Technical Whitepaper on MPLS L3VPN

1 Overview

1.1 Introduction to VPN

Virtual Private Network (VPN) is not a dedicated physical network, but it is able to realize the function of a private network. The so-called "virtual" means that the user does not need to have the actual long-distance data lines, but employs the data line of a SP's off-the-shelf network (through tunnel technology).

Without network adjustment, VPN can flexible allocate user bandwidth as needed to reduce CAPEX. It looks like a user has a private network to enjoy the Internet. Physical channels are shared between VPNs and between PTN and other Internet services, and QoS is affected by traffic jitter and delay. A carrier offers Diff-Serv according to service requirements and SLA.

VPN consists of L2 and L3 VPNs. QinQ, VPLS and VPWS belong to L2VPN, and IPSec and BGP/MPLS to L3VPN.

According to network layer, VPN can be divided into CE-Based VPN (e.g., IPSec) and PE-Based VPN (VPLS, VPWS and MPLS L3VPN). As CE-Based VPN needs to implement VPN function in a user network and high requirements for user equipment and maintenance capability, it is not a fit for large-scale networking. PE-Based VPN is managed by carriers, so it has low user requirements. Currently, VPLS, VPWS and BGP/MPLS L3VPN are widely used for carrier metro network to provide VPN service for enterprise users. As BGP/MPLS L3VPN has powerful protocol support and all routes are managed by carrier networks, it is widely deployed at metro network core layer. As L2VPN L2VPN transparently transmits services via carrier networks and support flexible deployment with low CAPEX, it is frequently used at metro network access and convergence layers.

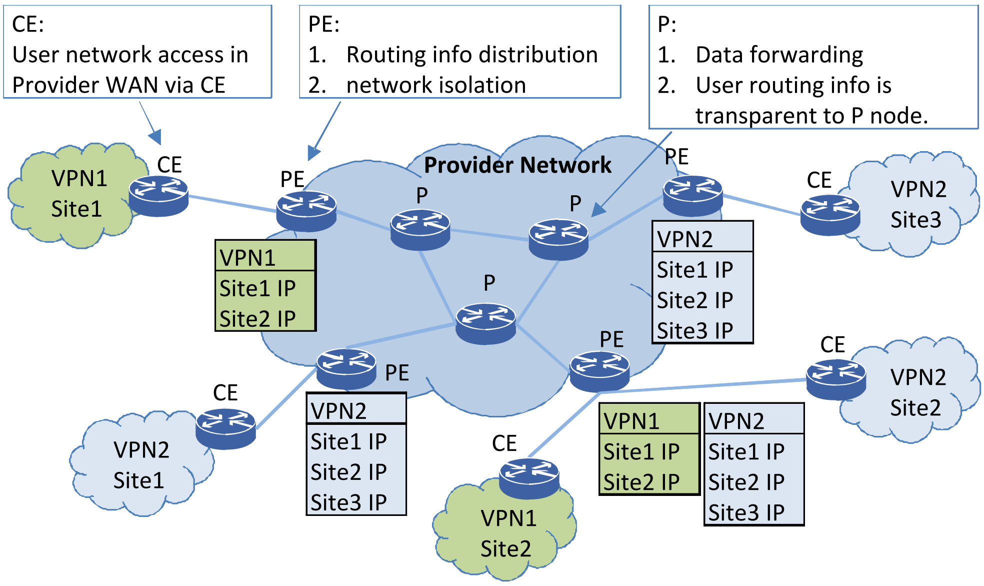
With MPLS tunnel technology, BGP/MPLS L3VPN has high security, powerful QoS and good scalability.

L3VPN services are essentially forwarded via a router. PE searches for use routes to transparently transmit data from one Site to the other one. Therefore, L3VPN bears IP data alone, while L2VPN supports a variety of services such as ATM, TDM and IP. L3VPN needs carriers to manager user routes by allocating them between PE and CE and between PE and PE and synchronizing them.

BGP/MPLSL3VPN is also known as MPLS L3VPN, and external tunnels include MPLS tunne and IP/GRE tunnel. L3VPN supporting MPLS tunnel is also called "conventional" MPLS L3VPN. IP/GRE tunnel, the useful supplement of the MPLS, is used for the networking in which carrier networks do not support MPLS. The MPLS L3VPN in this paper is the "conventional" MPLS L3VPN.

1.2 MPLS L3VPN structure

Figure 1-1 MPLS L3VPN basic structure


CE:
User network access in Provider WAN via CE
PE:
1. Routing info distribution
2. network isolation
P:
1. Data forwarding
2. User routing info is transparent to P node.
Provider Network
VPN1 Site1
CE
PE
P
P
PE
CE
VPN2 Site3
VPN1 Site1 IP Site2 IP
VPN2 Site1 IP Site2 IP Site3 IP
P
PE
CE
VPN2 Site2
PE
CE
VPN2 Site1
VPN2 Site1 IP Site2 IP Site3 IP
CE
VPN1 Site1 IP Site2 IP
VPN2 Site1 IP Site2 IP Site3 IP
VPN1 Site2

MPLS L3VPN consists of CE router, PE router and provider router. The network structure is shown in Figure 1-1. PE and provider routers, located in carrier networks, are interconnected in Full-mesh and hierarchical network topologies. PE is at carrier network edge layer and provider router inside the network. A user network is composed of VPN sites at different geographical locations. Each VPN Site is connected to carrier networks via CE router, and CE accesses PE via single or dual links, and interconnects VPN sites at different locations via carrier networks. PE allocates user routes to VPN PE and CE routers. In carrier networks, PE router maintains user routes and a VRF table for each VPN, and provider router does not maintain user routes. However, PE and provider routers both maintain a public network route table.

User sites at different locations are usually allocated to one VPN to make them communicate with each other. However, MPLS L3VPN can also allocate different departments of a user to different VPNs to isolate services, or allocate one department to several VPNs for VPN mutual access. MPLS L3VPN has strong user isolation flexibility to meet the requirements of different users in service security and flexible networking.

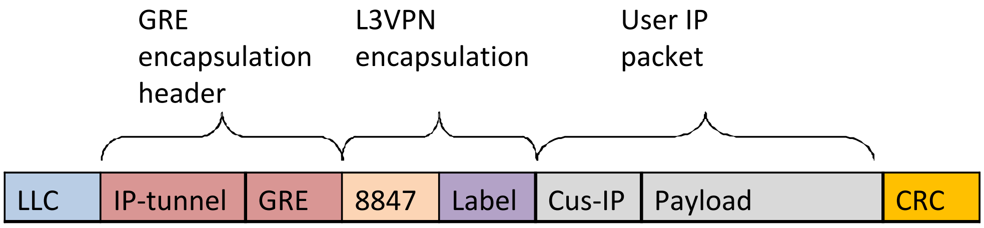
1.3 MPLS L3VPN tunnel

MPLS L3VPN uses MPLS tunnel and GRE/IP tunnel. A tunnel isolates a user route from a provider router. Provider router is only concerned with a public network route rather than a user route. Transparent transport of user data via tunnel can reduce route performance requirements for provider routers. With powerful LDP and RSVP-TE support, it is easy for MPLS to create and maintain a tunnel. With weak protocol support, tunnel management is complex for GRE. An IP network not supporting MPLS can transport VPN service via a GRE/IP tunnel to avoid the upgrade to put a cost pressure on the entire network. A network supporting MPLS can enable LDP, RSVP-TE or static tunnel for MPLS L3VPN functions.

1.3.1 GRE tunnel

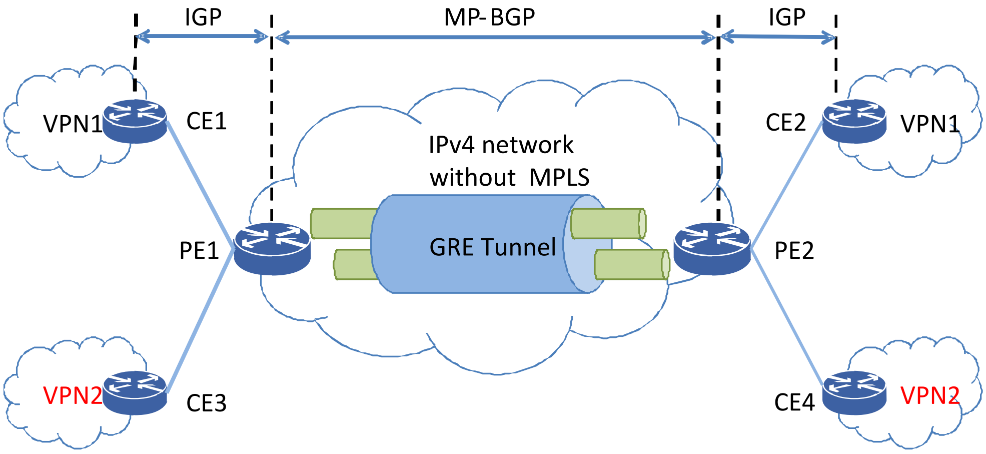
A GRE tunnel-based service packet is shown in Figure 1-2:

Figure 1-2 GRE tunnel-based L3VPN packet encapsulation format



- LLC: Link-Layer Communication header. It is a MAC header for Ethernet. The 12-byte header complies with Ethernet-II specifications.
- IP-tunnel: IP header in a GRE tunnel. An IPV4 tunnel has 20 bytes.
- GRE: GRE information header. 8~16 bytes based on encapsulation.
- Label: VPN label. A MP-BGP label is allocated together with a route. Each route, VPN and interface can be labeled. With the MPLS ID 884, it has 6 bytes.
- Customer-IP: User IP packet header.
- Payload: User IP packet data.

Figure 1-3 BGP L3VPN over GRE network model



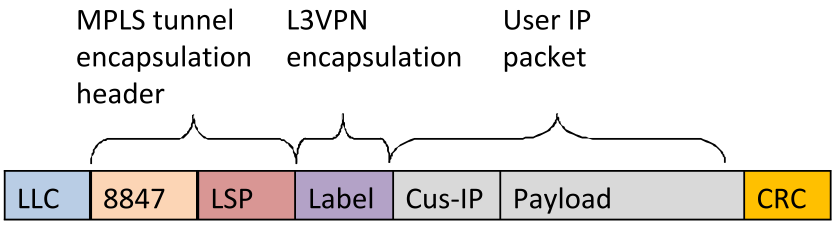
VPN over GRE network model is shown in Figure 1-3. A GRE tunnel is created between PE-1 and PE-2. When CE-1 sends an IP packet to CE-2, PE-1 searches for the VRF route table, inserts a VPN label to the packet, encapsulate it with a GRE tunnel, and transparently transmit the VPN data to PE-2 via a GRE tunnel. The middle routers forward the route according to the IP address in GRE encapsulation header. After receiving the packet, PE-2 strips the GRE encapsulation, obtains the VRF instance according to the VPN label, searches for the VPN1 VRF route table, and forwards the packet to CE-2.

In VPN over GRE, MP-BGP allocates inter-PE VPN routes and the allocated VPN route has a private network label.

1.3.2 LSP tunnel

LSP tunnel-based VPN service packet is shown in Figure 1-4:

Figure 1-4 LSP tunnel-based L3VPN packet encapsulation format



- LLC: Link-Layer Communication header. It is a MAC header for Ethernet. The 12-byte header complies with Ethernet-II specifications.

- LSP: A public network label in a MPLS tunnel. The length is 4 byte. With the MPLS ID, it has 6 byte.

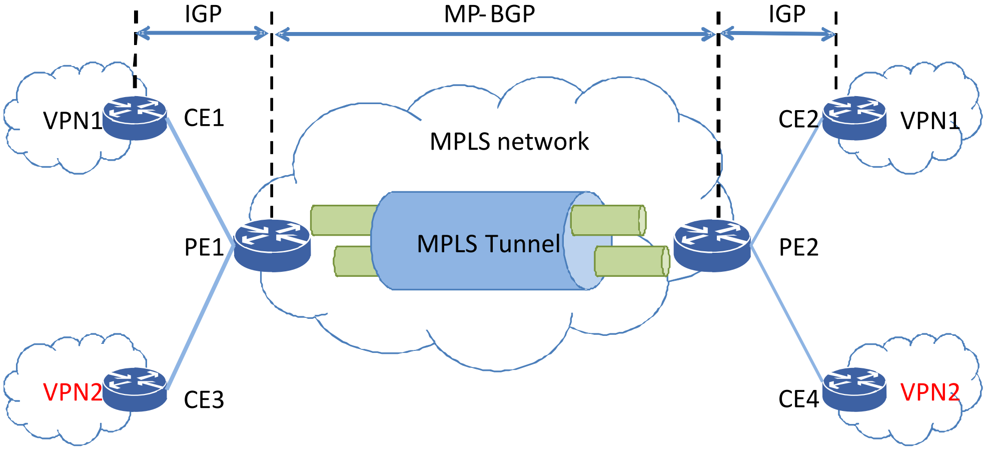
- Label: VPN label. A MP-BGP label is allocated together with a route. Each route, VPN and interface can be labeled. It has 4 bytes.

