable, meeting services' requirements on delay and convergence for the backbone network.
High security	• Securely isolate different types of services, and adopt different assurance policies based on service types. • Meet the technical requirements of graded protection of information security for all levels of security and achieve regulatory compliance.
Easy maintenance	• Design the entire network architecture in a unified and standard manner and use mature and open network technologies to reduce maintenance complexity. • Improve the automation capability, network quality, fault monitoring capability, and service traffic prediction capability.
Foresight	• Based on the industry's best practices, comply with mainstream technology development in the industry. • Open and evolvable network capabilities meet future service development requirements.

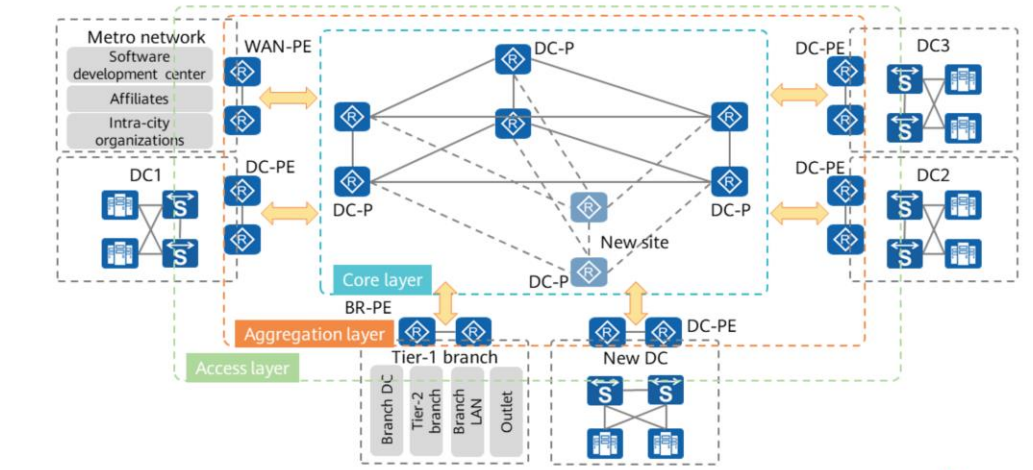
Development Trends of Financial Core Backbone Networks

- The backbone network architecture of financial enterprises undergoes three phases based on different service requirements and technology development.
  - Phase I: The network has a DC-rooted, tree-shaped architecture. Services are mainly processed in head office and tier-1 branch DCs, and users are distributed both inside and outside the banks.
  - Phase II: Network construction enters the phase of three centers in two cities. An independent core bearer network that consists of the core layer, aggregation layer, and access layer is widely built by large banks. MPLS VPN technology is deployed to carry services in a unified manner, making the network architecture clearer.
  - Phase III: Network construction evolves from three data centers in two cities to multiple data centers in multiple cities. The capacity of the core bearer network can be flexibly expanded, and core and aggregation nodes are added to implement multi-center interconnection. In addition, SDN and SRv6 technologies are introduced. Traditional networks work with the SDN controller to implement functions such as intelligent network management and traffic optimization.



HUAWEI

Typical Architecture of the Financial Cloud Backbone Network



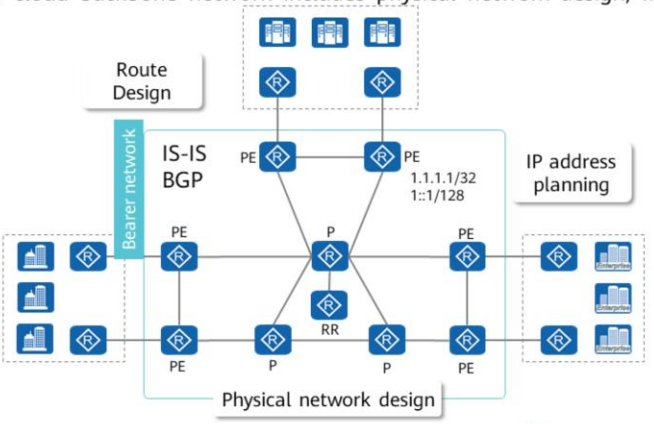
HUAWEI

- The backbone network functions as a financial transmission network to provide stable, reliable, and high-speed forwarding of enterprises' various service traffic.
- This target architecture model is suitable for most financial customers and can meet their main requirements for the backbone network. It is applicable to both three-center, two-city and multi-center, multi-city scenarios.
- In terms of the architecture, DCs are loosely coupled with the WAN, hierarchical networking is used, and services can flexibly access the network. This architecture supports smooth evolution to multi-center multi-city networking.
- Core layer: The full-mesh+dual-plane architecture is constructed as the top-layer traversal area to provide high-speed service access.
- Aggregation layer: aggregates different types of service traffic based on physical locations and connects to the core backbone network.
- Access layer: provides flexible and standard access based on the homing or region attributes of services.

Contents

Basic Design Overview for the Financial (Bank) Cloud Backbone Network

- Basic design for the financial (bank) cloud backbone network includes physical network design, IP address planning, and routing design.

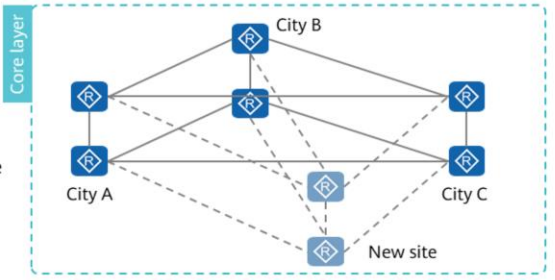


HUAWEI

Physical Network Design | IP Address Planning | Route Design

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.

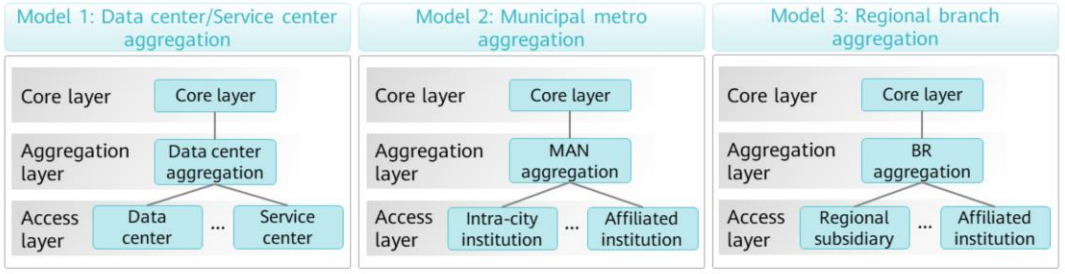


HUAWEI

- 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.

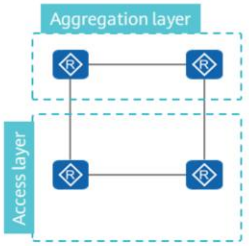
Aggregation Layer Design

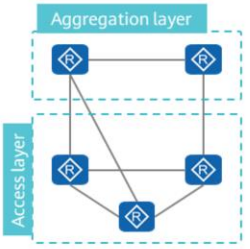
- Aggregation layer design must take into account the service types and scale on each aggregation node. In the initial phase of bearer network planning and construction, the aggregation layer can be planned based on existing enterprise services. The following three aggregation modes are available:
  - 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.



Access Layer Design

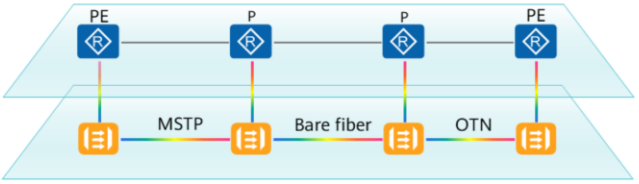
- When designing the access layer, you need to consider factors such as the bandwidth required by access services, required private lines, and private line prices. Moreover, you need to consider access reliability and traffic load balancing. For example:
  - You can use single-homed networking at the 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.





Private Line Selection for the Financial Cloud Backbone Network

- Currently, the transmission private lines provided by carriers for the financial cloud backbone network 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.



