

HUAWEI NetEngine40E Universal Service Router Product Documentation

Product Version: V800R012C10

Library Version: 05

Date: 2021-10-15



For any question, please [contact us](#).

[Copyright © Huawei Technologies Co., Ltd. 2021. All rights reserved.](#)

VPN

Contents

- 1 [VPN](#)
 - 1.1 [About This Document](#)
 - 1.2 [VPN Basics Description](#)
 - 1.2.1 [Overview of VPN Basics](#)
 - 1.2.1.1 [Classification](#)
 - 1.2.1.2 [Architecture](#)
 - 1.2.1.3 [Typical Networking](#)
 - 1.2.2 [Understanding VPN Basics](#)
 - 1.2.2.1 [Tunneling](#)
 - 1.2.2.2 [Implementation Modes](#)
 - 1.2.2.3 [Features Related to VPN Implementation](#)
 - 1.3 [GRE Description](#)
 - 1.3.1 [Overview of GRE](#)
 - 1.3.2 [Understanding GRE](#)
 - 1.3.2.1 [GRE Fundamentals](#)
 - 1.3.2.2 [Keepalive Detection](#)
 - 1.3.2.3 [Security Mechanism](#)
 - 1.3.3 [Application Scenarios for GRE](#)
 - 1.3.3.1 [Enlarging the Operation Scope of the Network with Limited Hops](#)
 - 1.3.3.2 [Connecting Discontinuous Sub-networks to Establish a VPN](#)
 - 1.3.3.3 [CEs Connecting to the MPLS VPN over GRE Tunnels](#)
 - 1.3.3.4 [Application of GRE on an ERSPAN Network](#)
 - 1.3.4 [Appendix](#)
 - 1.4 [DSVPN Description](#)
 - 1.4.1 [Overview of DSVPN](#)
 - 1.4.2 [Understanding DSVPN](#)
 - 1.4.2.1 [Basic Concepts](#)
 - 1.4.2.2 [DSVPN Fundamentals](#)

- 1.4.2.3 [DSVPN NAT Traversal](#)
- 1.4.2.4 [DSVPN IPsec Protection](#)
- 1.4.2.5 [Dual Hubs in Active/Standy Mode](#)
- 1.4.3 [Application Scenarios for DSVPN](#)
 - 1.4.3.1 [DSVPN Deployment on a Small- or Medium-sized Network](#)
 - 1.4.3.2 [DSVPN Deployment on a Large-sized Network](#)
 - 1.4.3.3 [Deploying DSVPN in Hierarchical Hub Networking](#)
- 1.5 [L2TPv3 Description](#)
 - 1.5.1 [Overview of L2TPv3](#)
 - 1.5.2 [Understanding L2TPv3](#)
 - 1.5.2.1 [L2TPv3 Basic Concepts](#)
 - 1.5.2.2 [L2TPv3 Fundamentals](#)
 - 1.5.3 [Application Scenarios for L2TPv3](#)
 - 1.5.4 [Terminology for L2TPv3](#)
- 1.6 [Tunnel Management](#)
 - 1.6.1 [Overview of Tunnel Management](#)
 - 1.6.2 [Understanding Tunnel Management](#)
 - 1.6.2.1 [Tunnel Policy](#)
 - 1.6.2.2 [Tunnel Policy Selector](#)
- 1.7 [BGP/MPLS IP VPN Description](#)
 - 1.7.1 [Overview of BGP/MPLS IP VPN](#)
 - 1.7.2 [Understanding BGP/MPLS IP VPN](#)
 - 1.7.2.1 [Basic BGP/MPLS IP VPN Fundamentals](#)
 - 1.7.2.2 [Hub & Spoke](#)
 - 1.7.2.3 [MCE](#)
 - 1.7.2.4 [Inter-AS VPN](#)
 - 1.7.2.5 [Carrier's Carrier](#)
 - 1.7.2.6 [HVPN](#)
 - 1.7.2.7 [BGP/MPLS IP VPN Label Allocation Modes](#)
 - 1.7.2.8 [BGP SoQ](#)
 - 1.7.2.9 [Route Import Between VPN and Public Network](#)
 - 1.7.2.10 [VPN FRR](#)
 - 1.7.2.11 [VPN GR](#)
 - 1.7.2.12 [VPN NSR](#)
 - 1.7.2.13 [BGP/MPLS IPv6 VPN Extension](#)
 - 1.7.2.14 [VPN Dual-Stack Access](#)
 - 1.7.2.15 [VPN MPLS/VPN SRv6 Dual-Stack Tunnel](#)
- 1.7.3 [Application Scenarios for BGP/MPLS IP VPN](#)
 - 1.7.3.1 [Application of MCEs on a Campus Network](#)
 - 1.7.3.2 [Application of MCEs on a Data Center Network](#)
 - 1.7.3.3 [Application of HVPN on an IP RAN](#)
 - 1.7.3.4 [Application of Route Import Between VPN and Public Network in the Traffic Cleaning Networking](#)
- 1.8 [VPWS Description](#)
 - 1.8.1 [Overview of VPWS](#)
 - 1.8.2 [Understanding VPWS](#)
 - 1.8.2.1 [VPWS Basic Functions](#)
 - 1.8.2.2 [VPWS in CCC Mode](#)
 - 1.8.2.3 [LDP VPWS](#)
 - 1.8.2.4 [VPWS in SVC Mode](#)
 - 1.8.2.5 [VPWS in BGP Mode](#)
 - 1.8.2.6 [Heterogeneous VPWS](#)
 - 1.8.2.7 [ATM Cell Relay](#)
 - 1.8.2.8 [VCCV](#)
 - 1.8.2.9 [PW Redundancy](#)

