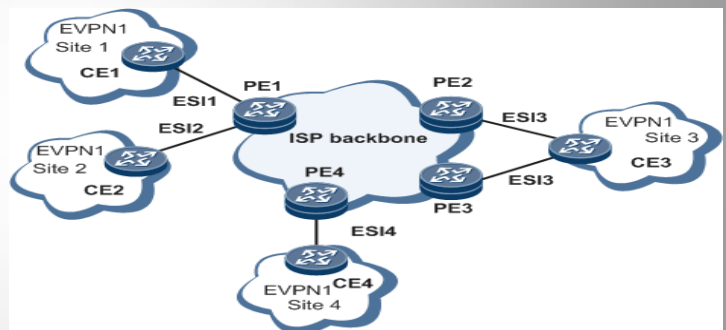


# Ethernet virtual private network (EVPN)



Usman Khan/Tech-Strategy & Planning/Lahore



Wateen Telecom(Pvt) Ltd.

+92-321-4915619

3/3/2020

# EVPN.

## Overview of EVPN:

### **Definition:**

Ethernet virtual private network (EVPN) is used for Layer 2 internetworking. EVPN is similar to BGP/MPLS IP VPN. Using extended BGP reachability information, EVPN implements MAC address learning and advertisement between Layer 2 networks at different sites on the control plane instead of on the data plane.

### **Purpose:**

As services grow rapidly, different sites have an increasingly strong need for Layer 2 interworking. VPLS, which is generally used for such a purpose, has the following shortcomings:

- Lack of support for load balancing: VPLS does not support traffic load balancing in multi-homing networking scenarios.
- High network resource usage: Interworking between sites requires all PEs serving these sites on the ISP backbone network to be fully meshed, with PWs established between any two PEs. If a large number of PEs exist, PW establishment will consume a significant amount of network resources. In addition, a large number of ARP messages must be transmitted for MAC address learning. These ARP messages not only consume network bandwidth but may also consume CPU resources on remote sites that do not need to learn the MAC addresses carried in them.

EVPN solves the preceding problems with the following characteristics:

- EVPN uses extended BGP to implement MAC address learning and advertisement on the control plane instead of on the data plane. This function allows a device to manage MAC addresses in the same way as it manages routes, implementing load balancing between EVPN routes with the same destination MAC address but different next hops.
- EVPN does not require PEs on the ISP backbone network to be fully meshed. PEs on an EVPN use BGP to communicate, and BGP provides the route reflection function. PEs can establish BGP peer relationships only with RRs deployed on the ISP backbone network, with RRs reflecting EVPN routes. This implementation significantly reduces network complexity and minimizes the number of network signaling messages.
- EVPN enables PEs to use ARP to learn the local MAC addresses and use MAC/IP address advertisement routes to learn remote MAC addresses and IP addresses corresponding to these MAC addresses, and store them locally. After receiving another ARP request, a PE searches the locally cached MAC address and IP address based on the destination IP address in the ARP request. If the corresponding information is found, the PE returns an ARP reply

packet. This prevents ARP request packets from being broadcast to other PEs, therefore reducing network resource consumption.

### **Benefits:**

EVPN offers the following benefits:

- Improved link usage and transmission efficiency: EVPN supports load balancing, fully utilizing network resources and reducing network congestion.
- Reduced network resource consumption: By deploying RRs on the public network, EVPN decreases the number of logical connections required between PEs on the public network. In addition, EVPN enables PEs to use locally stored MAC addresses to respond to ARP Request messages from connected sites, minimizing the number of broadcast ARP Request messages.

## **Understanding EVPN:**

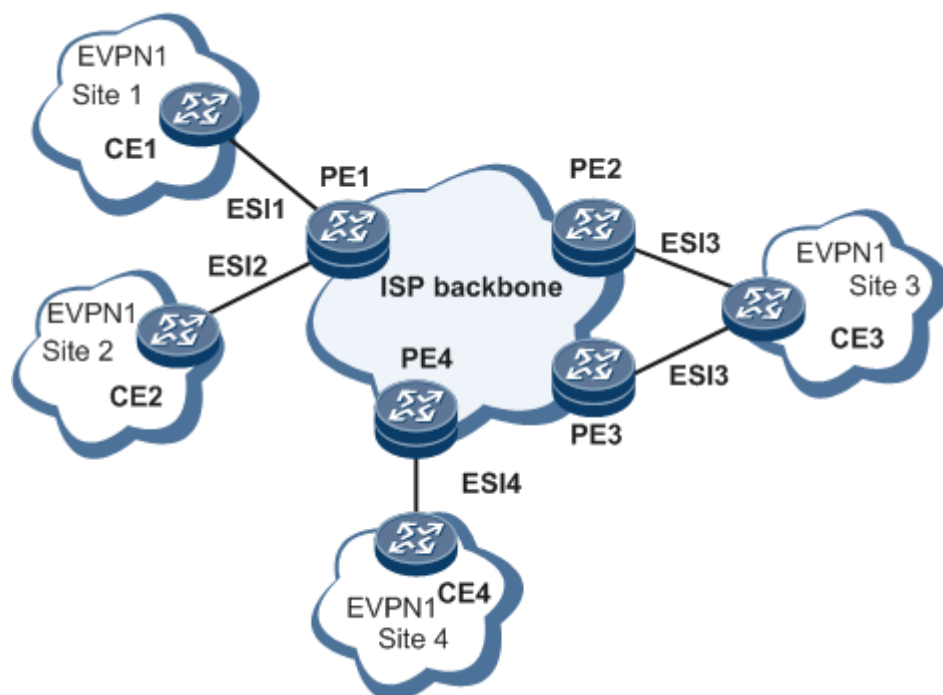
### **EVPN Networking:**

An EVPN has a similar network structure to a BGP/MPLS IP VPN. In EVPN networking, CEs at each site connect to PEs on the ISP backbone network. These PEs have EVPN instances configured and establish BGP EVPN peer relationships and MPLS/SR tunnels with each other. Unlike a BGP/MPLS IP VPN, an EVPN has its sites on Layer 2 networks. Therefore, the PEs learn MAC addresses but not IP routes from the CEs, and then advertise the learned MAC addresses to other sites using EVPN routes.

In EVPN networking, a CE can be single-homed to one PE or multi-homed to several PEs. On the network shown in Figure 12-1, CE1, CE2, and CE4 use the single-homing mode, whereas CE3 uses the multi-homing mode. Load balancing can be implemented in CE multi-homing networking.

EVPN defines Ethernet segment identifiers (ESIs) to identify links between PEs and CEs. Links connecting multiple PEs to the same CE have the same ESI, and links connecting multiple PEs to different CEs have different ESIs. PEs exchange routes that carry ESIs, so that a PE can discover other PEs connecting to the same CE as itself.

**Figure 12-1 EVPN networking**



### EVPN Routes:

To enable sites to learn MAC addresses from each other, EVPN defines a new type of BGP network layer reachability information (NLRI), called the EVPN NLRI. EVPN NLRI includes the following types of EVPN routes:

- Ethernet A-D route: carries the reachability of the local PE to the MAC addresses of its connected sites. PEs advertise Ethernet A-D routes after establishing a BGP EVPN peer relationship. Ethernet A-D routes can be classified as Ethernet A-D Per ES routes or Ethernet A-D Per EVI routes. Ethernet A-D Per ES routes are used in fast convergence, redundancy mode, and split horizon scenarios. Ethernet A-D Per EVI routes are used in alias scenarios. Figure 12-2 shows the NLRI packet format of Ethernet A-D routes.

**Figure 12-2 EVPN NLRI specific to the Ethernet A-D route**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
Ethernet Tag ID (4 bytes)
MPLS Label (3 bytes)

The meanings of fields are as follows:

- Route Distinguisher: In Ethernet A-D Per ES routes, the value of this field is the source IP address configured for a PE, such as X.X.X.X:0. In Ethernet A-D Per EVI routes, the value of this field is the RD value configured for an EVPN instance.
- Ethernet Segment Identifier: Uniquely identifies the connection between a PE and a CE.

- Ethernet Tag ID: In Ethernet A-D Per ES routes, the value of this field is all Fs. In Ethernet A-D Per EVI routes, this field identifies a sub-broadcast domain in an ES. If the value of this field is all 0s, the EVI has only one broadcast domain.
- MPLS Label: In Ethernet A-D Per ES routes, the value of this field is all 0s. In Ethernet A-D Per EVI routes, this field indicates the MPLS label used for EVPN unicast traffic load balancing.
- MAC/IP advertisement route: carries EVPN instance RD, ESI, and label information on the local PE. A PE uses MAC advertisement routes to advertise unicast MAC address reachability information to other PEs. For details, see Unicast MAC Address Transmission. Figure 12-3 shows the format of an EVPN NLRI specific to the MAC/IP advertisement route.

**Figure 12-3 EVPN NLRI specific to the MAC/IP advertisement route**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
Ethernet Tag ID (4 bytes)
MAC Address Length (1 byte)
MAC Address (6 bytes)
IP Address Length (1 byte)
IP Address (0, 4, or 16 bytes)
MPLS Label1 (3 bytes)
MPLS Label2 (0 or 3 bytes)

The description of each field is as follows:

- Route Distinguisher: an 8-byte field representing the RD value of an EVPN instance.
- Ethernet Segment Identifier: a 10-byte field that identifies links between PEs and CEs.
- Ethernet Tag ID: The value of this field is all zeros except that it is the same as the local service ID in an EVPN VPWS scenario or the same as the BD tag value in BD EVPN access in VLAN-aware mode.
- MAC Address Length: a 1-byte field representing the length of the MAC address advertised by the route.
- MAC Address: a 6-byte field representing the MAC address advertised by the route.
- IP Address Length: a field representing the mask length of the host IP address advertised by the route.
- IP Address: a field representing the host IP address advertised by the route.
- MPLS Label1: a field representing the label used for Layer 2 service traffic forwarding.
- MPLS Label2: a field representing the label used for Layer 3 service traffic forwarding.

This type of route plays the following roles on the control plane:

- **MAC address advertisement**

To implement Layer 2 service interworking between hosts connected to different PEs, the two PEs need to learn host MAC addresses from each other. The PEs function as BGP EVPN peers to exchange MAC/IP routes so that they can obtain the host MAC addresses. The MAC Address Length and MAC Address fields identify the MAC address of a host.

- **ARP advertisement**

A MAC/IP advertisement route can carry both the MAC and IP addresses of a host, and therefore can be used to advertise ARP entries between PEs. The MAC Address and MAC Address Length fields identify the MAC address of the host, whereas the IP Address and IP Address Length fields identify the IP address of the host. This type of MAC/IP route is called the ARP route.

- **IP route advertisement**

To implement Layer 3 service interworking between IPv4 hosts connected to different PEs, the two PEs need to learn host IPv4 routes from each other. After a BGP EVPN peer relationship is established between the PEs, they exchange MAC/IP advertisement routes to advertise host IPv4 addresses to each other. The IP Address Length and IP Address fields carried in the MAC/IP advertisement routes indicate the destination addresses of host IP routes, and the MPLS Label2 field must carry a label used for Layer 3 service traffic forwarding. In this case, MAC/IP advertisement routes are also called Integrate Routing and Bridge (IRB) routes.



**NOTE:**

An ARP route carries host MAC and IP addresses and a Layer 2 VNI. An IRB route carries host MAC and IP addresses, a Layer 2 VNI, and a Layer 3 VNI. Therefore, IRB routes carry ARP routes and can be used to advertise IP routes as well as ARP entries.

- **Host ND information advertisement**

A MAC/IP advertisement route can carry both the MAC and IPv6 addresses of a host, and therefore can be used to advertise ND entries between PEs. The MAC Address and MAC Address Length fields identify the MAC address of the host, whereas the IPv6 Address and IPv6 Address Length fields identify the IPv6 address of the host. This type of MAC/IP route is called the ND route.

- **IPv6 route advertisement**

To implement Layer 3 service interworking between IPv6 hosts connected to different PEs, the two PEs need to learn host IPv6 routes from each other. After a BGP EVPN peer relationship is established between the PEs, they exchange MAC/IP advertisement routes to advertise host IPv4 addresses to each other. The IP Address Length and IP Address fields carried in the MAC/IP advertisement routes indicate the destination addresses of host IPv6 routes, and the MPLS Label2 field must carry a label used for Layer 3 service traffic forwarding. In this case, MAC/IP advertisement routes are also called IRBv6 routes.

**NOTE:**

An ND route carries the following valid information: host MAC address, host IPv6 address, and Layer 2 VNI. An IRBv6 route carries the following valid information: host MAC address, host IPv6 address, Layer 2 VNI, and Layer 3 VNI. An IRBv6 route includes information about an ND route and therefore can be used to advertise both a host IPv6 route and host ND entry.

- Inclusive multicast route: After a BGP peer relationship is established between PEs, the PEs transmit inclusive multicast routes to each other. An inclusive multicast route carries the RD and RTs of the EVPN instance on the local PE, source IP address (usually the loopback address of the local PE), and Provider Multicast Service Interface (PMSI). The PMSI is used to carry the tunnel type (ingress replication or mLDP) and tunnel label information in multicast packets. The PMSI and RTs are carried in the route attribute information, and the RD and source IP address are carried in the NLRI. Figure 12-4 shows the format of the EVPN NLRI specific to an inclusive multicast route. In this situation, EVPN involves broadcast, unknown unicast, and multicast (BUM) traffic. A PE forwards the BUM traffic that it receives to other PEs in P2MP mode. A tunnel is established to transmit BUM traffic between PEs through inclusive multicast routes. For details, see BUM Packet Transmission.

**Figure 12-4 EVPN NLRI specific to the inclusive multicast route**

Route Distinguisher (8 bytes)
Ethernet Tag ID (4 bytes)
IP Address Length (1 byte)
Originating Router's IP Address (4 or 16 bytes)

The description of each field is as follows:

- Route Distinguisher: an 8-byte field representing the RD value of an EVPN instance.
  - Ethernet Tag ID: The value of this field is all zeros except that it is the same as the local service ID in an EVPN VPWS scenario or the same as the BD tag value in BD EVPN access in VLAN-aware mode.
  - IP Address Length: a 1-byte field representing the length of the source IP address configured on the local PE.
  - Originating Router's IP Address: a 4-byte or 16-byte field representing the source IP address configured on the local PE.
- Ethernet segment route: carries the EVPN instance RD and ESI information and source IP address on the local PE. PEs connecting to the same CE use Ethernet segment routes to discover each other. This type of route is used in Designated forwarder election. Figure 12-5 shows the format of an EVPN NLRI specific to the Ethernet segment route.

**Figure 12-5 EVPN NLRI specific to the Ethernet segment route**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
IP Address Length (1 byte)
Originating Router's IP Address (4 or 16 bytes)

The description of each field is as follows:

- Route Distinguisher: an 8-byte field representing a combination of the source IP address on the local PE and :0, such as X.X.X.X:0.
  - Ethernet Segment Identifier: a 10-byte field that identifies links between PEs and CEs.
  - IP Address Length: a 1-byte field representing the length of the source IP address configured on the local PE.
  - Originating Router's IP Address: a 4-byte or 16-byte field representing the source IP address configured on the local PE.
- IP prefix route: used to advertise a host IP address or the network segment to which the host IP address belongs, which has been received from the access network. Figure 12-6 shows the format of an EVPN NLRI specific to an IP prefix route.

**Figure 12-6 Format of EVPN NLRI specific to an IP prefix route**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
Ethernet Tag ID (4 bytes)
IP Prefix Length (bytes)
IP Prefix (4 or 16 bytes)
GW IP Address (4 or 16 bytes)
MPLS Label (3 bytes)

The description of each field is as follows:

- Route Distinguisher: an 8-byte field representing the RD value of an EVPN instance.
- Ethernet Segment Identifier: a 10-byte field that identifies links between PEs and CEs.
- Ethernet Tag ID: Currently, each bit of the field value must be 0.
- IP Prefix Length: IP prefix mask length carried in the route.
- IP Prefix: IP prefix address carried in the route.
- GW IP Address: default gateway address.
- MPLS Label: a field representing the label used for Layer 3 service traffic forwarding.

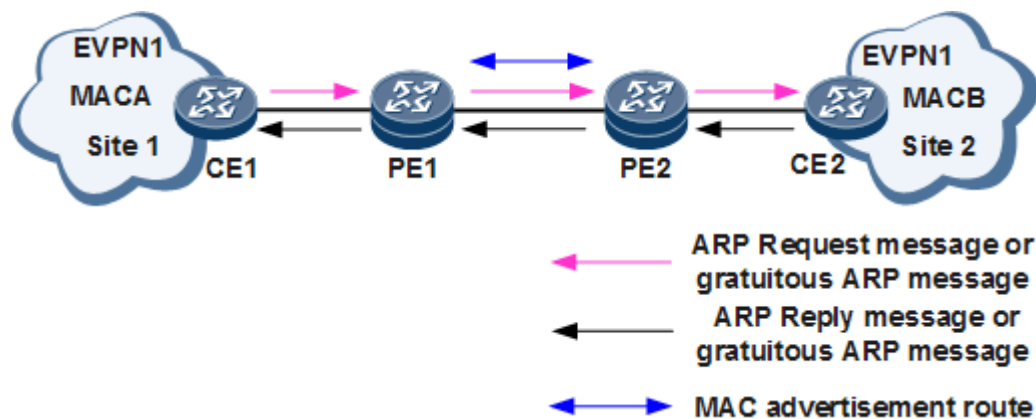
#### Unicast MAC Address Advertisement:



On the network shown in Figure 12-7, unicast MAC addresses are advertised as follows:

1. CE1 sends an ARP Request message or a gratuitous ARP message to advertise its MAC address (MAC A) and IP address to CE2. After the ARP Request message or gratuitous ARP message arrives at PE1, PE1 generates a MAC/IP advertisement route based on MAC A.
2. CE2 receives the ARP Request message or gratuitous ARP message from CE1 and responds with an ARP Reply message or a gratuitous ARP message carrying CE2's MAC address (MAC B) and IP address. After the ARP Reply message or gratuitous ARP message arrives at PE2, PE2 generates a MAC/IP advertisement route based on MAC B.
3. PE1 and PE2 exchange MAC/IP advertisement route that carry MAC addresses, next hops, and EVPN instance extended community attributes (such as RTs).
4. PE1 and PE2 construct EVPN instance forwarding entries based on the RTs carried in received MAC/IP advertisement route.

**Figure 12-7 Unicast MAC address advertisement networking**

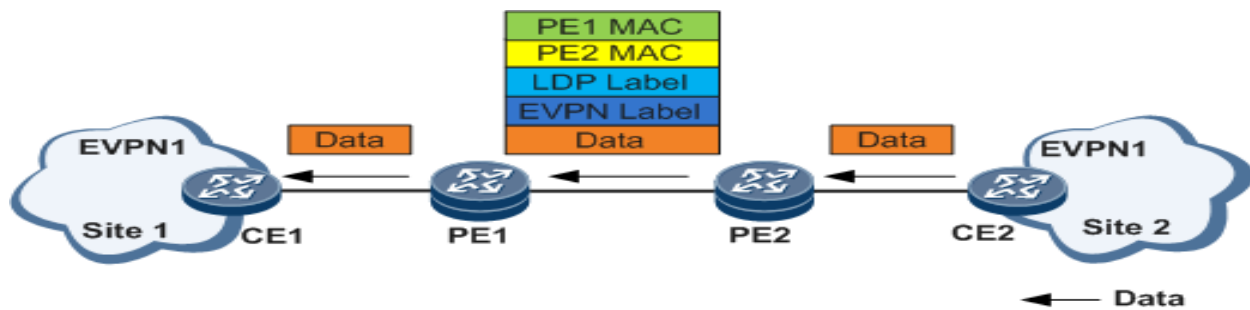


#### Unicast Packet Transmission:

After a PE connecting to a site has learned the MAC addresses of other sites and established public network tunnels, the PE can send unicast packets to other sites. On the network shown in Figure 12-8, unicast packets are transmitted as follows:

1. CE2 forwards unicast packets to PE2 at Layer 2.
2. Upon receipt of the unicast packets, PE2 encapsulates an EVPN label, a public network LDP LSP label, PE2's MAC address, and PE1's MAC address in sequence into the unicast packets. PE2 then forwards the encapsulated unicast packets to PE1.
3. PE1 decapsulates the received unicast packets and sends the unicast packets to the sites of the EVPN identified by the EVPN label carried in the packets.

**Figure 12-8 Unicast packet transmission networking**



### BUM Packet Transmission:

After two PEs establish a BGP EVPN peer relationship, they exchange inclusive multicast routes. A PE can discover PEs that belong to the same EVPN instance as itself based on RTs carried in the inclusive multicast routes it receives. The RTs identify the reachability information of these PEs. This PE then automatically establishes MPLS tunnels with these PEs. BUM packets can then traverse these LDP tunnels. On the network shown in Figure 12-9, BUM packets are transmitted as follows:

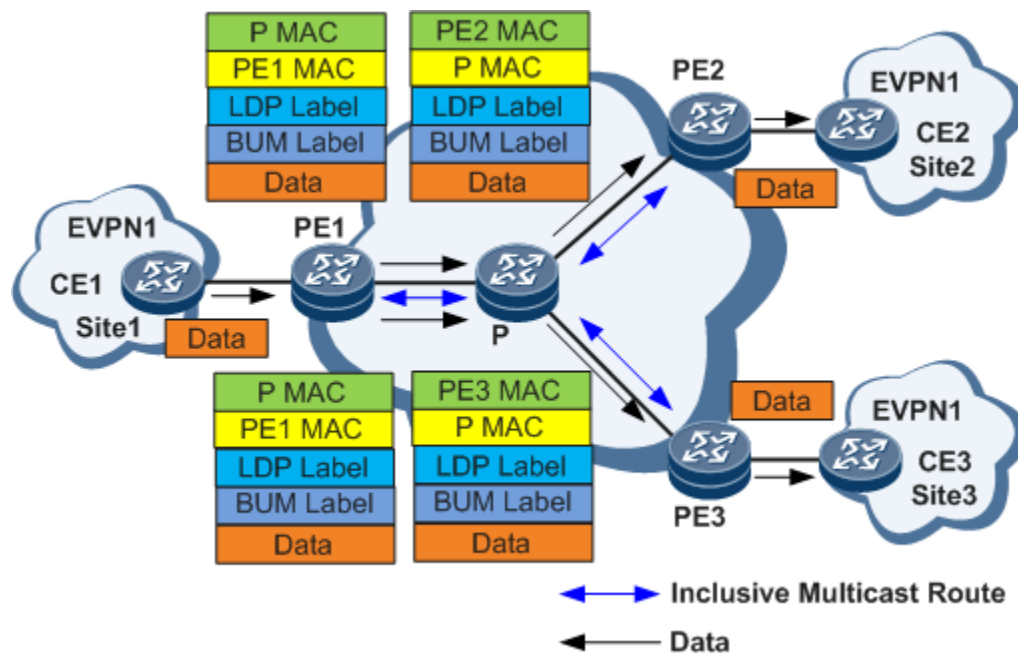
1. CE1 sends BUM packets to PE1.
2. Upon receipt of the BUM packets, PE1 forwards them to PE2 and PE3 that belong to the same EVPN. Specifically, PE1 replicates the received BUM packets and encapsulates the EVPN BUM label, public network Tunnel label, PE1's MAC address, and P's MAC address in sequence into these packets before sending them to PE2 and PE3.
3. Upon receipt of the BUM packets, PE2 and PE3 decapsulate the BUM packets and send the BUM packets to the sites of the EVPN identified by the EVPN BUM label carried in the packets.



#### NOTE:

In the case where a CE is dual-homed to two PEs, an EVPN ESI label will be encapsulated into the BUM packets exchanged between the two PEs to prevent loops.

**Figure 12-9 BUM packet transmission networking**



## EVPN – MPLS:

### EVPN Multi-Homing Technology:

#### *Related Concepts:*

- **Designated forwarder (DF) election:**

On the network shown in Figure 12-10, CE1 is dual-homed to PE1 and PE2, and CE2 sends BUM traffic to PE1 and PE2. In this scenario, CE1 receives the same copy of traffic from both PE1 and PE2, wasting network resources. To solve this problem, EVPN allows one PE to be elected to forward BUM traffic. The elected PE is referred to as the DF. If PE1 is elected, it becomes the primary DF, with PE2 functioning as the backup DF. The primary DF forwards BUM traffic from CE2 to CE1.

If a PE interface connecting to a CE is Down, the PE functions as a backup DF. If a PE interface connecting to a CE is Up, the PE and other PEs with Up interfaces elect a primary DF using the following procedure:

1. The PEs establish BGP EVPN peer relationships with each other and then exchange Ethernet segment routes.

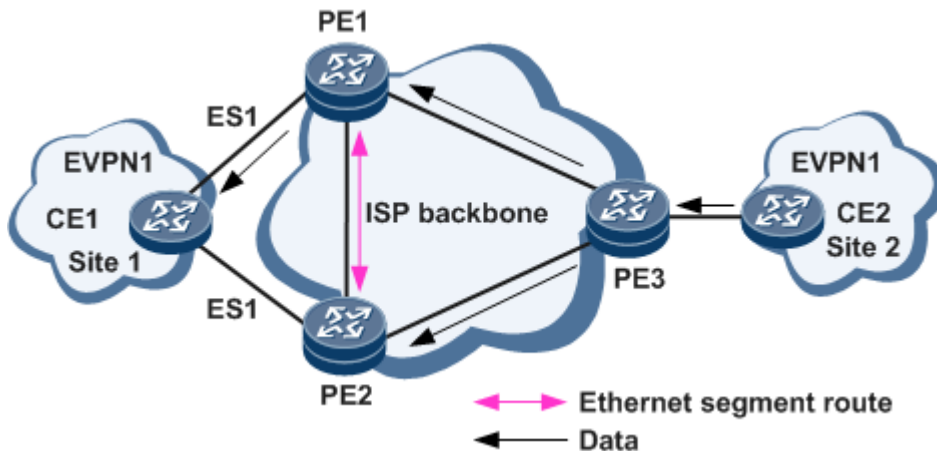
2. Upon receipt of the Ethernet segment routes, each PE generates a multi-homing PE list based on the ESIs carried in Ethernet segment routes. Each multi-homing PE list contains information about all PEs connecting to the same CE.
3. Each PE then sequences the PEs in each multi-homing PE list based on the source IP addresses carried in Ethernet segment routes. The PEs are numbered from 0.
4. If interface-based DF election is enabled, the PE with the smallest source IP address is elected to be the primary DF. If VLAN-based DF election is enabled, the PE with a specific sequence number is elected to be the primary DF. The sequence number is calculated using the following expression formula:  $(V \bmod N) = i$ , in which  $i$  indicates a PE's sequence number,  $N$  indicates the number of PEs to which a CE is multi-homed, and  $V$  indicates the VLAN ID over an Ethernet segment.



**NOTE:**

An Ethernet segment may have multiple VLANs configured. In this case, the smallest VLAN ID is used as the **V** value.

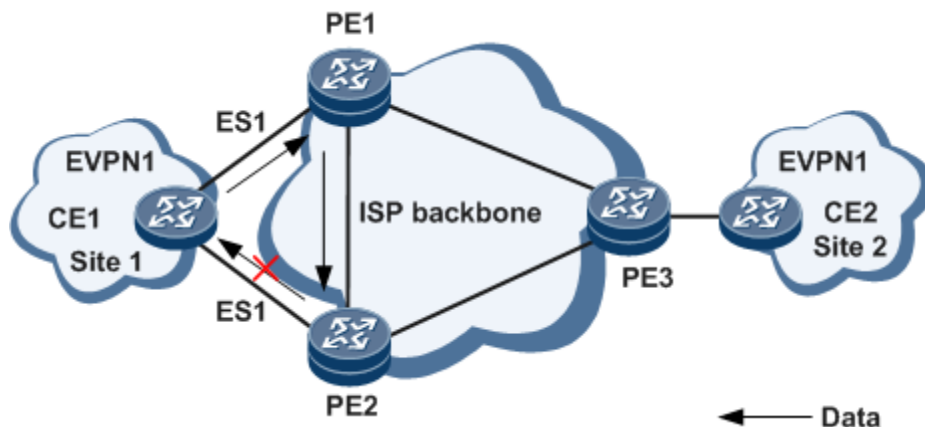
**Figure 12-10 DF election networking**



- **Split Horizon:**

On the network shown in Figure 12-11, CE1 is dual-homed to PE1 and PE2 and has load balancing enabled. If PE1 and PE2 have established a BGP EVPN peer relationship with each other, after PE1 receives BUM traffic from CE1, PE1 forwards the BUM traffic to PE2. If PE2 forwards BUM traffic to CE1, a loop will occur. To prevent this problem, EVPN uses split horizon. After PE1 forwards the BUM traffic to PE2, PE2 checks the EVPN ESI label carried in the traffic. If the ESI carried in the label equals the ESI for the link between PE2 and CE1, PE2 does not forward the traffic to CE1, preventing a loop.

**Figure 12-11 Split horizon networking**



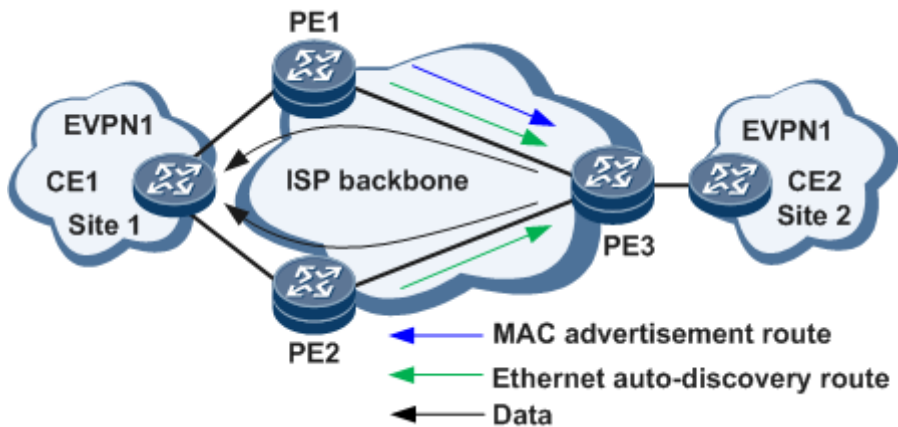
- **Redundancy mode and aliasing:**

If a CE is multi-homed to several PEs, a redundancy mode can be configured to specify the redundancy mode of PEs connecting to the same CE. The redundancy mode determines whether load balancing is implemented for unicast traffic in CE multi-homing scenarios. On the network shown in Figure 12-12, the transmission mode of unicast traffic sent by PE3 to CE1 varies according to the redundancy modes configured on PE1 and PE2.