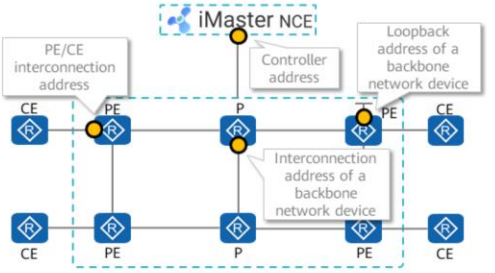
IPv4 Address Design Rules

- 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.

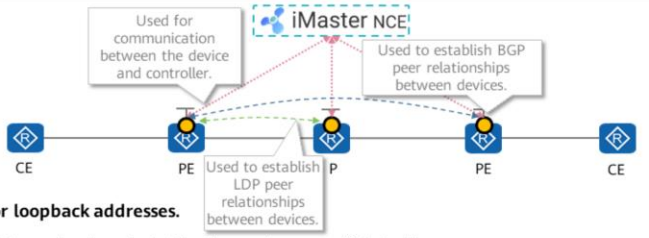
IPv4 address allocation range on the bearer WAN



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.

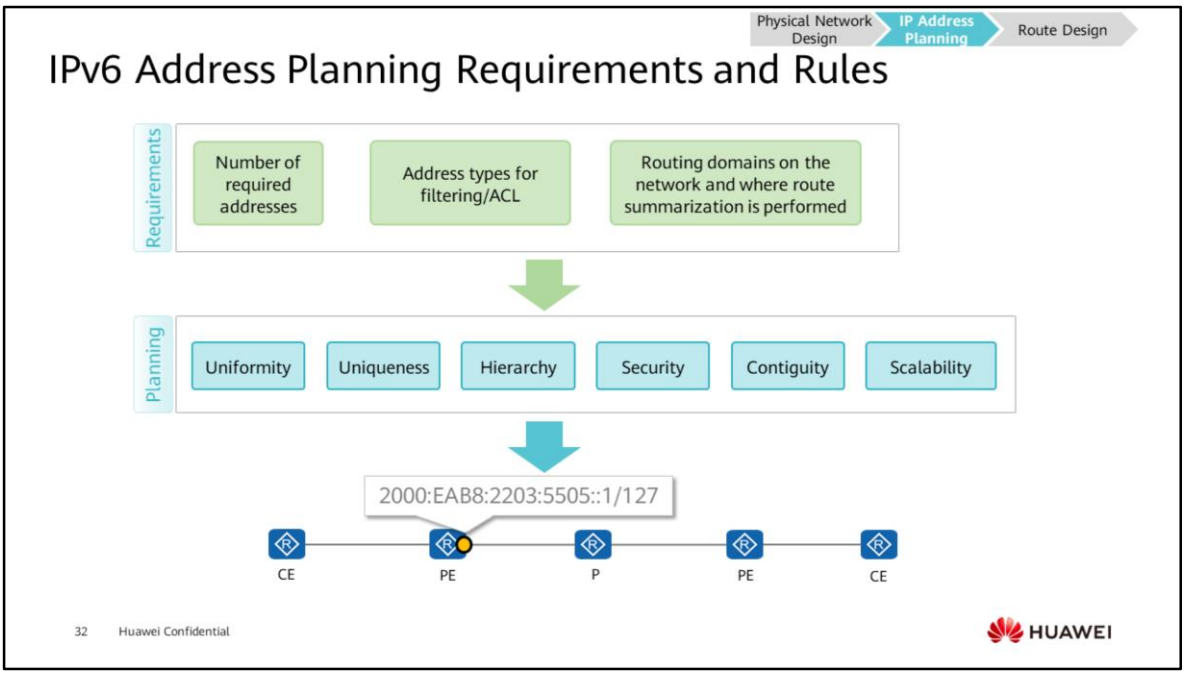
Loopback address design rules and application scenarios



- **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.

IPv4 Address Planning Suggestions

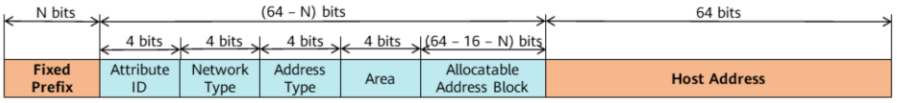
- It is recommended that the Financial Cloud Backbone Network use a dedicated subnet. The allocated addresses include the management addresses and internal interconnection addresses of devices and access addresses of the bearer network. The following table describes the address allocation plan:
  - Internal 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.



- 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.

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.

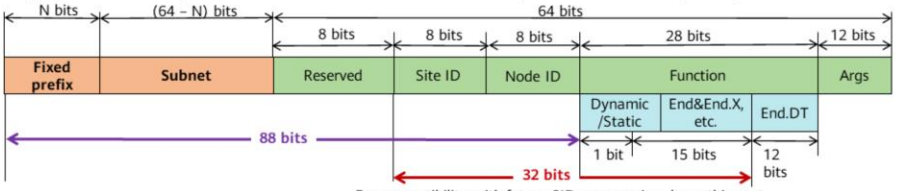


- Fixed Prefix: indicates a fixed-length prefix applied for by an enterprise from an address allocation organization.
- Subnet:
  - Attribute ID: is used to distinguish address types. It is used for level-1 address classification.
  - Network Type: identifies the type of a network.
  - Address Type: 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.

Physical Network Design | IP Address Planning | Route Design

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:



- 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.

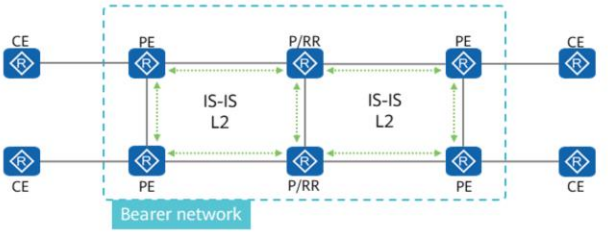
IGP Overview

- On a financial cloud backbone network, 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.	

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.

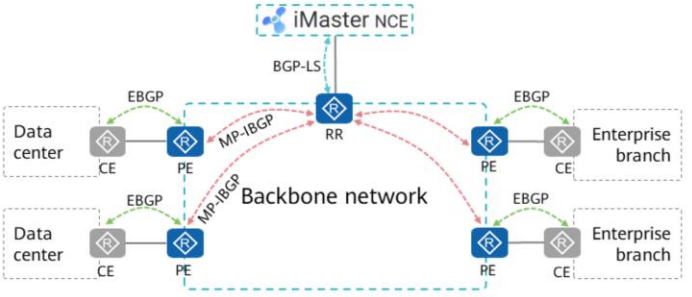
IGP Metric Planning

- The financial cloud backbone 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).

Physical Network Design > IP Address Planning > Route Design

Overall BGP Planning

- After a financial cloud backbone network 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.