- 1.8.2.10 [PW APS](#)
- 1.8.2.11 [Comparison of VPWS Implementation Modes](#)
- 1.8.2.12 [Comparison of LDP VPWS and BGP/MPLS IP VPN](#)
- 1.8.2.13 [Inter-AS VPWS](#)
- 1.8.2.14 [Flow-Label-based Load Balancing](#)
- 1.8.2.15 [Mutual Protection Between an LDP VC and a CCC VC](#)
- 1.8.2.16 [Multi-Segment PW Redundancy](#)
- 1.8.3 [Application Scenarios for VPWS](#)
 - 1.8.3.1 [Enterprise Leased Line Service Bearer Using PWE3](#)
 - 1.8.3.2 [HSI Service Bearer Using PWE3](#)
 - 1.8.3.3 [PW APS Application](#)
- 1.9 [IP Hard Pipe Description](#)
 - 1.9.1 [Overview of IP Hard Pipe](#)
 - 1.9.2 [Understanding IP Hard Pipe](#)
 - 1.9.2.1 [Centralized Management of IP Hard-Pipe-based Leased Line Services on the NMS](#)
 - 1.9.2.2 [Interface-based Hard Pipe Bandwidth Reservation](#)
 - 1.9.2.3 [AC Interface Service Bandwidth Limitation](#)
 - 1.9.2.4 [Hard-Pipe-based TE LSP](#)
 - 1.9.2.5 [Hard Pipe-based VPWS/VPLS](#)
 - 1.9.2.6 [Hard Pipe Reliability](#)
 - 1.9.2.7 [Hard Pipe Service Quality Monitoring](#)
 - 1.9.3 [Application Scenarios for IP Hard Pipe](#)
 - 1.9.3.1 [Hard-Pipe-based Enterprise Leased Line Application](#)
 - 1.9.3.2 [Hard-Pipe-based Enterprise Leased Line Protection](#)
 - 1.9.3.3 [Hard-Pipe-based Leased Line Services Implemented by Huawei and Non-Huawei Devices](#)
 - 1.9.4 [Terminology for IP Hard Pipe](#)
- 1.10 [VPLS Description](#)
 - 1.10.1 [Overview of VPLS](#)
 - 1.10.2 [Understanding VPLS](#)
 - 1.10.2.1 [VPLS Description](#)
 - 1.10.2.2 [VPLS Functions](#)
 - 1.10.2.3 [LDP VPLS](#)
 - 1.10.2.4 [BGP VPLS](#)
 - 1.10.2.5 [HVPLS](#)
 - 1.10.2.6 [BGP AD VPLS](#)
 - 1.10.2.7 [Inter-AS VPLS](#)
 - 1.10.2.8 [Flow-Label-based Load Balancing](#)
 - 1.10.2.9 [VPLS PW Redundancy](#)
 - 1.10.2.10 [Multicast VPLS](#)
 - 1.10.2.11 [VPLS Multi-homing](#)
 - 1.10.2.12 [VPLS Service Isolation](#)
 - 1.10.3 [Application Scenarios for VPLS](#)
 - 1.10.3.1 [Application of VPLS in Residential Services](#)
 - 1.10.3.2 [Application of VPLS in Enterprise Services](#)
 - 1.10.3.3 [VPLS PW Redundancy for Protecting Multicast Services](#)
 - 1.10.3.4 [VPLS PW Redundancy for Protecting Unicast Services](#)
 - 1.10.3.5 [Application of Multicast VPLS](#)
 - 1.10.3.6 [VPWS Accessing VPLS](#)
 - 1.10.3.7 [VPLS Multi-Homing Application](#)
- 1.11 [L2VPN Accessing L3VPN Description](#)
 - 1.11.1 [Overview of L2VPN Accessing L3VPN](#)
 - 1.11.2 [Understanding L2VPN Accessing L3VPN](#)
 - 1.11.2.1 [L2VPN Accessing L3VPN Fundamentals](#)
 - 1.11.2.2 [Classification of L2VPN Accessing L3VPN](#)