- Customer-IP: User IP packet header.

- Payload: User IP packet data.

When LSP tunnel bears VPN data flow, it adds 4-byte public network label. GRE tunnel needs to add at least 28 bytes, so LSP tunnel is more efficient. MPLS label nesting can easily implement VPN nesting network.

Figure 1-5 Network model based on BGP L3VPN over LSP tunnel



VPN over LSP network model is shown in Figure 1-5. A LSP tunnel is created between PE-1 and PE-2. When CE-1 sends an IP packet to CE-2, PE-1 searches for the VRF route table, inserts a VPN label to the packet, encapsulate it with a LSP tunnel, and transparently transmit it to PE-2. The middle routers forward the route according to the public network label in LSP encapsulation header. After receiving the packet, PE-2 strips the LSP encapsulation, obtains the VRF instance according to the VPN label, searches for the VPN1 VRF route table, and forwards the packet to CE-2.

Regardless of MPLS tunnel or GRE/IP tunnel, L3VPNs are the same in PE route allocation and VRF route management, and the difference is L3VPN data flow bearing mode. The former bears data via LSP tunnel and the latter via GRE/IP tunnel. LSP bears data packet more efficiently, uses smaller bandwidth, provides higher QoS, and supports overlay networking. This paper mainly introduces MPLS tunnel based L3VPN and is also called MPLS L3VPN.

1.4 MPLS L3VPN advantages

- Address overlay: Individual VRF has independent address space, and users employ one IP address to save IP address resources. Meanwhile, one VPN site can be allocated to multiple VPNs to ensure security between user departments and support cross-VPN interworking and flexible VPN control.

- Standardized protocol: MPLS L3VPN use IGP, MP-BGP and MPLS protocols to allocate routes and labels. These protocols are more mature.

- A variety of network topologies: L3VPN supports P2P, Full-mesh, HoVPN, VPN overlay and Hub-Spoke network topologies.

- Powerful QoS: L3VPN manages user bandwidth through TE tunnel technology and provides Diff-Serv QoS through EXP to support the refined QoS.

- High network scalability: L3VPN supports hierarchical VPN deployment and route protocol allocation on demand to reduce the requirements on the PE route storage capacity.

- Strong access: One VPN supports several access nodes as well as sub-interfaces to access more users. Each VPN and each route can be labeled to support more instances.

- Low user requirements: User IP addresses can be managed and planned uniformly by carrier PE routers, so there are low requirements for user network management capability.

- High reliability: L3VPN supports equipment, tunnel and service protections to meet the carrier-class reliability requirements.

2 Technical principle

2.1 Implementation principle

2.1.1 MPLS L3VPN basic principle

L3VPN forwarding instance model is shown in Figure 2-1. VPN A and VPN B access different PEs. PEs form two independent forwarding tables VFI of VPN A and VFI of VPN B which are logically isolated from each other. VPN A users forwards routes only in the VFI of VPN A to isolate services between VPNs.

PE connect left and right CEs via a hierarchical tunnel to transparently transmit user traffic via the tunnel. User flow contains two tunnel multiplexing: from IP to VPN tunnel, and from internal tunnel to MPLS tunnel. The internal VPN tunnel logically connects VFI instances and the external tunnel connects the pipe between PEs to multiplex from VPN tunnel to MPLS tunnel.

CE learns the routes of other CEs in one VPN through PE, and PE maintains a VRF route table for each VPN. VRF includes the local CE route and the routes of other CEs introduced through PE.

As shown in Figure 2-1, when the left CE sends data flows to the right one in one VPN, service forwarding is as follows:

1 CE searches DIP address and next hop and forwards IP packets to the left PE through the local connection.

2 PE finds the CE VFI instance according to the data flow receiving interface and obtains the DIP address, LSP and next hop of the right CE from the VFI instance. PE encapsulates IP packets with VPN tunnel and MPLS tunnel labels, modifies Ethernet header according to public network next hop and forwards the packets to VPN backbone network.

If a provider router is available in the backbone network, the router forwards the packets via LSP.

If LSP starts PHP, the VPN packet by the right PE only contains the VPN tunnel label, or external MPLS label (LER attribute label or explicit "0" label) and VPN tunnel label. No matter how many labels the received packet includes, PE finally finds the VPN tunnel label and obtains the VPN VFI instance according to the label. It searches the route forwarding table and next hop in the VFI instance, strips the label, recovers the IP packet, modifies the Ethernet header according to the next hop and sends it to the right CE.

Figure 2-1 MPLS L3VPN service model

```
               +----------+                  +----------+
+-----+        |PE device |                  |PE device |        +-----+
| CE  |        |          |                  |          |        | CE  |
| dev | Access | +------+ |                  | +------+ | Access | dev |
| of  |  conn. | |VFI of| |                  | |VFI of| |  conn. | of  |
|VPN A|----------|VPN A | |                  | |VPN A |----------|VPN A|
+-----+        | +------+\|      Tunnel      |/+------+ |        +-----+
               |          >==================<          |
+-----+ Access | +------+/|                  |\+------+ | Access +-----+
| CE  |  conn. | |VFI of| |                  | |VFI of| |  conn. | CE  |
| dev |----------|VPN B | |                  | |VPN B |----------| dev |
| of  |        | +------+ |                  | +------+ |        | of  |
|VPN B|        |          |                  |          |        |VPN B|
+-----+        +----------+                  +----------+        +-----+
```

MPLS L3VPN supports both P2P topology and MP2MP topology. Different sites in one VPN can be connected to one PE via different interfaces. Therefore, the above forwarding flow is just a general flow. In some cases, a packet can directly be forwarded to CE via PE.

2.1.2 MPLS L3VPN characteristics

1 MPLS VPN is structured by SP. In the network structure, SP provides a user with VPN service. The user cannot feel the presence of the public network just like he has independent network resources.

2 P router in SP backbone network is not directly connected to CE and does not know the presence of VPN. It just transmits data in the backbone network. It must support and enable the MPLS protocol.

3 PE structures, connects and manages all VPNs. It is located at the edge of a SP network. For PE, a connected user IP system is considered a site. Each site is connected to PE via CE. It is the basic unit of a VPN.

4 One VPN consists of several sites, and each site can belong to several VPNs at the same time. Two sites in one VPN are connected to each other via a SP public network. VPN data is transmitted via the public network and data transport must be private and secure. In other words, a site in a VPN can only forward a packet to another site in the same VPN instead of in other VPNs.

5 Meanwhile, any two VPNs without common sites can use an overlapping address space, namely, a user employs an independent address space in its private network and does not need to consider whether the space conflicts with address spaces of

other VPNs or public networks. All this depends on VRF (VPN Routing & Forwarding Instance).

2.1.3 Issues solved by MPLS L3VPN

The above MPLS L3VPN basic principle shows that relative to ordinary IPV4 management, MPLS L3VPN needs to solve the following issues:

1 VRF route table management: How to isolates different VRF route tables in one PE? A feature of VPN route is that a user can employ a private IP and different VPNs can employ the same IP address to save public network IP resources, thus a VPN route table needs to distinguish the same route of different VPNs in one PE.

2 VPN route allocation label: It can ensure that two same routes are transmitted in the network and the receiver can distinguish them from each other.

3 Packet forwarding: Even if route table conflict is resolved, when PE receives an IP packet, how should the PE know which VPN it sends the packet to? Because the only available information in the IP header is the destination address and the address may be available in many VPNs.

VPN allocates an separate table space to a VPN through VRF instances, and different VPN table spaces do not overlay. Public network IP and VRF are also independent spaces. VRF includes its own route protocol. It can learn the route released by local CE through IGP route protocol and the local VPN route synchronized by peer PE through MP-BGP to physically isolate different VPN route tables.

For a CE-released route, different VPNs have different ACs and independent IGP spaces. PE can distinguish which VRF a route is sent into. For a route released by PE to PE, all VPNs are in one MP-BGP domains. When the route is released, a RD ID identifying a VPN space is released together with IP address so that PE is able to use the RD to find out which VRF the IP comes from.

A forwarding plane also uses a field to identify a VRF address space. RD is an 8-byte ID. If RD directly identifies a VRF table space at the forwarding layer, the memory is wasted and the algorithm is complex. Control Plane (CP) allocates the locally unique VRF-ID to an instance and CE accesses VPN via an AC interface. One AC belongs to only one VPN, so VRF-ID can be saved into the AC interface table. The VPN packet encapsulation received at network side has only a VPN label. One VPN label belongs to only one VRF, so VRF-ID can be saved into the VPN label. In the forwarding, a VRF instance can be obtained from the AC or VPN label to search the route in the VRF route space for forwarding.

2.2 Route protocol

2.2.1 RD

PEs switch VPN routes via a public network. Ordinary IGP cannot distinguish overlapping IP addresses. Inter-PE route notification cannot adopt ordinary address structure and route protocol, so the concept of RD (Route Distinguisher) is introduced.

RD identifies VRF route space. Each VRF has a RD. RD is unique in a backbone network and different VPNs cannot use the same RD value to distinguish inter-VPN overlapping IP addresses. RD format is shown in Figure 2-2, and has 8 bytes. The value supports two formats, as shown in Table 2-1.

Figure 2-2 RD format definition

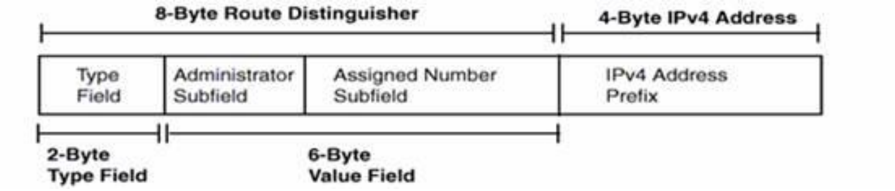


Table 2-1 Definition format of value domain

TYPE(2 byte)	Administrator Field	Assigned Number Field
0	2-byte AS No.	4-byte allocation No.
1	4-byte IP address	2-byte allocation No.

PE routers use BGP to release VPN routes between each other, and standard BGP installs and releases only one route for each IP prefix. Each VPN has its own address space, which means that one IP address can be used by several destination VPNs and the address in each VPN refers to a different system. Therefore, BGP should be allowed to install and release several routes for the same IP prefix of each VPN and use a specific policy to decide which route is used by which site. Thus, the multi-protocol BGP employs a new address family VPN-v4 address.

A VPN-v4 address has 12 bytes: 8 bytes for RD and 4 bytes for IP address. If two VPNs use the same IP address, PE router adds different RDs to convert it into the unique VPN-v4 address, avoiding the address space conflict.

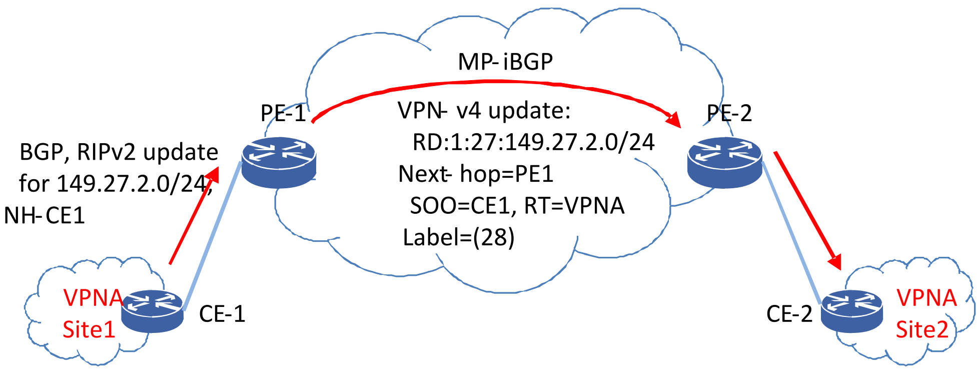
A VPN-v4 address can resolve the issue of address conflict when VPN route is transferred via a public network, but it is not the address structure of the original IP address family, and cannot be transmitted through an ordinary route protocol. Meanwhile, each user network is an independent system. It is obviously unsuitable to transfer and use IGP through SP route information. BGP should be extended to bear a new VPN-v4 address family route and transfer the RT attribute attached to the route.

RD and VPN-IPv4 address resolve the problem of how the receiver distinguishes two same routes transferred via the networks.

2.2.2 RT (Route Target)

PEs switches VPN information between each other. BGP UPDATE packet bears VPN-v4 address family routes. It is the extended BGP: MP-BGP (Multi-Protocol BGP). MP-BGP bears not only IPv4 route but also VPN, IPv6 and multicast routes. MP-BGP specifies the next hop and NLRI (network-layer information) of a specific network-layer protocol for routes.