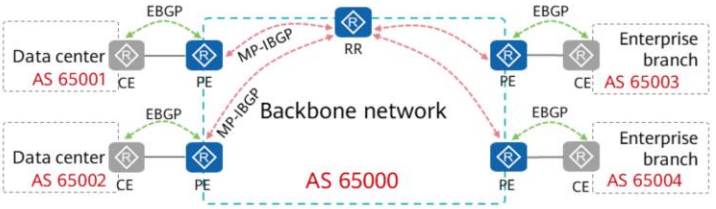


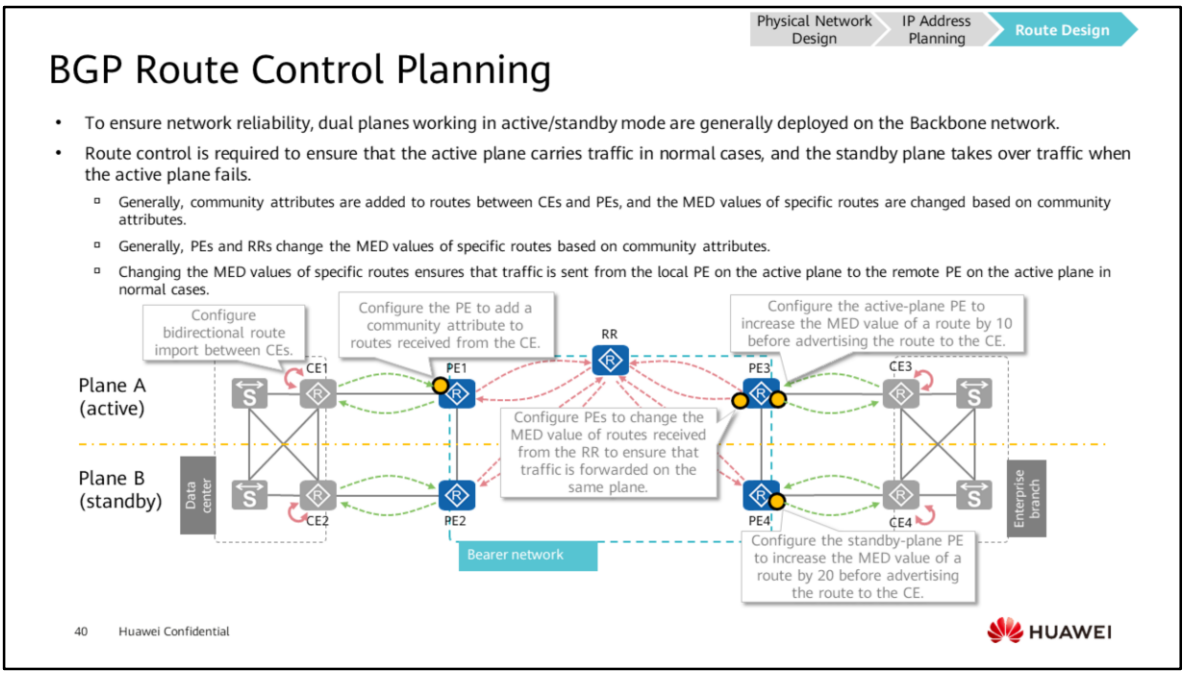
HUAWEI

- 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

- A financial cloud backbone 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.

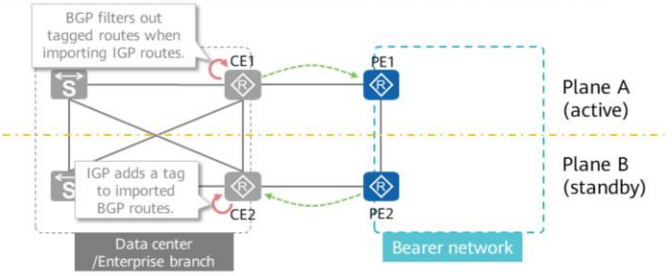




- PE3 changes the MED value to 100 for the route whose next hop is PE1 (a PE on the same plane) and changes the MED value to 200 for the route whose next hop is PE2 (a PE on a different plane).
- PE4 changes the MED value to 100 for the route whose next hop is PE2 (a PE on the same plane) and changes 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

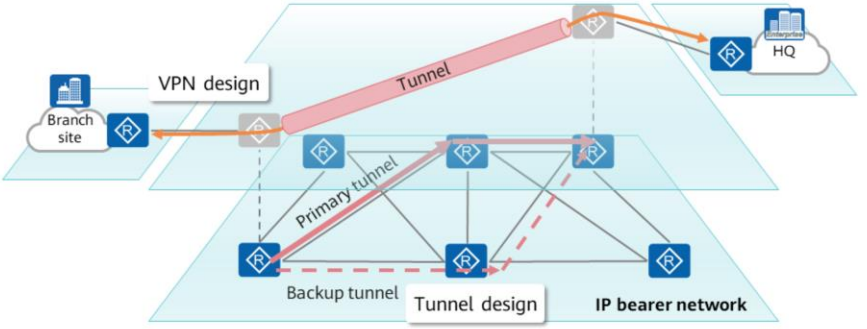
- Generally, data center networks and branch networks do not run BGP. Their IGPs need to import routes learned by EBGP.
- When an IGP imports EBGP routes, it needs to add a tag to the routes, so that the IBGP peer can filter out imported BGP routes based on tags when importing local IGP routes to BGP.



Contents

Tunnel and VPN Design Overview for the Financial Cloud Backbone Network

- Bank 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.



Tunnel Design | VPN Design

SR-MPLS/SRv6 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.

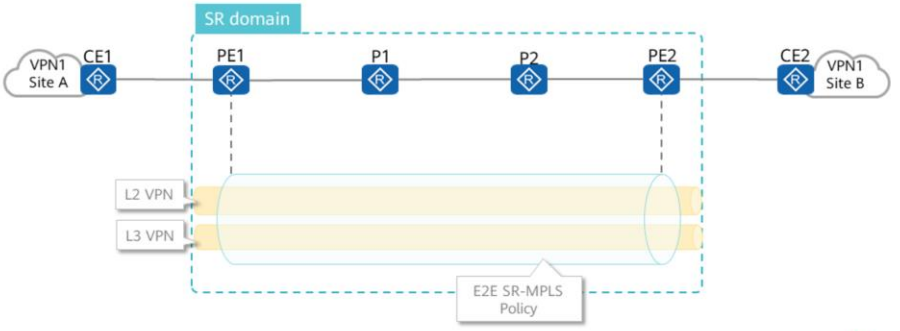
HUAWEI

- 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.

Tunnel Design | VPN Design

SR-MPLS/SRv6 Policy Deployment Design

- Generally, a financial cloud backbone 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.

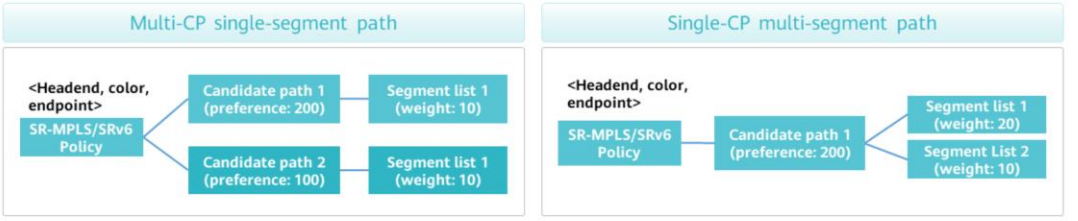


HUAWEI

- 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.

SR-MPLS/SRv6 Policy Path Planning

- SR-MPLS/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 SR-MPLS/SRV6 Policy paths are planned through the controller or static configuration, the following two modes are available (CP stands for candidate path):
  - Single-CP multi-segment path
  - Multi-CP single-segment path

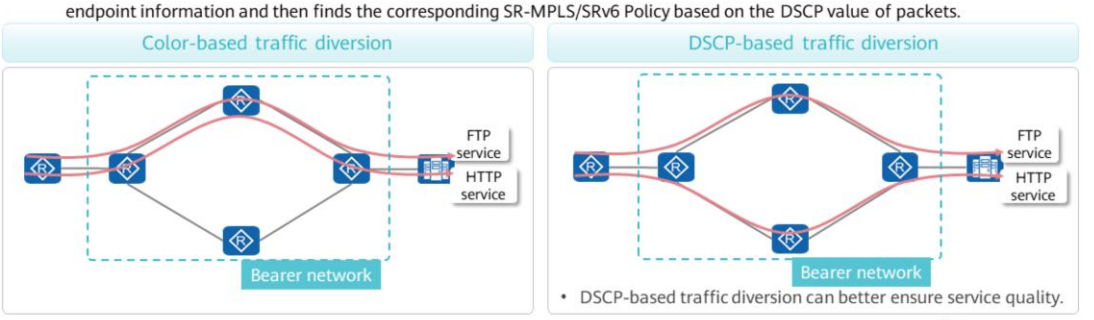


HUAWEI

- 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.

SR-MPLS/SRv6 Policy Traffic Diversion Mode Design

- SR-MPLS/SRv6 Policies need to steer service traffic into tunnels for forwarding. This process is called traffic diversion. Currently, SR-MPLS/SRv6 Policies support the following traffic diversion modes:
  - Color-based traffic diversion: In this mode, the headend steers traffic into an SR-MPLS/SRv6 Policy through route recursion implemented based on the color value and destination address in the route.
  - DSCP-based traffic diversion: In this mode, the headend searches for a matching SR-MPLS/SRv6 Policy group based on specific endpoint information and then finds the corresponding SR-MPLS/SRv6 Policy based on the DSCP value of packets.