- 1.11.3 [Application Scenarios for L2VPN Accessing L3VPN](#)
- 1.11.3.1 [VPWS Accessing L3VPN](#)
- 1.11.3.2 [VPLS Accessing L3VPN](#)
- 1.11.4 [Terminology for L2VPN Accessing L3VPN](#)
- 1.12 [EVPN Feature Description](#)
- 1.12.1 [Overview of EVPN](#)
- 1.12.2 [EVPN Fundamentals](#)
- 1.12.3 [EVPN-MPLS](#)
- 1.12.3.1 [EVPN Multi-Homing](#)
- 1.12.3.2 [Fundamentals of EVPN Seamless MPLS](#)
- 1.12.3.3 [EVPN Service Modes](#)
- 1.12.4 [EVPN-VXLAN](#)
- 1.12.4.1 [EVPN VXLAN Fundamentals](#)
- 1.12.5 [EVPN VPWS](#)
- 1.12.5.1 [EVPN VPWS Fundamentals](#)
- 1.12.6 [PBB-EVPN](#)
- 1.12.6.1 [PBB-EVPN Fundamentals](#)
- 1.12.6.2 [Migration from an HVPLS Network to a PBB-EVPN](#)
- 1.12.7 [EVPN E-Tree](#)
- 1.12.8 [MAC Duplication Suppression for EVPN](#)
- 1.12.9 [EVPN ORF](#)
- 1.12.10 [IGMP Snooping over EVPN MPLS](#)
- 1.12.11 [Application Scenarios for EVPN](#)
- 1.12.11.1 [Using EVPN to Interconnect Other Networks](#)
- 1.12.11.2 [EVPN L3VPN HVPN](#)
- 1.12.11.3 [EVPN 6VPE](#)
- 1.12.11.4 [EVPN Interworking Scenarios](#)
- 1.12.11.5 [Inter-AS EVPN Option C](#)
- 1.12.11.6 [DCI Scenarios](#)
- 1.12.11.7 [NFVI Distributed Gateway \(SR Tunnels\)](#)
- 1.12.11.8 [NFVI Distributed Gateway Function \(BGP VPNv4/v6 over E2E SR Tunnels\)](#)
- 1.12.11.9 [NFVI Distributed Gateway Function \(BGP EVPN over E2E SR Tunnels\)](#)
- 1.12.11.10 [Application Scenarios for EVPN E-LAN Accessing L3VPN](#)
- 1.13 [PBB VPLS Description](#)
- 1.13.1 [Overview of PBB VPLS](#)
- 1.13.2 [Understanding PBB VPLS](#)
- 1.13.2.1 [PBB VPLS Fundamentals](#)
- 1.13.3 [Application Scenarios for PBB VPLS](#)
- 1.13.3.1 [PBB VPLS Application](#)
- 1.14 [Proactive Loop Detection Description](#)
- 1.14.1 [Overview of Proactive Loop Detection](#)
- 1.14.2 [Understanding Proactive Loop Detection](#)
- 1.14.2.1 [Proactive Loop Detection](#)
- 1.14.2.2 [Loop Detection Packet Format](#)
- 1.14.3 [Application Scenarios for Proactive Loop Detection](#)
- 1.14.3.1 [AC Interface Receiving a Loop Detection Packet](#)
- 1.14.3.2 [PW Side Receiving a Loop Detection Packet](#)