Figure 2-3 Application of RT extension attribute in route allocation



BGP extension community attribute extends community attribute by expanding value domain and specifying internal structure. An extension community attribute is a transitional optional attribute. It consists of a set of extension communities. RT attribute is represented by BGP extension community attribute.

A RT attribute has 8 bytes and has two formats, as shown in Table 2-2.

Table 2-2 RT format

TYPE(2 byte)	**Administrator Field**	**Assigned Number Field**
0x0002	2-byte AS No.	4-byte allocation No.
0x0102	4-byte IP address	2-byte allocation No.

- RT attribute

VPN membership is obtained through a RT attribute. A VPN instance of each PE has one or several RT attributes, and RT attribute varies with VPN PEs. RT indicates which VRF receives the route and whose VRF route is received.

A route with such an attribute must be sent to all PE routers specified by RT. After PE receives the route including the attribute, if the route specified by the attribute is consistent with its own route, the route is added into the route table.

RT is in nature a way in which each VRF expresses its route preferences. It consists of Export Target and Import Target. The former refers to the attribute of the route sent, and the latter to the route interested. For example:

Site-A: The route sent is red, and only the red one is received.

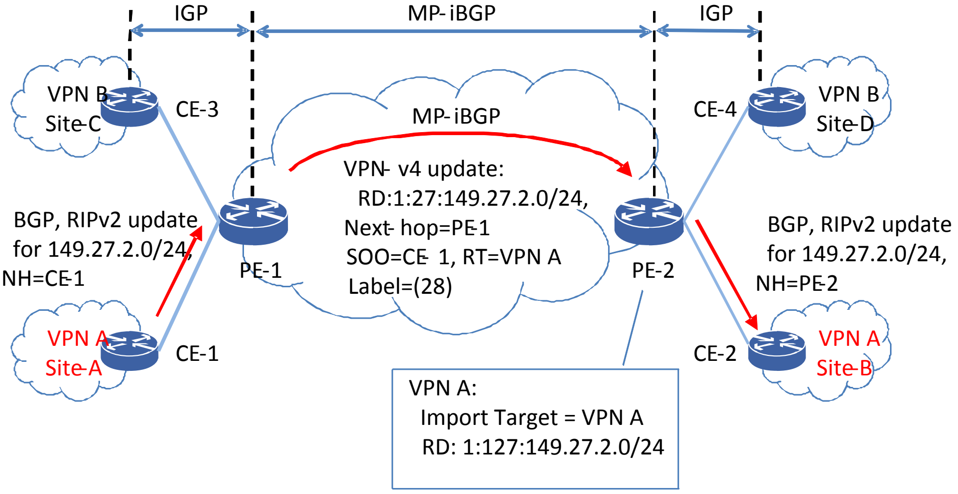
Site-B: The route sent is red, and only the red one is received.

Site-C: The route sent is black, and only the black one is received.

Site-D: The route sent is black, and only the black one is received.

Thus, site-A and site-B only have their own routes which support the mutual access. Site-C and site-D are similar. Site-A and site-B can be called the VPN-A, while site-C and site-D the VPN-B, as shown in Figure 2-4.

Figure 2-4 PE compares Export Target and local VPN Import Target receiving VPN route



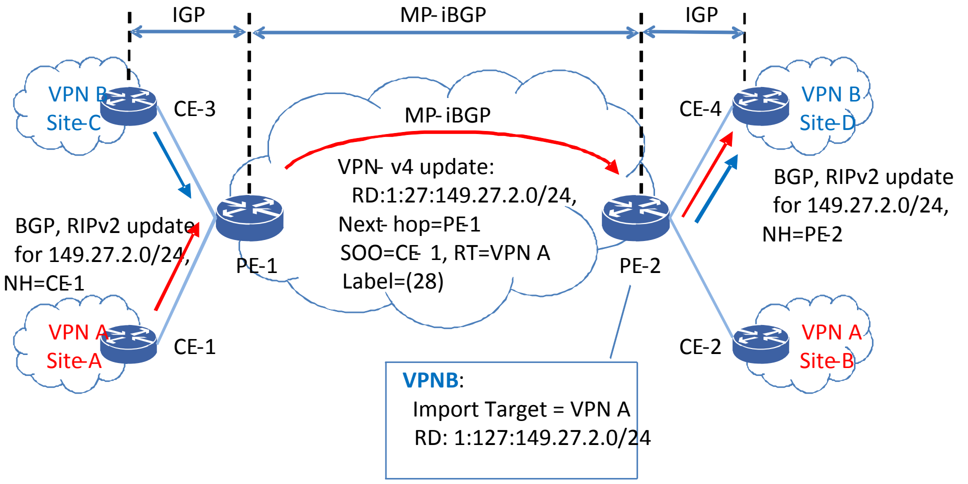
For a PE, the attribute collection of one RT is attached to the route received from a site, so the RT is called Export RT. The attribute collection of the other RT decides which routes can be introduced into the route table of the site, so the RT is called Import RT. They are different collections, and their combination can structure a VPN of any topology.

In PE, each VRF has an Import RT list. Only when Export RT of a route is consistent with the VRF Import RT list, can the route be introduced into the route table of the VRF.

- Flexible use of RT

VRF Export Target and Import Target can be configured with several attribute values, for example, I'm interested in the red or blue route. As it is a "or" operation, red, blue and red & blue routes can be received. Thus VPN access can be controlled flexibly. Refer to site-A sending a route to site-D in Figure 2-5.

Figure 2-5 RT control cross-VPN route import



2.2.3 Private network label

When released, a route has a RD. RD can theoretically identify a VPN data flow, but a RD has 64 bits which reduces the forwarding efficiency, so only a shorter, fixed-length distinguisher is needed.

The VPN data flow encapsulation is transparently transmitted via a VPN tunnel. Refer to the above VPN service model. The VPN tunnel can totally isolate data flows of different VPNs from each other to make data safer and QoS mechanism more flexible and provide pipe service for users without user QoS.

There are a variety of tunnels such as GRE, IP tunnel, MPLS tunnel and L2TP tunnel. The following factors should be taken into account to select a kind of tunnel:

- Whether the payload capacity is large and the introduced tunnel encapsulation header is as small as possible.
- Whether QoS capability is strong.
- Whether tunnel egress can easily get VRF information from tunnel information.
- Whether a tunnel can be created easily, protocol support is enough, and it is easy to fulfill P2P and MP2MP networking.
- Whether hierarchical tunnels are easy to implement for tunnel nesting and complex network topology.

- Whether tunnel scalability is high.

Taking these factors, MPLS tunnel technology has more advantages to meet the requirements of the VPN tunnel. MPLS L3VPN introduces a MPLS label to identify a VPN tunnel. The label is a private network label and the label for LSP is a public network label. The private network label is allocated by MP-BGP along with a VPN route, and the private network label and MPLS public network label share a label space.

- A private network label is allocated in three ways:

1. A label per VRF: PE allocates only one private network label to each VRF, and the routes from one PE are multiplexed into one private network label. The way saves label resources and is conducive to the separation of route and label. If VPN routes are separated from labels in FRR, better convergence is available and there is the multilayer QoS management for different VPN users according to VRF.

2. A label per AC: A private network label is allocated according to the logic interface bound to VRF. The way forwards data according to the label to avoid searching an IP address. However, once the user interface changes, the private network label needs to be reallocated, and user side and network side combine closely to lower the scalability.

3. A label per route: When MP-BGP allocates routes to other PEs, each subnet route has a label, and label and subnet route are in one-to-one correspondence. Thus label resources are wasted. Once network side changes, VRF routes need to be reconverged, which slows the route convergence speed.

2.2.4 MP-BGP

BGP and IGP are different in controlling route transport and selecting the best route instead of finding and calculating a route. VPN utilizes a public network to transfer VPN data, while the public network employs IGP to find and calculate a route. The key to create a VPN is to control VPN route transport and select the best inter-PE route.

In order to support multiple network-layer protocols, IETF extends BGP-4 by MP-BGP (Multi- protocol Extensions for BGP-4). MP-BGP enhances BGP-4 functions to enable

BGP to offer route information for multiple route protocols, including the support to IPv6 (BGP4+), multicast and VPN.

MP-BGP information is added with the extended NLRI (Network Layer Reachability Information). NLRI is added with the description about address family, private network label and RD. Refer to for them, Table 2-3 and Table 2-4 for RT list.

The BGP using the extended attribute MP_REACH_NLRI is called MP-BGP.

Table 2-3 NLRI definition

MP_REACH_NLRI:	
address-family :	VPN-IPV4 address family
next-hop:	PE router. It is usually a loopback address.
NLRI:	
label:	It has 24 bits. It has no TTL, the same as a MPLS label.
prefix:	RD:64bit+ip prefix

RT list : (Export Target attribute list)

Table 2-4 RT list

Extended_Communities (RT1)
Extended_Communities (RT2)
Extended_Communities (RT3)

2.3 Route allocation and learning

As MPLS L3VPN adopts the dual-label structure, P router exchange no route information with VPN. Through route exchange between CE and PE and between PE and PE, VPN sites form the network topology information belonging to a VPN.

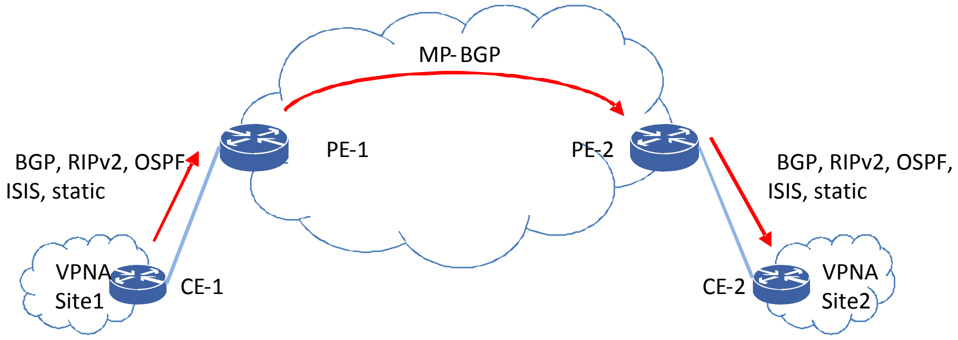
There are three steps here. Refer to Figure 2-6 network topology:

1 CE to PE route switching

2 PE to PE route switching

3 PE to CE route switching

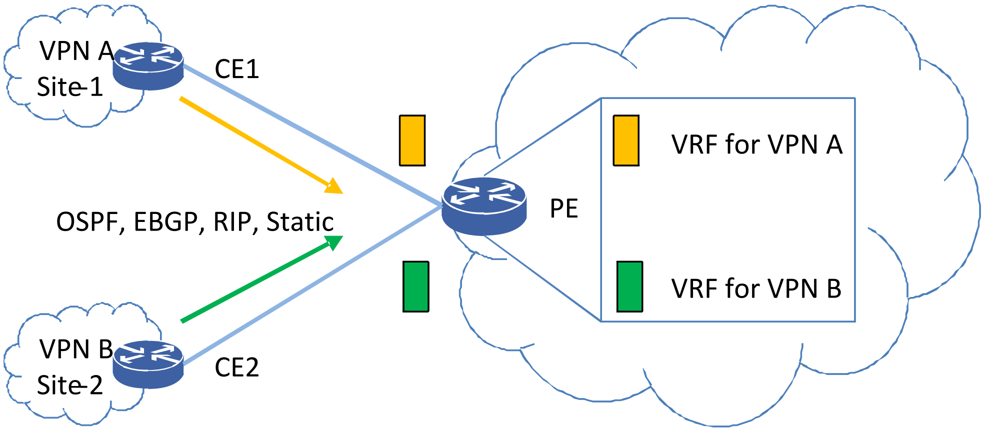
Figure 2-6 MPLS L3VPN route allocation protocol



CE and PE notify routes between them through IGP (OSPF, RIPv2, ISIS) or EGP (BGP), or statically configure them. PE and PE notify routes between them through MP-BGP (IBGP or EBGP).