- If PE1 and PE2 are both configured to work in All-Active mode, after PE1 and PE2 send Ethernet auto-discovery route carrying the redundancy mode to PE3, PE3 sends unicast traffic destined for CE1 to both PE1 and PE2 in load balancing mode.
- If either PE1 or PE2 or both PE1 and PE2 are configured to work in Single-Active mode, after PE1 and PE2 send Ethernet auto-discovery route carrying the redundancy mode to PE3, PE3 uses the optimal received route as the primary route and the second optimal received route as the backup route to implement FRR.

EVPN also supports aliasing, which is the ability of a PE to signal that it has reachability to an EVPN instance on a given Ethernet segment even when it has learned no MAC addresses from that Ethernet segment. In the case where a CE is multi-homed to several PEs, it is possible that only a single PE learns a set of the MAC addresses associated with traffic transmitted by the CE. Aliasing enables remote PEs to learn the reachability of CE-side MAC addresses based on the ESIs carried in Ethernet auto-discovery route received from multi-homing PEs. On the network shown in Figure 12-12, only PE1 sends MAC/IP advertisement routes that carry CE-side MAC addresses to PE3, but PE3 can learn from Ethernet auto-discovery route that PE2 is also reachable to CE1. As a result, PE3 load-balances traffic destined for CE1 between PE1 and PE2.

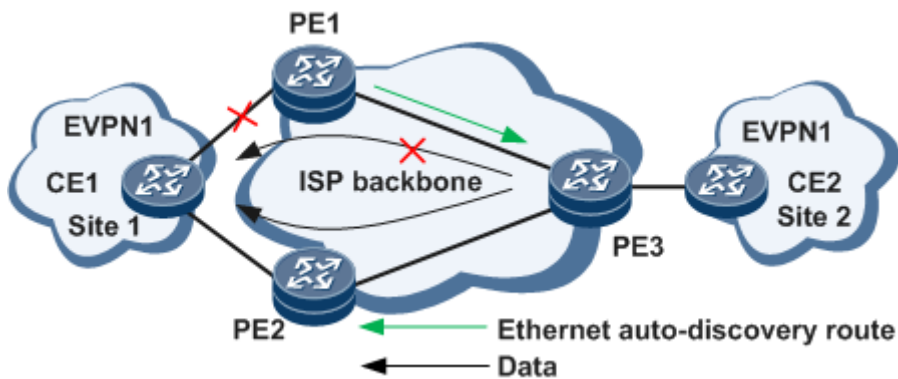
**Figure 12-12 Redundancy mode and aliasing networking**



- **Fast route convergence:**

On the network shown in Figure 12-13, if the link between CE1 and PE1 fails, PE1 advertises an Ethernet auto-discovery route to PE3, informing PE3 that PE1 has become unreachable to Site 1. Upon receipt of the route, PE3 withdraws the corresponding routes and sends traffic to Site 1 only through PE2, implementing fast route convergence.

**Figure 12-13 Fast route convergence networking**



## **EVPN Seamless MPLS Fundamentals:**

Seamless MPLS achieves end-to-end service transmission along an LSP traversing the access, aggregation, and core layers. Therefore, service traffic can be transmitted between any two points on an LSP. The seamless MPLS network architecture maximizes service scalability using the following functions:

- Allows access nodes to signal all services to an LSP.
- Uses the same transport layer convergence technology to switch all services to backup paths in case of network-side faults, ensuring proper service transmission.

### ***Background:***

The use of MPLS networks increases requirements for service scalability of network architecture. Different MANs of a service provider or collaborative backbone networks of different service providers often span multiple ASs.

### ***Implementation:***

Through seamless MPLS networking, all services (support inter-AS Option C) are signaled to an LSP only by access nodes, and all network-side faults are rectified using the same transport layer convergence technology, which does not affect service transmission.

### ***Usage Scenario:***

Seamless MPLS supports the following networking solutions:

- Intra-AS seamless MPLS: The access, aggregation, and core layers are deployed within a single AS. This solution mainly applies to mobile bearer networks.
- Inter-AS seamless MPLS: The access and aggregation layers are deployed within a single AS, whereas the core layer is in another AS. This solution mainly applies to enterprise services.

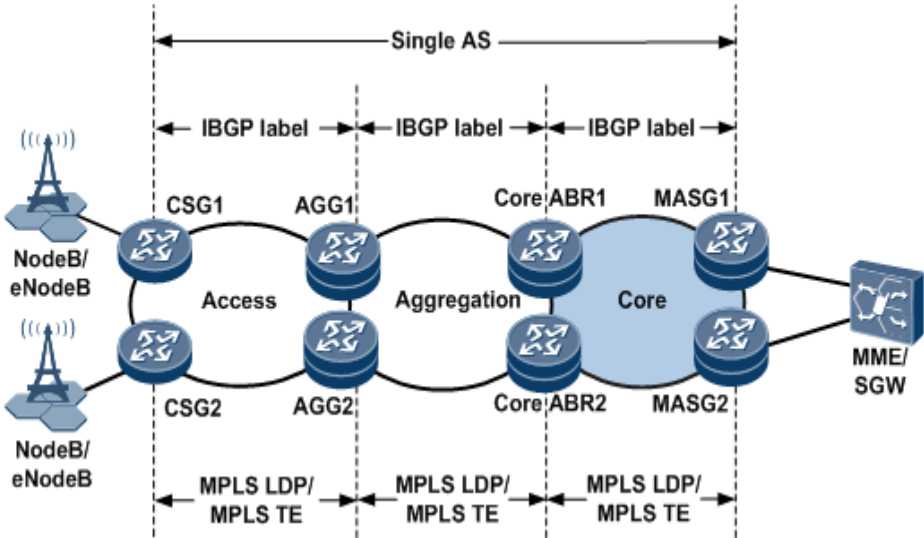
### ***EVPN Intra-AS Seamless MPLS:***

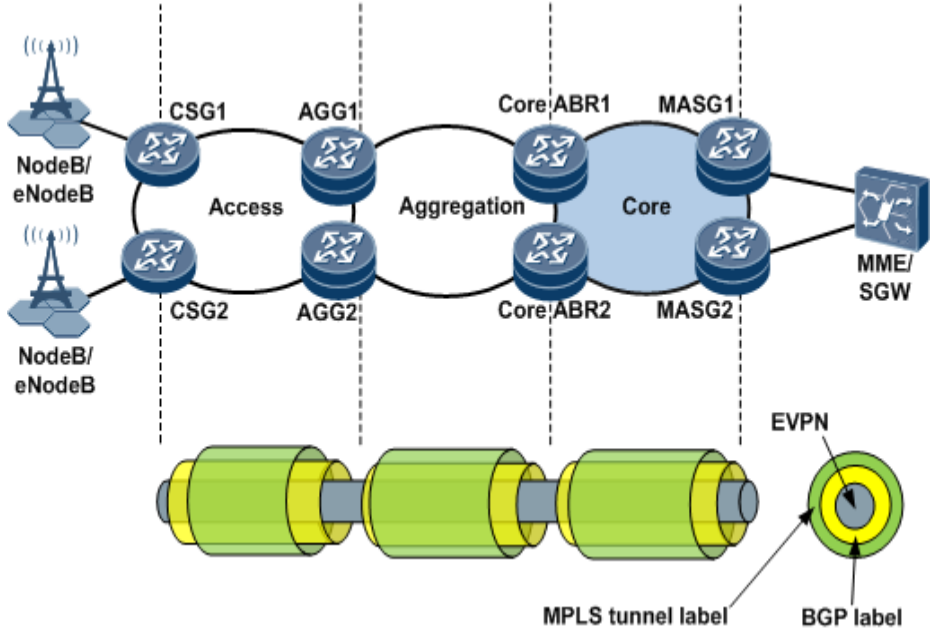
**Table 12-1 EVPN intra-AS seamless MPLS networking.**

Network Deployment		Description
Control plane	Deploy routing protocols	<p>In <u>Figure 12-14</u>, routing protocols are deployed on devices as follows:</p> <ul style="list-style-type: none"><li>• An IGP (IS-IS or OSPF) is enabled on devices at the access,</li></ul>

Network Deployment	Description
	<p>aggregation, and core layers to establish connectivity within the AS.</p> <ul style="list-style-type: none"> <li>An IBGP peer relationship is established between each of the following pairs of devices: <ul style="list-style-type: none"> <li>CSG and AGG</li> <li>AGG and core ABR</li> <li>Core ABR and MASG</li> </ul> </li> </ul> <p>An AGG and core ABR are configured as route reflectors (RRs) so that a CSG and MASG can obtain routes destined for each other's loopback address.</p> <ul style="list-style-type: none"> <li>The next hop addresses in BGP routes are set on the AGG and core ABR to the devices' own addresses to prevent advertising unnecessary IGP area-specific public routes.</li> </ul> <p><b>Figure 12-14 Routing protocol deployment for intra-AS seamless MPLS networking</b></p>
Deploy tunnels.	<p>In <a href="#">Figure 12-15</a>, tunnels are deployed as follows:</p> <ul style="list-style-type: none"> <li>A public network tunnel is established using LDP, TE, or LDP over TE in each IGP area.</li> <li>An IBGP peer relationship is established between each of the following pairs of devices: <ul style="list-style-type: none"> <li>CSG and AGG</li> </ul> </li> </ul>



Network Deployment	Description
	<ul style="list-style-type: none"> <li>▪ AGG and core ABR</li> <li>▪ Core ABR and MASG</li> </ul> <p>These devices are enabled to advertise labeled routes and assign labels to BGP routes that match a specified routing policy. After the devices exchange labeled BGP routes, an end-to-end BGP LSP is established between the CSG and MASG.</p> <p><b>Figure 12-15 Tunnel deployment for intra-AS seamless MPLS networking</b></p> 
Forwarding plane	<p><u>Figure 12-16</u> illustrates the forwarding plane of intra-AS seamless MPLS networking. Seamless MPLS is mainly used to transmit EVPN packets. The following example demonstrates how EVPN packets, including labels and data, are transmitted from a CSG to an MASG along the path CSG1-&gt;AGG1-&gt;core ABR1-&gt;MASG1.</p> <ol style="list-style-type: none"> <li>1. The CSG pushes a BGP LSP label and an MPLS tunnel label in sequence into each EVPN packet and forwards the packets to the AGG.</li> <li>2. Upon receipt, the AGG removes the access-layer MPLS tunnel labels from the packets and swaps the existing BGP LSP labels for new labels. The AGG then pushes an aggregation-layer MPLS tunnel label into each packet and proceeds to forward the packets to the core ABR. If the penultimate hop popping (PHP)</li> </ol>

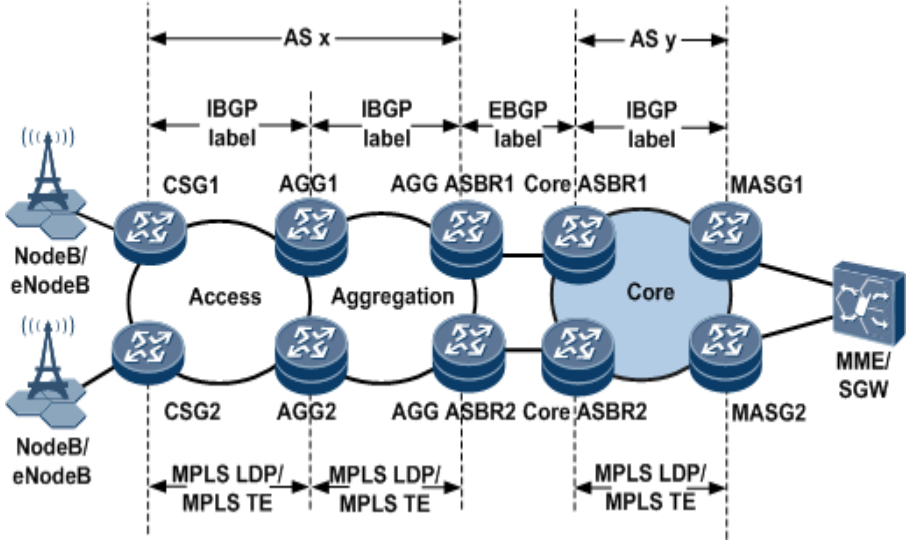
Network Deployment	Description
	<p>function is enabled on the AGG, the CSG has removed the MPLS tunnel labels from the packets, and therefore, the AGG receives packets without MPLS tunnel labels.</p> <ol style="list-style-type: none"> <li>3. Upon receipt, the core ABR removes aggregation-layer MPLS tunnel labels from the EVPN packets and swaps the existing BGP LSP labels for new labels. The AGG pushes a core-layer MPLS tunnel label to each packet and forwards the packets to the MASG.</li> <li>4. Upon receipt, the MASG removes MPLS tunnel labels and BGP LSP labels from the EVPN packets. If the PHP function is enabled on the MASG, the core ABR has removed the core-layer MPLS tunnel labels from the packets, and therefore, the MASG receives packets without MPLS tunnel labels. The EVPN packet transmission along the intra-AS seamless MPLS LSP is complete.</li> </ol> <p><b>Figure 12-16 Forwarding plane for intra-AS seamless MPLS networking</b></p> 

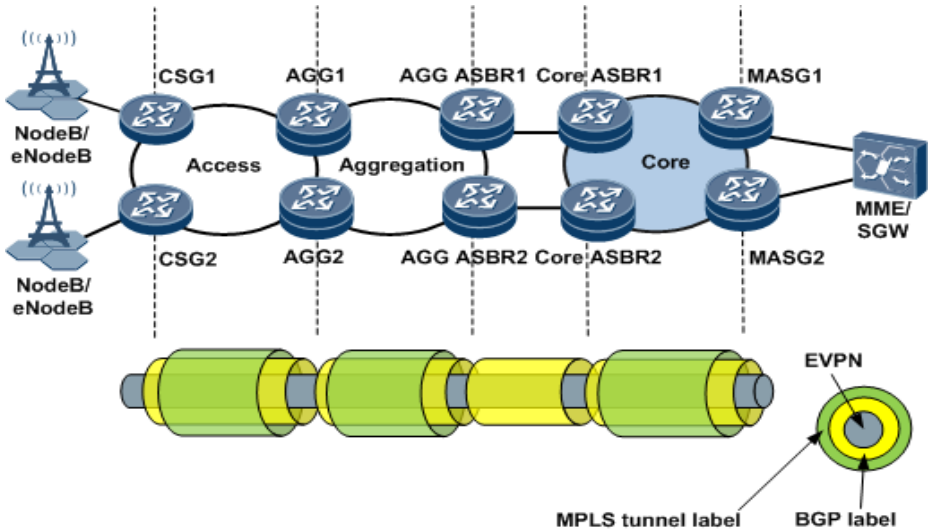
## ***EVPN Inter-AS Seamless MPLS:***

**Table 12-2 EVPN inter-AS seamless MPLS networking**

Network Deployment		Description
Control plane	Deploy routing protocols	<p>In <a href="#">Figure 12-17</a>, routing protocols are deployed on devices as follows:</p> <ul style="list-style-type: none"><li>• An IGP (IS-IS or OSPF) is enabled on devices at the access, aggregation, and core layers to establish connectivity within the AS.</li><li>• A BGP peer relationship is established between each of the following pairs of devices:<ul style="list-style-type: none"><li>▪ CSG and AGG</li><li>▪ AGG and AGG ASBR</li><li>▪ AGG ASBR and core ASBR</li><li>▪ Core ASBR and MASG</li></ul></li></ul> <p>An EBGP peer relationship between an AGG ASBR and a core ASBR is established, and IBGP peer relationships between other devices are established.</p> <ul style="list-style-type: none"><li>• The AGG is configured as an RR so that IBGP peers can exchange BGP routes, and the CSG and MASG can obtain BGP routes destined for each other's loopback addresses.</li><li>• If the AGG ASBR and core ASBR are indirectly connected, an IGP neighbor relationship between them must be established to implement connectivity between ASs.</li></ul> <p><b>Figure 12-17 Routing protocol deployment for inter-AS seamless MPLS networking</b></p>

Network Deployment	Description
	<div data-bbox="532 300 1459 787" data-label="Diagram"> </div> <p data-bbox="345 863 451 932">Deploy tunnels.</p> <p data-bbox="532 863 1162 890">In <a href="#">Figure 12-18</a>, tunnels are deployed as follows:</p> <ul data-bbox="532 932 1433 1801" style="list-style-type: none"> <li>• A public network tunnel is established using LDP, TE, or LDP over TE in each IGP area. An LDP LSP or a TE LSP is established if more than one hop exists between the AGG ASBR and core ASBR.</li> <li>• The CSG, AGG, AGG ASBR, and core ASBR are enabled to advertise labeled routes and assign labels to BGP routes that match a specified routing policy. After the devices exchange labeled BGP routes, a BGP LSP is established between the CSG and core ASBR.</li> <li>• Tunnel deployment in the core area is as follows:             <ul style="list-style-type: none"> <li>▪ A BGP LSP between the core ASBR and MASG is established. This BGP LSP and the BGP LSP between the CSG and core ASBR are combined into an end-to-end BGP LSP. The route to the MASG's loopback address is imported into the BGP routing table and advertised to the core ASBR using the IBGP peer relationship. The core ASBR assigns a label to the route and advertises the labeled route to the AGG ASBR.</li> <li>▪ No BGP LSP is established between the core ASBR and MASG. The core ASBR runs an IGP to learn the route destined for the MASG's loopback address and imports the route to the routing table. The core ASBR assigns a BGP label to the route and associates the route with an intra-AS LSP. The BGP LSP between the CSG and core ASBR and the MPLS LSP in the core area are combined into an end-to-end tunnel.</li> </ul> </li> </ul> <p data-bbox="532 1818 1333 1850"><b>Figure 12-18 Tunnel deployment for inter-AS seamless MPLS</b></p>

Network Deployment	Description
	
Forwarding plane	<p><u>Figure 12-19</u> illustrates the forwarding plane of the inter-AS seamless MPLS networking with a core-layer BGP LSP established. Seamless MPLS is mainly used to transmit EVPN packets. The following example demonstrates how EVPN packets, including labels and data, are transmitted from a CSG to an MASG along the path CSG1-&gt;AGG1-&gt;AGG ASBR1-&gt;core ASBR1-&gt;MASG1.</p> <ol style="list-style-type: none"> <li>1. The CSG pushes a BGP LSP label and an MPLS tunnel label in sequence into each EVPN packet and forwards the packets to the AGG.</li> <li>2. Upon receipt, the AGG removes the access-layer MPLS tunnel labels from the packets and swaps the existing BGP LSP labels for new labels. The AGG then pushes an aggregation-layer MPLS tunnel label into each packet and proceeds to forward the packets to the AGG ASBR. If the PHP function is enabled on the AGG, the CSG has removed the MPLS tunnel labels from the packets, and therefore, the AGG receives packets without MPLS tunnel labels.</li> <li>3. Upon receipt, the AGG ASBR removes the MPLS tunnel labels from the EVPN packets and swaps the existing BGP LSP label for a new label in each packet. It then forwards the packets to the core ASBR. If the PHP function is enabled on the AGG ASBR, the AGG has removed the MPLS tunnel labels from the packets, and</li> </ol>

Network Deployment	Description
	<p>therefore, the AGG ASBR receives packets without MPLS tunnel labels.</p> <ol style="list-style-type: none"> <li>4. Upon receipt, the core ASBR swaps a BGP LSP label for a new label and pushes a core-layer MPLS tunnel label into each packet. It then forwards the packets to the MASG.</li> <li>5. Upon receipt, the MASG removes MPLS tunnel labels, BGP LSP labels, and VPN labels from the packets. If the PHP function is enabled on the core ASBR, the core ASBR has removed the MPLS tunnel labels from the packets, and therefore, the MASG receives packets without MPLS tunnel labels.</li> </ol> <p>The EVPN packet transmission along the inter-AS seamless MPLS LSP is complete.</p> <p><b>Figure 12-19 Forwarding plane for the inter-AS seamless MPLS networking with a BGP LSP established in the core area</b></p>  <p>Figure12-20 illustrates the forwarding plane for the inter-AS seamless MPLS networking without a BGP LSP established in the core area. The process of transmitting EVPN packets on this network is similar to that on a network with a BGP LSP established. The difference is that without a BGP LSP in the core area, the core ASBR removes BGP labels from packets and pushes MPLS tunnel labels into these packets.</p> <p><b>Figure 12-20</b> Forwarding plane for the inter-AS seamless MPLS</p>

Network Deployment	Description
	networking without a BGP LSP established in the core area

### Reliability:

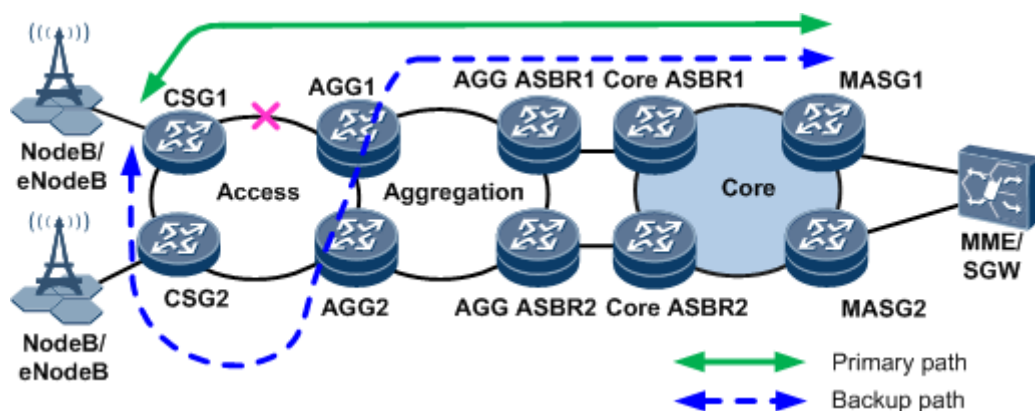
Seamless MPLS network reliability can be improved using various functions. If a network fault occurs, a device immediately detects the fault and switch traffic to a standby link.

The following examples demonstrate reliability functions on an inter-AS seamless MPLS network.

- **A fault occurs on a link between a CSG and an AGG.**

On the inter-AS seamless MPLS network shown in [Figure 12-21](#), the active link along the primary path between CSG1 and AGG1 fails. After BFD for LDP LSP or BFD for CR-LSP detects the fault, the BFD module uses LDP FRR, TE hot-standby, or BGP FRR to switch traffic from the primary path to the backup path.

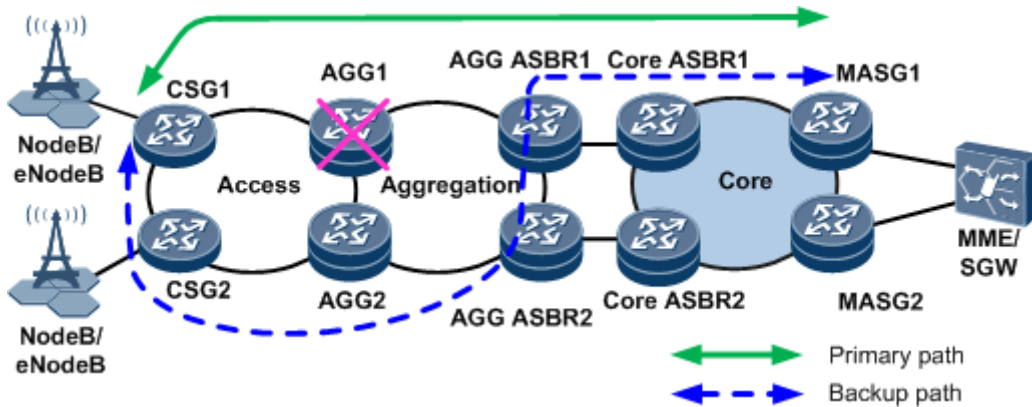
**Figure 12-21 Traffic protection triggered by a fault on the link between the CSG and AGG on the inter-AS seamless MPLS network**



- **A fault occurs on an AGG.**

On the inter-AS seamless MPLS network shown in [Figure 12-22](#), BGP auto FRR is configured on CSGs and AGG ASBRs to protect traffic on the BGP LSP between CSG1 and MASG1. If BFD for LDP or BFD for TE detects AGG1 failure, the BFD module instructs CSG1 to switch traffic from the primary path to the backup path.

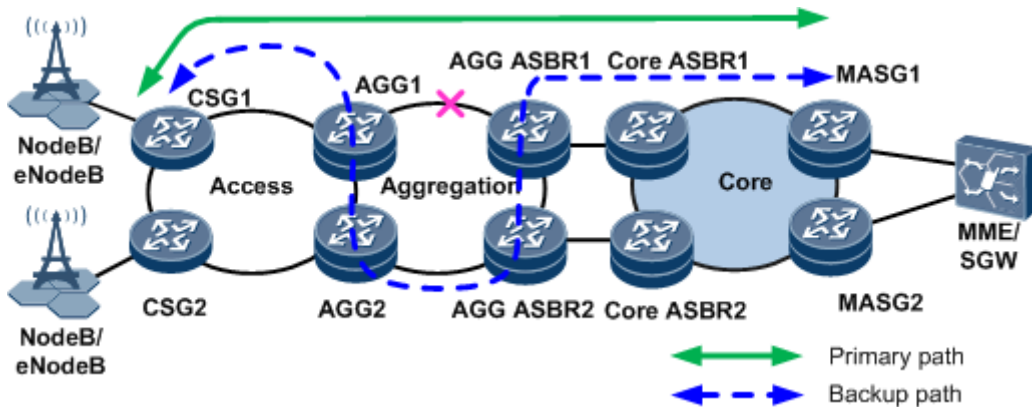
**Figure 12-22 Traffic protection triggered by a fault on an AGG on the inter-AS seamless MPLS network**



- **A fault occurs on the link between an AGG and an AGG ASBR.**

On the inter-AS seamless MPLS network shown in [Figure 12-23](#), a fault occurs on the link between AGG1 and AGG ASBR1. After BFD for LDP LSP or BFD for CR-LSP detects the fault, the BFD module uses LDP FRR, TE hot-standby, or BGP FRR to switch traffic from the primary path to the backup path.

**Figure 12-23 Traffic protection triggered by a fault on the link between an AGG and an AGG ASBR on the inter-AS seamless MPLS network**

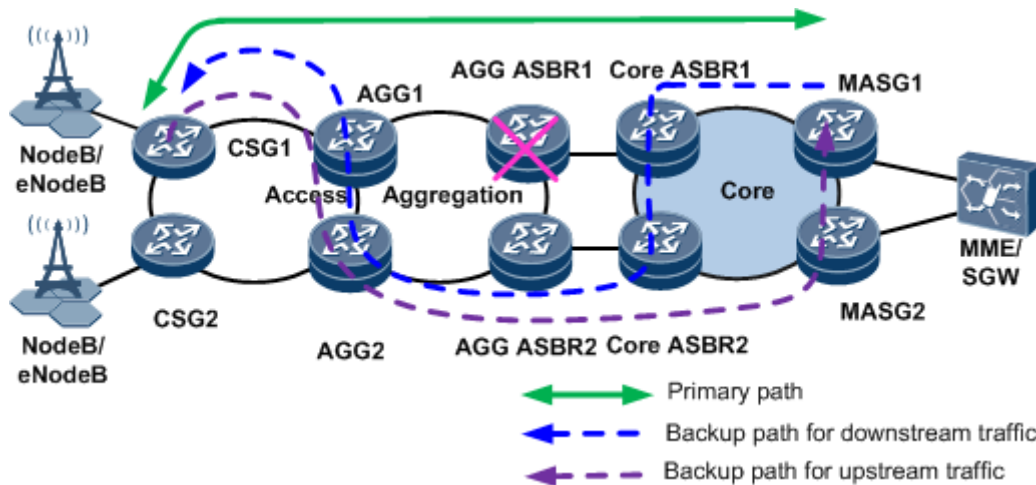


- **A fault occurs on an AGG ASBR.**

On the inter-AS seamless MPLS network shown in [Figure 12-24](#), BFD for LDP or BFD for TE is configured on AGG1, and BFD for interface is configured on core ASBR1. If AGG ASBR1 fails, the BFD modules on AGG1 and core ASBR1 detect the fault and trigger BGP auto FRR. BGP auto FRR switches both upstream and downstream traffic from the primary path to backup paths.



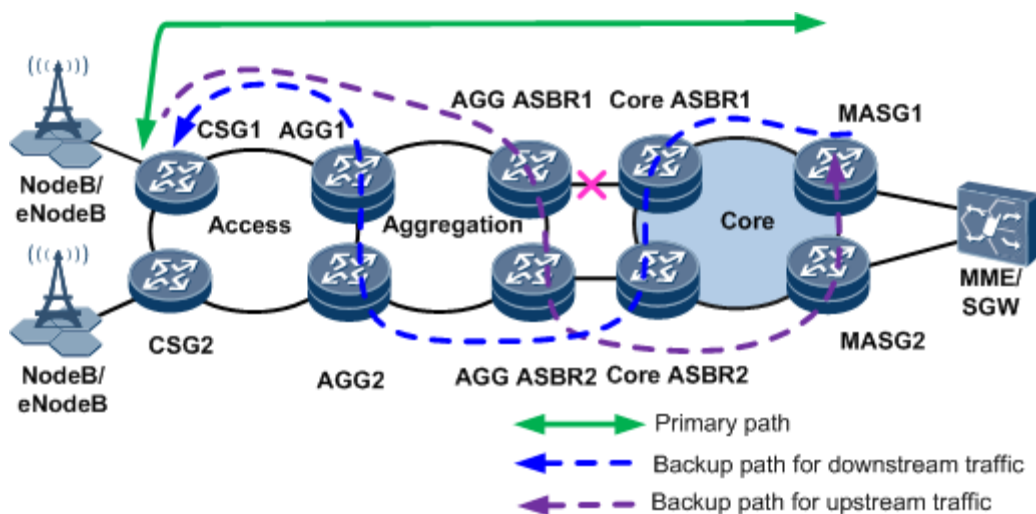
**Figure 12-24 Traffic protection triggered by a fault on an AGG ASBR on the inter-AS seamless MPLS network**



- **A fault occurs on the link between an AGG ASBR and a core ASBR.**

On the inter-AS seamless MPLS network shown in [Figure 12-25](#), BFD for interface is configured on AGG ASBR1 and core ASBR1. If the BFD module detects a fault on the link between AGG ASBR1 and core ASBR1, the BFD module triggers BGP Auto FRR. BGP auto FRR switches both upstream and downstream traffic from the primary path to backup paths.