1 VPN

[VPN Basics Description](#)

This chapter describes the background, classification, networking, and fundamentals for implementing virtual private network (VPN) services.

[GRE Description](#)

[DSVPN Description](#)

[L2TPv3 Description](#)

[Tunnel Management](#)

[BGP/MPLS IP VPN Description](#)

[VPWS Description](#)

[IP Hard Pipe Description](#)

[VPLS Description](#)

[L2VPN Accessing L3VPN Description](#)

[EVPN Feature Description](#)

[PBB VPLS Description](#)

[Proactive Loop Detection Description](#)

This chapter describes the basic concepts, principles, and applications of proactive loop detection.

Parent Topic: [Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.1 About This Document

Purpose

This document describes the VPN feature in terms of its overview, principles, and applications.

Related Version

The following table lists the product version related to this document.

Product Name	Version
HUAWEI NetEngine40E	V800R012C10
iMaster NCE-IP	V100R020C00SPC101

Intended Audience

This document is intended for:

- Network planning engineers

- Commissioning engineers
- Data configuration engineers
- System maintenance engineers

Security Declaration

- Encryption algorithm declaration

The encryption algorithms DES/3DES/RSA (with a key length of less than 2048 bits)/MD5 (in digital signature scenarios and password encryption)/SHA1 (in digital signature scenarios) have a low security, which may bring security risks. If protocols allowed, using more secure encryption algorithms, such as AES/RSA (with a key length of at least 2048 bits)/SHA2/HMAC-SHA2 is recommended.

- Password configuration declaration

- Do not set both the start and end characters of a password to "%^%#". This causes the password to be displayed directly in the configuration file.
- To further improve device security, periodically change the password.

- Personal data declaration

- Your purchased products, services, or features may use users' some personal data during service operation or fault locating. You must define user privacy policies in compliance with local laws and take proper measures to fully protect personal data.
- When discarding, recycling, or reusing a device, back up or delete data on the device as required to prevent data leakage. If you need support, contact after-sales technical support personnel.

- Feature declaration

- The NetStream feature may be used to analyze the communication information of terminal customers for network traffic statistics and management purposes. Before enabling the NetStream feature, ensure that it is performed within the boundaries permitted by applicable laws and regulations. Effective measures must be taken to ensure that information is securely protected.
- The mirroring feature may be used to analyze the communication information of terminal customers for a maintenance purpose. Before enabling the mirroring function, ensure that it is performed within the boundaries permitted by applicable laws and regulations. Effective measures must be taken to ensure that information is securely protected.
- The packet header obtaining feature may be used to collect or store some communication information about specific customers for transmission fault and error detection purposes. Huawei cannot offer services to collect or store this information unilaterally. Before enabling the function, ensure that it is performed within the boundaries permitted by applicable laws and regulations. Effective measures must be taken to ensure that information is securely protected.