2.3.1 CE to PE

Figure 2-7 CE releases a route to PE

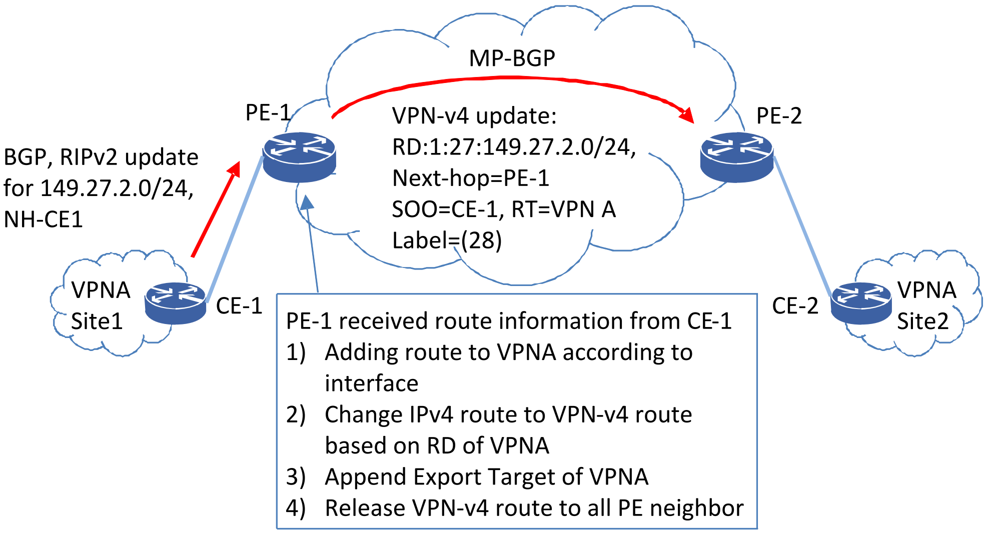


CE releases a route to PE through route protocols, or a static route to CE is manually configured on PE. Refer to Figure 2-7 for details. The specific mode depends on CE route scale and networking mode. CE and PE can employ intra-IGP or inter-BGP to notify a route, and all AC interfaces between PE and CE need to separately start independent route protocols.

CE notifies the standard V4 route including no VPN information. When receiving the route notification from an interface, PE determines which VRF the route is sent into according to the AC interface receiving the packet and route protocol. A static route should be assigned with a VPN instance in the configuration.

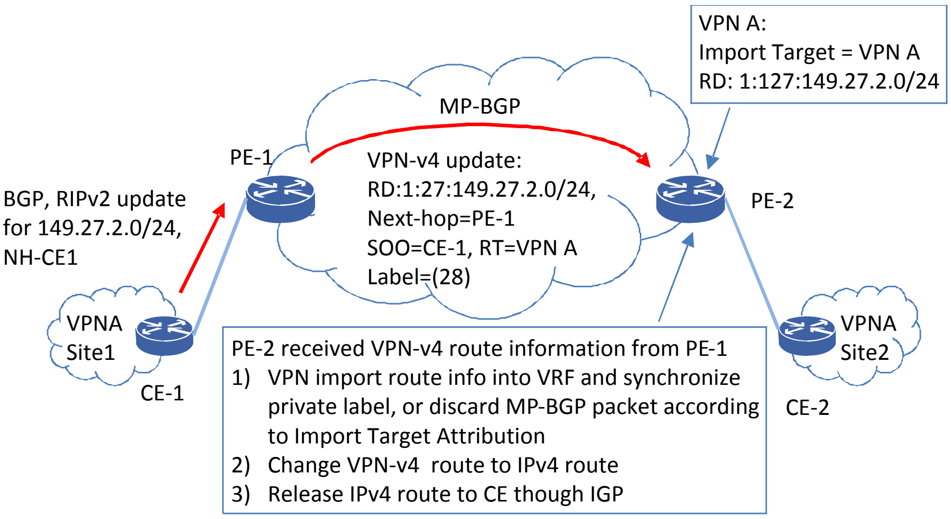
2.3.2 PE to PE

Figure 2-8 PE releases VPN route



PE-1 received route information from CE-1
1) Adding route to VPNA according to interface
2) Change IPv4 route to VPN-v4 route based on RD of VPNA
3) Append Export Target of VPNA
4) Release VPN-v4 route to all PE neighbor

Figure 2-9 PE receives VPN route



PE-2 received VPN-v4 route information from PE-1
1) VPN import route info into VRF and synchronize private label, or discard MP-BGP packet according to Import Target Attribution
2) Change VPN-v4 route to IPv4 route
3) Release IPv4 route to CE though IGP

PE to PE routes are released through MP-BGP to control which routes are released or received. MP-BGP uses RT attributes to control route releasing and receiving. When a route is released, some information is included such as private network label, next hop and RD information. All the information is included in MP-BGP protocol fields NLRI and RT.

A PE to PE VPN route is called a VPN-IPV4 address. Compared with an ordinary V4 address, it is added with RD attribute describing VRF address space. Refer to Section 2.2.1 for details. While receiving V4 routes from CE and sending them into VRF, PE converts these addresses by adding RD fields and translating them into VPN-IPV4 addresses.

When releasing a route, PE allocates a private network label to VPN-IPV4 through MP-BGP and writes a loopback address into NLRI header. Meanwhile, it adds the Export Target allocation list of VRF instance. PE allocates VPN-IPV4 routes in such a way. Refer to Figure 2-8 for more information.

When receiving a MP-BGP route allocation packet, PE compares the Export target in RT list with the Import Target attributes. If they are consistent, PE adds the VPN-IPV4 route into the VRF route table. If not, it discards the packet.

Import Target and Export Target attributes of a VPN instance may be different, and can be set to several values. Several PEs in one VPN may be different in instance RT. RT can help control the inter-PE route releasing in a VPN and even introduces the routes of other VPNs for cross-VPN interworking. RT import and export can be flexibly configured to control VPN route allocation. RT has the following basic application modes: traditional, Hub-Spoke and Extranet mode.

- Traditional mode

In the mode, Import Target attribute configuration is the same as Export Target, and all VPN instances are configured with the same value. RD is usually used as the identification attribute of RT. Thus, the routes from a PE are received by other PEs, each VPN PE has the same VPN route, and VPN sites interwork with each other, as shown in Figure 2-10. In such a mode, PE needs a number of route resources.

- Hub-Spoke mode

The mode can control the communication between VPN sites, and sites are interconnected with each other via VPN central control site. The central site is Hub-CE (The PE connecting Hub-CE is Hub-PE), and other sites are Spoke-CEs (PEs connecting Spoke-CEs are Spoke-PEs). Hub site can access all other nodes. Spoke sites cannot be interconnected with each other, but they can do it through the Hub-CE site. Spoke-PE only releases local routes to Hub-PE, and Hub-PE also releases to Spoke-PE the routes learnt from Hub-CE, but Spoke-PEs cannot directly release local routes to each other through MP-BGP. They learn routes from each other only through Hub-CE so that Spoke-CEs are not directly interconnected with each other, as shown in Figure 2-10.

Spoke-PE RT attribute Import Target is different from Export Target of other Spoke-PEs, but the same as Export Target of Hub-PE.

Hub-PE RT attribute Import Target contains Export Target attribute of other Spoke-PEs.

Hub-PE needs to create two VPNs: VPN-up and VPN-down (According to data flow, the flow to central site is the UP flow and the opposite is the DOWN one). VPN-down receives local routes from Spoke-PE and Hub-CE, and a VPN-down instance includes the routes from all Spoke-PEs for the central site to send the traffic to individual Spoke-CE. Meanwhile, VPN-down notifies Spoke-CE routes to Hub-CE through interface route protocol. VPN-up allows Hub-PE to release the routes learnt from Hub-CE to all Spoke-PEs. Hub-PE VPN-up only includes the routes notified by Hub-CE (As Hub-CE notifies the routes of other sites notified by VPN-down and Hub-CE routes to Hub-PE VPN-up instances through IGP/EGP, VPN-up instances also include the routes route of individual Spoke-CE), aiming to receive the traffic from Spoke-PE and send them to the central site.

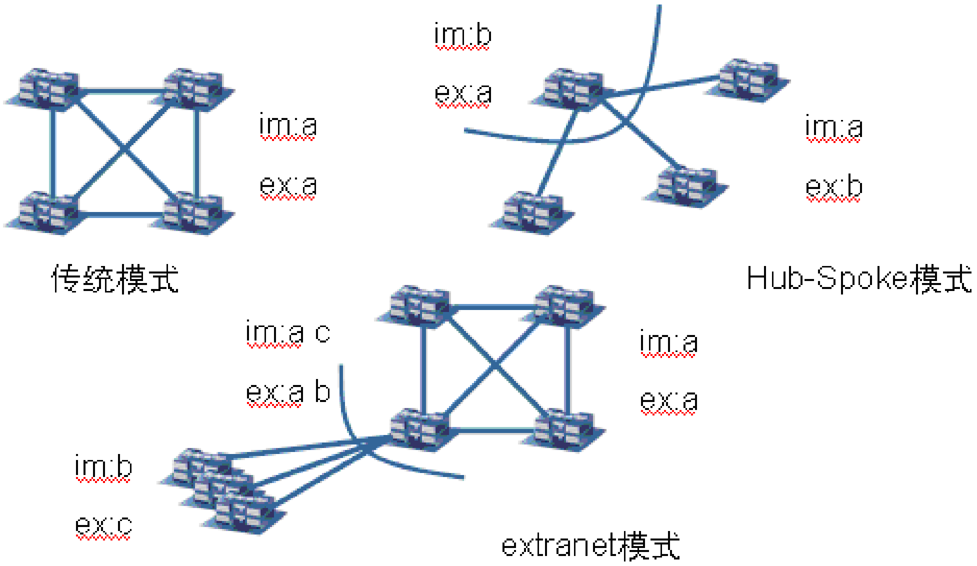
In the mode, Spoke-PE learns the routes of VPN sites through Hub-PE VPN-up instances, but all route paths point to Hub-CE for VPN site interworking under the centralized control of Hub-CEs.

- Extranet mode

The mode aims to interwork different VPN sites. For example, some sites of two VPNs belong to two VPN users at the same time, and other sites can access the site equipment. There is another case. Enterprise sites need to control the access. Some sites can access all internal sites and some sites need to be isolated from each other. These sites can be divided into several intersectant groups, thus each group creates a separate VPN and the sites accessible to different groups belong to several VPNs, namely, one VPN site belongs to several VPNs.

Multi-VPN sites need to be bound to VPN instances, and introduces the routes of other VPNs into their VRF table through RT. Refer to Figure 2-10. PE at central node imports through RT a and c the routes of the sites attached to PEs at both ends, and left PE imports through RT=b the routes of the sites attached to central site PE. Thus PEs on both ends do not learn the routes of the sites attached to the opposite end, but the routes of the middle sites. VRFs of middle PEs contain the routes of the sites attached to PEs at both ends, thus the sites attached to middle PEs can communicate with other sites, but the sites attached to PEs at both ends cannot interwork with each other.

Figure 2-10 PE uses RT to flexibly control VPN route releasing and import



2.3.3 PE to CE

When receiving routes, PE converts the VPN-IPv4 address of the best route into the IPv4 address (removing RD of the address) and import it into VRF. The private network label remains and is recorded into the forwarding table. Then the route is transferred to local

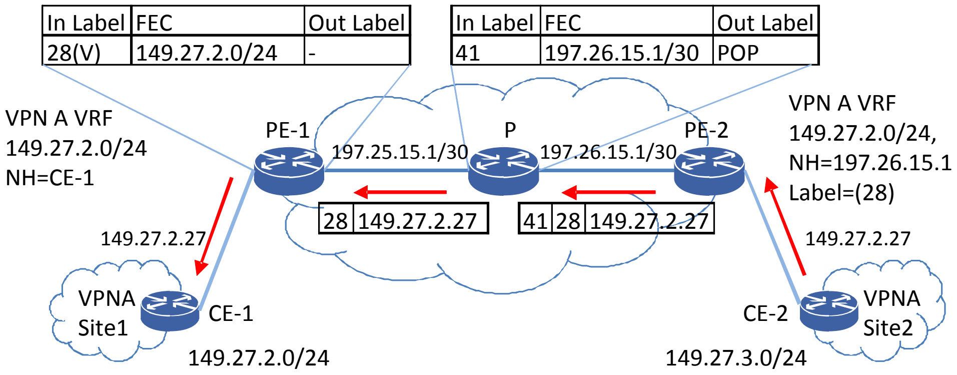
CE through VRF route protocol, and meanwhile the next hop is modified into the interface address of the receiving PE. Thus VPN routes are released to CE.

CE can also be configured with default routes or static routes pointing to PE.

2.4 Data forwarding

Take CE-1 and CE-2 for examples to describe the MPLS L3VPN forwarding. CE-1 network segment is 149.27.2.0/24 and CE-2 network segment 149.27.3.0/24. CE-2 transmits IP flow whose address is 149.27.2.27, as shown in Figure 2-11.

Figure 2-11 MPLS L3VPN data forwarding



2.4.1 CE-2 to PE-2

CE-2 route table has the PE-2-released route whose subnet segment is 149.27.2.0/24. As VPN is invisible to CE-2, CE-2 is an ordinary router. The IP traffic received by CE-2 is forwarded via an ordinary IP route. The forwarding is as follows:

Search DIP :149.27.2.27/24 and obtain the next hop pointing to PE-2.