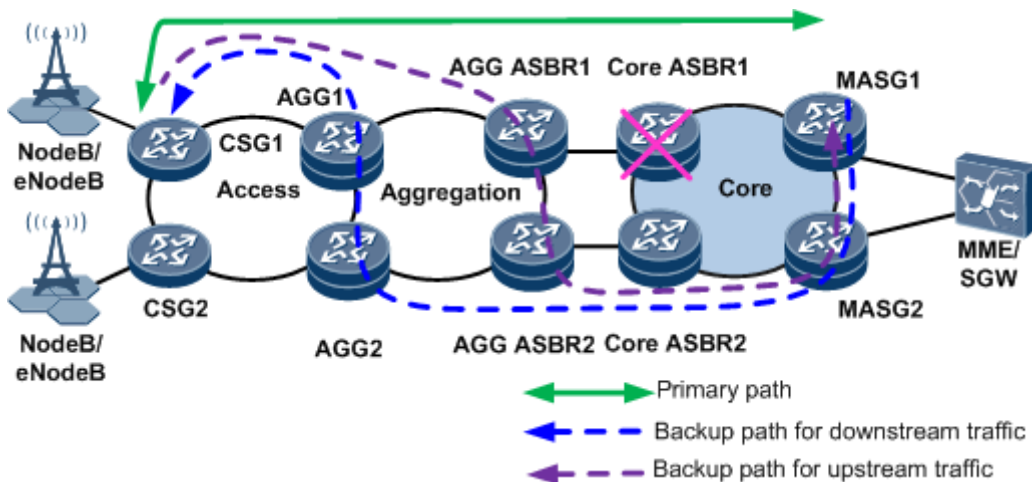
**Figure 12-25 Traffic protection triggered by a fault on the link between an AGG ASBR and a core ASBR on the inter-AS seamless MPLS network**



- **A fault occurs on a core ASBR.**

On the inter-AS seamless MPLS network shown in [Figure 12-26](#), BFD for interface and BGP auto FRR are configured on AGG ASBR1. BGP auto FRR and BFD for LDP (or for TE) are configured on MASGs to protect traffic on the BGP LSP between CSG1 and MASG1. If the BFD module detects a fault on core ASBR1, it instructs AGG ASBR1 to switch both upstream and downstream traffic from the primary path to backup paths.

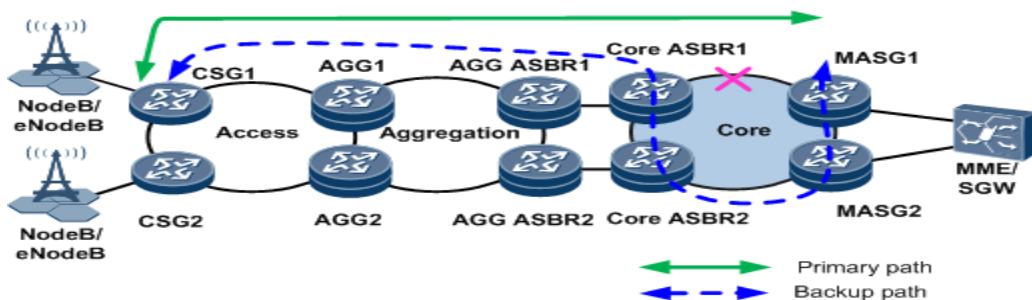
**Figure 12-26 Traffic protection triggered by a fault on a core ASBR on the inter-AS seamless MPLS network**



- **A link fault occurs in the core area.**

On the inter-AS seamless MPLS network shown in [Figure 12-27](#), BFD for LDP or BFD for TE is configured on core ASBR1. If the BFD module detects a fault on the link between core ASBR1 and MASG1, it triggers the LDP FRR, TE hot-standby, or BGP FRR function. The reliability function switches both upstream and downstream traffic from the primary path to the backup path.

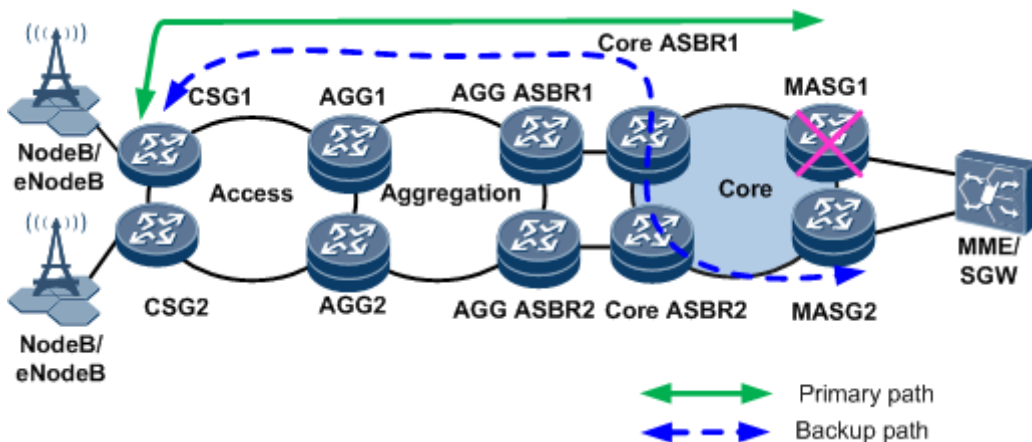
**Figure 12-27 Traffic protection from a link fault in a core area on the inter-AS seamless MPLS network**



- **A fault occurs on an MASG.**

On the inter-AS seamless MPLS network shown in [Figure 12-28](#), BFD for BGP tunnel is configured on CSG1. BFD for BGP tunnel is implemented in compliance with relevant standards "Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)." BFD for BGP tunnel monitors end-to-end BGP LSPs, including a BGP LSP connected to an LDP LSP. If MASG1 functioning as a PE fails, BFD for BGP LSP can rapidly detect the fault and trigger VPN FRR switching so that both upstream and downstream traffic are switched from the primary path to the backup path.

**Figure 12-28 Traffic protection triggered by a fault on an MASG on the inter-AS seamless MPLS network**

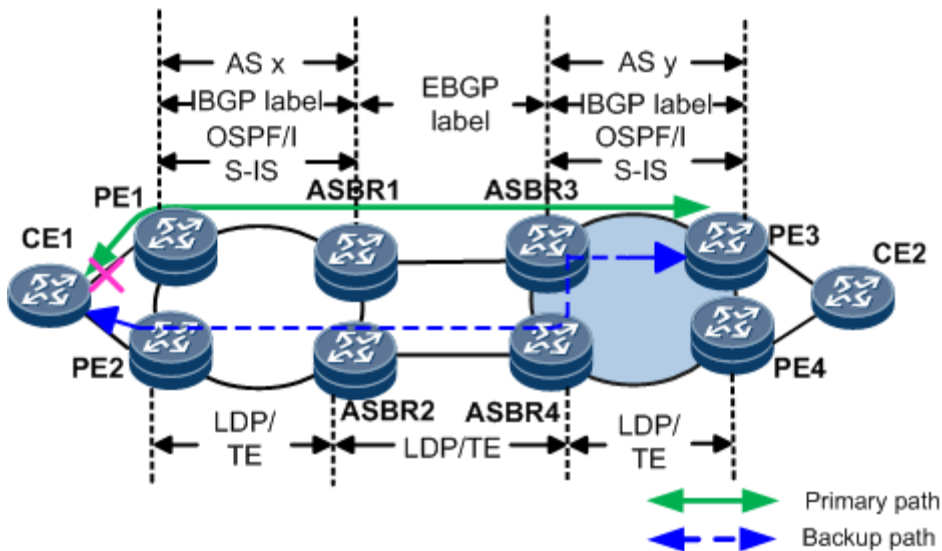


- **An access-side link fails.**

On the inter-AS seamless MPLS network shown in [Figure 12-29](#), if an E-Trunk in Single-Active redundancy mode detects the link failure, and the E-Trunk switches traffic from the primary path to the backup path and disables interface blocking on the link between CE1 and PE2. Then upstream traffic on CE1 is forwarded to PE2. For BUM traffic on the network side, PE1 sends a Per ES-AD-withdraw message to PE2, and PE2 is elected as the primary DF to forward BUM traffic. For unicast traffic, PE3 receives a MAC route advertised by PE2 and forwards the traffic to PE2.

If an E-Trunk in Active-Active redundancy mode detects the link failure, PE1 sends a Per ES-AD route-withdraw message to PE3, and PE3 forwards unicast traffic to PE2.

**Figure 12-29 Traffic protection triggered by an access-side link fault in a core area on the inter-AS seamless MPLS network**



- **A PE on the access side fails.**

If PE1 fails, the reliability implementation is similar to that in the access-side link failure scenario. The other PEs detect PE1 failure and switch traffic from the primary path to backup paths without withdrawing routes.

## EVPN's Service Modes:

Multiple Ethernet VPN instances (EVI) can be configured on PEs at the edge of an EVPN. Each EVI connects to one or more user networks. EVIs access user networks in various service modes:

- Port-based
- VLAN-based
- VLAN bundle
- VLAN-aware bundle

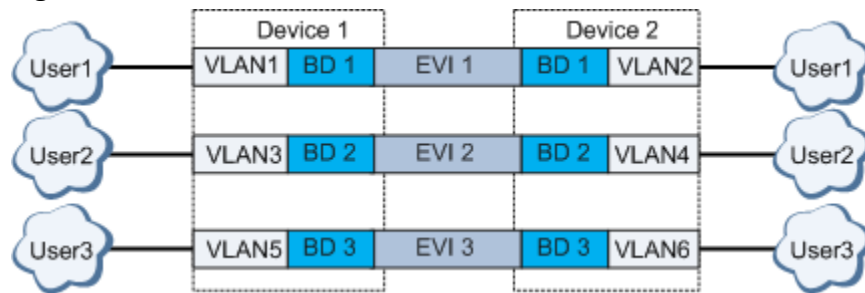
### ***Port-based Mode:***

In port-based mode, an interface is used to access a user service. Specifically, the physical interface connected to a user network is directly bound to a common EVI (neither an EVI in BD mode nor an EVI in VPWS mode) and has no sub-interfaces created. This service mode is used only to carry Layer 2 services.

### ***VLAN-based Mode:***

On the network shown in [Figure 12-30](#), in VLAN-based mode, the physical interfaces connected to user networks each have different sub-interfaces created. Each sub-interface is associated with a unique VLAN and added to a specific bridge domain (BD). Each BD is bound to a specific EVI. In this service mode, the sub-interface, VLAN, BD, and EVI are exclusive for a user to access a network, and a separate MAC forwarding table is used on the forwarding plane for each user. Therefore, this mode effectively ensures service isolation. However, an EVI is required per user, consuming numerous EVI resources. This service mode is used to carry Layer 2 or Layer 3 services.

**Figure** 12-30 VLAN-based mode



### ***VLAN Bundle:***

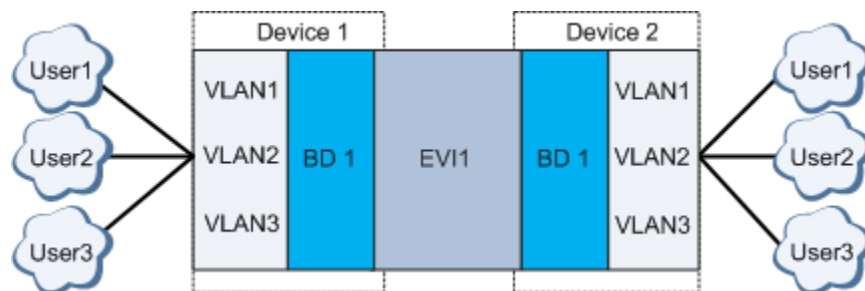
On the network shown in [Figure 12-31](#), in VLAN bundle mode, an EVI connects to multiple users, who are divided by VLAN, and the EVI is bound to a BD. In this service mode, the users connected to the same EVI share a MAC forwarding table, requiring each user on the network to have a unique MAC address. This service mode is used to carry Layer 2 or Layer 3 services.



#### **NOTE:**

In a VLAN bundle scenario, only termination EVC sub-interfaces support both Layer 2 and Layer 3 interfaces. Non-termination EVC sub-interfaces support only Layer 2 services.

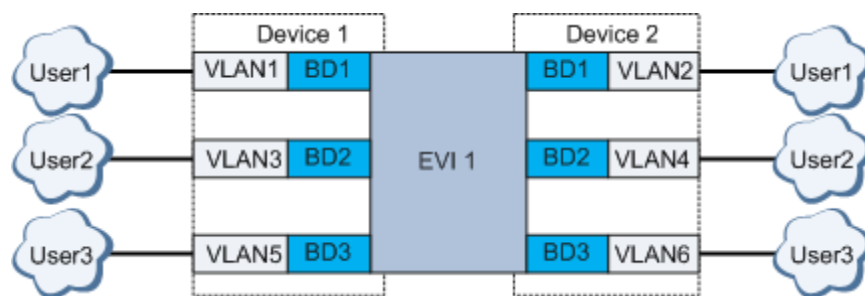
**Figure 12-31 VLAN-bundle mode**



### ***VLAN-Aware Bundle:***

On the network shown in [Figure 12-32](#), in VLAN-aware bundle mode, an EVI connects to multiple users, who are divided by VLAN. Additionally, the EVI can be bound to multiple BDs, in which case, the EVI must have different BD tags configured. When EVPN peers send routes to each other, a BD tag is encapsulated into the Ethernet Tag ID field of an Ethernet auto-discovery route, MAC/IP advertisement route, and inclusive multicast route. In this service mode, users connected to the same EVI use separate forwarding entries. During traffic forwarding, the system uses the BD tag carried in user packets to locate the corresponding MAC forwarding table and searches the table for a forwarding entry based on a MAC address.

**Figure 12-32 VLAN-aware bundle mode**



Unlike the other service mode, in VLAN-aware bundle mode, load balancing, designated forwarder (DF), host migration, and route re-origination are implemented based on a BD:

- Load balancing: In VLAN-aware bundle mode, load balancing can be implemented only if a MAC/IP advertisement route and Ethernet auto-discovery route have the same Ethernet segment identifier (ESI) and the same BD tag. If the BD tags are inconsistent, the routes belong to different BDs, preventing load balancing from being implemented.
- **DF election:**
  - For interface-based DF election, the system chooses the first interface to go Up in a BD for DF election.
  - During DF election after an AC interface is enabled to influence DF election, a PE cannot participate in DF election if the system does not receive the Ethernet auto-discovery route advertised by the PE. If the VLAN-aware bundle mode is enabled in this scenario, an Ethernet auto-discovery route is generated for each BD tag. The PE can participate in DF election only if the system receives Ethernet auto-discovery routes in all BDs bound to a specified EVI.
- Host migration: When the system generates a local MAC/IP advertisement route, the system checks whether it has received a MAC/IP advertisement route from the remote end. If the system has received such a route, the MAC address transfer attribute is added to the locally generated route, or the value of the sequence field in the MAC address transfer attribute is

incremented by 1. In VLAN-aware bundle mode, a BD tag is the prefix key of a MAC/IP advertisement route. The system compares the BD tags carried in the received MAC/IP advertisement route and the locally generated one, preventing MAC address conflict between different BDs from causing a host migration failure.

- Route re-origination: In the Data Center Interconnect (DCI) solution, a DCI-PE re-originate a MAC/IP advertisement route received from a peer device and then sends the new route to the peer device. When the VLAN-aware bundle mode is enabled on the DCI-PE, a MAC/IP advertisement route can be re-originated only if its Ethernet tag ID is consistent with the BD tag.

## **EVPN – VXLAN:**

### **EVPN VXLAN Fundamentals:**

#### ***Introduction:***

Ethernet virtual private network (EVPN) is a VPN technology used for Layer 2 internetworking. EVPN is similar to BGP/MPLS IP VPN. EVPN defines a new type of BGP network layer reachability information (NLRI), called the EVPN NLRI. The EVPN NLRI defines new BGP EVPN routes to implement MAC address learning and advertisement between Layer 2 networks at different sites.

VXLAN does not provide the control plane, and VTEP discovery and MAC addresses learning are implemented by traffic flooding on the data plane, resulting in high traffic volumes on DC networks. To address this problem, VXLAN uses EVPN as the control plane. EVPN allows VTEPs to exchange BGP EVPN routes to implement automatic VTEP discovery and host information advertisement, preventing unnecessary traffic flooding.

EVPN uses extended BGP and defines new BGP EVPN routes to transmit VTEP addresses and host information. As such, the application of EVPN on VXLANs moves VTEP discovery and host information learning from the data plane to the control plane.

### **BGP EVPN Routes:**

EVPN NLRI defines the following BGP EVPN route types applicable to the VXLAN control plane:

#### **Type 2 route—MAC/IP route:**

Figure 12-33 shows the format of MAC/IP routes.

**Figure 12-33 MAC/IP route**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
Ethernet Tag ID (4 bytes)
MAC Address Length (1 byte)
MAC Address (6 bytes)
IP Address Length (1 byte)
IP Address (0, 4, or 16 bytes)
MPLS Label1 (3 bytes)
MPLS Label2 (0 or 3 bytes)

The Table 12-3 describes the fields.

**Table 12-3 The fields of MAC/IP route**

Field	Description
Route Distinguisher	RD value of an EVPN instance
Ethernet Segment Identifier	Unique ID for defining the connection between local and remote devices
Ethernet Tag ID	VLAN ID configured on the device
MAC Address Length	Length of the host MAC address carried in the route
MAC Address	Host MAC address carried in the route
IP Address Length	Mask length of the host IP address carried in the route
IP Address	Host IP address carried in the route
MPLS Label1	Layer 2 VNI carried in the route



Field	Description
MPLS Label2	Layer 3 VNI carried in the route

MAC/IP routes function as follows on the VXLAN control plane:

- MAC address advertisement

To implement Layer 2 communication between intra-subnet hosts, the source and remote VTEPs must learn the MAC addresses of the hosts. The VTEPs function as BGP EVPN peers to exchange MAC/IP routes so that they can obtain the host MAC addresses. The **MAC Address Length** and **MAC Address** fields identify the MAC address of a host.

- ARP advertisement

A MAC/IP route can carry both the MAC and IP addresses of a host, and therefore can be used to advertise ARP entries between VTEPs. The **MAC Address** and **MAC Address Length** fields identify the MAC address of the host, whereas the **IP Address** and **IP Address Length** fields identify the IP address of the host. This type of MAC/IP route is called the ARP route.

- IP route advertisement

In distributed VXLAN gateway scenarios, to implement Layer 3 communication between inter-subnet hosts, the source and remote VTEPs that function as Layer 3 gateways must learn the host IP routes. The VTEPs function as BGP EVPN peers to exchange MAC/IP routes so that they can obtain the host IP routes. The **IP Address Length** and **IP Address** fields identify the destination address of the IP route. In addition, the MPLS Label2 field must carry the Layer 3 VNI. This type of MAC/IP route is called the integrated routing and bridging (IRB) route.



#### NOTE:

An ARP route carries host MAC and IP addresses and a Layer 2 VNI. An IRB route carries host MAC and IP addresses, a Layer 2 VNI, and a Layer 3 VNI. Therefore, IRB routes carry ARP routes and can be used to advertise IP routes as well as ARP entries.

- Host IPv6 route advertisement

In a distributed gateway scenario, to implement Layer 3 communication between hosts on different subnets, the VTEPs (functioning as Layer 3 gateways) must learn host IPv6 routes from each other. To achieve this, VTEPs as EVPN peers exchange MAC/IP routes to advertise host IPv6 routes to each other. The **IP Address Length** and **IP Address** fields carried in the MAC/IP routes indicate the destination addresses of host IPv6 routes, and the **MPLS Label2** field must carry a Layer 3 VNI. MAC/IP routes in this case are also called IRBv6 routes.



#### NOTE:

An ND route carries the following valid information: host MAC address, host IPv6 address, and Layer 2 VNI. An IRBv6 route carries the following valid information: host MAC address, host IPv6 address, Layer 2 VNI, and Layer 3 VNI. It can be seen that an IRBv6 route includes

information about an ND route and therefore can be used to advertise both a host IPv6 route and host ND entry.

### Type 3 route—inclusive multicast route

An inclusive multicast route comprises a prefix and a PMSI attribute. [Figure 12-34](#) shows the format of inclusive multicast routes.

**Figure 12-34 Format of an inclusive multicast route**

<b>Prefix</b>	
Route Distinguisher (8 bytes)	
Ethernet Tag ID (4 bytes)	
IP Address Length (1 byte)	
Originating Router's IP Address (4 or 16 bytes)	
<b>PMSI attribute</b>	
Flags (1 byte)	
Tunnel Type (1 byte)	
MPLS Label (3 bytes)	
Tunnel Identifier (variable)	

The [Table 12-4](#) describes the fields.

**Table 12-4 The fields of an inclusive multicast route**

Field	Description
Route Distinguisher	RD value of an EVI
Ethernet Tag ID	VLAN ID The value is all 0s in this type of route.
IP Address Length	Mask length of the local VTEP's IP address carried in the route
Originating Router's IP Address	Local VTEP's IP address carried in the route
Flags	Flags indicating whether leaf node information is required for

Field	Description
	the tunnel This field is inapplicable in VXLAN scenarios.
Tunnel Type	Tunnel type carried in the route The value can only be 6, representing Ingress Replication in VXLAN scenarios. It is used for BUM packet forwarding.
MPLS Label	Layer 2 VNI carried in the route
Tunnel Identifier	Tunnel identifier carried in the route This field is the local VTEP's IP address in VXLAN scenarios.

This type of route is used on the VXLAN control plane for automatic VTEP discovery and dynamic VXLAN tunnel establishment. VTEPs that function as BGP EVPN peers transmit Layer 2 VNIs and VTEPs' IP addresses through inclusive multicast routes. The Originating Router's **IP Address** field identifies the local VTEP's IP address; the **MPLS Label** field identifies a Layer 2 VNI. If the remote VTEP's IP address is reachable at Layer 3, a VXLAN tunnel to the remote VTEP is established. If the remote VNI is the same as the local VNI, an ingress replication list is created for subsequent BUM packet forwarding.

#### Type 5 route—IP prefix route

The [Figure 12-35](#) shows the format of IP prefix routes.

**Figure 12-35 IP prefix route**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
Ethernet Tag ID (4 bytes)
IP Prefix Length (bytes)
IP Prefix (4 or 16 bytes)
GW IP Address (4 or 16 bytes)
MPLS Label (3 bytes)

The [Table 12-5](#) describes the fields.

**Table 12-5 The fields of IP prefix route**

Field	Description
Route Distinguisher	RD value of an EVI
Ethernet Segment Identifier	Unique ID for defining the connection between local and remote devices
Ethernet Tag ID	VLAN ID configured on the device
IP Prefix Length	Length of the IP prefix carried in the route
IP Prefix	IP prefix carried in the route
GW IP Address	Default gateway address This field is inapplicable in VXLAN scenarios.
MPLS Label	Layer 3 VNI carried in the route

The **IP Prefix Length** and **IP Prefix** fields in an IP prefix route can identify a host IP address or network segment.

- If the **IP Prefix Length** and **IP Prefix** fields in an IP prefix route identify a host IP address, the route is used for IP route advertisement in distributed VXLAN gateway scenarios, which functions the same as an IRB route on the VXLAN control plane.
- If the **IP Prefix Length** and **IP Prefix** fields in an IP prefix route identify a network segment, the route allows external network access.

## **EVPN – VPWS:**

### **EVPN VPWS Fundamentals:**

#### **Overview:**

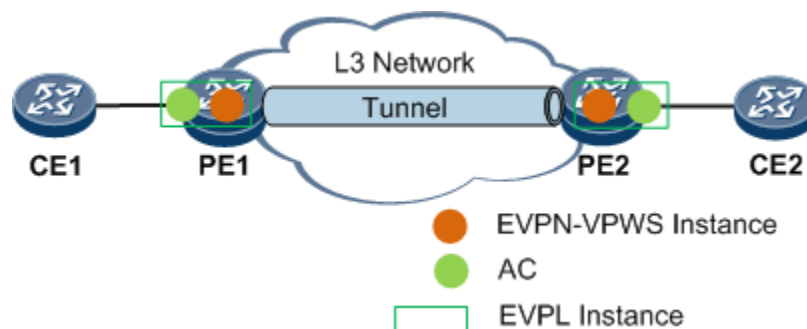
Ethernet virtual private network (EVPN) virtual private wire service (VPWS) provides a P2P L2VPN service solution based on the EVPN service architecture. This solution simplifies the EVPN technology by using MPLS tunnels over a backbone network to provide Layer 2 packet forwarding between access circuits (ACs) with no need of searching for MAC address entries.

As shown in Figure 12-36, the basic architecture of EVPN VPWS consists of the following parts:

- AC: is an independent link or circuit that connects a CE to a PE. An AC interface can be a physical or logical interface. AC attributes include the encapsulation type, maximum transmission unit (MTU), and interface parameters of a specified link type.
- EVPL instance: An EVPLS instance maps to an AC. Each EVPL instance has a service ID. The EVPL instance of the local PE maps to that of the peer PE. PEs exchange EVPN routes carrying a service ID to construct forwarding entries that are used to forward or receive service traffic from different ESs, achieving point-to-point interworking.
- EVPN VPWS instance: An EVPN VPWS instance is deployed on an edge PE and contains services that have the same access-side or network-side attributes. Routes are transmitted based on the RD and RT configured in each EVPN VPWS instance in a BGP EVPN address family.
- Tunnel: indicates a network-side MPLS tunnel or SR tunnel.

Compared with the traditional L2VPN VPWS (PWE3 and CCC/SVC) solution, the EVPN VPWS solution simplifies the control and data models and uses BGP as the control plane where the BGP route selection and the BGP next hop recursion are used to choose traffic paths over backbone networks. This eliminates the need of specifying PWs.

**Figure 12-36 EVPN VPWS networking**



#### **Routes Used by EVPN VPWS:**

On the basis of BGP, EVPN defines a new type of network layer reachability information (NLRI), which is called EVPN NLRI. EVPN VPWS supports the following types of EVPN NLRIs:

- Ethernet Auto-Discovery (Ethernet AD) routes: include Ethernet Auto-Discovery Per EVI routes and Ethernet Auto-Discovery Per ES routes.
  - Ethernet Auto-Discovery Per ES routes: are sent by PEs on an EVPN VPWS network to notify the peer device of whether the local redundancy mode is single-active or all-active.

- Ethernet Auto-Discovery Per EVI routes: are exchanged between PEs on an EVPN VPWS network to guide through Layer 2 traffic forwarding. Figure 12-37 shows the NLRI format of EVI Ethernet AD routes.

**Figure 12-37 NLRI format of EVI Ethernet AD routes**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
Ethernet Tag ID (4 bytes)
MPLS Label (3 bytes)

The description of each field is as follows:

- RD: can be either the RD value of an EVPN instance or a combination of the source IP address configured on a PE and :0, such as X.X.X.X:0.
- Ethernet Segment Identifier: uniquely identifies a connection between a PE and a CE.
- Ethernet Tag ID: indicates the local service ID of the EVPL instance on the local PE.
- MPLS Label: indicates the EVPL label assigned for each EVI Ethernet Auto-Discovery route.

In addition to NLRI, an EVI Ethernet Auto-Discovery route carries the Layer 2 extended community attribute that includes the following control fields:

- C: identifies a control word. If this control field is set to 1, packets sent by the local PE must carry control information.
  - P: is used to identify whether the local PE is the master PE. In all-active scenarios, this control field must be set to 1.
  - B: is used to identify whether the local PE is the backup PE in dual-homing single-active scenarios.
- Ethernet Segment (ES) route: An ES route carries the RD, Ethernet segment identifier (ESI) value, and source IP address of the local PE to allow automatic discovery and DF election between PEs connecting to the same CE. Figure 12-38 shows the NLRI format of ES routes.

**Figure 12-38 NLRI format of ES routes**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
IP Address Length (1 byte)
Originating Router's IP Address (4 or 16 bytes)

The description of each field is as follows:

- Route Distinguisher: is a combination of the source IP address on the local PE and :0, such as X.X.X.X:0.
- Ethernet Segment Identifier: uniquely identifies a connection between a PE and a CE.
- IP Address Length: indicates the length of the source IP address configured on the local PE.
- Originating Router's IP Address: indicates the source IP address configured on the local PE.

***Packet Exchange Process in Single-Homing Scenarios:***

Figure 12-36 shows the packet exchange process in EVPN VPWS single-homing scenarios.

1. PE1 and PE2 are each configured with an EVPL instance and an EVPN VPWS instance. The EVPL instance must be bound to an AC interface and an EVPN VPWS instance, and each EVPL instance must be configured with a local service ID and a remote service ID. After the configuration, the local PE generates the forwarding entries indicating the association between the AC interface and EVPL instance.
2. PE1 and PE2 each send EVI Ethernet AD routes to the peer device. The EVI Ethernet AD routes carry the RD, RT, next-hop information, local service ID, and EVPL label.
3. PE1 and PE2 each receive EVI Ethernet AD routes from the peer device and match the RTs of the corresponding EVPN VPWS instance. PE1 and PE1 then select an MPLS or SRv4 tunnel to perform traffic recursion based on the next-hop information. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, the forwarding entries indicating the association between the MPLS or SR tunnel and local EVPL instance are generated.