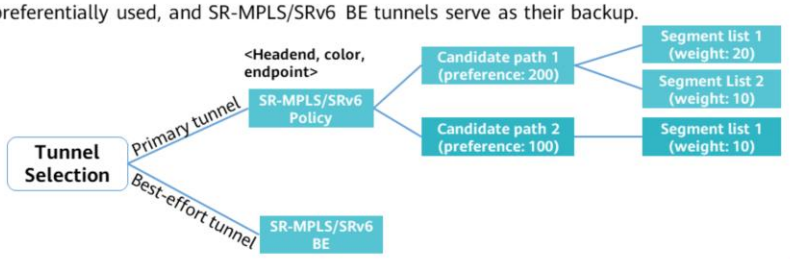


- In color-based traffic diversion, different tunnels (including primary and backup tunnels) can only be selected based on endpoints. If different service traffic (such as HTTP and FTP traffic) is destined for the same address, color-based traffic diversion diverts the traffic to the same tunnel. As 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.

Tunnel Design | VPN Design

Best-Effort Forwarding Design for SR-MPLS/SRv6 Tunnels

- VPN services can be carried over either SR-MPLS/SRv6 Policies or SR-MPLS/SRv6 BE tunnels.
- An SR-MPLS/SRv6 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/SRv6 Policy, and the valid candidate path with the second highest preference functions as the HSB path.
- If all SR-MPLS/SRv6 Policies fail, VPN services can be carried over SR-MPLS BE/SRv6 tunnels.
- It is recommended that both SR-MPLS/SRv6 Policies and SR-MPLS/SRv6 BE tunnels be deployed. SR-MPLS/SRv6 Policies are preferentially used, and SR-MPLS/SRv6 BE tunnels serve as their backup.

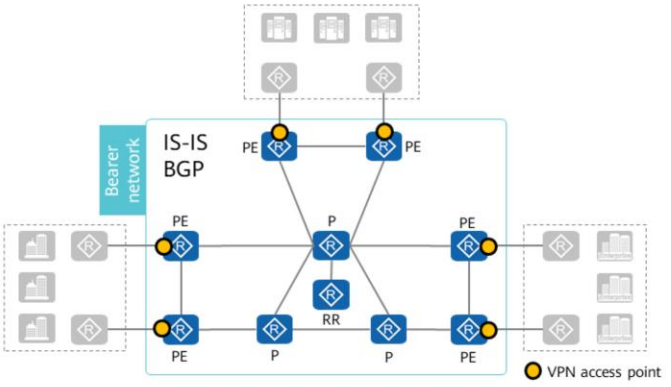


HUAWEI

- 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.

VPN Classification

- Different financial institutions have different requirements for VPN division. Most banks use one VPN for major services (including production and office services), one VPN for test services, one VPN for each branch, and one VPN for external services. If there are Internet services and public cloud services, banks will also use one VPN for each service type.



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

IPv6 VPN Type Selection

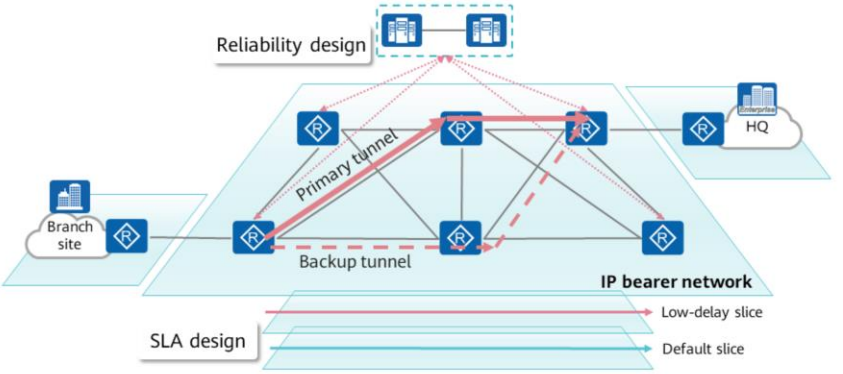
- When SR-MPLS is used to carry IPv6 L3VPN traffic, the MP-VPNv6 or BGP EVPN address family can be used to transmit VPN routes.
- When SRv6 is used to transmit IPv6 services, it is recommended that the BGP EVPN address family be used to transmit VPN routes.

	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

Contents

SLA and Reliability Design Overview for the Financial Cloud Backbone Network

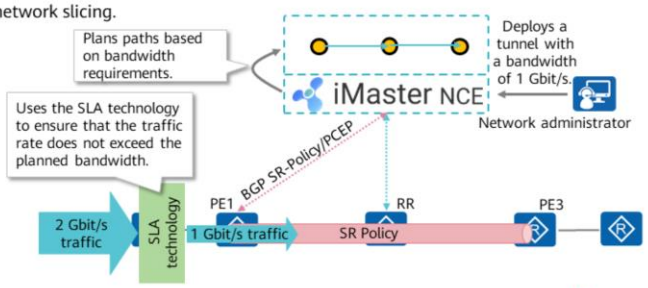
- The financial cloud backbone network uses reliability and SLA technologies to effectively ensure the quality of carried services. Therefore, reliability and SLA design are very important.



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 financial cloud backbone network needs to provide differentiated service quality assurance for different services. QoS planning ensures that various services are properly forwarded on the cloud backbone 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.

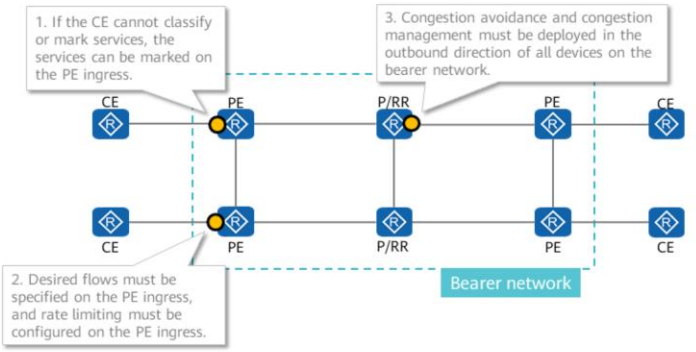
QoS Design Suggestions

- 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

QoS Deployment Design

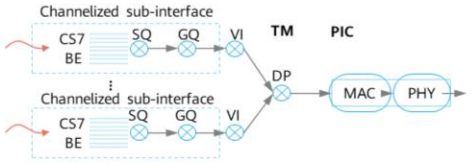
- QoS planning requires that QoS features be properly deployed at different locations on the network. The following figure shows the deployment of different features on the network:



Network Slicing Overview

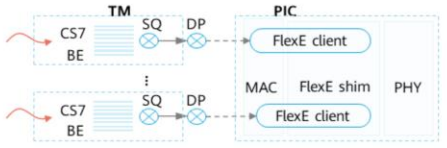
- Network slicing can be used to allocate dedicated network resources on a network to carry high-value service traffic.
- Network slicing and SR tunnels apply to different network layers. Network slicing reserves resources on the Layer 1.5 or Layer 2 network and can be used together with SR tunnels.
- 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.

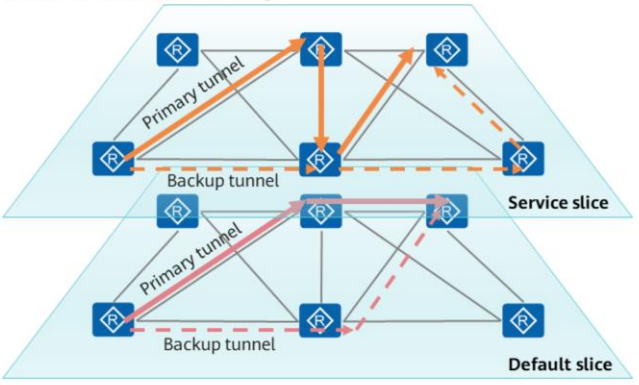
FlexE



- 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.