Modify LLC according to the next hop to send the IP packet to PE-2.

In the actual networking, CE can also be configured with default route or static route to directly point the IP traffic to PE-2, and there is no need to configure dynamic route allocation protocol between PE-2 and CE-2.

2.4.2 PE-2 to PE-1

- PE-2 forwarding

PE-2 receives an IP packet from CE-2. As AC interface is bound to VPN, the packet needs to be forwarded via L3VPN. The forwarding is as follows:

1 Search the VRF instance bound according to the AC interface receiving the packet.

2 Use DIP :149.27.2.27 to search the VRF virtual route forwarding table according to the instance VRF-ID. As PE-1 notifies subnet route 149.27.2.0/24 and a private network label to PE-2 through MP-BGP and sends them into the VRF instance, VRF saves the route 149.27.2.0/24 and DIP matches with it.

3 Obtain from the subnet route 149.27.2.0/24 table entry the private network label 28, the multiplexed LSP and public network next hop 197.26.15.1.

4 Obtain the public network label 41 according to LSP.

5 Obtain egress gateway MAC and PE-2 interface according to the next hop 197.26.15.1.

6 Use private network label 28, public network label 41, next hop gateway MAC and egress interface VLAN to encapsulate the IP packet into a dual-MPLS-label packet and send it to provider router.

- Provider router forwarding

As a provider router receives a label packet with a public network label 41, as the label 41 is a LSR label, the provider router directly swaps the label like a LSR to forward the MPLS packet. The next hop corresponding to label 41 points to 197.25.15.1 and PE-1, and the egress label is a POP label.

In the networking in Figure 3-11, the provider router is the PHP router and the egress label corresponding to label 41 is a POP label, so label 41 directly pops out and the packet sent to PE-1 is the MPLS packet only with label 28.

2.4.3 PE-1 to CE-1

After PE-1 receives the MPLS packet from a provider router, as label 28 is the VPN attribute label, the packet needs to be forwarded via L3VPN. The forwarding is as follows:

1. Search the external label 28. Label 28 is a private network label, and saves the VRF-ID identifying the location of data flow.
2. Take packet DIP :149.27.2.27 and search the subnet route in the VRF instance route table to find the matched 149.27.2.0/24 subnet route table entry and obtain the next hop pointing to CE-1.
3. As PE-1 is the VPN downlink, the private network label 28 is stripped to restore it into the user IP data flow which will be encapsulated with a CE-1 gateway MAC and an egress AC interface VLAN and then forwarded to CE-1.

3 ZTE solution

MPLS L3VPN networking applications support the following networking modes to meet networking needs in different scenarios:

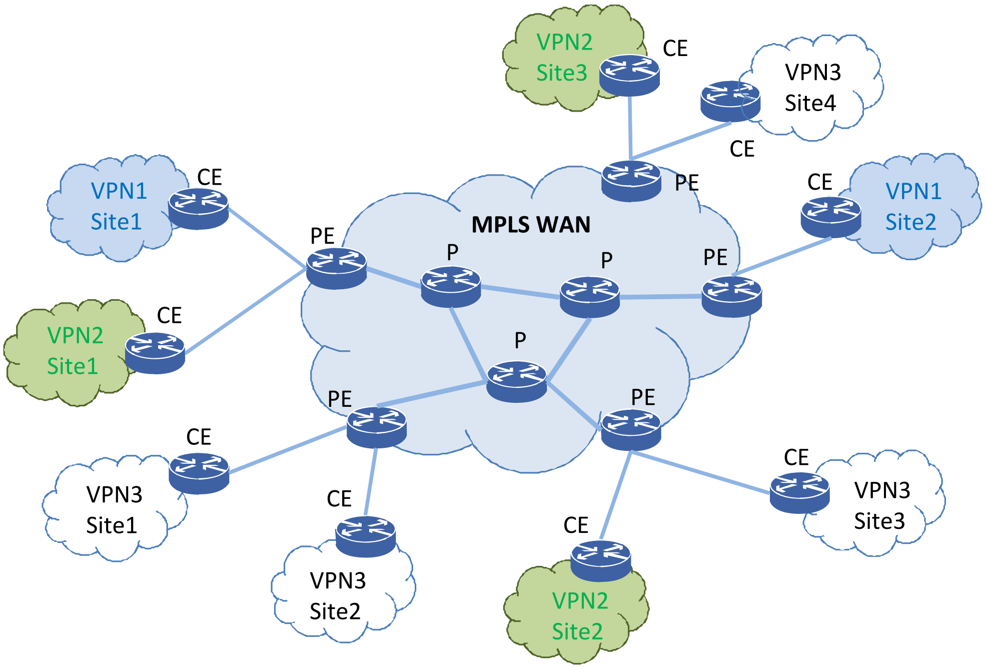
1. Full-mesh VPN networking
2. VPN cross-area Option A/B
3. Hub-Spoke VPN networking
4. half-duplex VRF networking
5. HoVPN networking
6. MCE networking
7. Cross-VPN mutual access
8. L2VPN access to L3VPN networking
9. VPN redundancy

This section describes the examples of these networking modes, and a user selects the proper networking mode as needed.

3.1 Full-mesh networking

MPLS L3VPN supports Full-mesh networking and P2P networking. Both of them have no essential difference in control protocol and service forwarding. In L2VPN, VPLS and VPWS are different in protocol and forwarding. Full-mesh VPN networking is shown in Figure 3-1.

Figure 3-1 Full-mesh networking



A carrier network consists of P and PE which form a Full-mesh network. PE is at the edge of MPLS network and accesses users, and P is in the MPLS network and accesses no user. A user network CE is connected to a carrier network PE. The networks have the following characteristics:

1 Each VPN user has at least 2 sites located at different locations, which are connected via a carrier network.

2 Each PE accesses several VPN users and each VPN user is connected to the PE via a separate logic interface.

3 P is in the carrier network and accesses no user.

4 All PEs are interconnected in the entire network and run MP-IBGP.

In the above networking, VPN1 has 2 sites, VPN2 3 sites, and VPN3 4 sites. They are interconnected via a MPLS network, and two VPN3 sites are connected to one PE.

- Full-mesh networking has the following advantages:
  - The configuration is simple in small networks.
  - The configuration is similar when user topologies are the same.
  - Two different interfaces in one VPN are connected to one PE for local interworking.
- Full-mesh networking has the following disadvantages::
  - All VPN PEs need to create BGP connections among them. Each new PE node needs to be added into the connections.
  - PEs need to create BGP neighborhoods, which increase PE expansion and configuration workload.
  - PE needs to create the routes to all VPN sites. A large network has high requirements for PE route performance.
- Application scenario

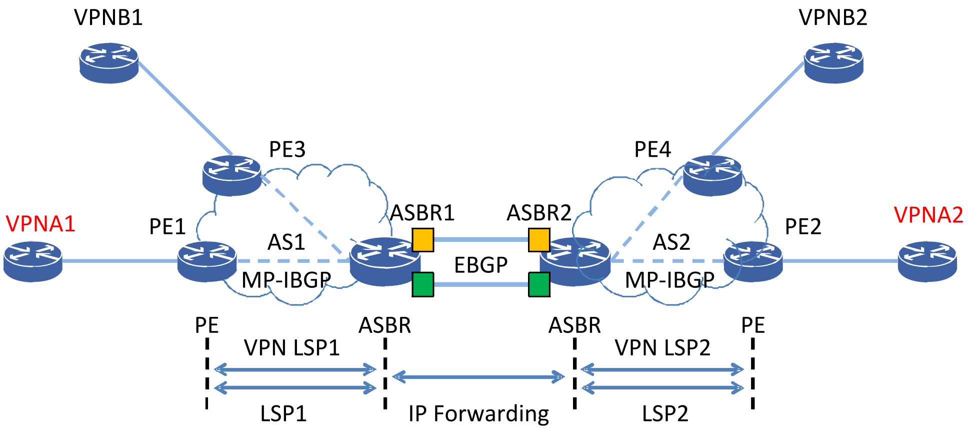
A full-mesh network fits a small and stable topology. The full-mesh VPN networking is usually at the core layer in a metro network. A large L3VPN has a large pressure of network expansion. If a network often changes, HoVPN networking can be taken into account. Therefore, VPN at the core layer adopts the full-mesh networking and L3VPN at convergence and access layers generally prefers HoVPN or MCE networking.

3.2 VPN cross-area networking

In the actual networking, several sites in a user VPN may be connected to several SPs using different ASs, or to several ASs of one SP. The VPN crossing multiple ASs is known as the cross-area VPN. The cross-area VPN needs to resolve the issue of notifying cross-area VPN-IPV4 routes. There describe three cross-area VPN networking modes: VPN Option A, Option B, and Option C. Supports of Option A and Option B rather than Option C is preferred because a MP-EBGP connection needs to be created between remote PEs, label and route allocation is complex, and route convergence is affected by several factors.

- Option A, which is also called VRF-VRF

Figure 3-2 L3VPN Option A networking



Option A networking mode is shown in Figure 3-2. PE is between two AS areas which are connected to each other via ASBR routers. ASBR (Autonomous System Border Router) is also a PE router.

PE and ASBR in one AS transfer VPN route information between them through MBGP, and ASBRs transfer VPN route information between them through a route between PE and CE.

1 VPNA2 transfers route information to PE2 through IGP.

2 PE2 transfers VPNA2 information to ASBR2 through MBGP.

3 As ASBR1 CE, ASBR2 transfers VPNA2 information to ASBR1 through IGP/EGP.

4 ASBR1 transfers VPNA2 information to PE1 through MBGP.

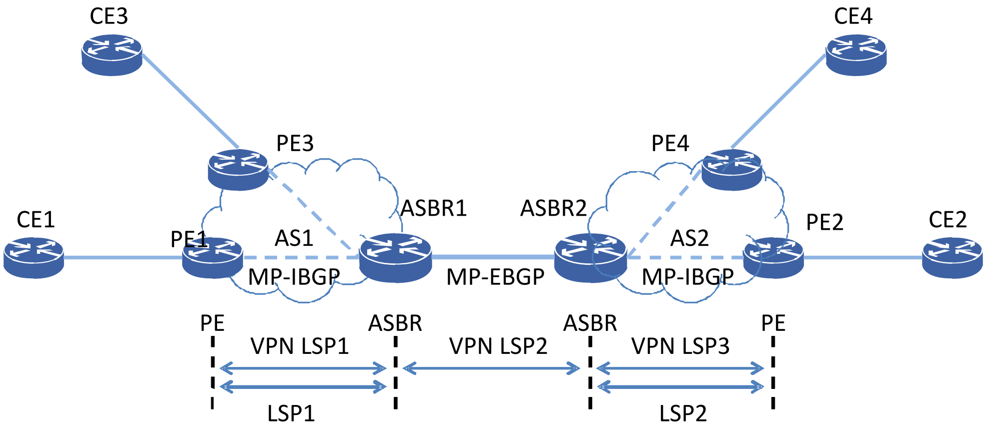
5 PE1 transfers VPNA2 information to VPNA1 through IGP. Thus, VPN information transfer is completed.

VPNB route allocation is the same as VPNA.

Individual AS creates a separate dual-label LSP tunnel from PE to ASBR. The internal label stands for VPN information, and the external label for the public network label to the target VPN route next hop PE. Like single AS creates a LSP tunnel, the data is forwarded between ASBR and ASBR through Native IP, without a LSP tunnel.

- Option B, which is also called single-hop MP-EBGP cross-area

Figure 3-3 L3VPN Option B networking



Option B networking mode has no difference from Option A topology. Their difference is the cross-area route allocation mode. See Figure 3-3.

AS transfers VPN information and creates LSP tunnel through MPLS/BGP in the system. ASs transfers VPN information and creates LSP tunnel through single-hop MP-EBGP between each other route.

1 CE2 transfers private network information to PE2 through IGP.

2 PE2 transfers VPN information to ASBR2 through MP-IBGP.

3 ASBR2 transfers VPN route information to ASBR1 through single-hop MP-EBGP.

4 ASBR1 transfers VPN information to PE1 through MP-IBGP.

5 Thus, CE1 has the information of the route to CE2.

Option B and Option A route allocations are different in inter-ABSR route allocation. Option B ABSRs allocate VPN routes through MP-EBGP, and routes allocated between ASBRs include such attributes as private network label and RT.

If the information crosses several ASs, ASs transfer the information through MP-IBGP in the systems and ASBRs transfer it through single-hop MP-EBGP between each other.

In Option B cross-area transfer, when ASBR2 transfers VPN route information to ASBR1, the next hop must change to ASBR2 and ASBR2 reallocates a label to VPN. When ASBR1 transfers VPN route information to PE1, there are two ways:

- Way 1: When ASBR1 transfers VPN route information to PE1, the next hop changes to ASBR1.