***Packet Exchange Process in Dual-Homing Single-Active Scenarios (with an E-Trunk Deployed):***

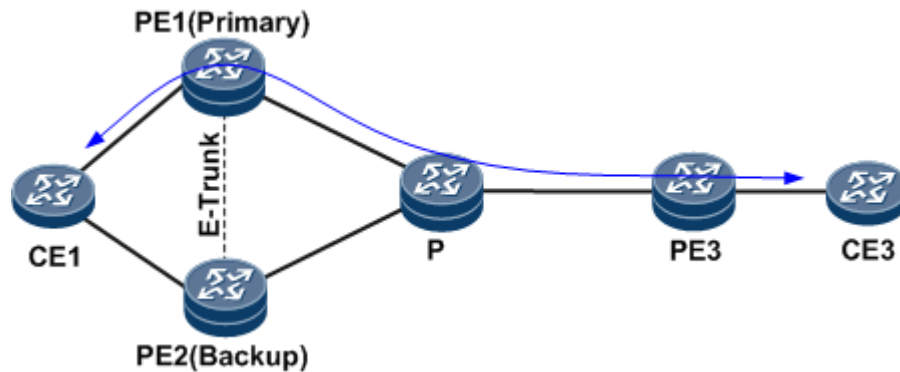
On the CE dual-homing network shown in Figure 12-39, PE1 and PE2 work in single-active mode and an E-Trunk is deployed between them. In this case, DF election is not triggered, and the

master/backup relationship between PE1 and PE2 is determined by the E-Trunk configured between PE1 and PE2. The packet exchange process in this scenario is as follows:

1. Each PE is configured with an EVPL instance and an EVPN VPWS instance. The EVPL instance must be bound to an AC interface and an EVPN VPWS instance, and each EVPL instance must be configured with a local service ID and a remote service ID. After the configuration, the local PE generates the forwarding entries indicating the association between the AC interface and EVPL instance. The access-side interfaces on PE1 and PE2 must be configured with the same ESI.
2. PE1 and PE2 send ES routes that carry the RD, RT, ESI, and source IP address. DF election is not triggered between PE1 and PE2 upon receipt of ES routes. The master/backup relationship between PE1 and PE2 are determined based on the E-Trunk deployed between them. In this example, PE1 is the master PE, and PE2 is the backup PE.
3. PE1 and PE2 send PE3 the ES Ethernet AD routes that carry the RD, RT, next-hop information, and single-active mode information.
4. The PEs send each other the EVI Ethernet AD routes that carry the RD, RT, next-hop information, local service ID, EVPL label, and master/backup role.
5. Upon receipt of EVI Ethernet AD routes from PE3, PE1 and PE2 match RTs of the corresponding EVPN VPWS instance and select an MPLS or SRv4 tunnel to perform traffic recursion based on the next-hop information. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, the forwarding entries indicating the association between the MPLS or SR tunnel and local EVPL instance are generated.
6. Upon receipt of EVI Ethernet AD routes from PE1 and PE2, PE3 matches RTs of the corresponding EVPN VPWS instance and select an MPLS or SRv4 tunnel to perform traffic recursion based on the next-hop information. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, the FRR entries indicating the association between the MPLS or SR tunnel and local EVPL instance are generated. The entries destined for PE1 are the master entries, and the entries destined for PE2 are the backup entries.
7. PE1 and PE2 each receive EVI Ethernet AD routes from the peer device and match the RTs of the corresponding EVPN VPWS instance. PE1 and PE1 then select an MPLS or SRv4 tunnel to perform traffic recursion based on the next-hop information. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, the bypass entries indicating the association between the MPLS or SR tunnel and local EVPL instance are generated.



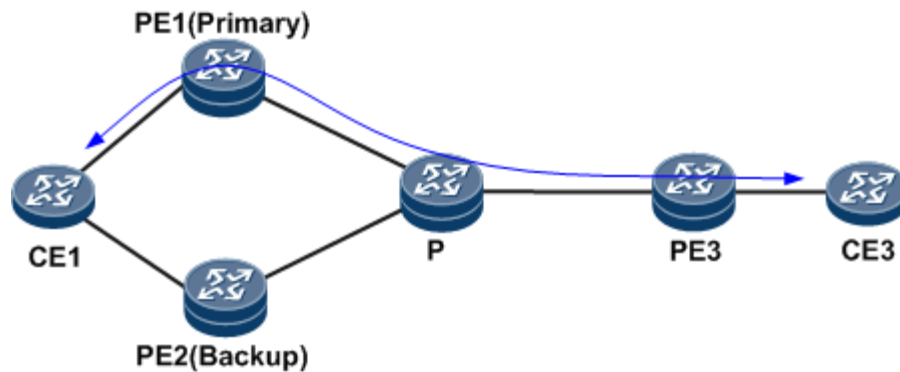
**Figure 12-39 EVPN VPWS dual-homing single-active networking (with an E-Trunk deployed)**



***Packet Exchange Process in Dual-Homing Single-Active Scenarios (with No E-Trunk Deployed):***

On the CE dual-homing network shown in Figure 12-40, PE1 and PE2 work in single-active mode and an E-Trunk is not deployed between them. The packet exchange process in this scenario is similar to that in the scenario with an E-Trunk deployed. The only difference is that the master/backup relationship between PE1 and PE2 in this scenario is determined by the DF election mechanism of EVPN VPWS.

**Figure 12-40 EVPN VPWS dual-homing single-active networking (with no E-Trunk deployed)**



By default, the PE with a smaller source IP address is elected as the master DF. This, however, causes all service traffic to travel through the same PE, which may lead to unbalanced network load. To address this problem, enable the service ID-based DF election. Taking Figure 12-41 as an example, the service ID-based DF election process is as follows:

1. PE1 and PE2 send each other the ES routes that carry the RD, RT, ESI, and source IP address.
2. Upon receipt of ES routes, PE1 and PE2 construct PE lists based on different ESIs. The PEs in a PE list are ordered by the source IP address in ascending order, and the system assigns an index starting at 0 to each PE in ascending order.

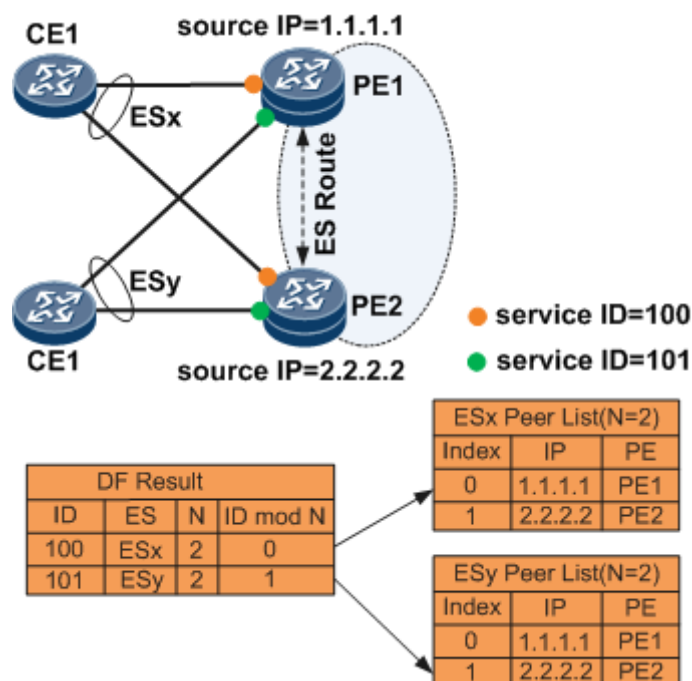
- Each ES corresponds to a local service ID. The system calculates the DF election result of each ES based on the formula "service ID mod N", where N indicates the number of PEs. As shown in Figure 12-41, the service ID corresponding to ESx is 100, the number of PEs (N) is 2, and the ESx is calculated out to be 0. The system searches the PE list of ESx for the index and finds that the DF election result of ESx is PE1. Similarly, the DF election result of ESy is PE2. This allows traffic from different Es to be transmitted over different PEs.



#### NOTE:

The DF election result determines the P control field in EVI Ethernet AD routes.

**Figure 12-41 DF election in EVPN VPWS**



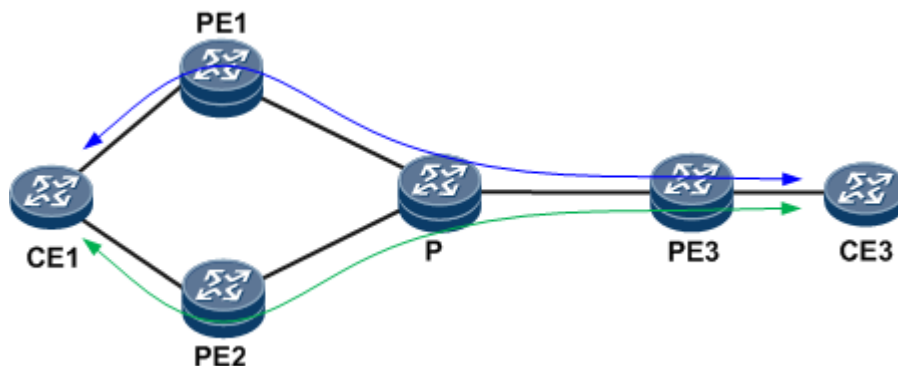
#### **Packet Exchange Process in Dual-Homing All-Active Scenarios:**

On the CE dual-homing network shown in Figure 12-42, PE1 and PE2 work in all-active mode. The packet exchange process in this scenario is as follows:

- Each PE is configured with an EVPL instance and an EVPN VPWS instance. The EVPL instance must be bound to an AC interface and an EVPN VPWS instance, and each EVPL instance must be configured with a local service ID and a remote service ID. After the configuration, the local PE generates the forwarding entries indicating the association between the AC interface and EVPL instance. PE1 and PE2 are configured to work in all-active mode and the access-side interfaces of PE1 and PEs are configured with the same ESI.

2. PE1 and PE2 send ES routes that carry the RD, RT, ESI, and source IP address. Upon receipt of ES routes, PE1 and PE2 do not trigger DF election. Both PE1 and PE2 are in the master DF state.
3. PE1 and PE2 send PE3 the ES Ethernet AD routes that carry the RD, RT, next-hop information, and all-active mode information.
4. The PEs send each other the EVI Ethernet AD routes that carry the RD, RT, next-hop information, local service ID, EVPL label, and master/backup role.
5. Upon receipt of EVI Ethernet AD routes from PE3, PE1 and PE2 match RTs of the corresponding EVPN VPWS instance and select an MPLS or SRv4 tunnel to perform traffic recursion based on the next-hop information. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, the forwarding entries indicating the association between the MPLS or SR tunnel and local EVPL instance are generated.
6. Upon receipt of EVI Ethernet AD routes from PE1 and PE2, PE3 matches RTs of the corresponding EVPN VPWS instance and select an MPLS or SRv4 tunnel to perform traffic recursion based on the next-hop information. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, load balancing entries of the MPLS or SR tunnel and local EVPL instance are generated.
7. PE1 and PE2 each receive EVI Ethernet AD routes from the peer device and match the RTs of the corresponding EVPN VPWS instance. PE1 and PE1 then select an MPLS or SRv4 tunnel to perform traffic recursion based on the next-hop information. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, the bypass entries indicating the association between the MPLS or SR tunnel and local EVPL instance are generated.

**Figure 12-42 EVPN VPWS networking in dual-homing all-active scenarios**



The data packets sent from AC-side interfaces are forwarded to the peer PE over the corresponding MPLS tunnel based on the forwarding entries indicating the association between tunnels and EVPL instances. Upon receipt of packets, the peer PE searches for the association

entries based on the label encapsulated in the packets and the forwards the packets to the corresponding AC interface based on the association entries.

## **PBB – EVPN:**

### **PBB-EVPN Fundamentals:**

#### ***PBB-EVPN Networking:***

PBB-EVPN is an L2VPN technology implemented based on MPLS and Ethernet technologies. PBB-EVPN uses BGP to exchange MAC address information between PEs on the control plane and controls the exchange of data packets among different sites across the MPLS network.

As shown in Figure 12-43, a PBB-EVPN has similar architecture as an EVPN. Compared with EVPN, PBB-EVPN introduces the following concepts:

- PBB: a technique defined in IEEE 802.1ah. PBB precedes C-MAC addresses with B-MAC addresses in a packet to completely separate the user network from the carrier network. This implementation enhances network stability and eases the pressure on the capacity of PEs' MAC forwarding tables.
- I-EVPN: accesses the user network by being bound to a PE interface connecting to a CE. After an I-EVPN instance receives a data packet from the user network, the I-EVPN instance encapsulates a PBB header into the packet.
- B-EVPN: accesses the backbone network. A B-EVPN instance manages EVPN routes received from other PEs.
- I-SID: uniquely identifies a broadcast domain. One I-EVPN instance corresponds to one I-SID. If two PEs share the same I-SID, the two PEs belong to the same BUM group.

**Figure 12-43 PBB-EVPN networking**

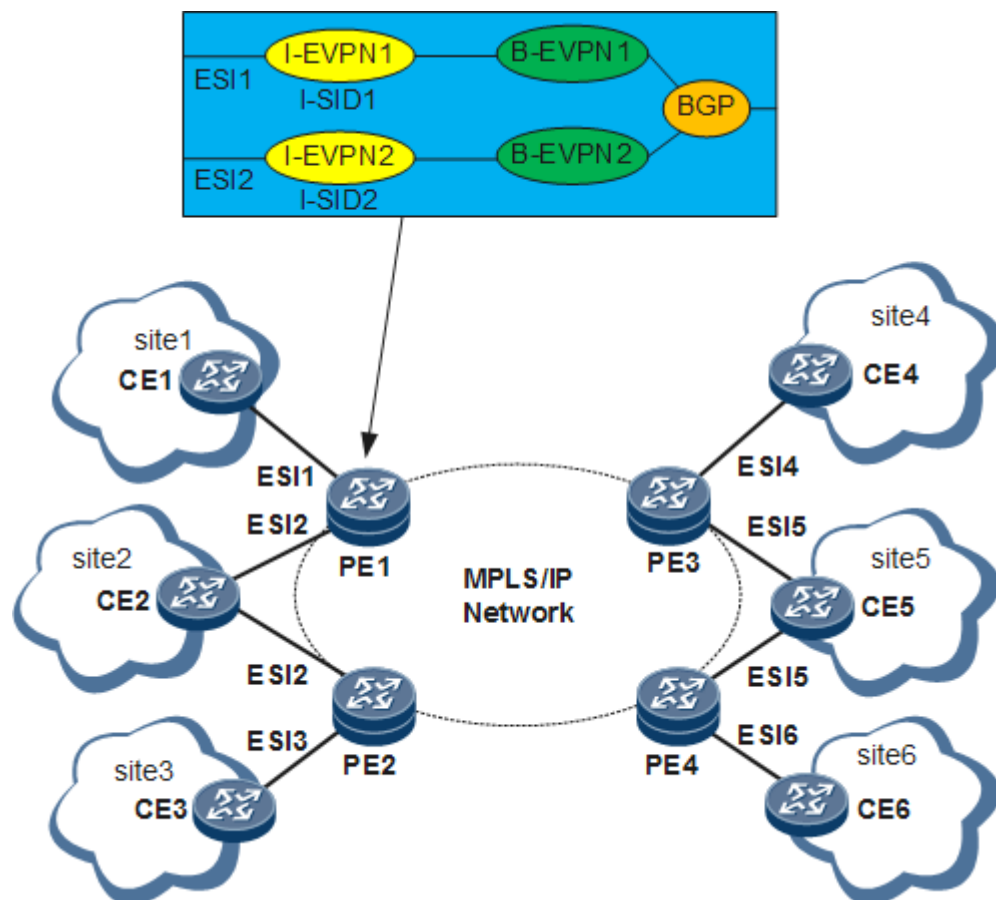


Table 12-6 describes the key points in PBB-EVPN implementation.

**Table 12-6 Key points in PBB-EVPN implementation**

Plane	Key Points in Implementation	Related Concepts
Control plane	PEs use BGP to exchange PBB-EVPN routes and use the B-MAC addresses learned from these routes for later data packet transmission.	Related PBB-EVPN routes: <ul style="list-style-type: none"> <li>MAC advertisement route</li> <li>Inclusive multicast route</li> </ul> Unicast MAC address advertisement BUM packet transmission
	PBB-EVPN supports fast convergence.	Fast convergence

Plane	Key Points in Implementation	Related Concepts
	On a multi-homing network, PBB-EVPN uses DF election to prevent bandwidth waste.	Ethernet segment routeDF election
	PBB-EVPN uses split horizon to prevent routing loops.	Split horizon
	On a multi-homing network, PBB-EVPN supports per-flow load balancing, but does not support per-ISID load balancing.	Redundancy mode
Data plane	PBB-EVPN supports the transmission of unicast and BUM packets.	Unicast packet transmission
		BUM packet transmission

#### **PBB-EVPN Routes:**

On a PBB-EVPN, PEs exchange the following types of routes:

- **MAC advertisement route:** carries B-EVPN instance RD, B-MAC address, and VPN label information on the local PE. Figure 12-44 shows the prefix format of a MAC advertisement route packet. A PE uses MAC advertisement routes to advertise B-MAC address reachability information to other PEs. When network topology changes due to a CE node failure or CE-PE link failure, the corresponding PE sends MAC advertisement routes to instruct other PEs to refresh C-MAC addresses corresponding to the specified B-MAC address, thereby achieving fast convergence.

**Figure 12-44 Prefix format of a MAC advertisement route packet**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
Ethernet Tag ID (4 bytes)
MAC Address Length (1 byte)
MAC Address (6 bytes)
IP Address Length (1 byte)
IP Address (0, 4, or 16 bytes)
MPLS Label1 (3 bytes)
MPLS Label2 (0 or 3 bytes)

The description of each field is as follows:

- Route Distinguisher: a field representing the RD of an EVPN instance.
- Ethernet Segment Identifier: a field of all 0s or Fs. Currently, this field can only be all Fs.
- Ethernet Tag ID: a field of all 0s for MAC advertisement routes.
- MAC Address Length: a field representing the length of the MAC address advertised by the route.
- MAC Address: a field representing the MAC address advertised by the route.
- IP Address Length: a reserved field.
- IP Address: a reserved field.
- MPLS Label1: a field that carries the ESI label.
- MPLS Label2: a reserved field.
- Inclusive multicast route: carries the EVPN instance RD and I-SID information and source IP address (loopback interface address) on the local PE. PEs exchange inclusive multicast routes after establishing an EVPN BGP peer relationship. Figure 12-45 shows the prefix format of an inclusive multicast route packet. PBB-EVPN involves BUM traffic. A PE forwards the BUM traffic that it receives to other PEs in P2MP mode. BUM traffic can be transmitted over MP2P or P2P tunnels established over inclusive multicast routes.

**Figure 12-45 Prefix format of an inclusive multicast route packet**

Route Distinguisher (8 bytes)
Ethernet Tag ID (4 bytes)
IP Address Length (1 byte)
Originating Router's IP Address (4 or 16 bytes)

The description of each field is as follows:

- Route Distinguisher: a field representing the RD of an EVPN instance.
- Ethernet Tag ID: a field representing the I-SID.
- IP Address Length: a field representing the length of the source IP address configured on the local PE.
- Originating Router's IP Address: a field representing the source IP address configured on the local PE.
- Ethernet segment route: carries the EVPN instance RD and ESI information and source IP address on the local PE. PEs connecting to the same CE use Ethernet segment routes to

discover each other. Ethernet segment routes are used in DF election. Figure 12-46 shows the prefix format of an Ethernet segment route packet.

**Figure 12-46 Prefix format of an Ethernet segment route packet**

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
IP Address Length (1 byte)
Originating Router's IP Address (4 or 16 bytes)

The description of each field is as follows:

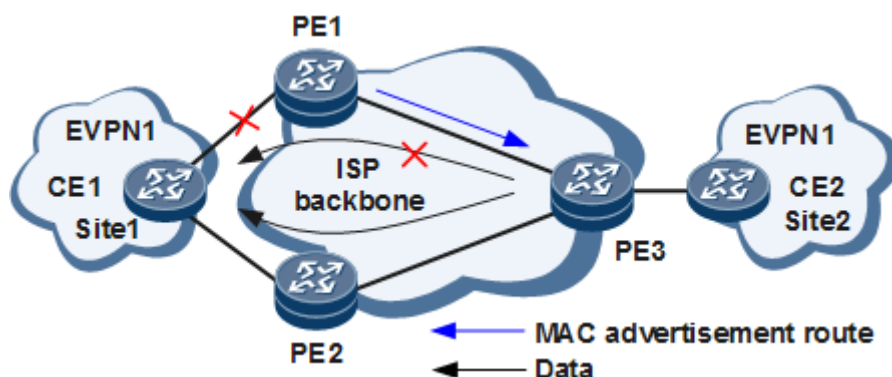
- Route Distinguisher: a field representing a combination of the source IP address on the local PE and :0, such as X.X.X.X:0.
- Ethernet Segment Identifier: a field that uniquely identifies links between PEs and CEs.
- IP Address Length: a field representing the length of the source IP address configured on the local PE.
- Originating Router's IP Address: a field representing the source IP address configured on the local PE.

#### **Other Concepts:**

- **Fast convergence:**

On the network shown in Figure 12-47, if the link between CE1 and PE1 fails, PE1 will send a MAC advertisement route that carries the MAC mobility extended community attribute to PE3, notifying PE3 that C-MAC addresses at Site1 are unreachable. Upon receipt of the route, PE3 sends traffic to Site1 only through PE2, implementing fast convergence.

**Figure 12-47 Fast convergence networking**





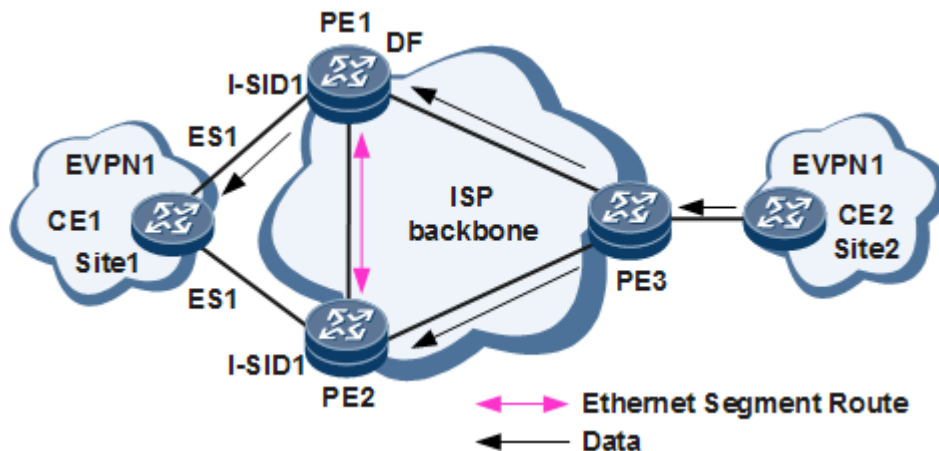
- **DF election:**

On the network shown in Figure 12-48, CE1 is dual-homed to PE1 and PE2, and CE2 sends BUM traffic to PE1 and PE2. In this scenario, CE1 receives the same copy of traffic from both PE1 and PE2, wasting network resources. To solve this problem, EVPN elects one PE as the DF to forward BUM traffic. If PE1 is elected, it becomes the primary DF, with PE2 functioning as the backup DF. The primary DF forwards BUM traffic from CE2 to CE1.

If a PE interface connecting to a CE goes Down, the PE functions as a backup DF. If a PE interface connecting to a CE goes Up, the PE and other PEs with Up interfaces elect a primary DF using the following procedure:

1. The PEs establish EVPN BGP peer relationships with each other and then exchange Ethernet segment routes.
2. Upon receipt of the Ethernet segment routes, each PE generates a multi-homing PE list based on the ESIs carried in Ethernet segment routes. Each multi-homing PE list contains information about all PEs connecting to the same CE.
3. Each PE then sequences the PEs in each multi-homing PE list based on the source IP addresses carried in Ethernet segment routes. The PEs are numbered from 0.
4. The primary DF is elected based on I-SIDs. Specifically, PBB-EVPN uses the formula of "I-SID modulo Number of PEs in the PE list corresponding to the I-SID" to calculate a number and then elects the PE with the same number as the calculated one as the primary DF.

**Figure 12-48 DF election networking**

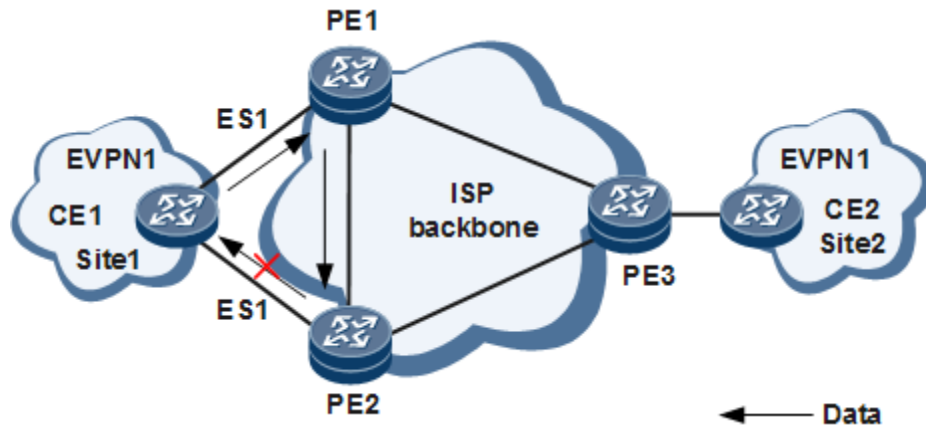


- **Split horizon:**

On the network shown in Figure 12-49, CE1 is dual-homed to PE1 and PE2. If PE1 and PE2 have established an EVPN BGP peer relationship with each other, after PE1 receives BUM traffic from CE1, it forwards the BUM traffic to PE2. If PE2 forwards BUM traffic to CE1, a loop will occur. To prevent this problem, EVPN uses split horizon. After PE1 forwards the BUM traffic to PE2, PE2 checks the B-SMAC address carried in the traffic. If the B-SMAC

address equals the B-MAC address configured on PE2, PE2 drops the traffic, preventing a routing loop.

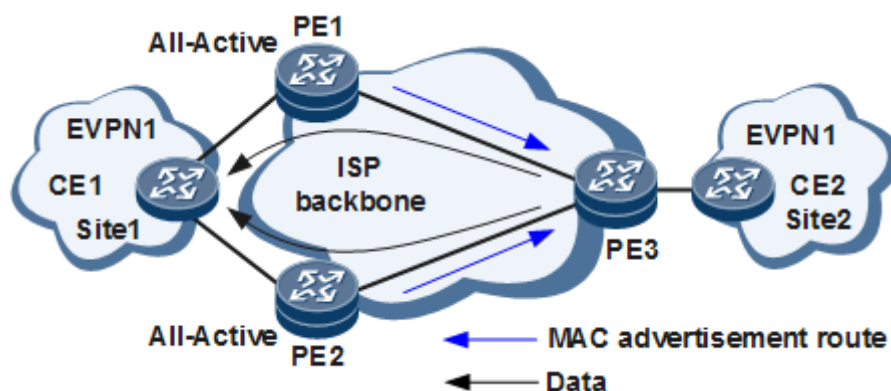
**Figure 12-49 Split horizon networking**



- **Redundancy mode:**

If a CE is multi-homed to several PEs, a redundancy mode can be configured to specify the redundancy mode of PEs connecting to the same CE. The redundancy mode determines whether load balancing is implemented for unicast traffic in CE multi-homing scenarios. On the network shown in Figure 12-50, if PE1 and PE2 are both configured to work in All-Active mode, after PE1 and PE2 send Ethernet segment routes carrying the redundancy mode information to PE3, PE3 sends unicast traffic destined for CE1 to both PE1 and PE2 in load balancing mode.

**Figure 12-50 Redundancy mode networking**

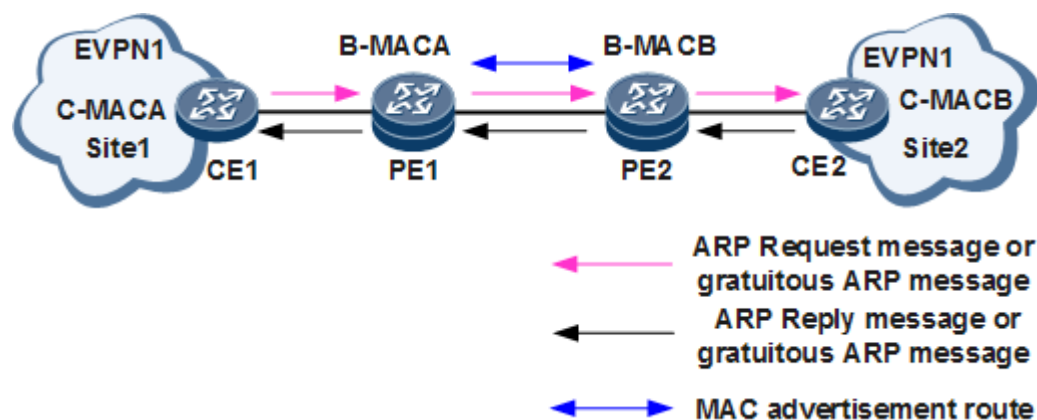


### **Unicast MAC Address Advertisement:**

On the network shown in Figure 12-51, unicast MAC addresses are advertised as follows:

1. Site1 sends an ARP request or gratuitous packet that carries Site1's C-MAC address C-MAC A and the corresponding IP address to Site2.
2. Upon receipt of the packet, Site2 returns an ARP reply or gratuitous packet that carries Site2's C-MAC address C-MAC B and the corresponding IP address to Site1.
3. PE1 and PE2 exchange MAC advertisement routes that carry B-MAC addresses, next hops, and EVPN instance extended community attributes (such as RTs).
4. PE1 and PE2 construct B-EVPN instance forwarding entries based on the RTs carried in received MAC advertisement routes.