- Reliability design declaration

Network planning and site design must comply with reliability design principles and provide device- and solution-level protection. Device-level protection includes planning principles of dual-network and inter-board dual-link to avoid single point or single link of failure.

Solution-level protection refers to a fast convergence mechanism, such as FRR and VRRP. If

solution-level protection is used, ensure that the primary and backup paths do not share links or transmission devices. Otherwise, solution-level protection may fail to take effect.

Special Declaration

- This document serves only as a guide. The content is written based on device information gathered under lab conditions. The content provided by this document is intended to be taken as general guidance, and does not cover all scenarios. The content provided by this document may be different from the information on user device interfaces due to factors such as version upgrades and differences in device models, board restrictions, and configuration files. The actual user device information takes precedence over the content provided by this document. The preceding differences are beyond the scope of this document.
- The maximum values provided in this document are obtained in specific lab environments (for example, only a certain type of board or protocol is configured on a tested device). The actually obtained maximum values may be different from the maximum values provided in this document due to factors such as differences in hardware configurations and carried services.
- Interface numbers used in this document are examples. Use the existing interface numbers on devices for configuration.
- The pictures of hardware in this document are for reference only.
- The supported boards are described in the document. Whether a customization requirement can be met is subject to the information provided at the pre-sales interface.
- In this document, public IP addresses may be used in feature introduction and configuration examples and are for reference only unless otherwise specified.
- The configuration precautions described in this document may not accurately reflect all scenarios.

Symbol Conventions

The symbols that may be found in this document are defined as follows.

Symbol	Description
 DANGER	Indicates a hazard with a high level of risk which, if not avoided, will result in death or serious injury.
 WARNING	Indicates a hazard with a medium level of risk which, if not avoided, could result in death or serious injury.
 CAUTION	Indicates a hazard with a low level of risk which, if not avoided, could result in minor or moderate injury.
 NOTICE	Indicates a potentially hazardous situation which, if not avoided, could result in equipment damage, data loss, performance deterioration, or unanticipated results. NOTICE is used to address practices not related to personal injury.
 NOTE	Supplements the important information in the main text. NOTE is used to address information not related to personal injury, equipment damage, and environment deterioration.

Change History

Changes between document issues are cumulative. The latest document issue contains all the changes made in earlier issues.

- **Changes in Issue 05 (2021-10-15)**

This issue is the fifth official release. The software version of this issue is V800R012C10SPC300.

- **Changes in Issue 04 (2020-12-10)**

This issue is the fourth official release. The software version of this issue is V800R012C10SPC300.

- **Changes in Issue 03 (2020-10-31)**

This issue is the third official release. The software version of this issue is V800R012C10SPC300.

- **Changes in Issue 02 (2020-08-31)**

This issue is the second official release. The software version of this issue is V800R012C10SPC100.

- **Changes in Issue 01 (2020-05-30)**

This issue is the first official release. The software version of this issue is V800R012C10.

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.2 VPN Basics Description

This chapter describes the background, classification, networking, and fundamentals for implementing virtual private network (VPN) services.

[Overview of VPN Basics](#)

[Understanding VPN Basics](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

< Previous topic [Next topic >](#)

1.2.1 Overview of VPN Basics

Definition

A virtual private network (VPN) is a virtual private communication network established on a public network, with the help of an Internet service provider (ISP) and a network service provider (NSP).

Background

An increasing number of information technologies are applied to modern enterprise businesses. For example, IP technologies are applied to enterprise resource programming, Voice over Internet Protocol (VoIP), video conference, and remote training. IP technologies allow an enterprise to achieve office automation and access information more easily. As the Internet economy develops, enterprises expand into new locations, cooperate with more and more partners, and require greater office mobility. Enterprises of this nature, therefore, need to interconnect their headquarters and branches with the help of carrier networks to form enterprise networks, so that staff can conveniently access the enterprise networks outside office buildings.