Network Slicing Design

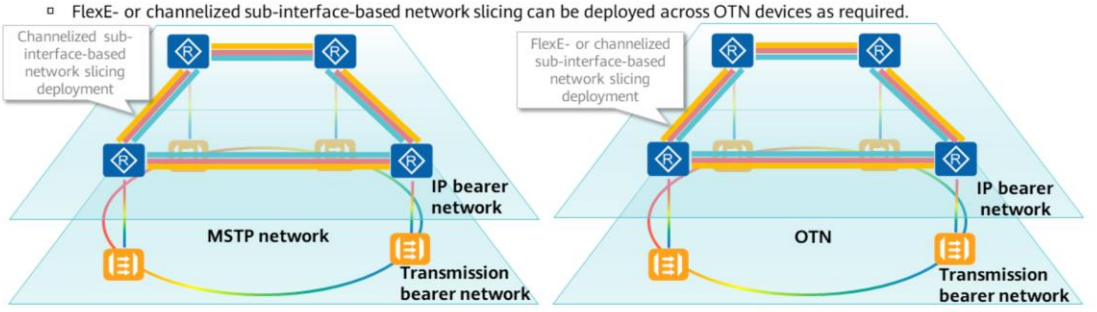
- Currently, existing devices mainly implement network slicing based on bandwidth.
- A slice with the corresponding bandwidth is created based on actual service requirements. An SR tunnel is then bound to the slice for bearing.



- 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.

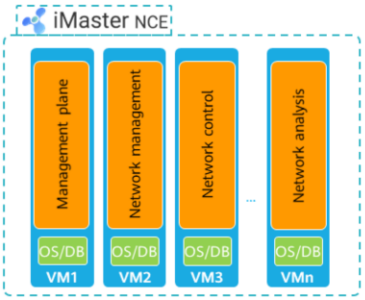
Network Slicing Application Scenarios

- 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.



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.

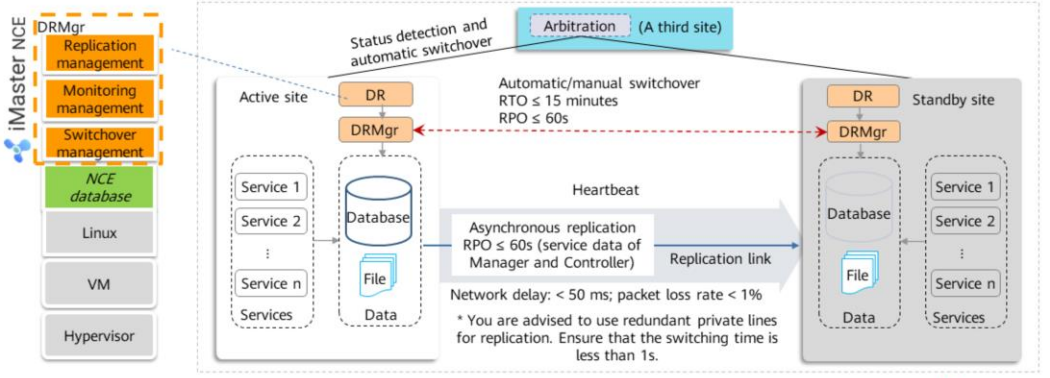


Protection Layer	Component	Remarks
Application layer	Manager	Protection mechanism: application active/standby protection Performance indicators: RPO: 0s; RTO: ≤ 5 minutes
	Controller	Protection mechanism: application active/standby protection Performance indicators: RPO: 0s; RTO: ≤ 60s
	Analyzer	Protection 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

- RPO: recovery point objective
- RTO: recovery time objective

Controller Reliability Design: Geographic Redundancy

- Two controllers with consistent hardware configuration, service schemes, and other configurations are respectively deployed at the active and standby sites. Data in each database at the active site is synchronized to the standby site in real time based on the corresponding synchronization policy. If the active site fails, the arbitration service automatically starts the standby site for controller service continuity. You can also manually start the standby site in this case.

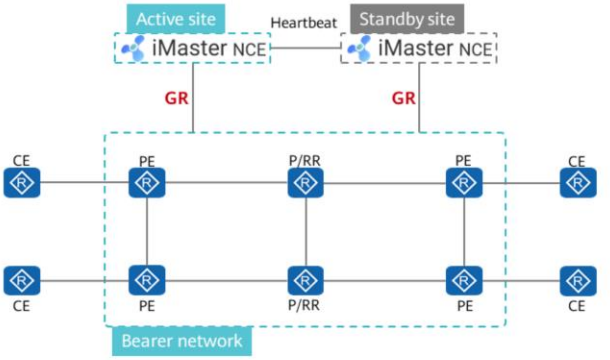


HUAWEI

- 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.

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)

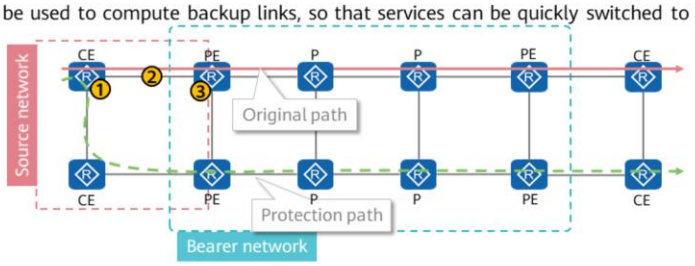
- NSR: non-stop routing

Network Reliability Solution Overview

- Multiple reliability technologies can be deployed on the financial cloud backbone 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.

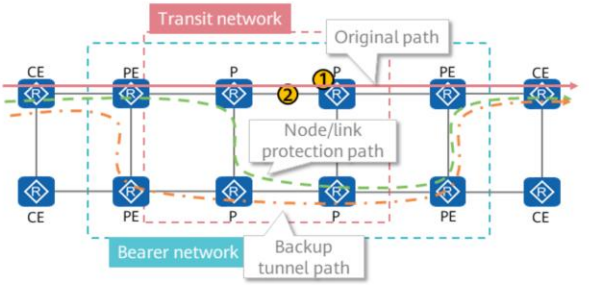
Source Network Reliability Design

- 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.



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.

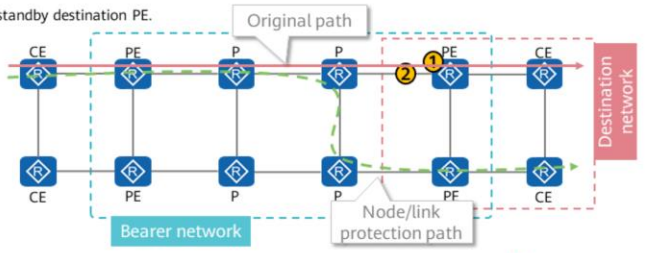


HUAWEI

- 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 (1)

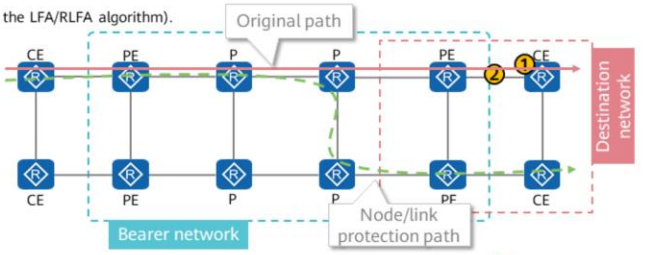
- The destination network mainly includes destination PEs, destination CEs, links between destination PEs and Ps, and links between destination PEs and destination CEs.
- The destination network may encounter the following faults:
  1. Destination PE fault
  2. Link fault between a destination PE and a P
- 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.



HUAWEI

Destination Network Reliability Design (2)

- 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 CE fault
  2. 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.
  - Compute backup paths through IP FRR (mainly based on the LFA/RLFA algorithm).