**Figure 12-51 Unicast MAC address advertisement networking**



### **BUM Packet Transmission:**

After two PEs establish an EVPN BGP peer relationship, they exchange inclusive multicast routes. PEs then form redundancy groups based on I-SIDs carried in received inclusive multicast routes, with PEs having the same I-SID belonging to the same redundancy group. On the network shown in Figure 12-52, BUM packets are transmitted as follows:

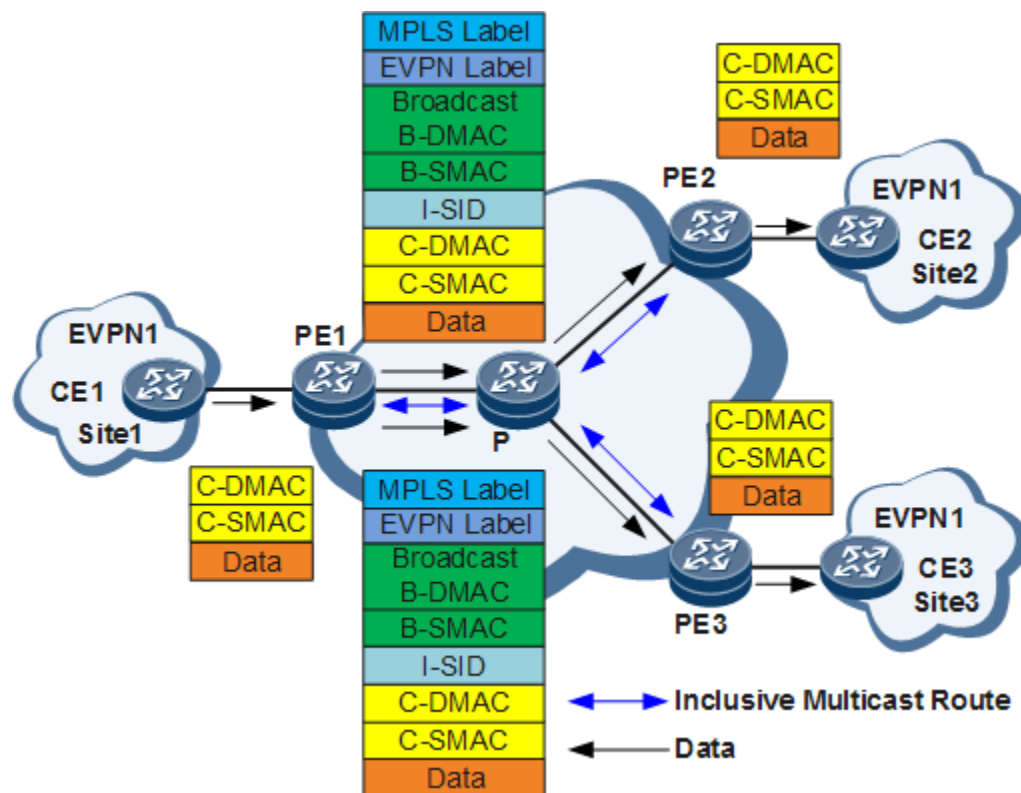
1. CE1 sends BUM packets to PE1.
2. Upon receipt of the packets, PE1 searches its C-MAC address table for the C-DMAC address carried in packets. If the C-DMAC address cannot be found, PE1 sends the BUM packets to all other PEs in the same redundancy group. Specifically, PE1 replicates a copy of received BUM packets, encapsulates the PBB header, public tunnel label, and VPN label into each copy, and sends the two copies of traffic to PE2 and PE3, respectively. The B-DMAC address carried in the PBB header is a broadcast MAC address.

- Upon receipt of the BUM packets, PE2 and PE3 decapsulate the BUM packets and send the BUM packets to the sites identified by the EVPN label carried in the packets.

**NOTE:**

Use the network shown in Figure 12-50 as an example. If PE1 and PE2 both work in Single-Active mode, the bidirectional BUM traffic between CE2 and CE1 will be dropped by the backup DF. If PE1 and PE2 both work in All-Active mode, only the BUM traffic from CE2 to CE1 will be dropped by the backup DF.

**Figure 12-52 BUM packet transmission networking**



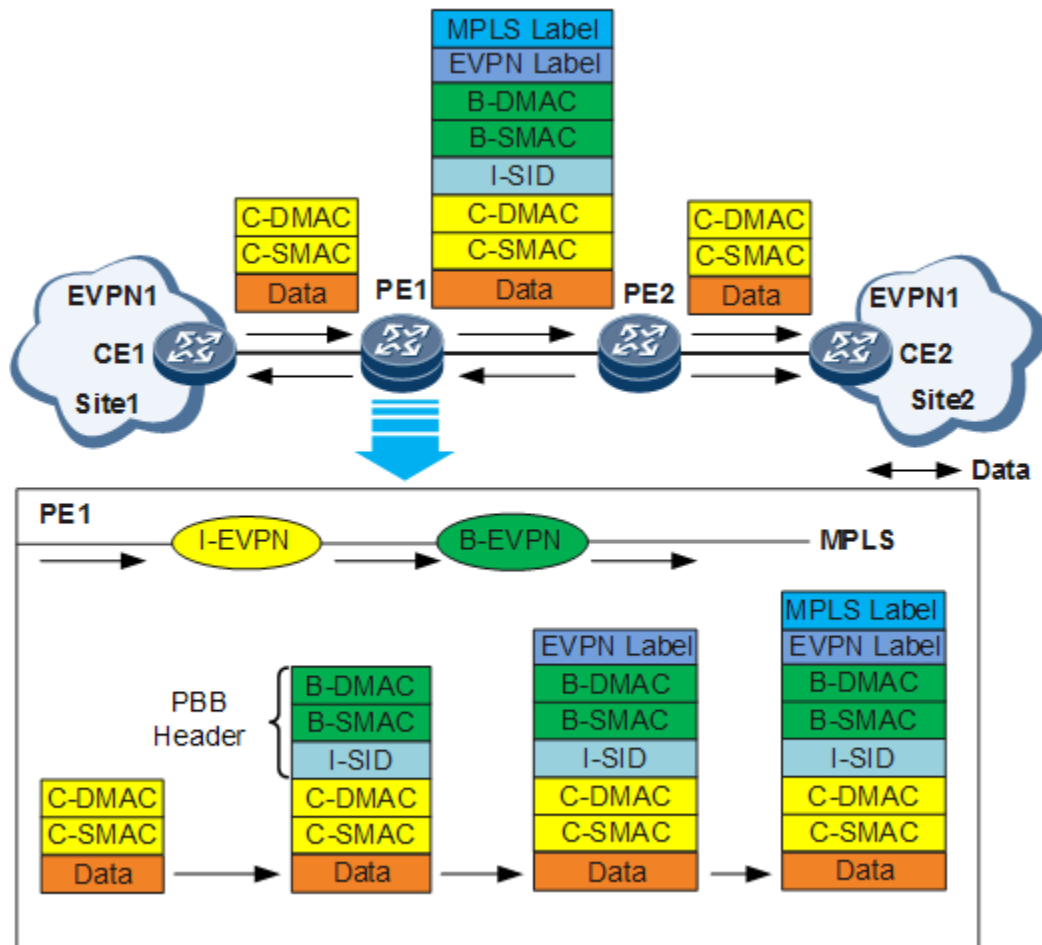
**Unicast packet transmission:**

On the network shown in Figure 12-53, unicast packets are transmitted as follows:

- CE1 forwards unicast packets that carry the source C-MAC (C-SMAC) and destination C-MAC (C-DMAC) addresses to PE1 at Layer 2.
- Upon receipt of the packets, the I-EVPN instance on PE1 searches its C-MAC address table for a matching forwarding entry. After finding such an entry, PE1 encapsulates a PBB header, a tunnel label, and a VPN label into these packets and forwards these packets to PE2. The PBB header carries the I-SID and B-SMAC address configured in the I-EVPN instance and the B-DMAC address obtained from the C-DMAC address table.

- Upon receipt of these packets, PE2 removes the tunnel label and PBB header, searches the local C-MAC address table for a matching forwarding entry, and forwards these packets to an outbound interface.

**Figure 12-53 Unicast packet transmission networking**



## EVPN E-Tree:

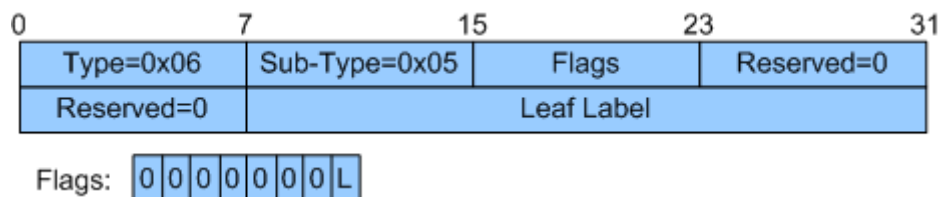
As the number of services carried on an EVPN increases, the number of user MAC addresses managed by the EVPN is also going Up. The user MAC addresses are flooded on the network through EVPN routes. As a result, all interfaces in the same broadcast domain can communicate with each other at Layer 2. However, broadcast, unknown unicast, multicast (BUM), and unicast traffic cannot be isolated for users who do not need to communicate with each other. To isolate interfaces that do not need to communicate with each other in the same broadcast domain, you can deploy the EVPN E-Tree function on the network.

EVPN E-Tree implements the E-Tree model defined by the Metro Ethernet Forum (MEF) by setting the root or leaf attribute for AC interfaces.

- A leaf AC interface and a root AC interface can send traffic to each other. However, flows between leaf AC interfaces are isolated from each other.
- A root AC interface can communicate with other root AC interfaces and with leaf AC interfaces.

To implement the preceding functions, an E-Tree extended community attribute is defined in a standard protocol. [Figure 12-54](#) shows the packet format of this attribute. The packet format includes the Leaf Label field and the Flags field. The Flags field contains eight bits, in which the first seven are all zeros and the last identifies whether an EVPN MAC route is from a leaf AC interface. If the MAC route comes from this interface, the value is set to 1. The extended community attribute can be advertised through Ethernet auto-discovery (A-D) per ES routes and MAC routes on an EVPN, so that known unicast traffic and BUM traffic on leaf AC interfaces are isolated from each other.

**Figure 12-54 Packet format of the extended community attribute used by EVPN E-Tree**



Take the network shown in [Figure 12-55](#) as an example. Known unicast traffic is isolated through the following process:

1. PE1 and PE2 transmit AC-side MAC addresses to each other through MAC routes. Take the MAC address (MAC1) of the AC interface on CE2 as an example. Because the AC interface has the leaf attribute, a MAC route carrying the MAC1 address also carries the extended community attribute of EVPN E-Tree. All bits in the Leaf Label field of the attribute are set to 0, and the L bit in the Flags field is set to 1. PE1 then sends this MAC route to PE2.
2. Upon receipt, PE2 checks the L bit in the Flags field. Because this bit is set to 1, PE2 marks the entry corresponding to MAC1 in the local MAC routing table.
3. When PE2 receives traffic destined for CE2 from its own leaf AC interface, PE2 determines that the traffic needs to be sent to the remote leaf AC interface based on the flag in the local MAC routing table and discards the traffic. In this way, known unicast traffic is isolated between leaf AC interfaces.

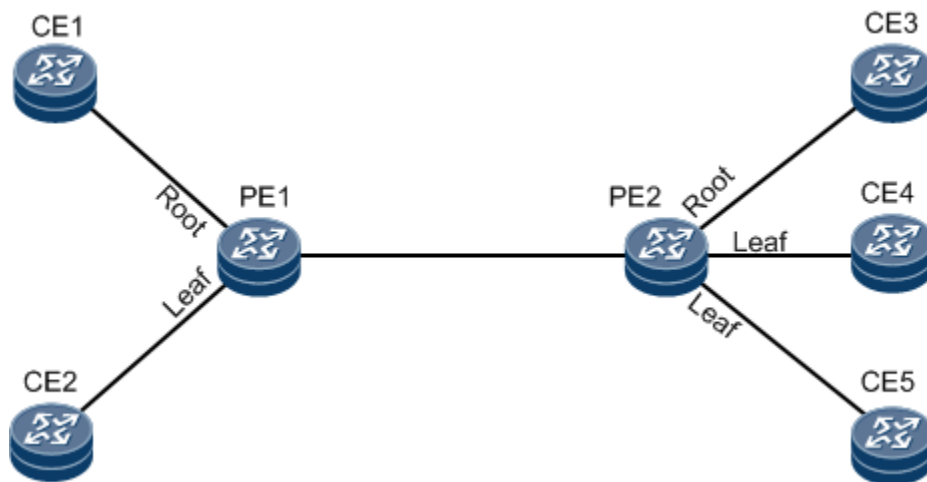
In the preceding example, BUM traffic is isolated through the following process:

1. After EVPN E-Tree is configured on the network, PE1 and PE2 send a special Ethernet A-D per ES route to each other. A regular Ethernet A-D per ES route carries the ESI attribute.

However, the ESI field in the Ethernet A-D per ES route used by EVPN E-Tree is set to all zeros, and the route carries the extended community attribute of EVPN E-Tree. The Leaf Label field of this attribute uses a label value, and the L bit in the Flags field is set to 0.

2. After PE1 receives the Ethernet A-D per ES route, it determines that the route is used to transmit the leaf label because the ESI field value is all zeros. PE1 then saves the label.
3. When PE1 needs to send BUM traffic from its leaf AC interface to PE2, PE1 encapsulates the saved leaf label into the BUM packets and then sends them to PE2.
4. Upon receipt, PE2 finds the locally allocated leaf label in the BUM packets. Therefore, PE2 does not send the traffic to CE4 and CE5. Instead, PE2 only sends the traffic to CE3, implementing BUM traffic isolation between leaf AC interfaces.

**Figure 12-55 Network with EVPN E-Tree deployed**



**NOTE:**

EVPN E-Tree supports the following types of AC interfaces: main interfaces bound to common EVPN instances, EVC Layer 2 sub-interfaces associated with BDs, and VLAN sub-interfaces.

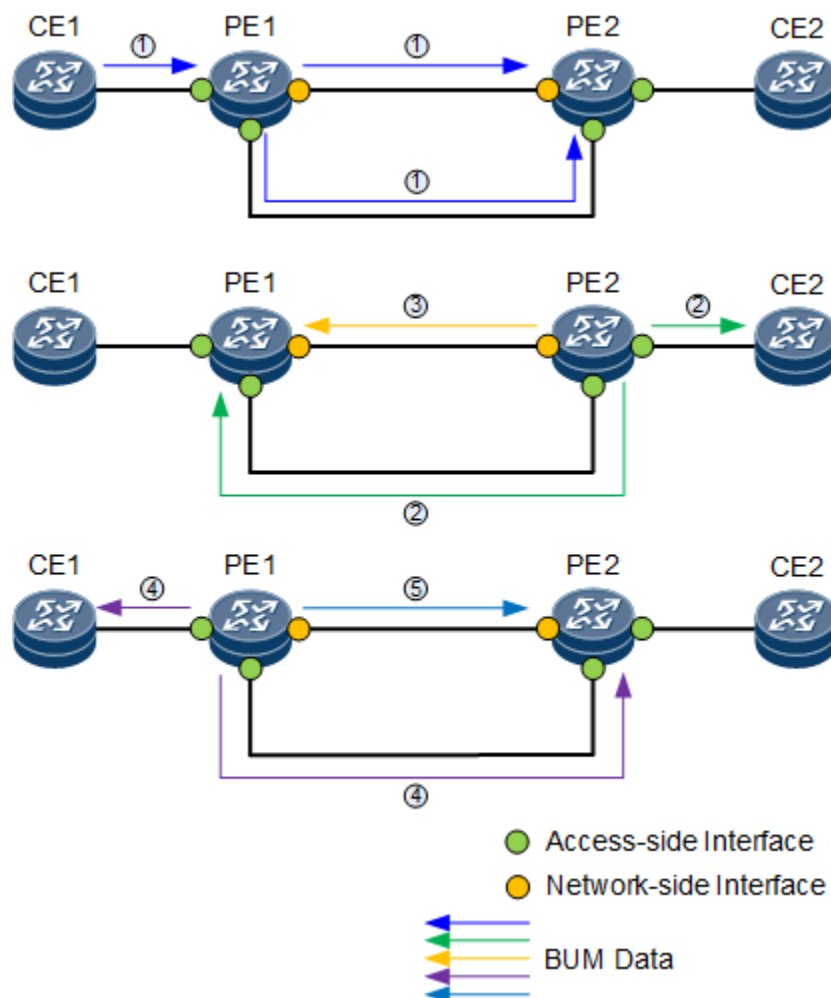
In a CE dual-homing scenario, ensure that the same root or leaf attribute is set for the same VLAN sub-interface in the same broadcast domain on two PEs. If the leaf attribute is set on both PEs, the Leaf label can replace the ESI label to implement split horizon.

Different root or leaf attributes can be set for a PE's interfaces or sub-interfaces that connect to different CEs.

# MAC Duplication Suppression for EVPN:

On an EVPN E-LAN, two PEs may be interconnected both through network-side and access-side links. If this is the case, a BUM traffic loop and MAC route flapping both occur, preventing devices from working properly. MAC duplication suppression for EVPN can resolve this problem.

**Figure 12-56 BUM traffic loop over an EVPN**



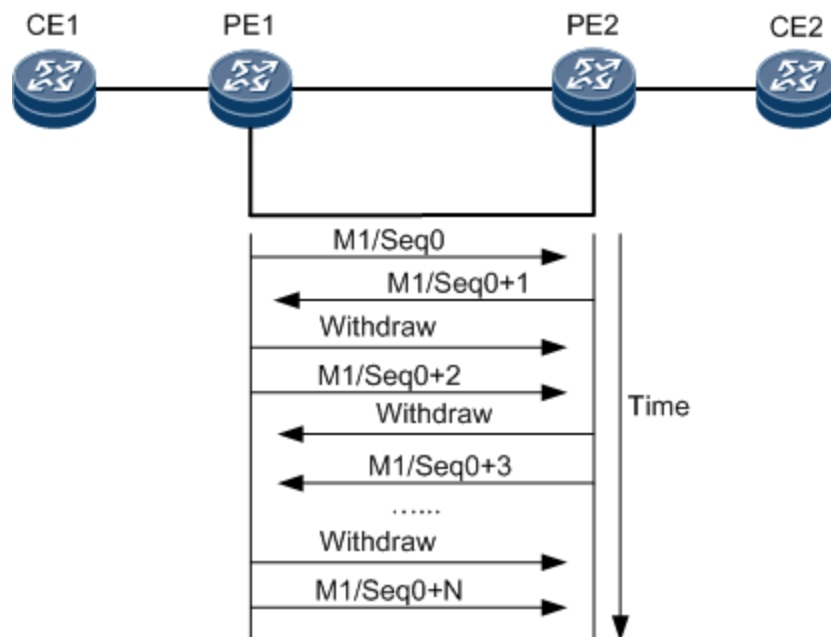
On the network shown in Figure 12-56, EVPN runs between PE1 and PE2. CE1 and CE2 access PE1 and PE2 respectively in one of the following ways: VLAN, QinQ, static or dynamic PW, or static VXLAN. PE1 and PE2 can communicate with each other both through network-side and access-side links, which induces a BUM traffic loop:

1. After PE1 receives BUM traffic from CE1, PE1 first replicates it, and then forwards it to both the network-side and access-side links (traffic 1 in Figure 12-56).



2. PE2 forwards the BUM traffic received from PE1 through the network-side link to the access-side link (traffic 2 in [Figure 12-56](#)). Equally, PE2 forwards the BUM traffic received from the access side to the network side (traffic 3 in [Figure 12-56](#)).
3. PE1 forwards the BUM traffic received from PE2 through the network-side link to the access-side link (traffic 4 in [Figure 12-56](#)). Equally, PE1 forwards the BUM traffic received from the access side to the network side (traffic 5 in [Figure 12-56](#)).
4. As steps 2 and 3 are repeated, BUM traffic is continuously transmitted between PE1 and PE2.

**Figure 12-57 Route flapping over the EVPN**



On the network shown in [Figure 12-57](#), in addition to a BUM traffic loop, route flapping also occurs:

1. After PE1 receives BUM traffic from CE1, PE1 learns CE1's MAC address (M1) from the source MAC address of the traffic. PE1 sends a MAC route with a destination address of M1 to PE2 by means of EVPN.
2. Upon receipt, PE2 matches the RT of the MAC route, imports the MAC route into a matching EVPN instance, generates a MAC entry, and iterates the MAC route to the network-side VXLAN or MPLS tunnel to PE1.
3. PE2 can also receive BUM traffic from PE1 through the direct access-side link between them. Upon receipt, PE2 also generates a MAC route to M1 based on the source MAC address of the traffic. In this case, PE2 considers M1 to have moved to its own access network. PE2 preferentially selects the MAC address received from the local access side.

PE2 therefore sends the MAC route destined for M1 to PE1. This route carries the MAC Mobility extended community attribute. The mobility sequence number is Seq0+1.

4. Upon receipt, PE1 matches the RT of the MAC route, and imports the MAC route into a matching EVPN instance. PE1 preferentially selects the MAC route received from PE2 because this route has a larger mobility sequence number. PE1 then generates a MAC entry and iterates the MAC route to the network-side VXLAN or MPLS tunnel to PE2. PE1 then sends a MAC Withdraw message to PE2.
5. After PE1 receives BUM traffic again from the access-side link, PE1 generates another MAC route to M1 and considers M1 to have moved to its own access network. PE1 preferentially selects the local MAC route to M1 and sends it to PE2. This route carries the MAC Mobility extended community attribute. The mobility sequence number is Seq0+2.
6. Upon receipt, PE2 matches the RT of the MAC route and imports the MAC route into a matching EVPN instance. PE2 preferentially selects the MAC route received from PE1 because this route has a larger mobility sequence number. PE2 then generates a MAC entry and iterates the MAC route to the network-side VXLAN or MPLS tunnel to PE1. PE2 then sends a MAC Withdraw message to PE1.
7. After PE2 receives BUM traffic again from PE1 through the direct access-side link between them, PE2 generates another MAC route to M1 and considers M1 to have moved to its own access network. PE2 preferentially selects the local MAC route and sends the MAC route destined for M1 to PE1. This route carries the MAC Mobility extended community attribute. The mobility sequence number is Seq0+3.
8. As steps 3 to 7 are repeated, the mobility sequence number of the MAC route is incremented by 1 continuously, causing route flapping on the network.

To prevent traffic loops and route flapping, the system starts the process of MAC duplication suppression. The system checks the number of times a MAC entry flaps within a detection period. If the number of MAC flaps exceeds the upper threshold, the system considers MAC route flapping to be occurring on the network and consequently suppresses the flapping MAC routes. The suppressed MAC routes cannot be sent to a remote PE through a BGP EVPN peer relationship.

In addition to suppressing MAC route flapping, you can also configure black-hole MAC routing and AC interface blocking:

- After black-hole MAC routing has been configured, the system sets the suppressed MAC routes to black-hole routes. If a PE receives traffic with the same source or destination MAC address as the MAC address of a black-hole MAC route, the PE discards the traffic.
- If AC interface blocking is also configured, that is, if the traffic comes from a local AC interface and the source MAC address of the traffic is the same as the MAC address of a black-hole MAC route, the AC interface is blocked. In this way, a loop can be removed quickly. Only BD-EVPN instances support AC interface blocking.

# EVPN ORF:

## Background:

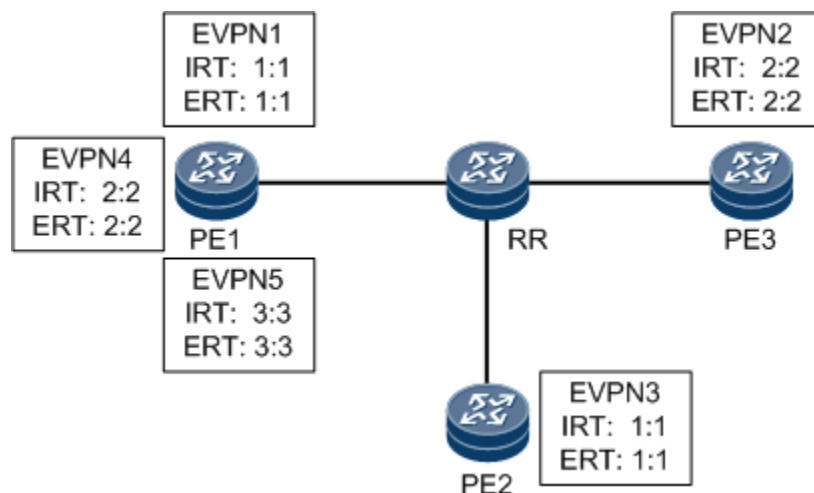
The growing number of services over EVPNs has triggered a proliferation of new users. As a result, BGP-EVPN peers on an EVPN are sending vast quantities of EVPN routes to each other. Even if the remote peer does not have an RT-matching EVPN instance, the local PE still sends it EVPN routes. To reduce network load, each PE needs to receive only desired routes. If a separate export route policy is configured for each user, the cost of O&M goes up. To address this issue, EVPN outbound route filtering (ORF) can be deployed.

## Implementation:

After EVPN ORF is configured, each PE on the EVPN sends the import VPN target (IRT) and original AS number of the local EVPN instance to the other PEs or BGP EVPN RRs that function as BGP-EVPN peers. The information is sent through ORF routes. Upon receipt, the peers construct export route policies based on these routes so that the local PE only receives the expected routes, which reduces the receiving pressure on the local PE.

Figure 12-58 shows the basic EVPN ORF network on which each device supports EVPN ORF. PE1, PE2, and PE3 establish BGP-EVPN peer relationships with the RR, and are also clients of the RR. An EVPN instance with a specific RT is configured on each PE.

**Figure 12-58 Basic usage scenario of EVPN ORF**



Before EVPN ORF is enabled, the RR advertises all the routes received from PE1's EVPN instances to PE2 and PE3. However, PE2 only needs routes with an export VPN target (ERT) of 1:1, whereas PE3 only needs routes with an ERT of 2:2. As a result, PE2 and PE3 discard unwanted routes upon receipt, which wastes device resources.

After EVPN ORF is enabled on all devices and BGP-EVPN peer relationships are established between the PEs and RR in the BGP-VT address family view, the BGP-EVPN peers negotiate the EVPN ORF capability. Each device sends the IRT of its local EVPN instance to the BGP-EVPN peers in the form of ORF routes. Each device then constructs an export route policy based on the received ORF routes. Upon construction, PE1 only sends EVPN1's and EVPN4's routes to the RR. The RR then only sends routes with an ERT of 1:1 to PE2 and those with an ERT of 2:2 to PE3.

The BGP-VT address family obtains the IRT configured on the local device regardless of the type of the instance that the IRT comes from. If EVPN ORF is enabled on a network that consists of devices that do not support EVPN ORF, the EVPN service cannot run properly. But the BGP-VT address family can resolve this problem.

On the network shown in Figure 12-58, PE1, PE2, and PE3 establish BGP-EVPN peer relationships with the RR. PE1, PE2, and PE3 are clients of the RR. Suppose that PE1, PE2, and the RR all support EVPN ORF but that PE3 does not, as it is running an early version. If EVPN ORF is enabled on the network and the BGP-VT peer relationships are established, PE3 does not send ORF routes to the RR, which means that PE1 does not receive the ORF routes with an ERT of 2:2 from the RR. As a result, PE1 does not send EVPN4's routes to the RR, thereby compromising the services between EVPN4 and EVPN2. Because the BGP-VT address family does not differentiate the type of instance the IRT belongs to, you can configure an L3VPN instance on PE3 and set both IRT and ERT to 2:2. This configuration allows PE3 to advertise an ORF route with an IRT of 2:2 to the RR, which then advertises this route to PE1. Upon receipt, PE1 modifies its export route policy so that it can advertise EVPN2's routes to the other PEs.

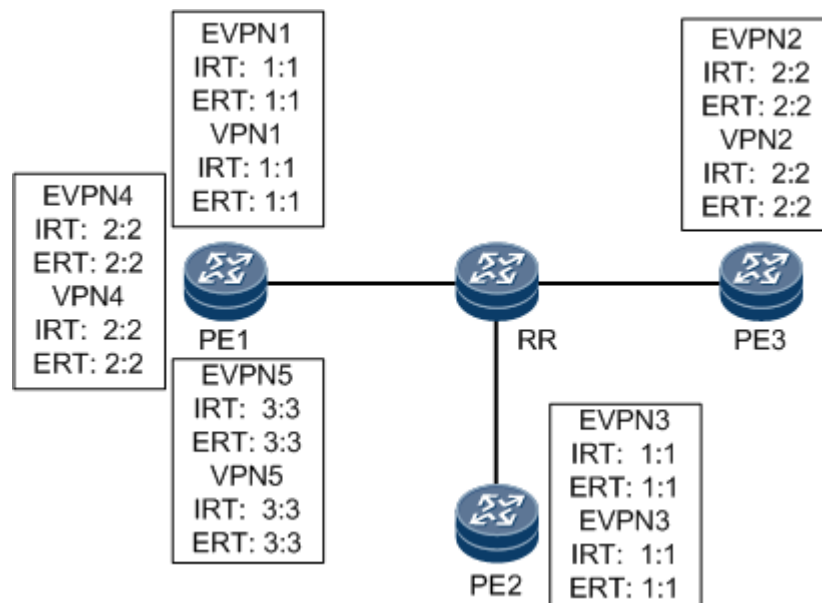


**NOTE:**

In addition to configuring an L3VPN instance, you can also configure the RR to advertise default ORF routes to PE1 and PE3 and delete the BGP-VT peer relationship between the RR and PE3. After the configuration is complete, PE1, PE2, and PE3 advertise all routes to the RR. The RR then advertises routes with ERTs of 1:1 and 2:2 to PE1, routes with an ERT of 1:1 to PE2, and all routes to PE3.

If both EVPN and L3VPN services are deployed on the network in Figure 12-58, the preceding two ways cannot be used. If you use either of them, the L3VPN service cannot run properly. On the network shown in Figure 12-59, only PE3 does not support EVPN ORF. After EVPN ORF is enabled on the network, the EVPN service cannot run properly. If an L3VPN instance is created, the new L3VPN instance receives the other PEs' L3VPN routes from the RR, which compromises the L3VPN service. To resolve this issue, you can disable the RR from filtering routes based on the IRT for PE3, thereby ensuring that both EVPN and L3VPN services can run properly.

**Figure 12-59 An EVPN ORF network carrying both EVPN and L3VPN services**



### Benefits

- Bandwidth consumption is lowered (because the number of routes being advertised is smaller).
- System resources such as CPU and memory are saved.

## IGMP Snooping over EVPN MPLS:

If the Ethernet virtual private network (EVPN) function is deployed on a network to carry multicast services but no Internet Group Management Protocol (IGMP) snooping is configured on PEs, multicast data packets are broadcast on the network. The devices that do not need to receive the multicast data packets also receive these packets, which wastes network bandwidth resources. To resolve this issue, deploy IGMP snooping over EVPN Multiprotocol Label Switching (MPLS). After IGMP snooping over EVPN MPLS is deployed, IGMP snooping packets are transmitted on the network through EVPN routes, and multicast forwarding entries are generated on devices. Multicast data packets from a multicast source are advertised only to the devices that need these packets, saving network bandwidth resources.

For details about EVPN routes used by IGMP snooping over EVPN MPLS, see [Related Routes](#). For details about route advertisement and traffic forwarding, see [Route Advertisement and Traffic Forwarding](#).