- Way 2: When ASBR1 transfers private network route information to PE1, the next hop does not change, in other words, the next hop is still ASBR2.

When the route next hop is ASBR1, ASBR1 reallocates a label to VPN. VPN path from PE1 to PE2 is PE1->ASBR1->ASBR2->PE2. AS1 creates a dual-layer LSP tunnel from PE1 to ASBR1 in the system. A VPN label (allocated by ASBR1) is at the internal layer, and a public network tunnel from PE1 to ASBR1 is at the external layer. ASBR creates a single-layer LSP tunnel between each other, and only a VPN label (allocated by ASBR2) is available. AS2 creates a dual-layer LSP tunnel in the system. A VPN label (allocated by PE2) is at the internal layer, and a public network tunnel from ASBR2 to PE2 is at the external layer. As VPN labels are reallocated at ASBRs, the lowest-layer labels are swapped at the ASBRs. The swapping connects VPN tunnels of two ASs.

If the VPN route next hop does not change, the VPN route next hop received by PE1 is ASBR2 AS2. The VPN path from PE1 to PE2 is PE1->ASBR2->PE2. A dual-layer LSP tunnel needs to be created from PE1 to ASBR2. A VPN label (allocated by ASBR2) is at the internal layer, and a public network tunnel from PE1 to ASBR2 is at the external layer. A dual-layer LSP tunnel also needs to be created from ASBR2 to PE2. A VPN label (allocated by PE2) is at the internal layer, and a public network tunnel from ASBR2 to PE2 is at the external layer. In such a case, ASBR2 and ASBR1 need to run a label allocation protocol between them to allocate ASBR2 public network labels. In addition, internal and external labels of ASBR2 LSP are swapped to convert two LSPs into a ETE LSP tunnel.

No matter whether a ASBR route is allocated in Way one or Way two, ASBR forwards data by searching route on the data forwarding layer, but the data can be encapsulated with a label at the egress.

- Application scenario

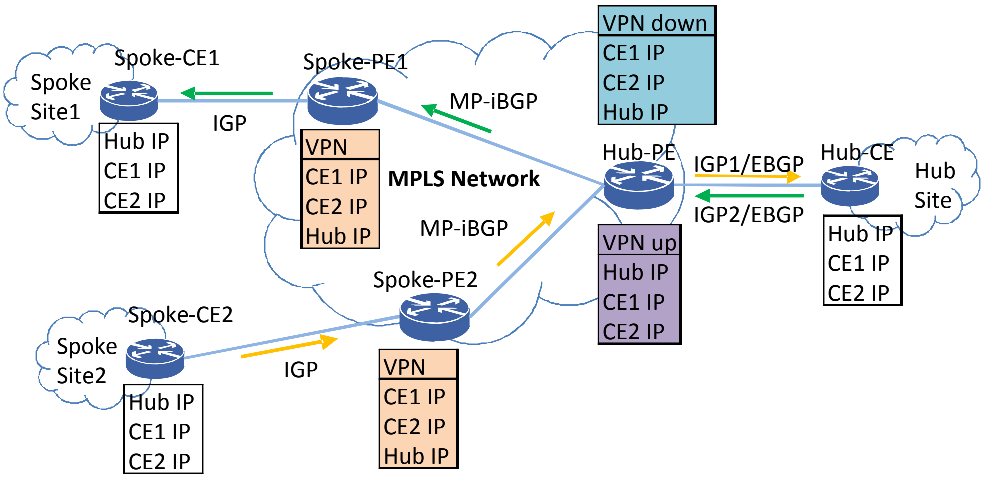
In Option A, ASBRs need to be configured with VRF instance between them, and each VRF needs to be allocated with a sub-interface. IP packet is transferred between areas. This option wastes interface resources, which is not conducive to QoS planning, and has large service expansion and configuration workload, which is not suitable for large-scale networking.

In Option B, ASBRs also need to be configured with VRF instance between them, but each VRF does not need to be allocated with a sub-interface. MPLS packet is transferred between areas. This option has strong QoS and small service expansion and configuration workload, so it is suitable for large-scale networking.

3.3 Hub-Spoke networking

Hub-Spoke network topology is shown in Figure 3-4. It can control Spoke-CE site communication in one VPN. All sites must communicate with each other via Hub-CE. For VPN configuration and route allocation, refer to Section 2.3.2.

Figure 3-4 Hub-Spoke networking and traffic model



PE route table is shown in Figure 3-4. The arrow indicates the mutual access service flow between Spoke-CEs. The flow from Hub-PE to Hub-CE searches the VPN-up virtual route table, and the flow from Hub-CE searches the VPN-down virtual route table. Spoke-CE supports mutual access only through Hub-CE.

- Application scenario

Hub-Spoke networking monitors site communication in one VPN. It should be noticed that the route protocol between CE and PE should be configured reasonably to ensure normal route notification.

3.4 Half-duplex VRF networking

When Spoke-PE accesses several sites or CEs at different locations in one city are accessed to Spoke-PE after L2 convergence, Hub-Spoke networking allows direct CE connection via PE to disable central control CE to monitor the traffic.

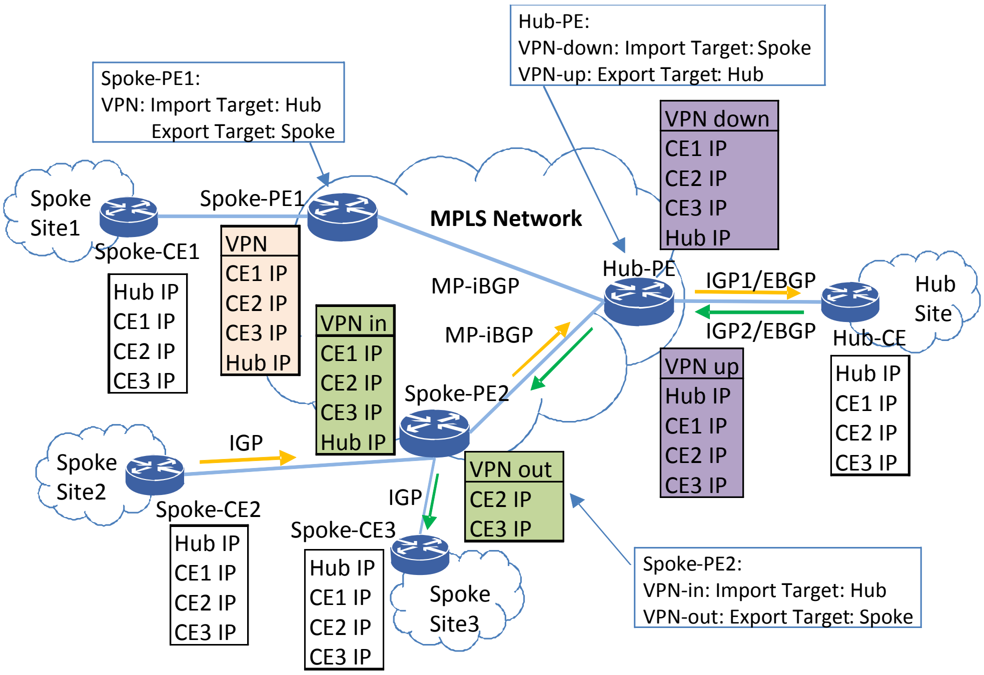
The half-duplex VRF networking can resolve the above issue, as shown in Figure 3-5. Spoke-PE2 accesses two user CEs: Spoke-CE2 and Spoke-CE3. In Hub-Spoke networking, CE2 and CE3 routes learn one VRF table, which allows direct service interconnection via Spoke-PE2 to disable Hub-CE to monitor the traffic. According to the current networking, CE2 and CE3 has the following networking modes:

1 They are accessed to PE2 via different VLANs and physical interfaces.

2 They use the same VLAN and different physical interfaces.

3 They uplinks PE2 directly via the same VLAN and physical interface after L2 convergence.

Figure 3-5 Half-duplex VRF networking



For the first two modes, PE2 can create two VPNs to resolve the issue of local interworking, but the method does not work for the third mode. Half-duplex VPN means that VPN service traffic is forwarded only in one direction to avoid local interworking. PE2 creates two VPNs: VPN-out and VPN-in (According to VPN flow direction, VPN-in traffic is sent to carrier network, and VPN-out to user network). And interface and VRF binding relationship should be improved so that the interface binds two half-duplex VRF instances which are allocated with two VRF-IDs to control route allocation, but the interface on the forwarding layer only saves VPN-in VRF-ID for uplink forwarding.

VPN-out allows PE to receive a IGP route from CE and notifies the CE route to Hub-PE through MP-IBGP. A VPN-in user receives the VPN route from Hub-PE and PE2 notifies the VPN-in route to CE through IGP.

Take the traffic from CE2 to CE3 for example to describe the traffic forwarding. PE2 receives data flow and obtains VPN-in instance ID. As VPN-in route is released by Hub-PE, it is forwarded directly to Hub-PE. As the route is notified via VPN-out, the service flow from Hub-PE to PE2 searches the private network label to obtain the VPN-out instance ID. The VPN-out instance table saves the route notified through local IGP, so the data is directly forwarded to CE3. Hub-PE and Hub-CE forwarding is the same as Hub-Spoke. Thus the traffic is controlled centralizedly.

- Application scenario

The half-duplex VRF is the supplement to Hub-Spoke. It resolves the issue of Spoke CE local networking in the Hub-Spoke networking.

3.5 HoVPN networking

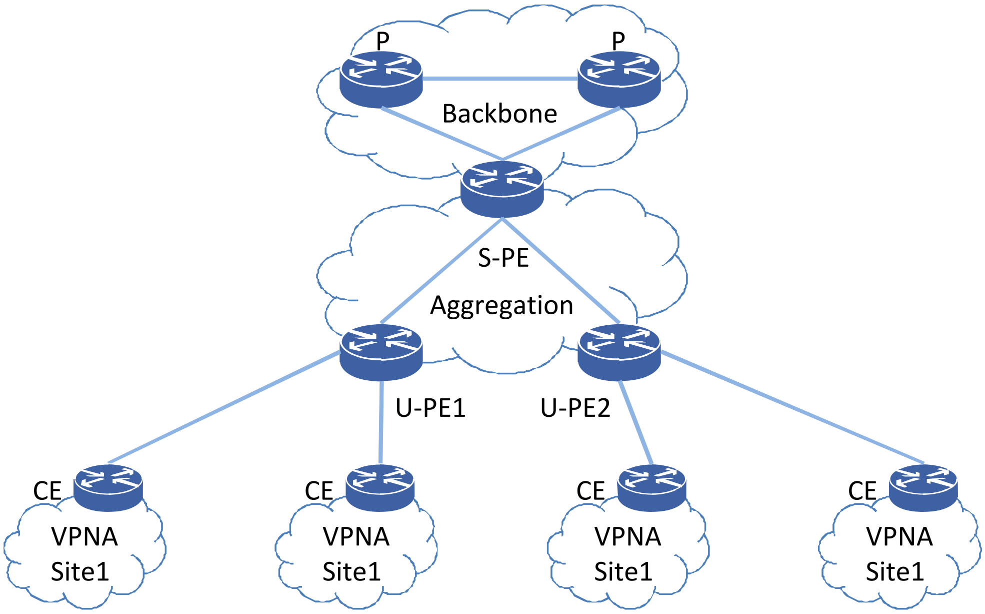
In the Full-mesh networking, PEs need to create the BGP neighbor relationship between each other, which has high PE route requirement and low network scalability.

Currently, carrier network adopts the classic hierarchical structure. For example, the classic structure of metro network is a 3-layer model: core layer, convergence layer and access layer. Equipment performance declines and network scale expands from core layer to access layer.

MPLS L3VPN is a plane model, and has the same performance requirements for all PEs. When the performance and scalability of some PEs have problems, the performance and scalability of the entire network are affected. As MPLS L3VPN plane model is inconsistent with the classic hierarchical network model, the scalability issue exists when PEs are deployed at different layers, which have an effect on large-scale VPN deployment.

HoVPN can meet the actual networking requirements. Access-layer VPN uses the small-capacity, high-performance router, and core-layer VPN the large-capacity router to save the performance requirement of core-layer VRF interface.

Figure 3-6 HoVPN networking



PE interconnected with core/convergence-layer VPN is S-PE, and PE accessing users at convergence/access layer is U-PE. U-PE accesses user and S-PE converges routes and controls hierarchical allocation of VPN routes.

Inter-S-PE route allocation in backbone network is the same as Full-mesh network. S-PE maintains all VPN routes. It sends the default route or convergence route to U-PE instead of remote sites. U-PE maintains the local and default routes, or the convergence route to other sites. U-PE route capacity is greatly reduced.

U-PE supports normal VPN forwarding, and S-PE switches and forwards routes between VPN tunnels through VRF routing.