During the initial stages of information technologies, telecom carriers used leased lines to provide Layer 2 connections for enterprises. The disadvantages of leased lines are as follows:

- Constructing leased lines takes a long period.
- Leased lines require huge investments.
- Leased lines are difficult to manage.

After the emergence of Asynchronous Transfer Mode (ATM) and Frame Relay (FR) technologies, telecom carriers began to use virtual circuits (VCs) to provide point-to-point (P2P) Layer 2 connections for clients. Clients can set up Layer 3 networks and transmit IP data over the P2P Layer 2 connections. Compared with leased lines, VCs are less expensive and can be constructed within a short period. In addition, VCs enable users of different private networks to share the same carrier's network.

Despite their advantages over leased lines, VCs also have their disadvantages:

- VCs are dependent on media such as ATM or FR. To provide VPN services based on ATM or FR, carriers must construct ATM networks covering all service areas. This implementation results in heavy capital expense.
- The speed of ATM or FR networks is lower than that required by the Internet.
- The deployment of ATM or FR networks is complex. To add a site to an existing ATM or FR network, you must modify the configurations of the edge nodes that connect to the site.

Traditional private networks help to boost enterprise profits, but do not meet the requirements for flexibility, security, economy, and scalability. To solve these problems, VPNs, emulated private networks carried over IP networks, have been introduced as a substitution to traditional private networks.

VPNs are virtual communication channels set up over public networks by Internet service providers (ISPs) or network service providers (NSPs).

Characteristics

A VPN has the following characteristics:

- Privacy

VPNs and traditional private networks make no difference to users in terms of privacy. VPN resources are separated from bearer network resources and are exclusive to VPN users. In addition, VPNs offer sufficient security measures to protect internal information against external interference.

- Virtuality

VPN users communicate with each other over public networks, which are used by non-VPN users at the same time. A VPN is only a logical private network. A public network that carries a VPN is called a VPN backbone network.

The VPN technology can flexibly segment an existing IP network into several logically isolated networks. This feature allows an enterprise to flexibly interconnect or isolate different departments or branches. This feature also facilitates service provisioning. For example, creating a VPN for the IP phone service can solve the problem of inadequate IP addresses although; whereas guaranteeing quality of service (QoS).

VPNs, especially Multiprotocol Label Switching (MPLS) VPNs, are highly valued by carriers in terms of providing interworking between enterprises and providing other enhanced services. VPNs have, as never before, become an important means for carriers to provide value-added services (VASs) over IP networks.

Benefits

VPNs offer the following benefits to users:

- Guaranteed data security

A VPN provides reliable connections between remote users, branches, business partners, suppliers, and company headquarters to ensure data transmission security. High security is becoming increasingly important as e-business and financial networks converge with communication networks.

- High cost-effectiveness

An enterprise can connect its headquarters with branches, personnel on business, and business partners over public networks at low costs.

- Increased office mobility

Enterprise employees can access the enterprise network from anywhere and at any time, meeting the increasing demand for office mobility.

- QoS guarantee

A QoS-capable VPN, such as an MPLS VPN, can provide users with different levels of QoS guarantee.

VPNs offer the following benefits to carriers:

- Easy operation

VPNs increase carriers' profits by improving resource utilization.

- Flexible configuration

Carriers can add or delete VPN users by means of software configurations without hardware modifications.

- Diversified services

In addition to basic VPN interworking services, carriers can also provide enhanced services, such as network outsourcing, service outsourcing, and customized services.

VPNs allow enterprises to direct less attention to network operation and maintenance and more attention to the achievement of their business goals. This feature enables VPNs to be increasingly popular with enterprises. A carrier can provide multiple types of services, such as best-effort IP services, VPNs, traffic engineering, and differentiated services (DSs), over only one network, reducing network construction, maintenance, and operation costs.

VPNs improve the scalability and flexibility of networks in addition to providing security, reliability, and manageability. Users can enjoy VPN services provided that; if they have Internet access, regardless of their location.