## Related Routes:

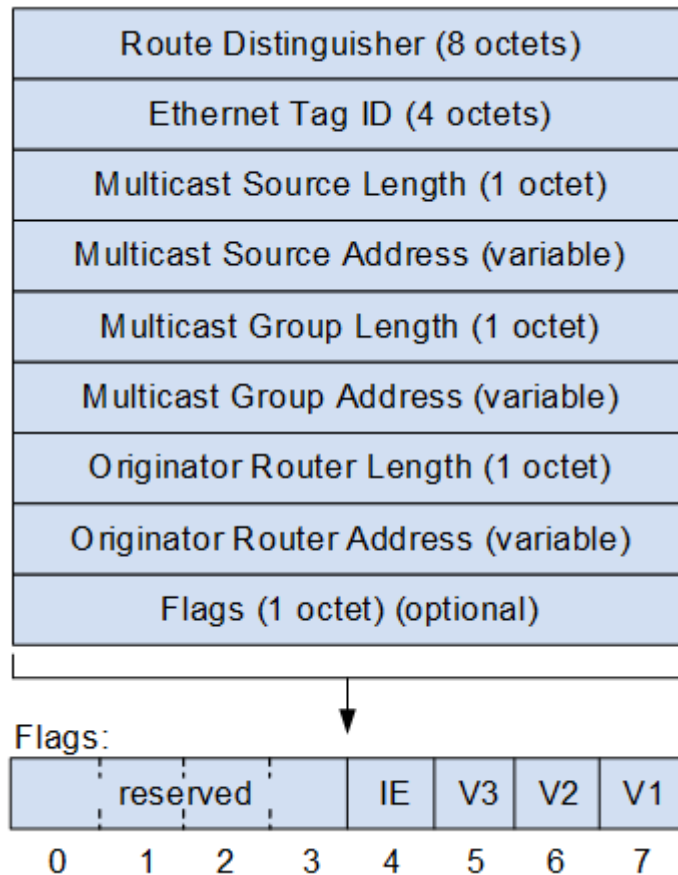
EVPN routes used by IGMP snooping over EVPN MPLS include Selective Multicast Ethernet Tag (SMET), IGMP Join Synch, and IGMP Leave Synch routes.

- SMET route

SMET routes are used to transmit multicast group information between BGP EVPN peers. A device that receives an SMET route can construct local (\*, G) or (S, G) entries based on the routing information. As shown in Figure 12-60, the fields in the routing information are described as follows:

- Route Distinguisher: route distinguisher (RD) set in an EVPN instance.
- Ethernet Tag ID: This field is set to 0 when the VLAN-based or VLAN bundle service mode is used to access a user network.
- Multicast Source Length: length of a multicast source address. This field is set to 0 for any multicast source.
- Multicast Source Address: address of a multicast source. Packets do not contain this field for any multicast source.
- Multicast Group Length: length of a multicast group address.
- Multicast Group Address: address of a multicast group.
- Originator Router Length: address length of the device that generated the SMET route.
- Originator Router Address: address of the device that generated the SMET route.
- Flags: This field contains eight bits. The first four most significant bits are reserved, and the last three least significant bits are used to identify the IGMP version. For example, if bit 5 is set to 1, the IGMP version of the multicast entry carried in the route is IGMPv3. Only one of the last three least significant bits can be set to 1. Bit 4 indicates the filtering mode of group records in IGMPv3. The values 0 and 1 indicate Include and Exclude, respectively.

**Figure 12-60 SMET route format**



- IGMP Join Synch route

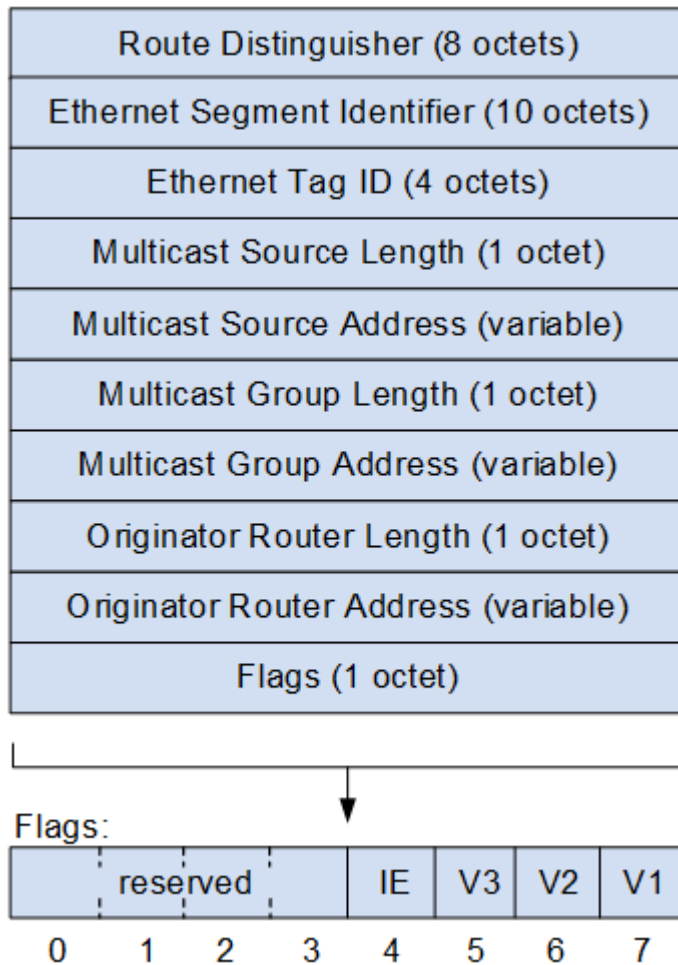
IGMP Join Synch routes are used to synchronize multicast group join information between dual-homed devices on the access side. A device that receives an IGMP Join Synch route can add member entries to the local (S, G) entries based on the routing information, ensuring that the local entries are the same as those on the device connected to the same user network. As shown in [Figure 12-61](#), the fields in the routing information are described as follows:

- Route Distinguisher: route distinguisher (RD) set in an EVPN instance.
- Ethernet Segment Identifier: unique identifier defined for a device to connect to the access network.
- Ethernet Tag ID: This field is set to 0 when the VLAN-based or VLAN bundle service mode is used to access a user network.
- Multicast Source Length: length of a multicast source address. This field is set to 0 for any multicast source.

- Multicast Source Address: address of a multicast source. Packets do not contain this field for any multicast source.
- Multicast Group Length: length of a multicast group address.
- Multicast Group Address: address of a multicast group.
- Originator Router Length: address length of the device that generated the IGMP Join Synch route.
- Originator Router Address: address of the device that generated the IGMP Join Synch route.
- Flags: This field contains eight bits. The first four most significant bits are reserved, and the last three least significant bits are used to identify the IGMP version. For example, if bit 5 is set to 1, the IGMP version of the multicast entry carried in the route is IGMPv3. Only one of the last three least significant bits can be set to 1. Bit 4 indicates the filtering mode of group records in IGMPv3. The values 0 and 1 indicate Include and Exclude, respectively.



**Figure 12-61 IGMP Join Synch route format**



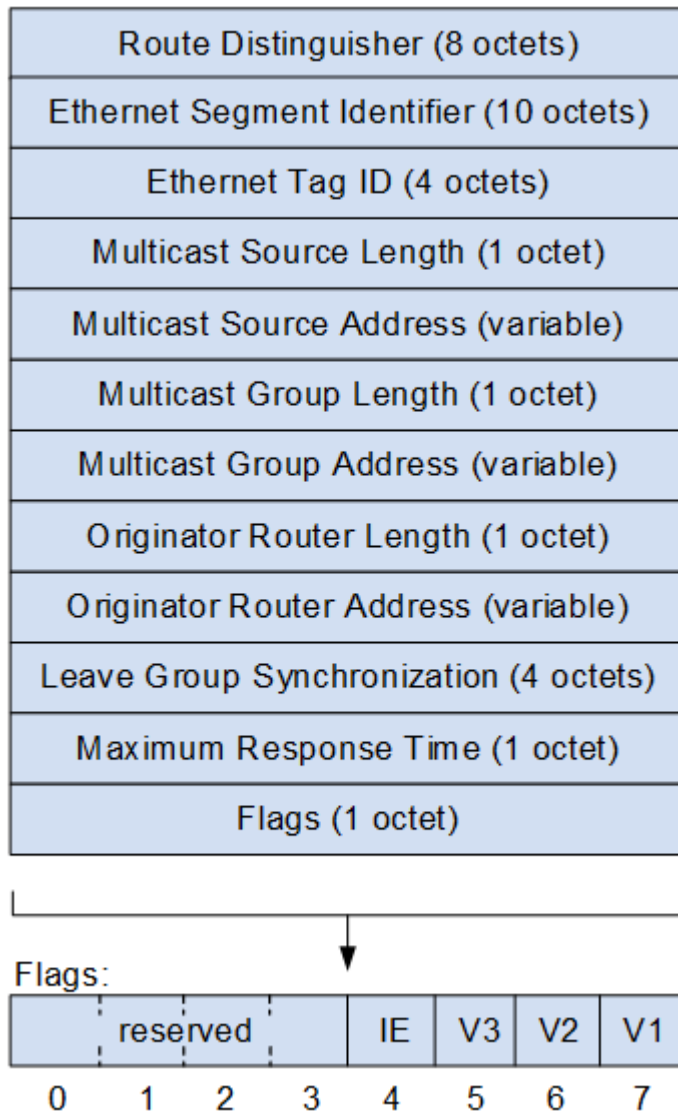
- IGMP Leave Synch route

IGMP Leave Synch routes are used to synchronize multicast group leave information between dual-homed devices on the access side. A device that receives an IGMP Leave Synch route can delete member entries from the local (S, G) entries based on the routing information, ensuring that the local entries are the same as those on the device connected to the same user network. As shown in [Figure 12-62](#), the fields in the routing information are described as follows:

- Route Distinguisher: route distinguisher (RD) set in an EVPN instance.
- Ethernet Segment Identifier: unique identifier defined for a device to connect to the access network.
- Ethernet Tag ID: This field is set to 0 when the VLAN-based or VLAN bundle service mode is used to access a user network.

- Multicast Source Length: length of a multicast source address. This field is set to 0 for any multicast source.
- Multicast Source Address: address of a multicast source. Packets do not contain this field for any multicast source.
- Multicast Group Length: length of a multicast group address.
- Multicast Group Address: address of a multicast group.
- Originator Router Length: address length of the device that generated the IGMP Leave Synch route.
- Originator Router Address: address of the device that generated the IGMP Leave Synch route.
- Leave Group Synchronization: sequence number of the process in which a multicast member leaves a specified multicast group. The sequence number increases each time the device starts the process.
- Maximum Response Time: longest time for a querier to wait for responses from downstream hosts. A host must respond to the querier before the specified longest time expires if it wants to receive traffic of a multicast group.
- Flags: This field contains eight bits. The first four most significant bits are reserved, and the last three least significant bits are used to identify the IGMP version. For example, if bit 5 is set to 1, the IGMP version of the multicast entry carried in the route is IGMPv3. Only one of the last three least significant bits can be set to 1. Bit 4 indicates the filtering mode of group records in IGMPv3. The values 0 and 1 indicate Include and Exclude, respectively.

**Figure 12-62 IGMP Leave Synch route format**



### Route Advertisement and Traffic Forwarding:

IGMP snooping over EVPN MPLS supports single- and dual-homing access.

#### Single-homing access for IGMP snooping over EVPN MPLS:

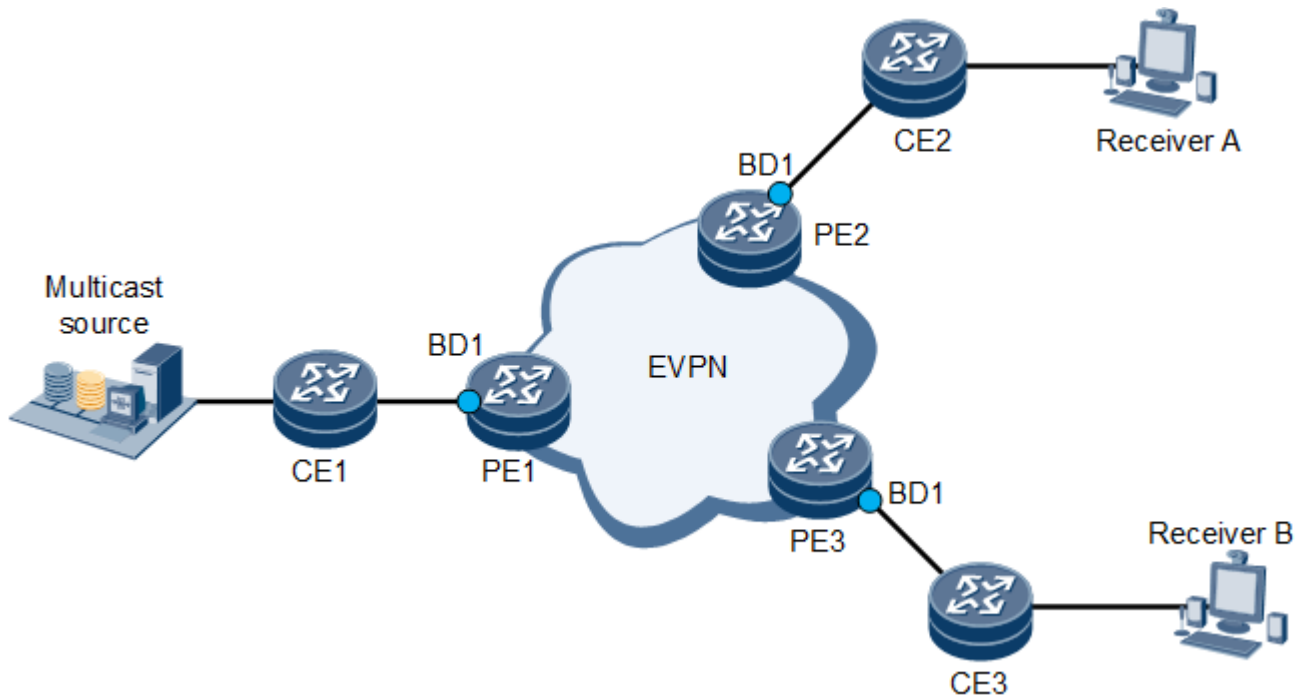
Figure 12-63 shows single-homing access for IGMP snooping over EVPN MPLS. Configure an EVPN instance on PE1, PE2, and PE3, and bind a BD to the EVPN instance. Establish BGP EVPN peer relationships between the PEs, and deploy EVPN IGMP proxy on each PE. Deploy PE1 as a sender PE, and deploy PE2 and PE3 as receiver PEs. Configure IGMP snooping and IGMP proxy on BD1 bound to the EVPN instance on PE1, PE2, and PE3. Connect BD1 on PE1, PE2, and PE3 to

CE1, CE2, and CE3 through VLAN dot1q sub-interfaces, respectively. Configure PIM-SM and IGMP on CE1's interface connected to PE1.

The process of IGMP snooping over EVPN MPLS (single-homing access) is described as follows:

1. PE1, PE2, and PE3 periodically send IGMP Query messages to the access side in BD1.
2. Receiver A and Receiver B send IGMP Report messages to CE2 and CE3, respectively. For example, Receiver A sends an IGMPv3 (S, G) Report message, and Receiver B sends an IGMPv2 (\*, G) Report message.
3. After receiving the corresponding IGMP Report messages, PE2 and PE3 establish (S, G) and (\*, G) entries of IGMP snooping in BD1 and add the interfaces connected to CE2 and CE3 as outbound interfaces, respectively.
4. PE2 sends a BGP EVPN SMET route to other PEs through BGP EVPN peer relationships. The route carries (S, G) entries, and the Flags field in the route is set to IGMPv3 and Include.
5. PE3 sends a BGP EVPN SMET route to other PEs through BGP EVPN peer relationships. The route carries (\*, G) entries, and the Flags field in the route is set to IGMPv2.
6. After receiving the corresponding BGP EVPN SMET routes, PE1 establishes (S, G) and (\*, G) entries of IGMP snooping in BD1 and adds the mLDP tunnel interfaces of the corresponding EVPN instances as outbound interfaces.
7. PE1 sends IGMPv3 (S, G) Report and IGMPv2 (\*, G) Report messages to CE1. CE1 establishes IGMP and PIM entries and forwards multicast traffic to PE1.
8. After receiving the multicast traffic, PE1 forwards the traffic to PE2 and PE3 through the mLDP tunnel interfaces based on the (S, G) and (\*, G) entries in BD1.
9. After receiving the multicast traffic, PE2 and PE3 forward the traffic to Receiver A and Receiver B based on the (S, G) and (\*, G) entries, respectively.

**Figure 12-63 Single-homing access for IGMP snooping over EVPN MPLS**



**Dual-homing access for IGMP snooping over EVPN MPLS on the multicast source side:**

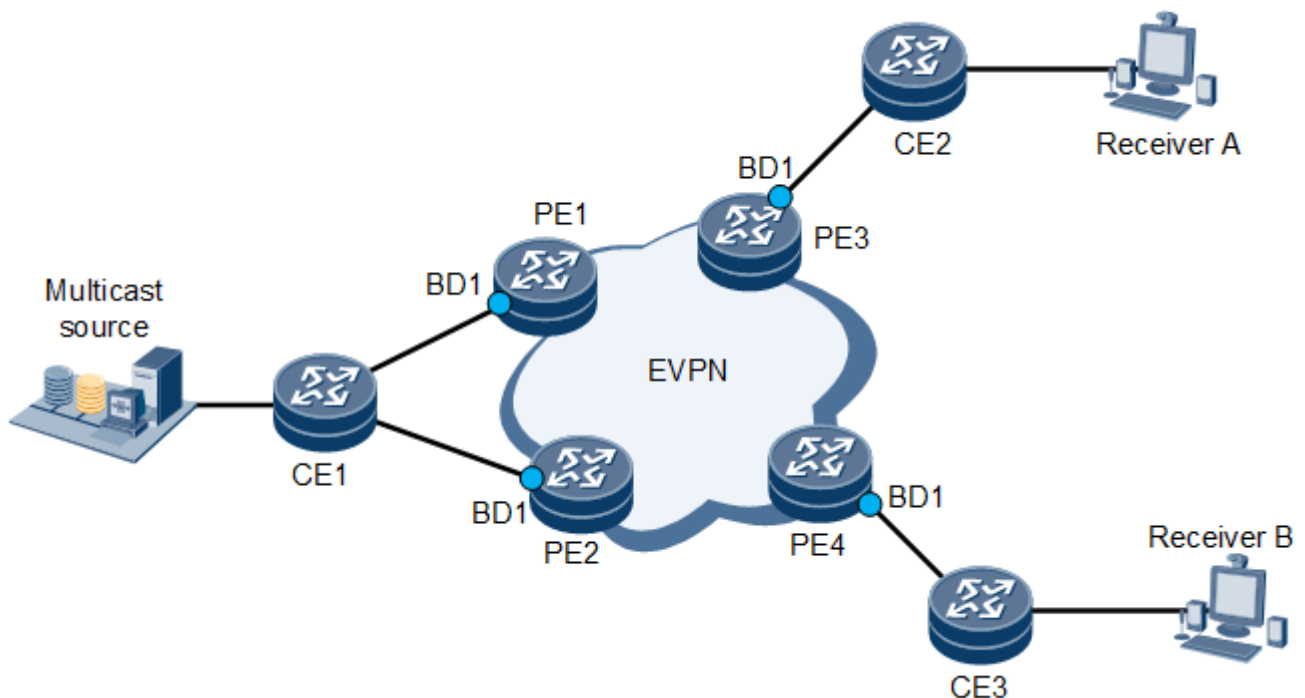
Figure 12-64 shows dual-homing access for IGMP snooping over EVPN MPLS on the multicast source side. Configure an EVPN instance on PE1, PE2, PE3, and PE4, and bind a BD to the EVPN instance. Establish BGP EVPN peer relationships between the PEs, and deploy EVPN IGMP proxy on each PE. Deploy PE1 and PE2 as sender PEs, and deploy PE3 and PE4 as receiver PEs. Connect BD1 on PE3 and PE4 to CE3 and CE4 through VLAN dot1q sub-interfaces, respectively. Connect CE1 to PE1 and PE2 through Eth-Trunk interfaces, and configure PIM-SM and IGMP on the interfaces. Bind the Eth-Trunk interfaces of CE1 to an E-Trunk on PE1 and PE2. Configure static router interfaces, and set the same ESI. Configure the E-Trunk to work in dual-active mode, and ensure that the Eth-Trunk interfaces on PE1 and PE2 are both Up.

The process of IGMP snooping over EVPN MPLS (dual-homing access on the multicast source side) is described as follows:

1. PE3 and PE4 periodically send IGMP Query messages to the access side in BD1.
2. Receiver A and Receiver B send IGMP Report messages to CE2 and CE3, respectively. For example, Receiver A sends an IGMPv3 (S, G) Report message, and Receiver B sends an IGMPv2 (\*, G) Report message.
3. After receiving the corresponding IGMP Report messages, PE3 and PE4 establish (S, G) and (\*, G) entries of IGMP snooping in BD1 and add the interfaces connected to CE2 and CE3 as outbound interfaces, respectively.

4. PE3 sends a BGP EVPN SMET route to other PEs through BGP EVPN peer relationships. The route carries (S, G) entries, and the Flags field in the route is set to IGMPv3 and Include.
5. PE4 sends a BGP EVPN SMET route to other PEs through BGP EVPN peer relationships. The route carries (\*, G) entries, and the Flags field in the route is set to IGMPv2.
6. After receiving the corresponding BGP EVPN SMET routes, PE1 and PE2 establish (S, G) and (\*, G) entries of IGMP snooping in BD1 and adds the mLDP tunnel interfaces of the corresponding EVPN instances as outbound interfaces.
7. The Eth-Trunk interface of CE1 periodically sends IGMP Query messages to BD1 of PE1 or PE2 based on hash rules. PE1 or PE2 periodically sends IGMP Report messages to CE1.
8. After receiving an IGMP Report message, CE1 creates IGMP and PIM entries and forwards multicast traffic to PE1.
9. CE1 forwards the multicast traffic from the multicast source to BD1 of PE1 or PE2 based on hash rules. PE1 or PE2 forwards the multicast traffic to PE3 and PE4 through the mLDP tunnel interfaces based on the (\*, G) and (S, G) entries of BD1.
10. After receiving the multicast traffic, PE2 and PE3 forward the traffic to Receiver A and Receiver B based on the (S, G) and (\*, G) entries, respectively.

**Figure 12-64 Dual-homing access for IGMP snooping over EVPN MPLS on the multicast source side**



#### **Dual-homing access for IGMP snooping over EVPN MPLS on the access side:**

Figure 12-65 shows dual-homing access for IGMP snooping over EVPN MPLS on the access side. Configure an EVPN instance on PE1, PE2, and PE3, and bind a BD to the EVPN instance. Establish BGP EVPN peer relationships between the PEs, and deploy EVPN IGMP proxy on each PE.

Deploy PE1 as a sender PE, and deploy PE2 and PE3 as receiver PEs. Configure IGMP snooping and IGMP proxy on BD1 bound to the EVPN instance on PE1, PE2, and PE3. Connect BD1 on PE1, PE2, and PE3 to CE1, CE2, and CE3 through VLAN dot1q sub-interfaces, respectively. Configure PIM-SM and IGMP on CE1's interface connected to PE1, and connect CE2 to PE2 and PE3 through Eth-Trunk interfaces. Bind the Eth-Trunk interfaces of CE2 to an E-Trunk and configure the same ESI on PE2 and PE3. Configure the E-Trunk on PE2 and PE3 to work in single-active mode, select PE2 as the master device, and ensure that the Eth-Trunk interface of PE2 is Up.



**NOTE:**

IGMPv3 is not supported in access-side dual-homing access scenarios.

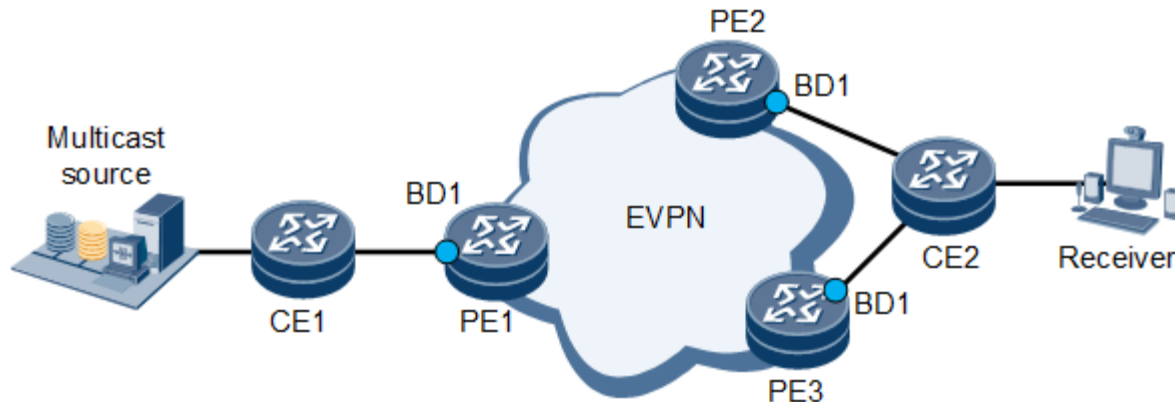
The process of IGMP snooping over EVPN MPLS (dual-homing access on the access side) is described as follows:

1. PE2 periodically sends IGMP Query messages to the access side in BD1.
2. The receiver sends an IGMP Report message, for example, IGMPv2 (\*, G) Report message, to CE2.
3. After receiving an IGMP Report message, PE2 establishes (\*, G) entries of IGMP snooping, adds the Eth-Trunk interface to CE2 as the outbound interface, and sends the IGMP Join Synch route of BGP EVPN to other PEs. The route carries the access-side ESI of PE2 and contains the IGMP version and V2 source filtering mode.
4. After receiving the IGMP Join Synch route, PE3 creates the corresponding (\*, G) entries of IGMP snooping in BD1. PE3 does not need to send a BGP EVPN SMET route, because it is a non-DF. Additionally, PE3 does not add the Eth-Trunk interface to CE2 as the outbound interface, because the Eth-Trunk interface is Down.
5. PE2 functioning as a DF sends a BGP EVPN SMET route based on (\*, G) entries of IGMP snooping.
6. After receiving the BGP EVPN SMET route from PE2, PE1 creates (\*, G) entries of IGMP snooping and sends an IGMP Report message to CE1.
7. After receiving an IGMP Report message, CE1 creates IGMP and PIM entries and forwards multicast traffic to PE1.
8. CE1 sends the multicast traffic received from the multicast source to PE1.
9. After receiving the multicast traffic, PE1 forwards the traffic to PE2 and PE3 through the mLDP tunnel interfaces based on the (\*, G) entries in BD1.
10. After PE2 and PE3 receive the multicast traffic, PE2 forwards the traffic to CE2 based on the (\*, G) entries, but PE3 does not. In this case, the receiver receives only one copy of multicast traffic.
11. If some receivers are disconnected or do not need to receive multicast traffic, PE2 updates the (\*, G) entries based on the IGMP Report messages received from CE2, and sends IGMP

Leave Synch routes to PE3. PE3 then deletes the receivers' entries to ensure that the local (\*, G) entries are the same as those on PE2.

12. If the access side of PE2 fails, the EVPN instance selects PE3 as a DF, and the Eth-Trunk interface of PE3 goes Up. PE3 then adds the Eth-Trunk interface to CE2 as the outbound interface, so that multicast traffic is forwarded from PE3 to CE2.

**Figure 12-65 Dual-homing access for IGMP snooping over EVPN MPLS on the access side**



## Application Scenarios for EVPN:

- Inter-AS EVPN Option C
- DCI Scenarios
- Migration from an HVPLS Network to a PBB-EVPN
- Using EVPN to Interconnect Other Networks
- EVPN Splicing
- Seamless Migration of VPLS to EVPN
- EVPN L3VPN HVPN

- **Inter-AS EVPN Option C:**

Inter-AS EVPN Option C implements Layer 2 interconnection between networks in different ASs.

**Background:**

With the wide application of MPLS VPN solutions, different MANs of a service provider or collaborative backbone networks of different service providers often span multiple ASs. Similar to L3VPN services, EVPN services running on an MPLS network must also have the capability of spanning ASs.



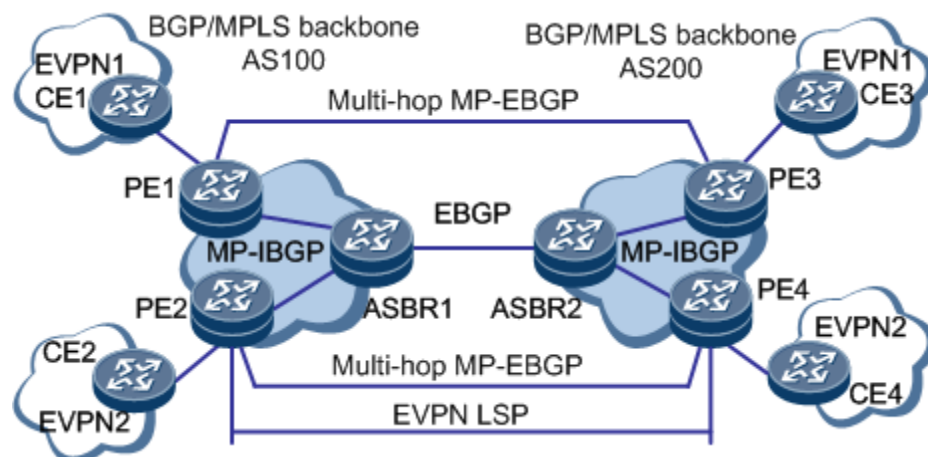
### Implementation:

By advertisement of labeled routes between PEs, end-to-end BGP LSPs can be established to carry Layer 2 traffic in BGP ASs (including inter-IGP areas) and inter-BGP ASs that only support Option C.