- HoVPN has the following advantages:
  - MPLS L3VPN can be deployed layer by layer. When UPE performance is not enough to meet the requirements, a SPE can be added and the UPE can be moved downward. When SPE access capacity cannot meet the need, a UPE can be added.
  - UPE and SPE are connected via a label, so only one (sub-)interface can be used for the connection to save the limited interface resource.

- If there is an IP/MPLS network between UPE and SPE, they are connected via such tunnel as LSP. In the hierarchical deployment, MPLS VPN has good scalability.

- UPE only maintains locally accessed VPN routes, and all remote routes are replaced by a default or convergence route to reduce the burden of the UPE.

- SPE and UPE switch and release routes through the dynamic route protocol MP-BGP. Each UPE needs to create only one MP-BGP peer, which has small protocol overhead and configuration workload.

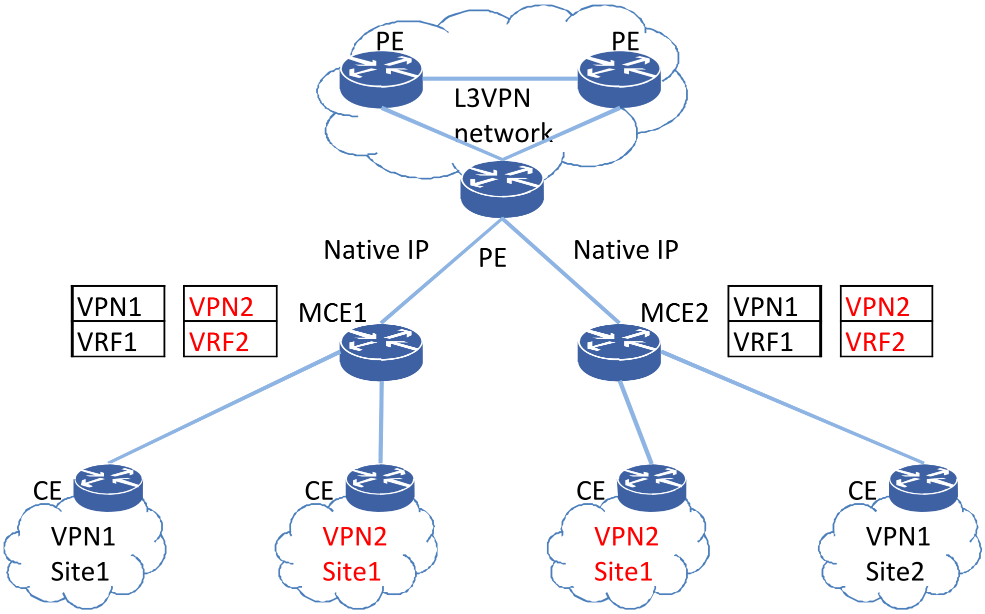
- Application scenario

In a large-scale deployment of MPLS L3VPN, HoVPN can decrease PE L3 interface number, reduce the inter-PE BGP full-connection requirements in VPN, and improve network scalability, so it is a useful supplement to the Full-mesh networking.

3.6 MCE networking

MPLS L3VPN can isolate user routes, but how can it isolate user services in the network not supporting MPLS or on the equipment not supporting MP-BGP routes? MCE can do that. The simplified MPLS L3VPN technology introduces a VRF virtual route table to isolate services, but it does not adopt the complex VPN route allocation protocol MP-BGP. It isolates services in a simple way. MCE networking is shown in Figure4-7.

Figure 3-7 MCE networking



User router is accessed to MCE router, and MCE and L3VPN are connected. MCE is configured with VRF instances to manage local route and PE-allocated route. PE can release default or convergence route alone. Each VRF occupies a L3 interface between MCE and PE. If a route is released dynamically, each L3 interface starts IGP.

When receiving a user packet, MCE searches the VRF route table like L3 VPN PE, but the forwarded packet to PE is the IP packet without an encapsulation label. The forwarding of the packet reaching PE at network side or user side is the same as the Full-mesh PE forwarding.

- Application scenario

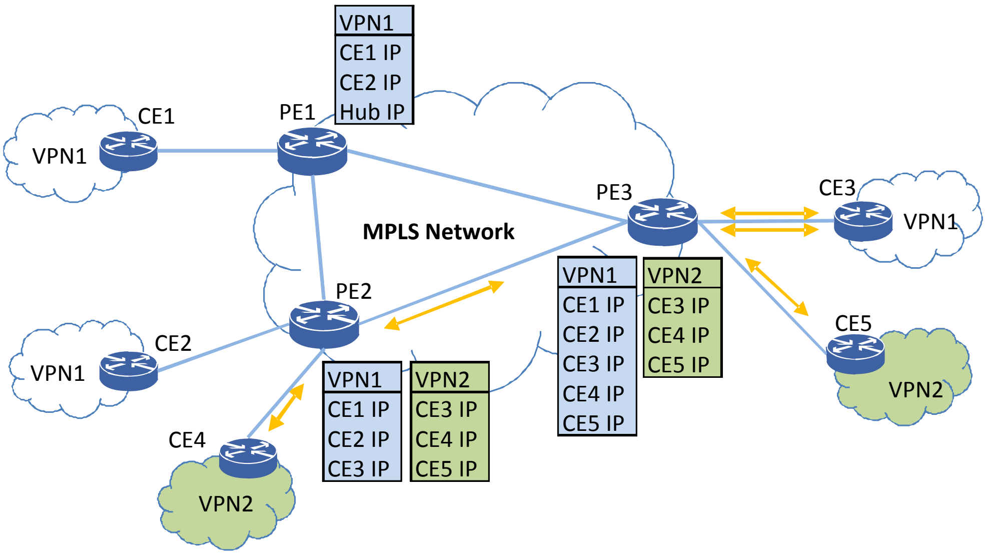
MCE isolates services at access layer. A access-layer router is weak in routing and MPLS. If MPLS is not required, MCE is a good VPN networking technology. However, more VRFs need uplink PE to offer more L3 interfaces and each L3 interface starts IGP, so large-scale VRF deployment is not suitable for MCE. HoVPN is the best choice in such as case. MCE should be deployed with only less than several dozens VPN users.

3.7 Cross-VPN mutual access

As an enterprise expands, different service departments need to be divided into different VPNs. Meanwhile, all departments need to access some service departments or two enterprise VPNs need to access one VPN site, cross-VPN mutual access can meet the requirements.

The cross-VPN mutual access uses MP-BGP RT to flexibly control the release and import of VPN routes. Refer to the analysis on Extranet mode in Section 3.3.2.

Figure 3-8 Cross-VPN mutual access networking



As shown in Figure 3-8, the network is configured with VPN1 and VPN2. VPN1 includes three sites and VPN2 two sites. It is required that all sites of the two VPNs can access CE3. Other sites cannot access each other via VPN, but they can interwork with VPN. The cross-VPN networking aims to control the route import.

- Route release:

PE3 VPN1 VR import CE1/CE2/CE3 routes, CE4 VPN routes released by PE2, and CE5 routes in local VPN2 (local routes may be converged and then imported). PE3 VPN2 VRF locally imports CE3 routes from VPN1 VRF, but not CE1/CE2 routes.

PE2 VPN2 instance imports CE3 VPN routes released by PE3, but not CE1 VPN routes released by PE1 and local CE2 routes.

Finally, PE VRF route allocation in VPN is shown in Figure 3-8:

When CE4 accesses CE3, data forwarding is as below :

1 PE2 forwarding: PE2 searches the VPN2 VRF route table and forwards it to PE3.

2 PE3 forwarding: When PE3 releases a CE3 VPN route, MPLS allocates a label to VPN1. When PE3 receives the packet from CE4, the instance in the label corresponds to VPN1. PE3 searches the route table in the VPN1 VRF route table and forwards it to CE3.

In addition, CE5 can directly access CE3. The process is as below:

1 Receiving the traffic from CE5, PE3 searches the VPN2 VRF route table.

2 VPN2 routes may be converged by VPN1 and then locally imported to the VPN2 VRF. According to the longest route matching requirement, PE3 needs to obtain the instance ID of VPN1 related to subnet table items in the VPN2 VRF table subnet, searches the VPN1 VRF route table to obtain the next hop to CE3, and forwards it to CE3.

- Application scenario

As the cross-VPN route import may make the routes of a site saved in different VRFs, the route import mode should be designed strictly to avoid wasting route table resources.

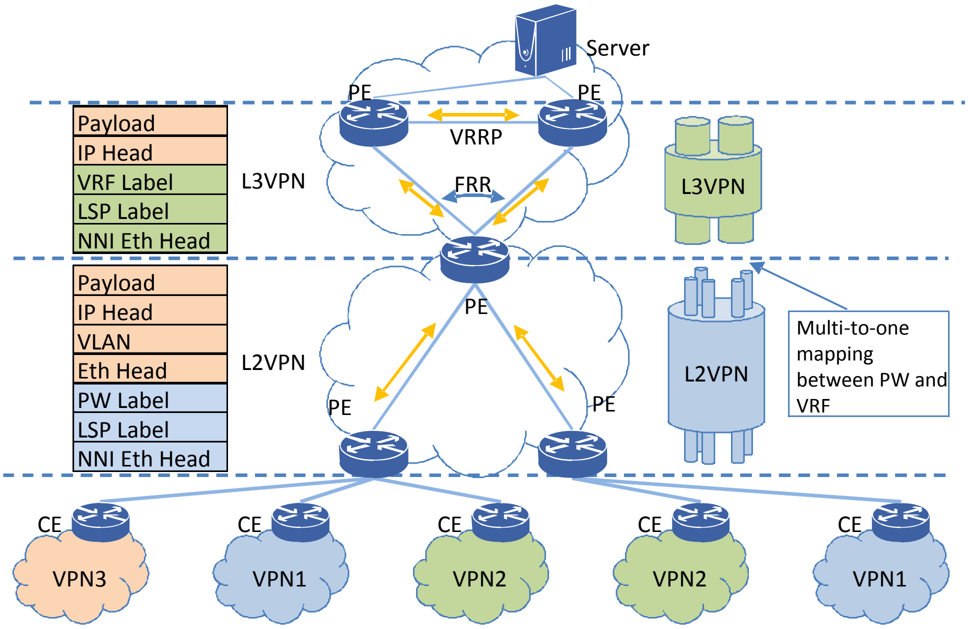
3.8 L2VPN access L3VPN

L2VPN includes VLL service and VPLS service, supports non-Ethernet service forwarding and has low requirements for carrier network protocols. Carrier network provides a transparent channel alone, and the change of user network has a small effect on carrier network. User network supports simple deployment and has strong networking capability, low OPEX and good scalability. Because of broadcast features and user-side route management, L2VPN is not suitable for large-scale networking. L2VPN is applied to network deployment at access/convergence layer and L3VPN to the deployment at backbone layer.

In traditional L2VPN and L3VPN networking, two PEs need to be interconnected, thus Native IP flow is transmitted between L2VPN and L3VPN and interconnected nodes are protected only through VRRP. Such a networking mode is not conducive to full-range MPLS deployment, QoS assurance and E2E management mechanism, and has weak protection.

PW access to L3VPN is another hierarchical VPN mechanism which provides a L2VPN and L3VPN interconnection mode. L2VPN PE and L3VPN PE are located in one PE, which is called the L2/L3 bridge. The networking is shown in Figure 3-9.

Figure 3-9 L2/L3VPN bridge networking



The L2/L3 bridge connects L2VPN and L3VPN through an internal virtual logic interface. The virtual interface is not only a L3VPN AC interface but also a L3VPN service interface. The interface supports IP address and route protocol and it has the same functions as an ordinary L3 interface.

The L2/L3 VPN bridge supports the full-range MPLS service forwarding. The bridge PE can switch a label from MPLS flow to MPLS data flow, which is like a ETE transport tunnel for the user.

The same as HoVPN, a L2/L3 bridge PE can converge access-layer services, and access-layer L2VPN PWs have a several-to-one relationship with VPN tunnel. The bridge PE can converge PWs. It learns user CE routes through IGP, and L3VPN allocates these routes to PEs of other L3VPNs through MP-BGP. PEs connecting users communicate with each other via the bridge PE. The data encapsulation format from L2VPN to L3VPN is shown in Figure 3-9.