[Classification](#)

[Architecture](#)

[Typical Networking](#)

Parent Topic: [VPN Basics Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.2.1.1 Classification

With the development of network technologies, the VPN technology is widely applied and many new VPN technologies emerge. VPNs can be divided into different types.

Classification Based on Applications

VPNs are divided into the following types based on applications:

- Intranet VPN

An intranet VPN connects the headquarters, branches, regional offices, and mobile personnel of an enterprise over public networks. Intranet VPNs are the extension to or substitute for traditional private networks or other enterprise networks.

Intranet VPNs can be used by banks and governments to construct their intranets.

Chain businesses, such as chain stores, storage and logistics companies, and gas station chains, are typical examples of enterprises using intranet VPNs.

- Extranet VPN

An extranet VPN extends selected resources and applications from an enterprise network to users outside the enterprise, such as suppliers, business partners, and clients. The extranet VPN is established between enterprises with common interests over public networks.

An extranet established with traditional leased lines requires complex network management and access control, or even the installation of compatible user-side network devices.

Although an extranet can be established in dialing mode, different extranet users must be configured respectively. In addition, an extranet in dialing mode is expensive to construct and maintain, especially if the business partners and customers are scattered far and wide. As a result, many enterprises have given up on extranets, which leads to complex and inefficient business processes between enterprises.

Extranet VPNs are a solution to the problems of extranets. Similar to intranet VPNs in terms of technical implementation, extranet VPNs are easy to construct and manage. Currently, enterprises generally use VPNs to construct extranets. Extranet VPNs provide better QoS guarantee and higher data transmission security than the Internet. In addition, the extranet VPN owner can configure the access rights of extranet VPN users using firewalls or by other means.

Parent Topic: [Overview of VPN Basics](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.2.1.2 Architecture

The VPN technology is much more complex than the P2P technology. VPN implementation requires construction of network connections between users, which includes network topology planning, route calculation, and maintenance of VPN users joining or leaving. The VPN architecture comprises the following parts:

- VPN tunnels
 - Establishment of tunnels
 - Management of tunnels
- VPN management
 - VPN configuration management
 - VPN member management
 - VPN attribute management: management of attributes of multiple VPNs on provider edges (PEs) and differentiation of VPN address spaces
- VPN signaling protocol
 - Exchange and share of VPN resources between customer edges (CEs) on a VPN
 - VPN member discovery in some applications

Parent Topic: [Overview of VPN Basics](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.2.1.3 Typical Networking

A typical VPN has the following layers:

- Access layer

The devices on the access layer provide access services for users. These devices do not need to implement many functions, but must provide many access interfaces. For metropolitan area networks (MANs) in big cities, the access layer needs to provide more functions besides the access function.

Generally, a CE is dual-homed or multi-homed to access nodes on the access layer. Dual homing can be either physical or logical. In physical dual homing, a CE accesses two nodes over two physical links; in logical dual homing, a CE accesses two nodes that reside on a ring.

- Convergence layer

The convergence layer has either a mesh topology or a ring topology.

- Backbone layer

The backbone layer must have a full-mesh topology and multi-level backup. The devices on the backbone layer are generally connected through high-speed interfaces.

Parent Topic: [Overview of VPN Basics](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.2.2 Understanding VPN Basics

[Tunneling](#)

[Implementation Modes](#)

[Features Related to VPN Implementation](#)

Parent Topic: [VPN Basics Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.2.2.1 Tunneling

The VPN technology is based on the idea of tunneling. Packets constructed in a specific protocol format can be encapsulated with carrier protocol headers and transparently transmitted over tunnels on the VPN backbone network.

The tunneling technology uses one protocol to encapsulate the packets of another protocol, and the carrier protocol itself can be encapsulated or carried by other protocols. From the perspective of a user, a tunnel is a logical extension of a public switched telephone network (PSTN) or integrated services digital network (ISDN) link and functions in the same way as a physical link.