In Option C mode, an autonomous system boundary router (ASBR) does not maintain or advertise EVPN routes. Instead, PEs exchange EVPN routes directly. EVPN routes include the following:

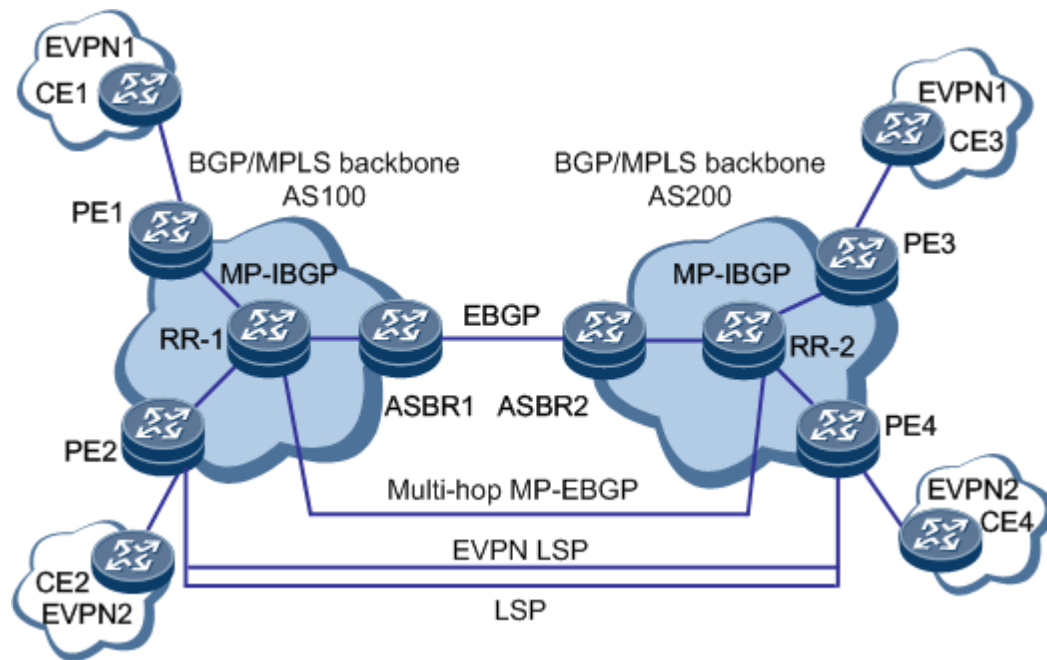
- Ethernet auto-discovery routes
- MAC and IP routes
- Inclusive multicast routes
- Ethernet segment routes
- ASBRs advertise labeled IPv4 routes to PEs in their respective ASs through MP-IBGP, and advertise labeled IPv4 routes received on PEs in the local AS to the ASBR peers in other ASs. ASBRs in the transit AS also advertise labeled IPv4 routes. Therefore, a BGP LSP can be established between the ingress PE and egress PE.
- PEs in different ASs establish multi-hop EBGP connections with each other and exchange EVPN routes.
- ASBRs do not store EVPN routes or advertise EVPN routes to each other.

**Figure 12-66 Inter-AS EVPN Option C networking where PEs advertise labeled EVPN routes**



To improve expansibility, you can specify a route reflector (RR) in each AS. An RR stores all EVPN routes and exchanges EVPN routes with PEs in the AS. RRs in two ASs establish MP-EBGP connections with each other and advertise EVPN routes.

**Figure 12-67 Inter-AS EVPN Option C networking with RRs**



Inter-AS EVPN Option C can be implemented using the following solutions:

- A local ASBR learns a labeled public network BGP route from the peer ASBR, assigns a label to this route based on a matching policy, and advertises this route to its IBGP peer. Then, a complete public network LSP is established.
- The IBGP peer relationship between a PE and ASBR in the same AS is not required. In this solution, a local ASBR learns a labeled public network BGP route from the peer ASBR and imports this route to an IGP to trigger LDP LSP establishment. Then, a complete LSP is established between the ingress and egress on the public network.

**Benefits:**

- EVPN routes are directly exchanged between an ingress PE and egress PE. The routes do not have to be stored and forwarded by intermediate devices.
- Only PEs exchange EVPN routing information. Ps and ASBRs forward packets only. The intermediate devices need to support only MPLS forwarding rather than MPLS VPN services. In such a case, ASBRs are no longer the performance bottlenecks. Inter-AS EVPN Option C, therefore, is suitable for an EVPN that spans multiple ASs.

- **DCI Scenarios:**

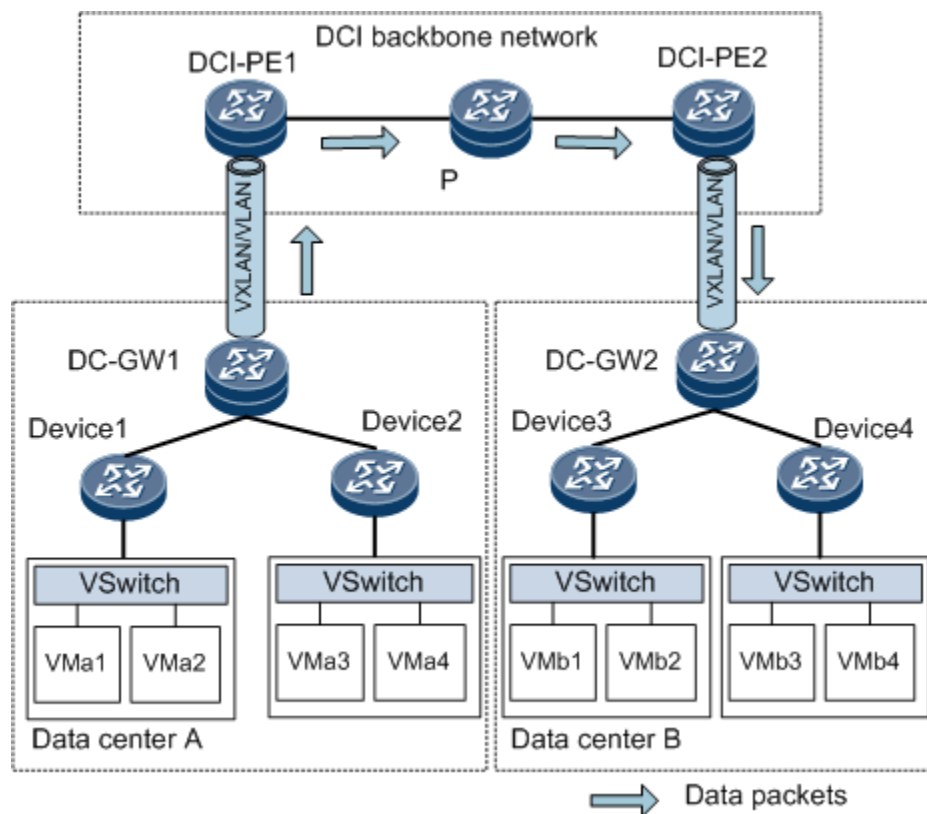
Data Center Interconnect (DCI) is a solution for communication between virtual machines (VMs) in different data centers (DCs). DCI runs on carriers' networks. It uses technologies such as Virtual eXtensible Local Area Network (VXLAN), Ethernet virtual private network (EVPN), and BGP/MPLS IP VPN to ensure secure and reliable transmission of packets from DCs, implementing communication between VMs in different DCs.

**Table 12-7** Basic DCI concepts

Concept	Description
Overlay network	<ul style="list-style-type: none"> <li>• An overlay network is a logical network established on a physical network and can be considered as a network connected through virtual or logical links.</li> <li>• The overlay network has an independent control plane and forwarding plane.</li> <li>• The overlay network deeply extends a physical network to a cloud-based and virtualized network and frees the cloud resource pool from the limitations of the physical network. This is the key to the convergence of the cloud network.</li> </ul>
Underlay network	An underlay network carries an overlay network and is usually a physical network at the underlying layer.
Individual deployment of DC-GWs and DCI-PEs	A DC-GW and a DCI-PE are different devices.
Integrated deployment of DCI-PEs and DC-GWs	A DC-GW and a DCI-PE are a single device, which applies to scenarios where carriers build their own DCs.

On the network shown in [Figure 12-68](#), gateways in the DCs (DC-GW1 and DC-GW2) can access the carrier's network edge devices (DCI-PE1 and DCI-PE2) in EVPN-VXLAN or VLAN mode. The L3VPN or EVPN-MPLS function can be deployed on the DCI backbone network to transmit Layer 2 or Layer 3 service traffic. When DC A and DC B exchange their tenant host IP addresses or MAC addresses, EVPN integrated routing and bridging (IRB) routes, EVPN IP prefix routes, BGP VPNv4 routes, EVPN MAC routes, or ARP routes are used. For details about these routes, see [Table 12-8](#).

**Figure 12-68 Basic DCI scenarios**



**Table 12-8 Route information**

Route	Function	Fields Carried in a Route
EVPN IRB route	Used to transmit a tenant's host IP address and MAC address on an EVPN.	<ul style="list-style-type: none"> <li>RD1: route distinguisher 1, indicating the route ID of an EVPN instance.</li> <li>VM-MAC: MAC address of a VM.</li> <li>VM-IP: IP address of a VM.</li> <li>Label 1: L2VNI of a VXLAN tunnel or Layer 2 MPLS label.</li> <li>Label 2: L3VNI of a VXLAN tunnel or Layer 3 MPLS label.</li> <li>NHP: next hop of a route, usually a local IP address</li> </ul>

Route	Function	Fields Carried in a Route
		<p>used to establish a BGP EVPN peer relationship.</p> <ul style="list-style-type: none"> <li>ExtCommunity: extended community attributes of a route, including the VXLAN encapsulation mode, Router-MAC, and export route target (ERT) of a route.</li> </ul>
EVPN IP prefix route	Used to transmit a tenant's host IP address or the address of the network segment to which the host IP address belongs on an EVPN.	<ul style="list-style-type: none"> <li>RD1: route distinguisher 1, indicating the route ID of an EVPN instance.</li> <li>IP: VM's IP address or address of the network segment to which a VM's IP address belongs.</li> <li>Label: L3VNI of a VXLAN tunnel or Layer 3 MPLS label.</li> <li>NHP: next hop of a route, usually a local IP address used to establish a BGP EVPN peer relationship.</li> <li>ExtCommunity: extended community attributes of a route, including the VXLAN encapsulation mode, Router-MAC, and ERT of a route.</li> </ul>
VPNv4 route	Used to transmit a tenant's host IP address or the address of the network segment to which the host IP address belongs on an L3VPN.	<ul style="list-style-type: none"> <li>RD2: route distinguisher 2, indicating the ID of a VPNv4 route.</li> <li>VM-IP: IP address of a VM.</li> <li>Label: VPN label carried in VPNv4 routes.</li> <li>NHP: next hop of a route, usually a local IP address used to establish a BGP</li> </ul>

Route	Function	Fields Carried in a Route
		VPNv4 peer relationship. <ul style="list-style-type: none"> <li>ExtCommunity: extended community attribute of a route, only the ERT attribute.</li> </ul>
EVPN MAC route or ARP route	Used to transmit a tenant's host MAC address or ARP information on an EVPN.	<ul style="list-style-type: none"> <li>RD1: route distinguisher 1, indicating the route ID of an EVPN instance.</li> <li>VM-MAC: MAC address of a VM.</li> <li>VM-IP: IP address of a VM. This field is carried only in ARP routes.</li> <li>Label: L2VNI of a VXLAN tunnel or Layer 2 MPLS label.</li> <li>NHP: next hop of a route, usually a local IP address used to establish a BGP EVPN peer relationship.</li> <li>ExtCommunity: extended community attributes of a route, including the VXLAN encapsulation mode and ERT of a route.</li> </ul>

### ***DCI Control Plane:***

The DCI control plane advertises both Layer 3 and Layer 2 routes:

- During Layer 3 route advertisement, a DC sends an IRB route or IP prefix route carrying a tenant's host IP address to a DCI-PE through the EVPN protocol. Upon receipt, the DCI-PE re-encapsulates the routing information into a BGP VPNv4 route if an L3VPN is deployed on the backbone network. Alternatively, if EVPN-MPLS is deployed on the backbone network, the DCI-PE re-encapsulates the received route into an IRB or IP prefix route. The re-encapsulated routes carry the VM's IP route and are transmitted to the remote DCI-PE through the backbone network.
- The process of Layer 2 route advertisement is that a DC uses EVPN to send packets carrying the host's MAC address or ARP entries to the local DCI-PE. The local DCI-PE then re-

generates the EVPN MAC/ARP routes that carry the MPLS encapsulation attribute. The regenerated routes that carry the VM's MAC address or ARP entries are transmitted to the remote DCI-PE.

Table 12-9 describes Layer 3 route advertisement and Layer 2 route advertisement.

**Table 12-9 Route advertisement**

Deployment Mode	Services	Advertisement Process		
		DC-GW1 to DCI-PE1	DCI-PE1 to DCI-PE2	DCI-PE2 to DC-GW2
L3VPN (VXLAN access)	Layer 3 services	DC-GW1 sends a tenant's host IP address to DCI-PE1 through an IRB route or IP prefix route. DCI-PE1 parses the tenant's host IP route from the received EVPN route. Then the system imports the tenant's route into the IP VPN instance based on RT matching between the EVPN route and the IP VPN instance and delivers information about VXLAN tunnel recursion to the VPN forwarding table.	DCI-PE1 re-encapsulates the EVPN route received from DC-GW1 into a BGP VPNv4 route, applying the following changes: <ul style="list-style-type: none"> <li>Changes the next hop to the local device's IP address used to establish a BGP VPNv4 peer relationship.</li> <li>Replaces the RD and RT values of the EVPN route with those of an L3VPN instance.</li> <li>Applies for and encapsulates a VPN label.</li> </ul> After re-	Upon receipt, DCI-PE2 imports the BGP VPNv4 route into the local IP VPN instance based on the route RT and delivers information about MPLS tunnel recursion to the VPN forwarding table. DCI-PE2 re-encapsulates the received BGP VPNv4 route into an IP prefix route, applying the following changes: <ul style="list-style-type: none"> <li>Changes the next hop to the VTEP address of DCI-PE2.</li> <li>Replaces the RD and RT</li> </ul>

Deployment Mode	Services	Advertisement Process		
		DC-GW1 to DCI-PE1	DCI-PE1 to DCI-PE2	DCI-PE2 to DC-GW2
			encapsulation, DCI-PE1 sends the route to DCI-PE2.	values of the BGP VPNv4 route with those of the L3VPN instance and pads the route with an L3VNI.  After re-encapsulation, DCI-PE2 sends the IP prefix route to DC-GW2.
EVPN-MPLS (VLAN access)	Layer 3 services	DC-GW1 sends routes destined for the network segment on which a tenant's host IP address resides to DCI-PE1 through an IGP or BGP route. Upon receipt, DCI-PE1 delivers these routes to the VPN forwarding table.	DCI-PE1 re-encapsulates the VPN route into an IP prefix route, applying the following changes: <ul style="list-style-type: none"> <li>Changes the next hop to the local device's IP address used to establish a BGP EVPN peer relationship.</li> <li>Adds the RD and RT attributes to the EVPN</li> </ul>	After receiving the EVPN route, DCI-PE2 imports the route into the local IP VPN instance based on the RT of the EVPN route, generates a VPN route forwarding entry, and advertises the EVPN route to DC-GW2 through a VPN IGP or BGP peer relationship.



Deployment Mode	Services	Advertisement Process		
		DC-GW1 to DCI-PE1	DCI-PE1 to DCI-PE2	DCI-PE2 to DC-GW2
			<p>route.</p> <ul style="list-style-type: none"> <li>Applies for and encapsulates a VPN label.</li> </ul> <p>After re-encapsulation, DCI-PE1 sends the route to DCI-PE2.</p>	
	Layer 2 services	<p>DCI-PE1 learns the source MAC address of service traffic received from DC-GW1. Then DCI-PE1 generates a local MAC forwarding entry and an EVPN MAC route.</p>	<p>DCI-PE1 generates an EVPN MAC route, applying the following changes:</p> <ul style="list-style-type: none"> <li>Changes the next hop to the local device's IP address used to establish a BGP EVPN peer relationship.</li> <li>Adds the RD and RT attributes to the EVPN route.</li> <li>Applies for and encapsulates a VPN label.</li> </ul> <p>After re-</p>	<p>Upon receipt, DCI-PE2 imports the MAC/IP advertisement route into the local EVPN instance based on the route RT and generates a local Layer 2 forwarding entry accordingly.</p>

Deployment Mode	Services	Advertisement Process		
		DC-GW1 to DCI-PE1	DCI-PE1 to DCI-PE2	DCI-PE2 to DC-GW2
			encapsulation, DCI-PE1 sends the route to DCI-PE2.	
EVPN-MPLS (VXLAN access)	Layer 3 services	DC-GW1 sends a tenant's host IP address to DCI-PE1 through an IRB route or IP prefix route. DCI-PE1 parses the tenant's host IP route from the received EVPN route. Then the system imports the tenant's route into the IP VPN instance based on RT matching between the local EVPN instance and the IP VPN instance and delivers information about VXLAN tunnel recursion to the VPN forwarding table.	DCI-PE1 re-encapsulates the route into an IRB or IP prefix route. The encapsulation mode changes from VXLAN to MPLS: <ul style="list-style-type: none"> <li>Changes the next hop to the local device's IP address used to establish a BGP EVPN peer relationship.</li> <li>Adds the RD and RT attributes to the EVPN route.</li> <li>Applies for and encapsulates a VPN label.</li> </ul> After re-encapsulation, DCI-PE1 sends the route to DCI-	Upon receipt, DCI-PE2 imports the IRB or IP prefix route into the IP VPN instance and delivers information about MPLS tunnel recursion to the VPN forwarding table. DCI-PE2 changes the L2 and L3 VPN labels in the route to L2 and L3 VNIs, re-encapsulates the route into an IRB or IP prefix route, and then sends the route to DC-GW2.

Deployment Mode	Services	Advertisement Process		
		DC-GW1 to DCI-PE1	DCI-PE1 to DCI-PE2	DCI-PE2 to DC-GW2
			PE2.	
	Layer 2 services	DC-GW1 sends a tenant's host MAC address to DCI-PE1 through a MAC/IP advertisement route. DCI-PE1 imports the MAC/IP advertisement route into the local EVPN instance based on RT matching and generates a MAC forwarding entry.	DCI-PE1 re-encapsulates the EVPN routes and change the next-hop IP address to the IP address of the locally established EVPN peer. The RD and RT attributes in the EVPN routes that carry the VXLAN encapsulation attribute are replaced with the RD and RT of the local EVPN instance. The MPLS label is requested. The re-encapsulated MAC/IP Advertisement routes are then advertised to DCI-PE2.	Upon receipt, DCI-PE2 imports the MAC/IP advertisement route into the local EVPN instance based on RT matching. DCI-PE2 re-encapsulates the EVPN route by changing the next hop to its own VTEP address, replacing the RD and RT values of the EVPN route with those of the local EVPN instance and padding the route with an L2VNI. Then DCI-PE2 sends the re-encapsulated MAC address advertisement route to DC-GW2.

### **DCI Data Plane**

Table 12-10 describes Layer 2 traffic forwarding and Layer 3 traffic forwarding.

**Table 12-10** Service traffic forwarding

Deployment Mode	Services	Forwarding Process		
		DC-GW2 to DCI-PE2	DCI-PE2 to DCI-PE1	DCI-PE1 to DC-GW1
L3VPN (VXLAN access)	Layer 3 services	DC-GW2 sends a data packet to DCI-PE2 through the VXLAN tunnel.	DCI-PE2 parses the VXLAN data packet to obtain the VNI and data packet. Based on the VNI, DCI-PE2 finds the corresponding VPN instance and, based on the tenant's host IP address for the MPLS tunnel to DCI-PE1, searches the corresponding VPN instance forwarding table. After encapsulating a VPN label and a public MPLS tunnel label into the data packet, DCI-PE2 sends the packet to DCI-PE1 through the MPLS tunnel.	Upon receipt, DCI-PE1 removes the public MPLS tunnel label, and, based on the VPN label, finds the corresponding VPN instance. Then, based on the tenant's host IP address for the VXLAN tunnel to DC-GW1, DCI-PE1 searches the corresponding VPN instance forwarding table. DCI-PE1 encapsulates the data packet with a VXLAN header and then sends the VXLAN packet to DC-GW1.
EVPN-MPLS (VLAN access)	Layer 3 services	DC-GW2 sends a data packet to DCI-PE2 through VPN forwarding.	DCI-PE2 searches the forwarding table of the VPN instance bound to the interface that receives the data packet and, based on the destination	Upon receipt, DCI-PE1 removes the public MPLS tunnel label, and, based on the VPN label, finds the corresponding VPN instance.

Deployment Mode	Services	Forwarding Process		
		DC-GW2 to DCI-PE2	DCI-PE2 to DCI-PE1	DCI-PE1 to DC-GW1
			address of the data packet, finds the MPLS tunnel to DCI-PE1. After encapsulating a VPN label and a public MPLS tunnel label into the data packet, DCI-PE2 sends the packet to DCI-PE1 through the MPLS tunnel.	Based on the tenant's host IP address, DC-PE1 searches the corresponding VPN instance forwarding table for the outbound interface to DC-GW1. Then, DC-PE1 sends the data packet to DC-GW1 through the outbound interface.
	Layer 2 services	DC-GW2 sends a data packet to DCI-PE2 through Layer 2 forwarding on the data plane.	DCI-PE2 searches the forwarding table of the EVPN instance bound to the interface that receives the data packet and, based on the destination address of the data packet, finds the MPLS tunnel to DCI-PE1. After encapsulating a VPN label and a public MPLS tunnel label into the data packet, DCI-PE2 sends the packet to	Upon receipt, DCI-PE1 removes the public MPLS tunnel label, and, based on the VPN label, finds the corresponding EVPN instance. Based on the MAC forwarding entry for the broadcast domain bound to the EVPN instance, DC-PE1 finds the corresponding outbound interface and sends the data packet to DC-

Deployment Mode	Services	Forwarding Process		
		DC-GW2 to DCI-PE2	DCI-PE2 to DCI-PE1	DCI-PE1 to DC-GW1
			DCI-PE1 through the MPLS tunnel.	GW1 through the outbound interface.
EVPN-MPLS (VXLAN access)	Layer 3 services	DC-GW2 sends a data packet to DCI-PE2 through the VXLAN tunnel.	DCI-PE2 parses the VXLAN data packet to obtain the VNI and data packet. Based on the VNI, DCI-PE2 finds the corresponding VPN instance and, based on the tenant's host IP address for the MPLS tunnel to DCI-PE1, searches the corresponding VPN instance forwarding table. After encapsulating a VPN label and a public MPLS tunnel label into the data packet, DCI-PE2 sends the packet to DCI-PE1 through the MPLS tunnel.	Upon receipt, DCI-PE1 removes the public MPLS tunnel label, and, based on the VPN label, finds the corresponding VPN instance. Then, based on the tenant's host IP address for the VXLAN tunnel to DC-GW1, DCI-PE1 searches the corresponding VPN instance forwarding table. DCI-PE1 encapsulates the data packet with a VXLAN header and then sends the VXLAN packet to DC-GW1.
	Layer 2 services	DC-GW2 sends a data packet to DCI-PE2 through the VXLAN tunnel.	DCI-PE2 parses the VXLAN data packet to obtain the VNI and data packet. Based on	Upon receipt, DCI-PE1 removes the public MPLS tunnel label and, based on the

Deployment Mode	Services	Forwarding Process		
		DC-GW2 to DCI-PE2	DCI-PE2 to DCI-PE1	DCI-PE1 to DC-GW1
			<p>the VNI, DCI-PE2 finds the corresponding broadcast domain. Based on the broadcast domain, DCI-PE2 finds the forwarding table of the corresponding EVPN instance. DCI-PE2 searches for the forwarding information corresponding to the destination address of the data packet, that is, information about the MPLS tunnel to DCI-PE1. After encapsulating a VPN label and a public MPLS tunnel label into the data packet, DCI-PE2 sends the packet to DCI-PE1 through the MPLS tunnel.</p>	<p>VPN label and BD ID, finds the corresponding broadcast domain, and then, based on the tenant's host destination MAC address, searches the broadcast domain for the VXLAN tunnel to DC-GW1. DCI-PE1 encapsulates the data packet with a VXLAN header and then sends the VXLAN packet to DC-GW1.</p>

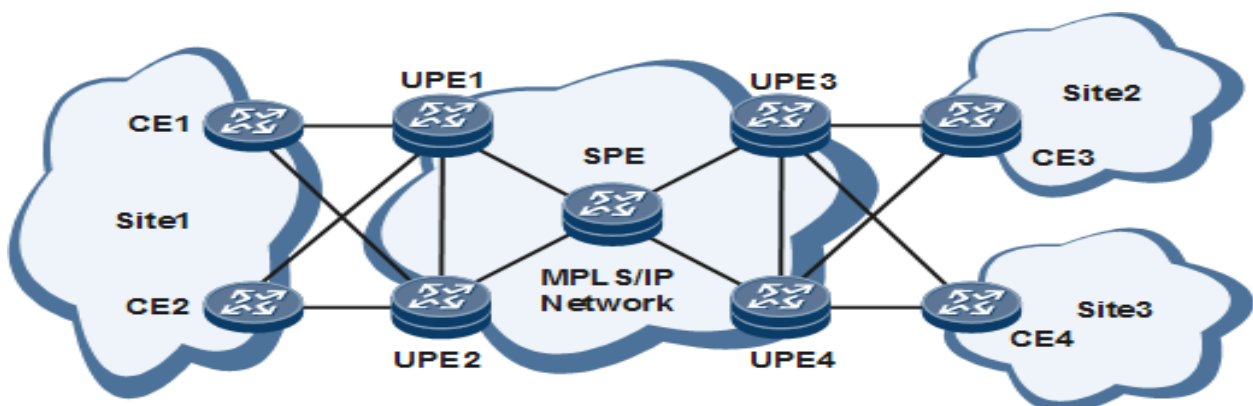
- **Migration from an HVPLS Network to a PBB-EVPN:**

On the network shown in [Figure 12-69](#), VPLS is deployed to allow services of the same private network to access VSIs over different PEs. To avoid establishment of full-mesh PWs, SPEs are deployed on the network to form an HVPLS.

After devices have PBB-EVPN enabled, the HVPLS network can migrate to a PBB-EVPN. Because this network has large numbers of devices, migration needs to be performed step by step and HVPLS and PBB-EVPN will temporarily coexist. The implementation process is as follows:

1. Configure a B-EVPN instance on the SPE and specify a unique B-MAC address for the B-EVPN instance.
2. Change the existing VSI on the SPE to be an MP2MP I-VSI and bind the I-VSI to the B-EVPN instance previously configured. The I-tag for the I-VSI must be the same as the I-tag for the B-EVPN instance. Otherwise, services cannot be forwarded.
3. Specify each UPE as an EVPN BGP peer for the SPE.
4. Configure a B-EVPN instance on each UPE and specify the SPE as an EVPN BGP peer for each UPE. Then, UPEs will learn B-MAC addresses from their EVPN BGP peers and the SPE will learn the B-MAC addresses of the entire network.
5. Change the existing VSI on each UPE to be an I-EVPN instance, bind the I-EVPN instance to the previously configured B-EVPN instance, and bind the AC interface on each UPE to the I-EVPN instance on that UPE. After all configurations are complete, the network becomes a PBB-EVPN.

**Figure 12-69 Typical networking**



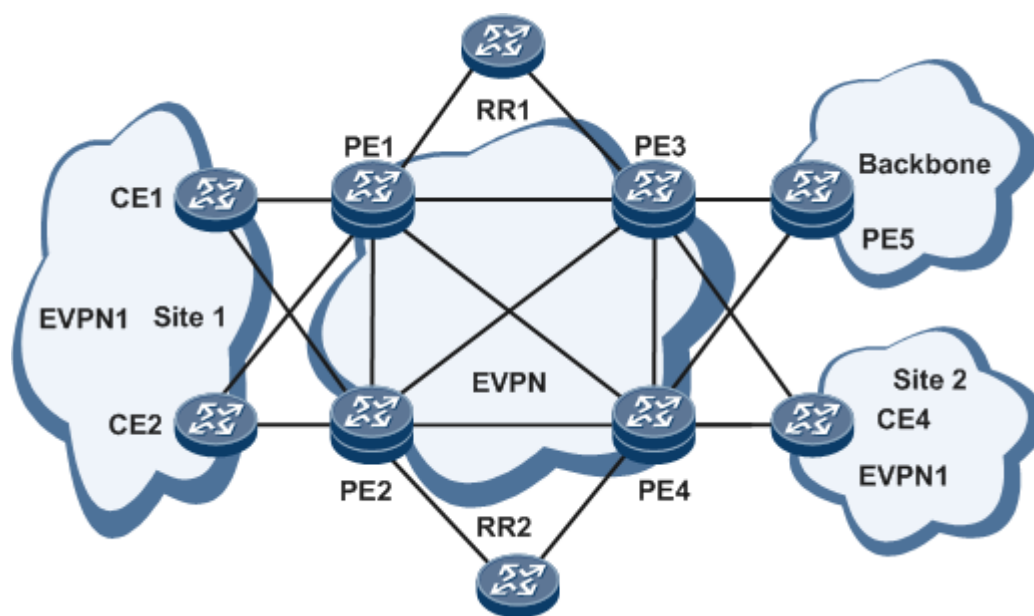


- **Using EVPN to Interconnect Other Networks:**

On the network shown in [Figure 12-70](#), to interconnect different sites through a public network, deploy EVPN by performing the following configurations:

- Configure a PE on the backbone network as an EVPN RR and the other PEs as RR clients. Establish BGP EVPN peer relationships between the RR and clients, but not between the clients. To improve reliability, you can configure two EVPN RRs, one as the master and the other as the backup.
- Create EVPN instances on PEs. Configure the same RT values for the PEs to allow EVPN route cross.
- Configure PE redundancy. If all PEs connecting to the same CE are configured to work in All-Active mode, these PEs load-balance traffic destined for the CE.

**Figure 12-70 EVPN application networking**



- **EVPN Splicing**

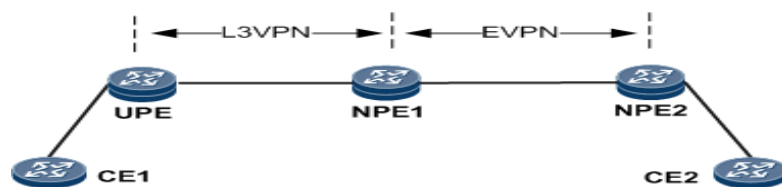
**Background:**

The current MAN is evolving into EVPN. However, because there are a large number of devices at the aggregation layer, it is difficult for the MAN to evolve into EVPN at a time. To allow traditional L3VPN, VPWS or VPLS to be still used at the aggregation layer and the core layer to evolve into EVPN first, splicing between EVPN and the traditional network must be supported.