- Application scenario

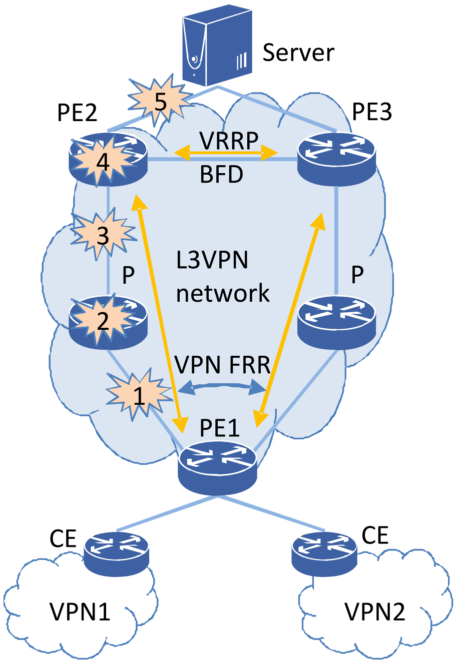
L2/L3 bridge supports the full-range MPLS pipe. L2VPN at convergence/access layer can flexibly transport services and improve network scalability, and L3VPN at core layer helps carriers manage user routes better. L2VPNs are interconnected with each other via L3VPN. Compared with HoVPN, L2/L3 bridge allows access layer to introduce L2VPN equipment so as to enhance network scalability.

3.9 VPN redundant networking

MPLSL3VPN usually depends on route and LDP reconvergence to restore services. Services are generally converged in seconds. Although IGP fast convergence shortens service restoration time, it cannot meet the carrier-class service protection requirements unless it adopts various fast detection and FRR mechanisms.

Carrier network can use TE-FRR, LDP-FRR, TE Host-Standby, VPN-FRR, IP-FRR and VRRP to fast restore services. The first three FRR mechanisms belong to LSP protection which protects inter-PE fault rather than PE node fault. PE node protection adopts VPN-FRR, IP-FRR, VRRP and IGP fast convergence to restore services. As IGP fast convergence actually depends on MP-BGP to re-release VPNV4 routes, the convergence time is related to the scale and the performance is poor. The networking mainly describes the BFD-triggered VPN-FRR technology. BFD is based on BGP or LSP. The former can detect BGP Down and the latter has good switching performance but service restoration relies on BGP reconvergence. Figure 3-10 shows two basic redundant protections of VPN: PE dual-homed redundant protection and PE VRRP redundant protection.

Figure 3-10 L3VPN redundant networking



PE1 is dual-homed to PE2. PE3 provides the VPN-FRR switching and supports the ETE service protection from PE to PE as well as the PE to CE dual-homed protection, in other words, the FRR at VPN user is similar to ordinary IP-FRR. In the VPN-FRR of Figure 3-10, PE2 and PE3 release server VPN route through MP-IBGP at the same time to form VPN-FRR at PE1. LSP between PE1 and PE2 starts CC/CV detection to detect the fault of faulty node 1, 2, 3 and 4. Once a LSP fault is detected, PE1 triggers VPN route fast switching which runs on LSP. When BGP routes are reconverged (changing from FRR to ordinary VPN route), the routes are just updated, which leads to no disconnection on the forwarding layer. When a LSP is recreated between PE1 and PE2, the routes is reconverged. As the routes are just updated, traffic switching does not lead to data disconnection.

If VPN route and VPN tunnel are separated in the forwarding table, VPN tunnel can be updated once (allocating private network label according to VPN label) to restore the services of the entire VRF. Service restoration is independent of VRF route scale to shorten the restoration time. However, if many VRFs are available, VPN tunnel may be updated many times, which leads to low switching performance. If active/standby VPN tunnels of all VRF instances are associated with LSP and LSP status decides VPN tunnel selection, the FRR status table managed by LSP can be switched once (or VPN route FRRs are multiplexed to LSP to avoid generating a FRR for each VPN) to switch the FRRs of all VPNs. The switching speed is independent of VRF instance and route scale, so services can be restored in a shorter time.

If forwarding-layer route table, VPN tunnel table, LSP tunnel table, public network next hop table and FRR switching table can be designed properly, service convergence will speed up. Otherwise, each route needs to be upgraded. As route scale grows, the switching time cannot meet the carrier-class requirements.

CE dual-homed redundant protection is available between server and PE2/PE3, so the VRRP+BFD networking can be adopted to protect faulty node 4 and 5 from fault. PE quickly sends free ARP to trigger server next-hop update to protect service traffic from CE to PE.

- Application scenario

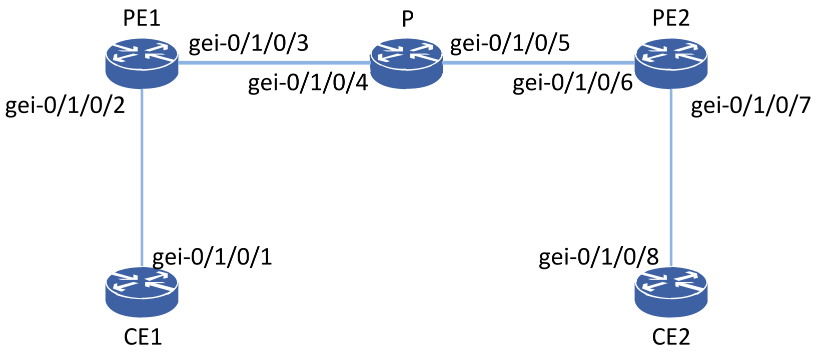
50ms service restoration is a major index to evaluate equipment reliability. In multiservice networking, high reliability is the priority of carrier network. VPN-FRR meets 50ms service protection requirement. It combines with LSP FRR protection, BFD, MPLS-TP fast detection and port down fast detection to provide high reliability for carrier network.

4 Typical configuration

4.1 MPLS L3VPN basic configuration

The topology in Figure 4-1 shows the configuration process of L3VPN basic functions. CE1 and CE2 are in one VPN. After proper configuration, CE1 and CE2 can learn routes from each other.

Figure 4-1 MPLS L3VPN basic configuration topology



- Configuration process:

1 Make basic configuration of IP and MPLS at CE, PE and P nodes.

2 Configure VRF instance and relative RD for each VPN at PEs connected to CE, e.g., PE1 and PE2 in Figure 4-1. L3 interface connected to CE is bound to VRF instance.

3 Configure MP-IBGP neighbor relationship between PE1 and PE2 and enable VPNV4 functions.

4 Configure route protocols between CE and PE to support static route, IGP route and BGP route.

5 Configure L3VPN instance route allocation control, RT, Export Target and Import Target, route convergence, route restriction and alarm at PE.

6 Check configuration results to confirm whether the routes of opposite CE/PE is available at local CE/PE. If yes, the configuration succeeds.

It should be noticed that this is a dynamic L3VPN configuration. The static L3VPN configuration and the process is as follows:

1 PE is configured with static routes to remote CE.

2 PE is configured with internal egress labels of public network next hops. The labels are allocated in a label per VRF mode.

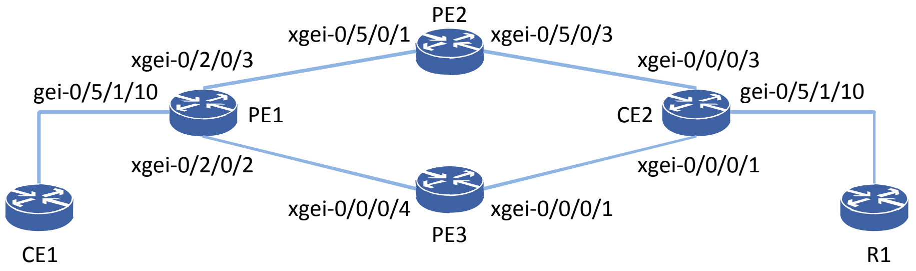
3 PE is configured with the association relationship between VRF public network next hops and internal egress labels. One next hop corresponds to one internal egress label.

4 Thus, PE finds the internal egress label according to VPNID and public network next hop addresses, and the public network external tunnel according to public network next hop.

4.2 MPLS L3VPN-FRR configuration

The topology in Figure 4-2 shows the configuration process of L3VPN FRR functions. PE1 learns the private network routes of the same network segment from two next hops PE2 and PE3, which form L3VPN FRR at PE1. When CE1 traffic is sent to CE2, active and standby private network routes form L3VPN FRR at PE1 to fast switch the traffic in the case of PE2 fault.

Figure 4-2 VPN FRR configuration topology



- Configuration process:

1 Make basic configuration of IP and MPLS at CE, PE and P nodes.

2 Make L3VPN basic configuration at CE and PE nodes. After configuration, PE1 can learn two routes to CE2 via PE2 and PE3 respectively.

3 Enable VPN FRR in PE1 VRF.

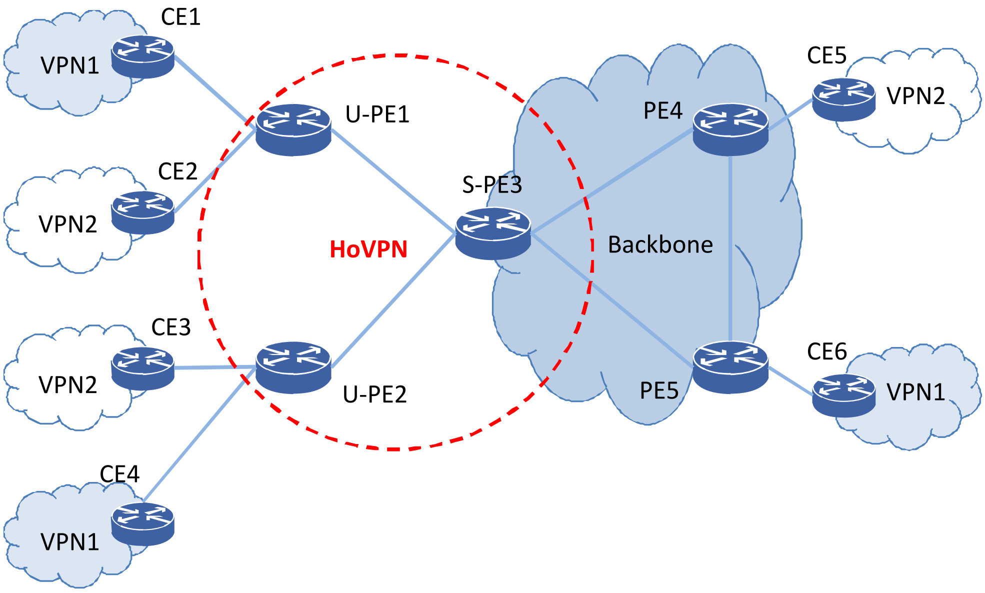
4 Check configuration results to confirm whether active and standby routes to CE2 are available at PE1. If yes, the configuration succeeds.

5 It should be noticed that this is a dynamic VPN FRR configuration. The static VPN FRR configuration is to configure two routes to CE2 for PE1 on the basis of L3VPN configuration (Next hops are PE2 and PE3 respectively. The route to PE2 is active, and that to PE3 is standby).

4.3 HoVPN configuration

The topology in Figure 4-3 shows the configuration process of hierarchical L3VPN functions. The left VPN is the hierarchical L3VPN and the right VPN is the regular L3VPN. UPE notifies local VPN routes to SPE, but SPE notifies a default route or some convergence routes to UPE. Thus, UPE1 and UPE2 only maintain local VPN routes and a default route to SPE3 or some convergence routes.

Figure 4-3 HoVPN configuration topology



- Configuration process:

1 Make basic configuration of IP address, route, MPLS and L3VPN at CE, PE (including UPE, SPE and regular PE), and P nodes.

2 Designate the neighbor UPE address as UPE of local node at SPE.

3 Configure a default route for neighbor UPE or all VPNs of the UPE at SPE, in other words, send to UPE the default route information of the designated/all VPNs with the route's next hop being SPE, and all packets of the UPE-designated/all VPNs are forwarded via SPE.

4 Configure UPE (UPE1 and UPE2) and regular PE (PE4 and PE5) regularly instead of specially.

5 Check configuration results. CE5 has the routes to CE1 and CE3, and CE6 has the routes to CE2 and CE4. CE1, CE2, CE3 and CE4 only have the default route whose next hop is UPE.

5 Abbreviations

Table 5-1 Abbreviations

Abbreviations	Full name
CE ()	Custom Edge
PE ()	Provider Edge
P Router	Provider Router
VPN Site	Virtual Private Network User Site
VRF	VPN Routing and Forwarding
VRF-ID	VPN Routing and Forwarding - Identify
VFI	VPN Forwarding Instance
MP-BGP	Multi-Protocol extensions for BGP-4
VPN	Virtual Private Network
QinQ (802.1Q Stacking)	VPN function based on VLAN stack
VPWS	Virtual Private Wire Service
VPLS	Virtual Private LAN Service
RT	Route Target
RD	Route Distinguisher
IBGP	Internal BGP
EBGP	External BGP

Abbreviations	Full name
VPN Tunnel	Logic channel connecting two PEs
PE-based VPN	PE-based VPN Virtual Private Network
CE-based VPN	CE-based Virtual Private Network
PHP	Penultimate Hop Popping
HoVPN	Hierarchy of VPN
MCE	Multi-VPN-instance CE
ASBR	Autonomous System Border Router