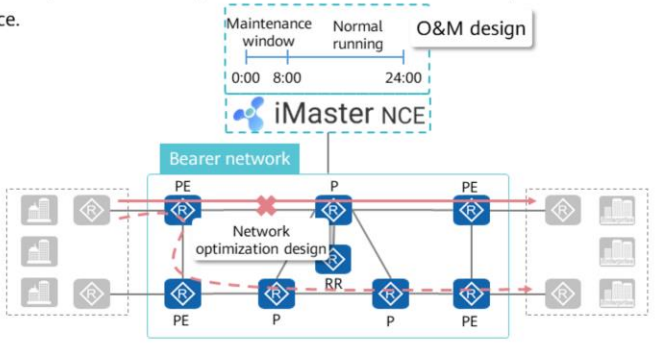


HUAWEI

Contents

Optimization and O&M Design Overview for the Financial Cloud Backbone Network

- 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.



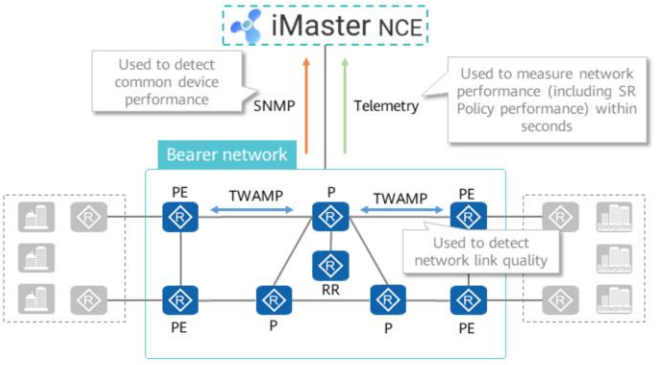
HUAWEI

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.

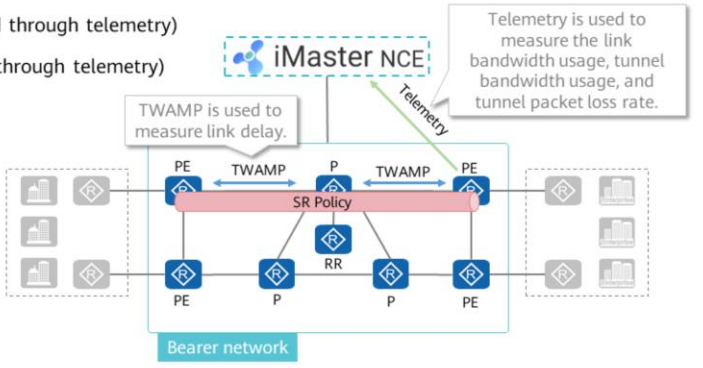
Application Scenarios of Network Performance Monitoring

- 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).



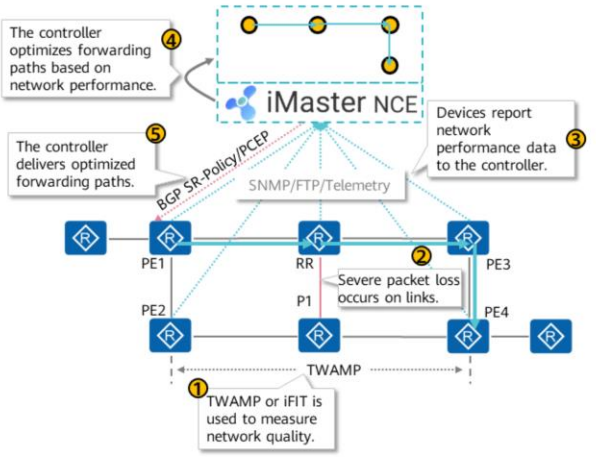
Network Performance Monitoring Design

- The major performance indicators for bank 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)



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/SRv6 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.

Automatic Tunnel Optimization Design

- If the network scale is too large for manual tunnel quality assurance, automatic optimization can be used to automatically analyze tunnel quality and optimize tunnel paths. This helps reduce manual maintenance costs and improve network optimization efficiency.
- Automatic tunnel optimization applies to the following scenarios:
  - Scheduled optimization
  - Automatic optimization upon bandwidth threshold crossing
  - Maintenance window-triggered optimization
- The optimization trigger mode can be Traffic, Delay, or Delay + Traffic.

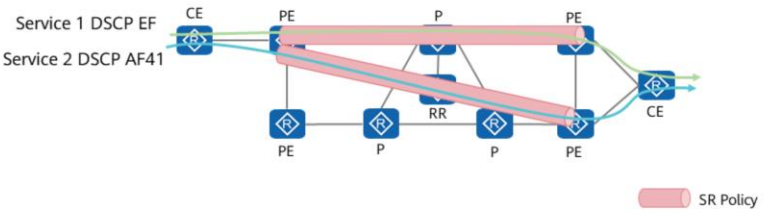
- Automatic optimization:
  - Scheduled optimization: You can set the interval for automatically optimizing network paths to 5 minutes or longer to ensure that the current service paths are optimal.
  - Automatic optimization upon bandwidth threshold crossing: You can set the link threshold. Then, when 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.

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.

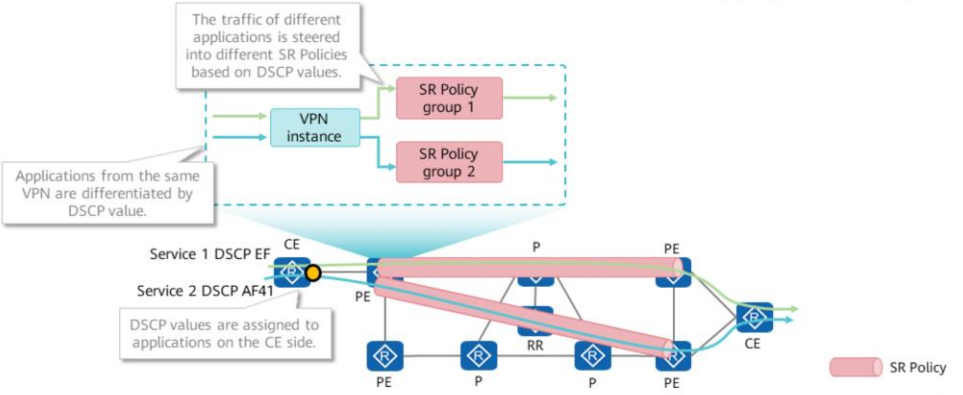
Application Optimization Overview

- 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.



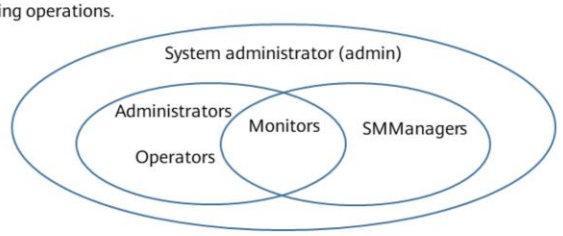
Application Optimization Design

- 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.



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.

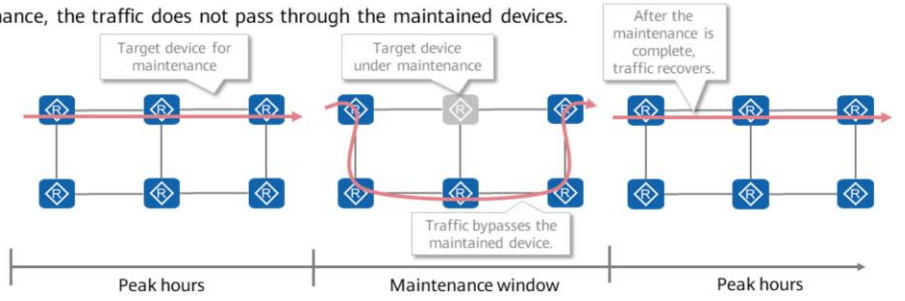


User Monitoring Overview

- 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.