### **L3VPN Accessing EVPN:**

The network between the UPE and NPE1 resides at the aggregation layer. The network between NPE1 and NPE2 resides at the core layer. An L3VPN is deployed at the aggregation layer, and EVPN-MPLS is deployed at the core layer. After receiving user routes from the access side, the UPE sends these routes to NPE1 through a BGP VPNv4 peer relationship. Both an EVPN instance and an L3VPN instance are configured on NPE1. After receiving BGP VPNv4 routes, NPE1 imports these routes into the L3VPN instance, encapsulates the routes into EVPN routes, and sends the EVPN routes to NPE2 through a BGP EVPN peer relationship. This implementation is L3VPN accessing EVPN as such.

**Figure 12-71 L3VPN accessing EVPN**



### **VLL Accessing EVPN:**

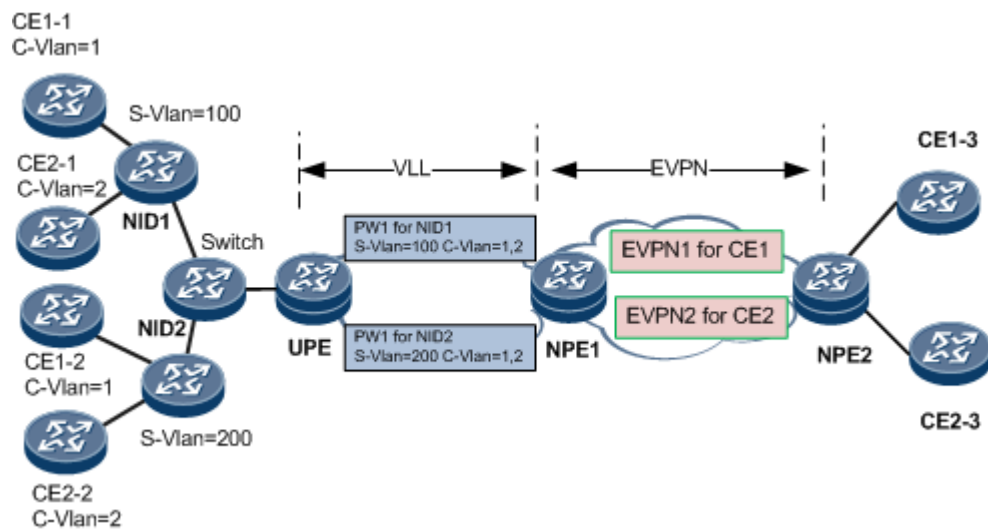
On a network with VLL accessing EVPN, CE1 and CE2 stand for two users. Each user has three sites: CE1-1, CE1-2, and CE1-3 for CE1, and CE2-1, CE2-2, and CE2-3 for CE2. NIDs, which function as aggregation devices on the user side, are attached to the user sites and access the aggregation network. When accessing the aggregation network, the CEs use the S-VLAN and C-VLAN tags. S-VLAN indicates an NID, and C-VLAN indicates the user site connected to the NID. The users access the VLL network, a Layer 2 network, through the NIDs. The UPE and NPE1 belong to the aggregation layer, at which an MPLS network is deployed. Services between the devices are carried using a VLL. NPE1 and NPE2 belong to the core layer, at which an MPLS network is deployed. Services between them are carried through an EVPN.

To allow communication between different sites of the same user, VLL accessing EVPN supports the following scenarios:

- **Single-homing scenario:**

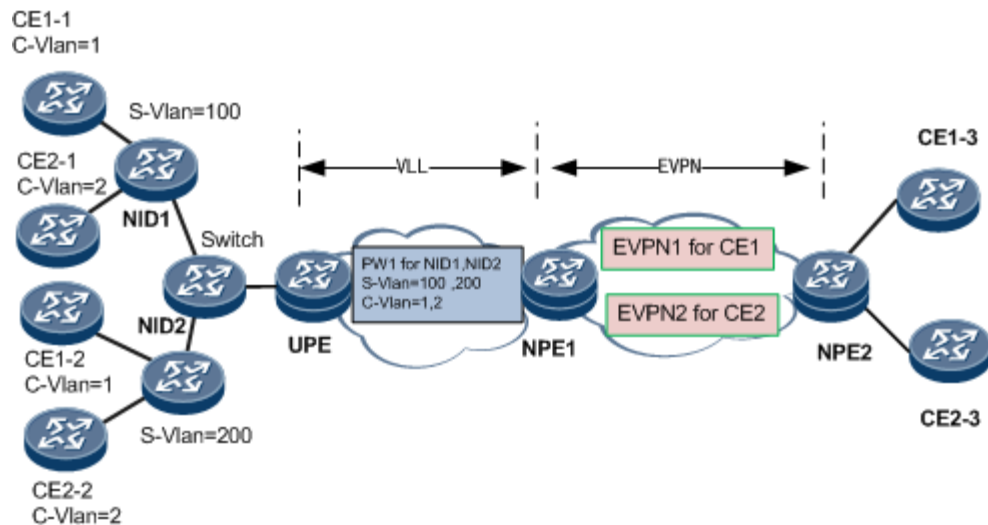
An NID on the access side can be single-homed to a UPE through a main interface. The UPE establishes a PW with NPE1 for each NID. On NPE1 and NPE2, an EVPN instance is created for each user. On NPE1, a VLL is connected to the EVPN through a PW VE interface. The VLL is bound to the PW VE interface, and the EVPN instances are bound to the PW VE sub-interfaces that are configured as QinQ VLAN tag termination sub-interfaces. In this manner, traffic of user packets is imported to different EVPN instances based on the S-VLAN and C-VLAN tags.

**Figure 12-72 Single-homing scenario 1 for VLL accessing EVPN (per NID per PW)**



Additionally, VLL accessing EVPN allows multiple NIDs to share a PW. In this scenario, multiple NIDs are aggregated to a switch, which then accesses a PW on a UPE.

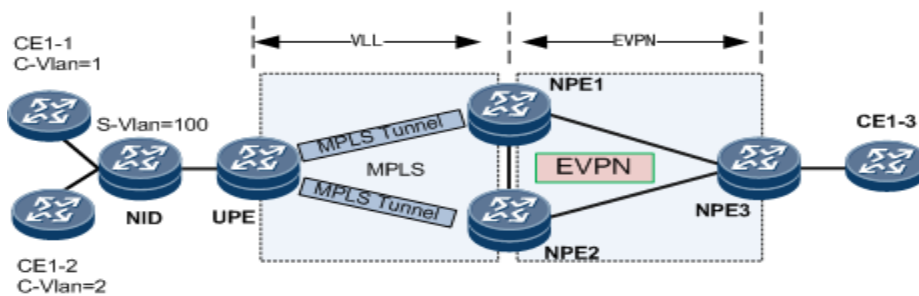
**Figure 12-73 Single-homing scenario 2 for VLL accessing EVPN (multiple NIDs per PW)**



- **Dual-homing scenario:**

A UPE is dual-homed to the master and slave NPEs through primary and secondary PWs respectively to improve access reliability. On the EVPN, the NPE1-NPE3 link and the NPE2-NPE3 link can be configured to work in single-active mode or in all-active mode, which allows for load balancing.

**Figure 12-74 Dual-homing scenario for VLL accessing EVPN**

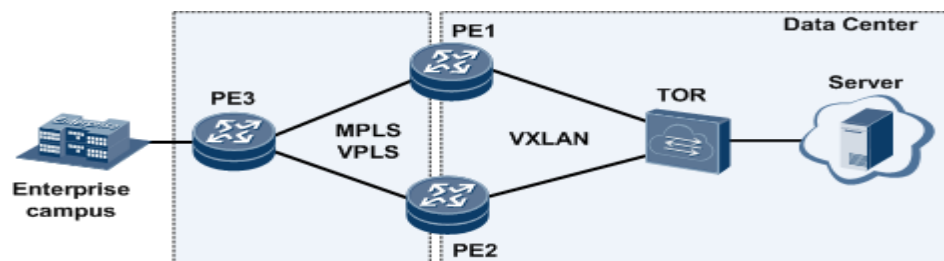


### Splicing VXLAN and VPLS:

When a DC with an EVPN VXLAN deployed interconnects to an enterprise campus through an MPLS L2VPN, splicing VXLAN and VPLS must be deployed.

On the network shown in [Figure 12-75](#), the TOR, which is a DC's gateway, accesses the backbone network through the egress routers PE1 and PE2 on the DC network. PE3, which is the egress router on the campus network, interconnects to PE1 and PE2 through the MPLS VPLS network. Splicing VXLAN and VPLS is configured on PE1 and PE2 to implement communication between the DC and campus network.

**Figure 12-75 Splicing VXLAN and VPLS**



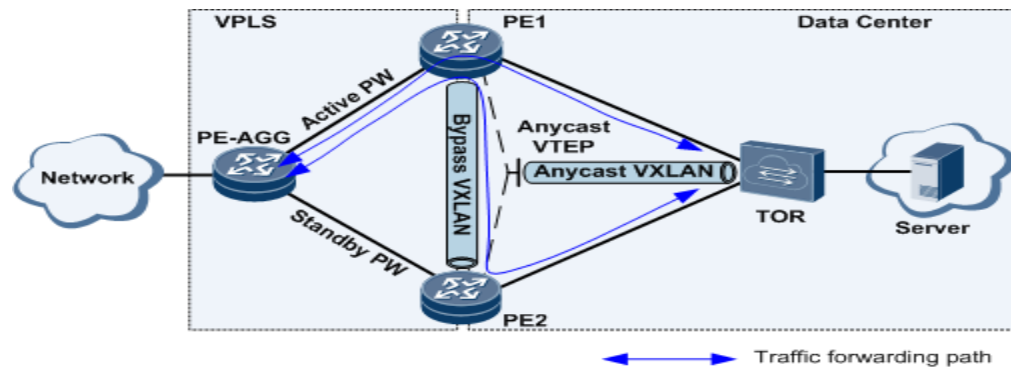
### Splicing Primary and Secondary PWs with an Anycast VXLAN Tunnel in an EVPN Active-Active Scenario:

On the network shown in [Figure 12-76](#), PE1 and PE2 are egress devices of the data center network. PE1 and PE2 work in active-active mode with a bypass VXLAN tunnel deployed between them. They use an anycast VTEP address to establish a VXLAN tunnel with the TOR. In this manner, PE1, PE2, and the TOR can communicate with each other. PE1 and PE2 communicate with the external network (an access network or the Internet) through the VPLS network. PW redundancy is deployed on the VPLS network. That is, the PE-AGG connects to PE1 and PE2 through primary and secondary PWs, respectively. In this example, the PW between the PE-AGG and PE1 is the primary PW.

Through the TOR, the server in the data center can send traffic to PE1 and PE2. Traffic received by PE1 is directly sent to the PE-AGG through the primary PW. Traffic received by PE2 is

forwarded to PE1 through the bypass VXLAN tunnel and then sent to the PE-AGG through the primary PW. Traffic from the PE-AGG to the server is transmitted along the reverse paths.

**Figure 12-76 Splicing primary and secondary PWs with an anycast VXLAN tunnel in an EVPN active-active scenario**



### **Splicing VPLS and MPLS EVPN:**

VPLS has inherent defects, such as a lack of support in load balancing and heavy consumption of network resources (MAC learning and ARP learning require packet broadcast on the entire network). As EVPN becomes widely used, VPLS networks are gradually evolving to EVPNs. However, such evolution cannot be implemented at a time due to complex network environments. Specifically, some devices may be deployed with VPLS and some other devices are deployed with EVPN. In this case, the function of VPLS splicing with MPLS EVPN can be deployed to ensure interworking on the entire network.

As shown in [Figure 12-77](#), VPLS is deployed between CSGs and ASGs, and CSGs are connected to ASGs through the primary and secondary PWs. EVPN is deployed between ASGs and RSGs. On ASG1 and ASG2, a BD is configured and bound to a VSI and an EVPN instance. In this manner, all PWs in the VSI can be connected to the EVPN through BDs. On the CSG dual-homed to ASG1 and ASG2, the same ESI is configured for the primary and secondary PW interfaces. The procedure for traffic forwarding is as follows:

1. Because an ESI is configured on ASG1's PW interface and the PW is in the Up state, ASG1 sends Ethernet A-D routes to the RSG.
2. After the Layer 2 packets sent by Site 1 reach ASG1 through the CSG, ASG1 generates MAC routes for the EVPN based on the MAC address of Site 1 in the Layer 2 packets. Such MAC routes are sent to the RSG based on the BGP EVPN peer relationship, and the RSG generates MAC forwarding entries based on the received Ethernet A-D routes. Similarly, the RSG sends MAC routes that carry the MAC address of Site 2 to ASGs and generate the corresponding MAC forwarding entries.

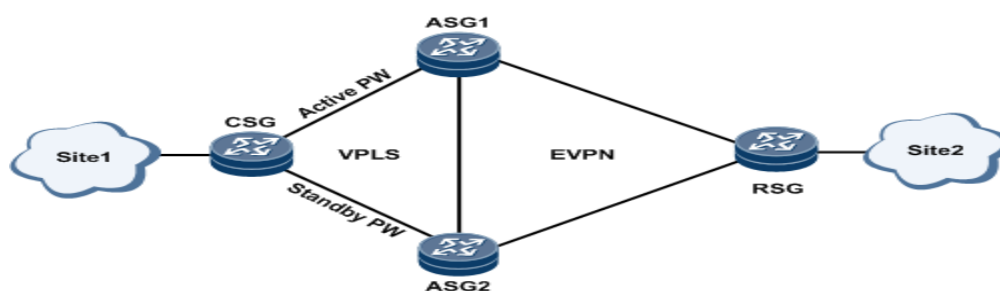
3. After the forwarding entries are successfully set up, these entries can guide through the forwarding of unicast traffic and BUM traffic. Taking the unicast traffic sent from Site 1 to Site 2 as an example, upon receipt of the traffic from the primary PW, ASG1 forwards the traffic to the RSG based on the MAC routes sent by the RSG. The RSG then forwards the traffic to Site 2.

**NOTE:**

Although ASG1 and ASG2 transmit Ethernet Segment routes to each other, DF election between ASG1 and ASG2 is implemented based on the PW status. The device (ASG1) connected to the primary PW is the primary DF, and the device (ASG2) connected to the secondary PW is the backup DF.

In BUM traffic forwarding scenarios, because the network is deployed with split horizon and the backup DF blocks traffic, loops or extra packets do not occur on the network.

**Figure 12-77 Networking of VPLS splicing with MPLS EVPN**



- **Seamless Migration of VPLS to EVPN:**

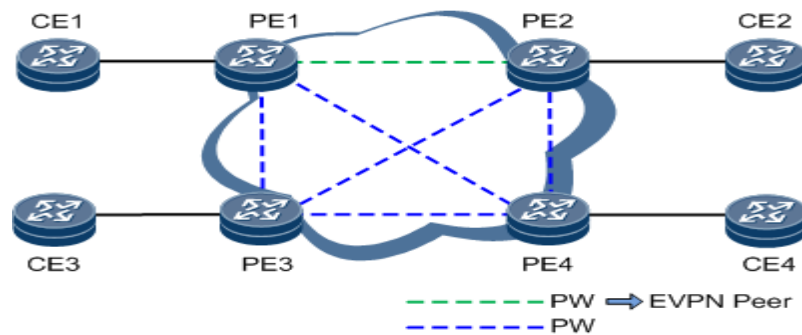
To convert each VPLS PE on a network into an EVPN device, you can configure the seamless migration of VPLS to EVPN function, which does not interrupt services during network running.

On the network shown in Figure 12-78, seamless migration of VPLS to EVPN involves the following process:

1. After EVPN is enabled on PE1, PE1 starts to advertise inclusive multicast routes to the other PEs. Because PE1 does not receive any inclusive multicast routes from the other PEs, traffic between PE1 and the other PEs continues to be forwarded through VPLS connections.
2. When EVPN continues to be enabled on another PE, for example PE2, PE2 starts to send inclusive multicast routes to the remaining EVPN-disabled PEs.
3. After PE1 and PE2 receive inclusive multicast routes from each other, they discover each other and disable the VPLS connection between them. The service between PE1 and PE2 is carried through an EVPN. Simultaneously, services between PE1/PE2 and the other PEs remain carried through the VPLS connections.

- The preceding process continues on the EVPN-incapable PEs one after another, implementing seamless migration of VPLS to EVPN.

**Figure 12-78 Seamless migration of VPLS to EVPN**



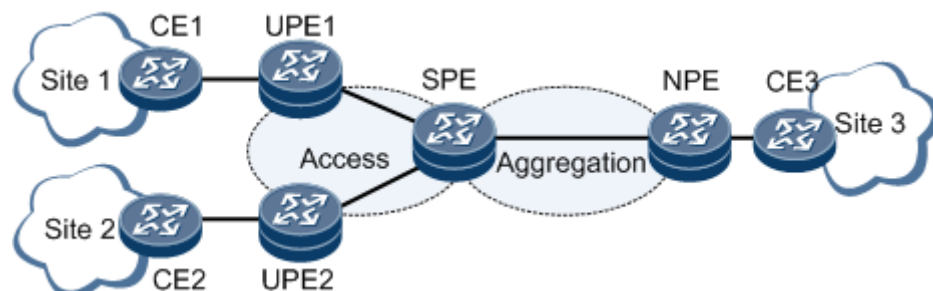
### EVPN L3VPN HVPN:

At present, the IP bearer network uses L2VPN and L3VPN (HVPN) to carry Layer 2 and Layer 3 services, respectively. The protocols are complex. EVPN can carry both Layer 2 and Layer 3 services. To simplify service bearer protocols, many IP bearer networks will evolve to EVPN. Specifically, L3VPN HVPN, which carries Layer 3 services, needs to evolve to EVPN L3VPN HVPN.

Figure 12-79 shows the basic architecture of an EVPN L3VPN HVPN consisting of mainly UPEs, SPE, and NPE:

- UPE: A UPE is a device that is directly connected to a user and is referred to as an underlayer PE or a user-end PE, therefore shortened as UPE. UPEs provide access services for users.
- SPE: An SPE is a superstratum PE or service provider-end PE, which is connected to UPEs and located at the core of a network. An SPE manages and advertises VPN routes.
- NPE: An NPE is a network provider-end PE that is connected to SPEs and located at the network side.

**Figure 12-79 Basic EVPN L3VPN HVPN architecture**



EVPN L3VPN HVPN is classified into EVPN L3VPN HoVPN or EVPN L3VPN H-VPN:



- **EVPN L3VPN HoVPN:** An SPE advertises only default routes or summarized routes to UPEs. UPEs do not have specific routes to NPEs and can only send service data to SPEs over default routes. As a result, route isolation is implemented. An EVPN L3VPN HoVPN can use devices with relatively poor route management capabilities as UPEs, reducing network deployment costs.
- **EVPN L3VPN H-VPN:** SPEs advertise specific routes to UPEs. UPEs function as RR clients to receive the specific routes reflected by SPEs functioning as RRs. This mechanism facilitates route management and traffic forwarding control.

As L3VPN HoVPN evolves towards EVPN L3VPN HoVPN, the following splicing scenarios occur:

- **Splicing between EVPN L3VPN HoVPN and common L3VPN:** EVPN L3VPN HoVPN is deployed between the UPEs and SPE, and L3VPN is deployed between the SPE and NPE. The SPE advertises only default routes or summarized routes to the UPEs. After receiving specific routes (EVPN routes) from the UPEs, the SPE encapsulates these routes into VPNv4 routes and advertises them to the NPE.
- **Splicing between L3VPN HoVPN and BD EVPN L3VPN:** L3VPN HoVPN is deployed between the UPEs and SPE, and BD EVPN L3VPN is deployed between the SPE and NPE. The SPE advertises only default routes or summarized routes to the UPEs. After receiving specific routes (L3VPN routes) from the UPEs, the SPE encapsulates these routes into EVPN routes and advertises them to the NPE.

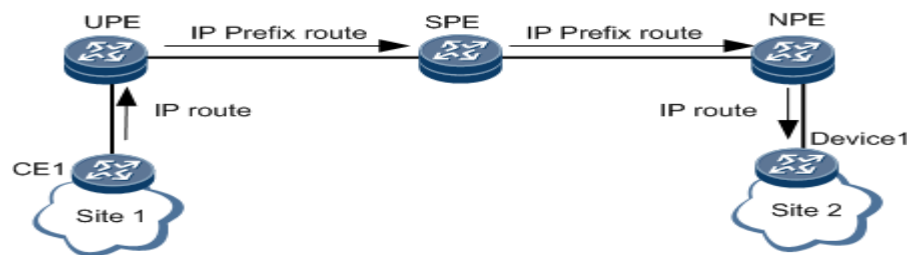
***Route Advertisement from CE1 to Device 1 on an EVPN L3VPN HoVPN or EVPN L3VPN H-VPN:***

Figure 12-80 shows route advertisement from CE1 to Device 1 on an EVPN L3VPN HoVPN or EVPN L3VPN H-VPN.

1. CE1 advertises an IPv4 route to the UPE using the IP protocol.
2. The UPE converts the IPv4 route into an IP prefix route with the next hop being the UPE and then sends the IP prefix route to the SPE through a BGP-EVPN peer relationship.
3. Upon receipt, the SPE advertises this route to the NPE in either of the following ways:
  - **Using RR:** Configure the SPE as an RR so that the RR directly reflects the received IP prefix route to the NPE, and change the next hop of the route to the SPE. An EVPN L3VPN H-VPN supports only this mode.
  - **Using re-encapsulation:** The SPE re-encapsulates the IP prefix route into a new IP prefix route with the next hop being the SPE. Then the SPE advertises the new route to the NPE through a BGP-EVPN peer relationship.
4. After receiving the IP prefix route, the NPE imports the route into its VRF table under the condition that the route's next hop is reachable.
5. The NPE advertises the IPv4 route to Device 1 using the IP protocol.



**Figure 12-80 Route advertisement from CE1 to Device 1 on an EVPN L3VPN HoVPN or EVPN L3VPN H-VPN**

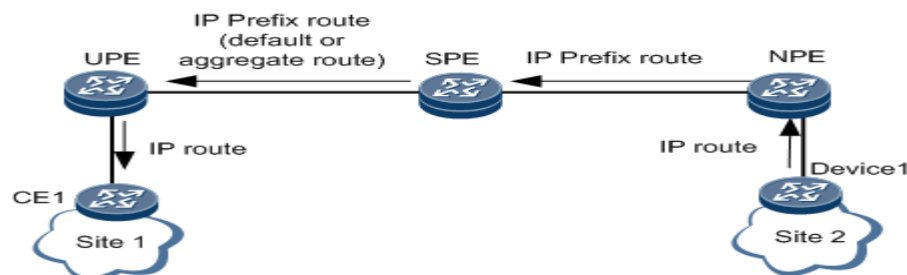


**Route Advertisement from Device 1 to CE1 on an EVPN L3VPN HoVPN:**

Figure 12-81 shows route advertisement from Device 1 to CE1 on an EVPN L3VPN HoVPN.

1. Device 1 advertises an IPv4 route to the NPE using the IP protocol.
2. The NPE converts the IPv4 route into an IP prefix route with the next hop being the NPE and then sends it to the SPE.
3. Upon receipt, the SPE converts the IP prefix route into an IPv4 route and imports it into its VRF table under the condition that the route's next hop is reachable.
4. The SPE imports a default route or summarized route into its VRF table, converts the default or summarized route into an IP prefix route with the next hop being the SPE, and then advertises the IP prefix route to the UPE.
5. Upon receipt, the UPE converts the IP prefix route into an IPv4 route and imports it into its VRF table under the condition that the route's next hop is reachable.
6. The UPE advertises the IPv4 route to CE1 using the IP protocol.

**Figure 12-81 Route advertisement from Device 1 to CE1 on an EVPN L3VPN HoVPN**

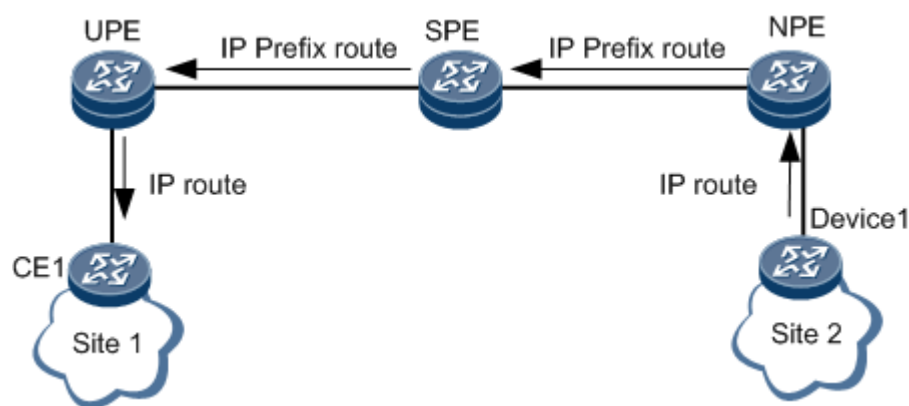


**Route Advertisement from Device 1 to CE1 on an EVPN L3VPN H-VPN:**

Figure 12-82 shows route advertisement from Device 1 to CE1 on an EVPN L3VPN H-VPN.

1. Device 1 advertises an IPv4 route to the NPE using the IP protocol.
2. The NPE converts the IPv4 route into an IP prefix route with the next hop being the NPE and then sends it to the SPE.
3. Upon receipt, the RR-enabled SPE advertises the IP prefix route to the UPE, and the route's next hop is changed to the SPE.
4. Upon receipt, the UPE converts the IP prefix route into an IPv4 route and imports it into its VRF table under the condition that the route's next hop is reachable.
5. The UPE advertises the IPv4 route to CE1 using the IP protocol.

**Figure 12-82 Route advertisement from Device 1 to CE1 on an EVPN L3VPN H-VPN**



***Route Advertisement from Device 1 to CE1 on an EVPN L3VPN HoVPN or EVPN L3VPN H-VPN:***

Packet forwarding from Device 1 to CE1 on an EVPN L3VPN HoVPN or EVPN L3VPN H-VPN is as follows:

1. Device 1 sends a VPN packet to the NPE.
2. After receiving the packet, the NPE searches its VPN forwarding table for a tunnel to forward the packet based on the destination address of the packet. Then, the NPE adds a VPN label (inner) and a tunnel label (outer) to the packet and sends the packet to the SPE over the found tunnel.
3. Upon receipt, the SPE removes the outer tunnel label, replaces the inner VPN label with a new one, and then adds the outer tunnel label to the packet. Then, the SPE forwards the packet to the UPE through the tunnel.
4. After receiving the packet, the UPE removes the outer tunnel label and searches for a VPN instance corresponding to the packet based on the inner VPN label. Then, the UPE searches the forwarding table of the found VPN instance for the outbound interface of the packet based on the destination address of the packet. The UPE sends the packet from the

corresponding outbound interface to CE1. The packet sent by the UPE is a pure IP packet with no label.

***Packet Forwarding from CE1 to Device 1 on an EVPN L3VPN HoVPN:***

Packet forwarding from CE1 to Device 1 on an EVPN L3VPN HoVPN is as follows:

1. CE1 sends a VPN packet to the UPE.
2. After receiving the packet, the UPE searches its VPN forwarding table for a tunnel to forward the packet based on the destination address of the packet (the UPE does so by matching the destination address of the packet against the forwarding entry for the default route or summarized route). Then, the UPE adds a VPN label (inner) and a tunnel label (outer) to the packet and sends the packet to the SPE over the found tunnel.
3. Upon receipt, the SPE removes the outer tunnel label and finds the corresponding VPN instance based on the inner VPN label. The SPE then removes the inner VPN label, searches the forwarding table of the VPN instance for a tunnel to forward the packet based on the destination address of the packet. Then, the SPE adds a new VPN label (inner) and tunnel label (outer) to the packet and sends the packet to the NPE through the found tunnel.
4. After receiving the packet, the NPE removes the outer tunnel label and searches for a VPN instance corresponding to the packet based on the inner VPN label. Then, the NPE searches the forwarding table of the found VPN instance for the outbound interface of the packet based on the destination address of the packet. The NPE sends the packet from the corresponding outbound interface to Device 1. The packet sent by the NPE is a pure IP packet with no label.

***Packet Forwarding from CE1 to Device 1 on an EVPN L3VPN H-VPN:***

Packet forwarding from CE1 to Device 1 on an EVPN L3VPN H-VPN is as follows:

1. CE1 sends a VPN packet to the NPE.
2. After receiving the packet, the UPE searches its VPN forwarding table for a tunnel to forward the packet based on the destination address of the packet (the UPE does so by matching the destination address of the packet against the forwarding entry for the specific route received from the SPE). Then, the UPE adds a VPN label (inner) and a tunnel label (outer) to the packet and sends the packet to the SPE over the found tunnel.
3. Upon receipt, the SPE removes the outer tunnel label, replaces the inner VPN label with a new one, and then adds the outer tunnel label to the packet. Then, the SPE forwards the packet to the NPE through the tunnel.
4. After receiving the packet, the NPE removes the outer tunnel label and searches for a VPN instance corresponding to the packet based on the inner VPN label. Then, the NPE searches the forwarding table of the found VPN instance for the outbound interface of the packet based on the destination address of the packet. The NPE sends the packet from the

corresponding outbound interface to Device 1. The packet sent by the NPE is a pure IP packet with no label.

Route advertisement and packet forwarding in scenarios where EVPN L3VPN HoVPN and common L3VPN are spliced or L3VPN HoVPN and BD EVPN L3VPN are spliced differ from those processes on an EVPN L3VPN HoVPN or L3VPN HoVPN only in re-encapsulation of BGP VPNv4 or IP prefix routes on the SPE:

- Splicing between EVPN L3VPN HoVPN and common L3VPN: After receiving the IP prefix route carrying CE1's specific route from the UPE, the SPE re-encapsulates the IP prefix route into a BGP VPNv4 route and advertises it to the NPE.
- Splicing between L3VPN HoVPN and BD EVPN L3VPN: After receiving the BGP VPNv4 route carrying CE1's specific route from the UPE, the SPE re-encapsulates the BGP VPNv4 route into an IP prefix route and advertises it to the NPE.

<https://support.huawei.com/>