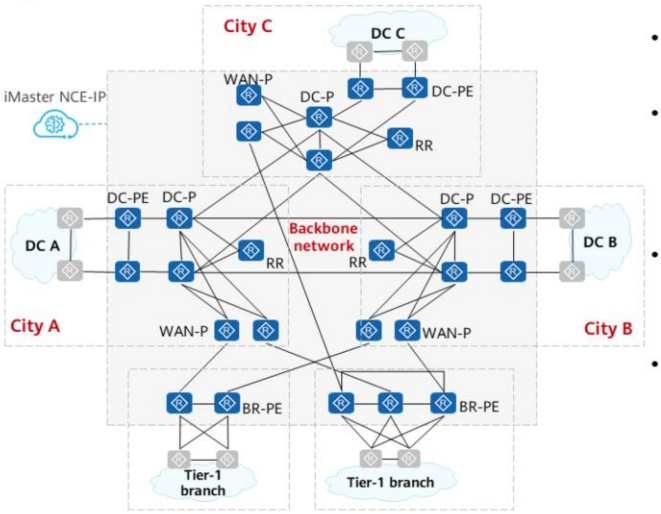
Maintenance Window Overview

- 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.



Contents

Physical Network Design for the Cloud WAN of a Bank

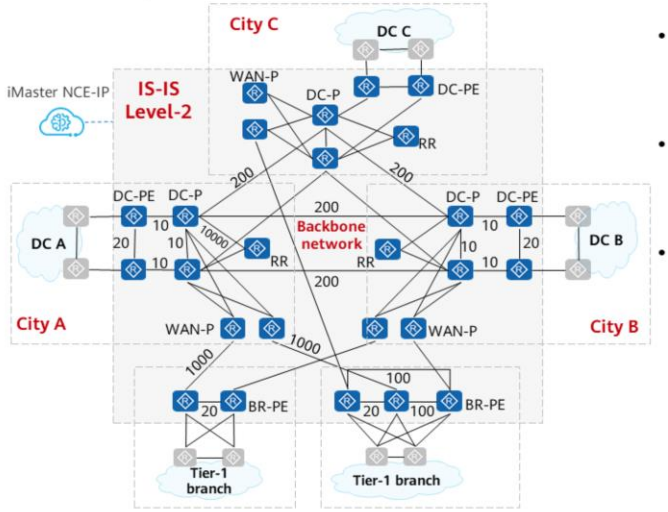


- WAN-Ps are deployed in the three DCs to aggregate uplink traffic of tier-1 branches.
- BR-PEs are deployed at tier-1 branches and connected to WAN-Ps of DCs through two or three WAN links in full-mesh mode.
- Standalone RRs are deployed for the DCs to reflect routes and work in redundancy mode.
- Dual planes are built on the backbone network to improve network reliability.

HUAWEI

- Backbone network core nodes are deployed in cities A, B, and C, and six P devices are used for high-speed interconnection.
- Two DC-PEs are deployed in cities A, B, and C, to aggregate services of DCs and intra-city organizations.
- 10 Gbit/s links are used for intra-DC interconnection. Intra-city DCs are directly connected using WDM devices and bare optical fibers. Inter-city DCs are interconnected through carrier MSTP/OTN links. Tier-1 branches are connected to DCs through MSTP links.
- IS-IS or OSPF and BGP are configured on routers on the backbone network.
- iMaster NCE-IP and tunneling technologies that can control paths are deployed to implement load optimization and visualized O&M for cloud backbone network traffic.

IGP Design for the Cloud WAN of a Bank

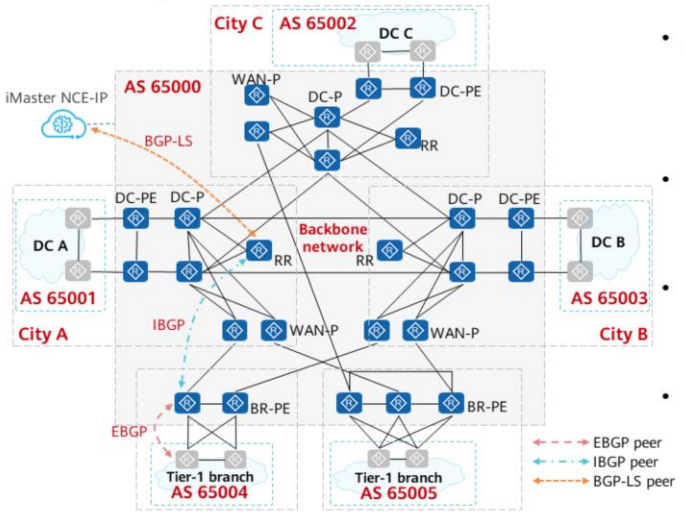


- IS-IS is used as the IGP on the backbone network, and the entire WAN is in an IS-IS Level-2 area.
- Route advertisement: Routes to the IP addresses of interconnection and loopback interfaces are advertised.
- Fast convergence: BFD is also used to implement second-level convergence.

HUAWEI

- IGP metric deployment rules:
  - Set the metric of LAN links to a value smaller than that of WAN links.
  - Set the metric of the WAN links between DCs to a value smaller than that of the WAN links between branches and DCs, preventing DC traffic detour.
  - Ensure that inter-plane traffic between DCs preferentially crosses planes on Ps.
  - Set the metric from the RR to the P to the largest value. (The RR is only responsible for reflecting route information and does not forward data.)

BGP Design for the Cloud WAN of a Bank

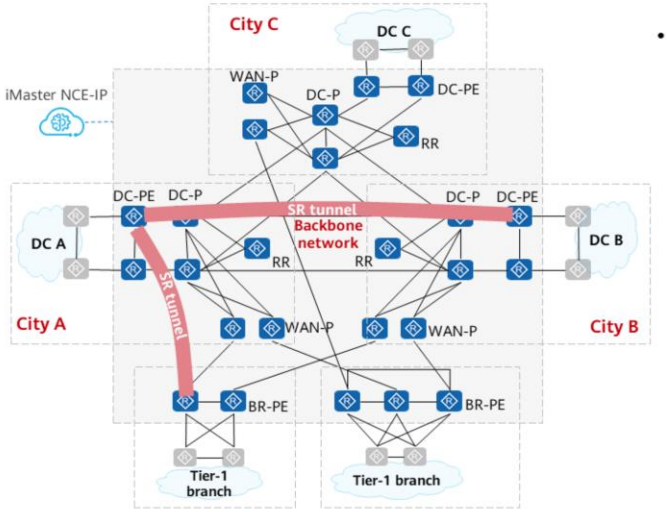


- An independent AS is assigned to the core backbone network and each DCN, service center network, tier-1 branch, and metro network.
- DC-PEs and BR-PEs function as RR clients and establish IBGP peer relationships with RRs.
- EBGP peer relationships are established between each DC/tier-1 branch and the core backbone network.
- A BGP-LS peer relationship is established between each RR and iMaster NCE-IP.

HUAWEI

- The same cluster ID is configured for RRs.
- No IBGP peer relationship is established between BGP RRs.
- With backup RRs, clients can receive multiple routes to the same destination from different RRs. The clients then use BGP routing policies to select the optimal route. The RR with a lower loopback IP address is preferred.

Tunnel Design for the Cloud WAN of a Bank

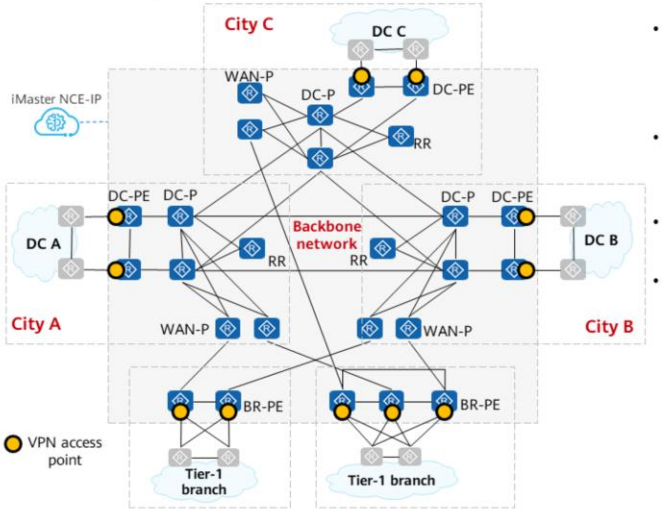


- To reduce the number of tunnels on the entire network, ensure that each DC-PE establishes SR tunnels only with the DC-PEs on the same plane or with the BR-PEs with the same odd/even number.

HUAWEI

- Number of SRv6 TE Policy tunnels:
  - Estimation model: There are three DCs, and each DC has two DC-PEs, totaling six DC-PEs. There are 40 branches, and each branch has two BR-PEs, totaling 80 BR-PEs. Two VPNs (the primary VPN is for production and the secondary VPN is for tests) require traffic optimization and path control. In the production VPN, 10 policy tunnels between two nodes are distinguished based on DSCP values in the unidirectional direction. In the test VPN, two policy tunnels between two nodes are distinguished based on DSCP values in the unidirectional direction. There are 12 policy tunnels in total for the two VPNs.
  - Number of SRv6 TE Policy tunnels between DC-PEs = 6 (local DC-PEs) x 2 (remote DC-PEs on the same plane) x 12 (tunnels in the Groups of two VPNs) = 144
  - Number of SRv6 TE Policy tunnels between DC-PEs and BR-PEs = 2 (local BR-PEs of a branch) x 3 (remote DC-PEs with the same odd/even number) x 12 (tunnels in the Groups of two VPNs) x 2 (bidirectional) x 40 (branches) = 5760

VPN Design for the Cloud WAN of a Bank



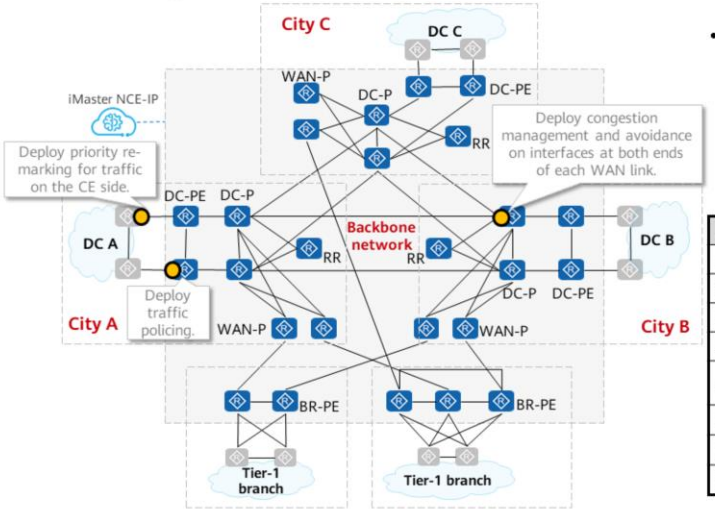
- Based on the existing services and future services on customer networks, VPNs can be planned for the following types of services: production, extranet, test, Internet, and hosting.
- DC-PEs and BR-PEs function as the access points for VPN services and provide access to VPNs for branch/DC service traffic.
- No VPN needs to be deployed on DC-Ps, which only forward services at a high speed.
- In the multi-VPN scenario, VPN routes are isolated by default. If mutual access is required, a policy can be used to control the import of routes between VPNs.

VPN Planning	VPN Type
Production service	MPLS L3VPN/SRv6 L3VPN
Extranet service	MPLS L3VPN/SRv6 L3VPN
Internet service	MPLS L2VPN/SRv6 L2VPN
Test service	MPLS L3VPN/SRv6 L3VPN
Hosting services	MPLS L2VPN/SRv6 L2VPN

HUAWEI

- Different RDs can be set for the same VPN for related purposes (for example, to implement VPN FRR).

QoS Design for the Cloud WAN of a Bank



- QoS deployment focuses on priority re-marking when external traffic is injected into the core network and QoS queue scheduling when traffic is transmitted on WAN links.

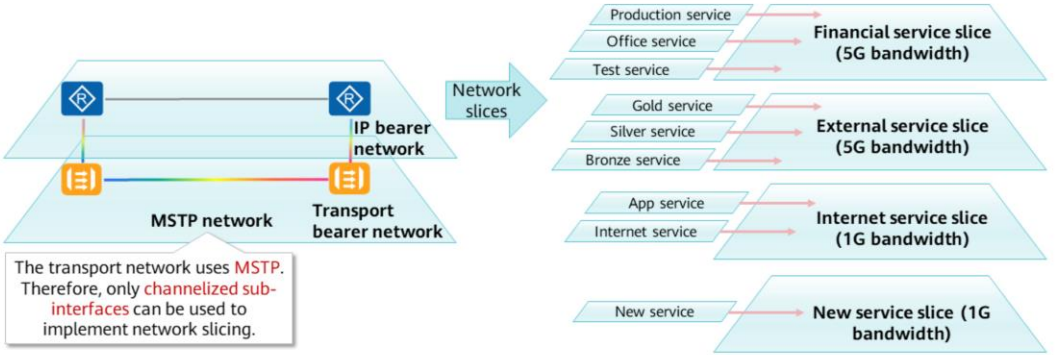
Priority	VPN Type	Scheduling Mode
CS7	Protocol packets	PQ
CS6	Protocol packets	PQ
EF	Core services	PQ
AF4	Video Services	PQ
AF3	Common services	WFQ
AF2	Office services	WFQ
AF1	Batch services	WFQ
BE	Test and others	WFQ

HUAWEI

- **Key QoS Planning Points**
  - Inherit the service importance definition of the live network and use different DSCP values to mark different types of services based on the consistency principle, facilitating refined QoS policy adjustment.
  - Keep the definitions and settings of the DSCP values of the original services as much as possible, and change the DSCP values on the core network as little as possible.
  - The IPv6 DSCP value is directly mapped to the IPv4 DSCP value.
  - Plan and design QoS based on templates to reduce the subsequent modification workload.
- **Priority re-marking:** Use multi-field classification to re-mark the QoS priority on the outbound interface of a CE based on the unified policy, and run the **trust upstream** command on other interfaces.
- **Queue scheduling:** Enable queue scheduling in the outbound direction of the WAN interfaces on Ps and PEs. The scheduling mode can be PQ, WFQ, or LPQ. PQ+WFQ is recommended.

Slice Design for the Cloud WAN of a Bank

- To meet the bearer requirements of different types of services, a physical backbone network can be divided into multiple service bearer networks through network slicing to implement hard isolation of bandwidth resources for services, such as financial services, external services, and Internet services.
- Core backbone network devices of the bank are interconnected using MSTP or OTN private lines. When MSTP private lines are leased, only channelized sub-interfaces can be used to implement network slicing.



HUAWEI

- The transparent transmission mode of OTN devices can be set to bit transparent transmission or MAC transparent transmission.
  - Bit transparent transmission mode: Each received packet is completely encapsulated (including information such as Idle/preamble) into an OTN frame.
  - MAC transparent transmission mode: Ethernet packets are compressed and information such as Idle/preamble is removed to improve transmission efficiency.
- When OTN lines are leased to connect core backbone network devices of the bank:
  - If bit transparent transmission mode is set on OTN devices, FlexE can be used to implement network slicing.
  - If MAC transparent transmission mode is set on OTN devices, only channelized sub-interfaces can be used to implement network slicing.

Quiz

1. (Multiple-answer question) Which of the following types of tunnels can be used to carry IPv6 services? ( )
   A. SR-MPLS BE tunnel
   B. SRv6 BE tunnel
   C. SR-MPLS Policy tunnel
   D. SRv6 Policy tunnel

- ABCD

Summary

- The basic network must be first designed for the financial CloudWAN, including the physical network, IPv4/IPv6 addresses, and an IGP.
- Tunnels can be established on the basic network. Such tunnels include SR-MPLS BE, SR-MPLS Policy, SRv6 BE, and SRv6 Policy tunnels. During tunnel design, pay attention to tunnel path planning, traffic diversion, and escape mode.
- After tunnels are established, they can carry VPN services. VPN planning must be based on enterprise service types and enterprise service requirements.
- Reliability and SLA assurance are also important for services. The design of QoS and network slicing can ensure service SLAs. In addition, the high reliability design of controllers, devices, and networks can ensure high reliability of services.
- After the network construction is complete, the network needs to be optimized and maintained. The optimization design needs to focus on performance monitoring and traffic optimization. The maintenance scope needs to be determined based on the user level, and the network needs to be maintained at a proper time.

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.