A VPN tunnel provides the following functions:

- Encapsulates user data.
- Establishes a link between two endpoints.
- Periodically checks link connectivity.
- Guarantees data transmission security.
- Provides QoS guarantee.

Parent Topic: [Understanding VPN Basics](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.2.2.2 Implementation Modes

The VPN technology can be implemented in the following modes.

Tunneling + VPN Management

In this mode, the VPN architecture comprises the following parts:

- VPN tunnels: establishment of tunnels
- VPN management
 - Deployment of network management
 - Accounting
 - QoS

Tunneling + VPN Management + VPN Signaling Protocol

In this mode, the VPN architecture comprises the following parts:

- VPN tunnels: establishment of tunnels
- VPN management
 - VPN configuration management
 - VPN member management
 - VPN attribute management
 - VPN automatic configuration
- VPN signaling protocol: exchange and share of VPN resources between CEs on a VPN

Instantiation

In instantiation mode, each VPN on Layer 2 and Layer 3 is instantiated, and instances of private forwarding information of each VPN are established. Besides tunnel management, an instantiated VPN also performs member discovery, member management, and automatic configuration.

This mode is adopted by L3VPNs based on the relevant standards.

NOTE

This chapter briefly describes VPN implementation. For more information, see the description of related features in other chapters of this document.

Parent Topic: [Understanding VPN Basics](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.2.2.3 Features Related to VPN Implementation

Operability

The VPN technology is generally used to share services between different departments of an enterprise over public networks. Nowadays, VPN users want to spend less time and energy on network maintenance, and require carriers to do this task. Therefore, when designing a VPN, consider network operability first.

Manageability

VPNs allow enterprises to seamlessly extend their network management from LANs to public networks, even to clients and business partners. After delegating nonessential network management tasks to the carrier, enterprises still need to fulfill many network management tasks. A complete VPN management system is absolutely necessary.

VPN management includes security management, equipment management, configuration management, access control list (ACL) management, and QoS management.

VPN management offers the following benefits:

- Reduced network risks

After an intranet is extended to a public network using the VPN technology, the intranet faces new security risks and monitoring challenges. VPN management can guarantee the integrity of data resources on an intranet although; whereas allowing branches, clients, and business partners to access the intranet.

- Increased scalability

VPN management can quickly adapt to the increased numbers of clients and partners, such as upgrading network hardware and software, guaranteeing network quality, and maintaining security policies.

- Improved cost-effectiveness

VPN management can control operation and maintenance expenses although; whereas ensuring service scalability.

- Enhanced reliability

VPNs are established over public networks. Compared with traditional wide area networks (WANs) established using leased lines, VPNs have lower controllability. VPN management must be performed to guarantee network stability and reliability.

Security

VPN implementation is simple, convenient, and flexible. However, network risks arise at the same time.

- A traditional IP VPN faces serious risks, such as data obtaining, data tampering, and access of unauthorized users. Extranet VPNs face even more serious risks.

The following solutions help to improve VPN security:

- Tunneling and tunnel encryption

The tunneling technology uses multi-protocol encapsulation to enhance VPN flexibility and provide P2P logical channels on connectionless IP networks. Tunnel encryption helps to protect data privacy and ensure that data is not illegally obtained or tampered with.

- Data authentication

On an insecure network, such as the public network used by a VPN, packets may be illegally obtained and tampered with. As a result, the receiver may receive incorrect packets. Data authentication helps receivers to determine the integrity and authenticity of received data.

- User authentication

User authentication allows a VPN to permit the access of authorized users and deny the access of unauthorized users. Authentication, Authorization and Accounting (AAA)-capable routers can authenticate users, authorize users for specific resources, and generate access records. User authentication greatly improves the security of access VPNs and extranet VPNs.

- Firewalls and attack detection

Firewalls help to filter packets and prevent unauthorized access. Attack detection helps to determine the validity of packets, implement security policies in real time, disconnect unauthorized sessions, and record unauthorized access.

 **NOTE**

For more information about tunnel encryption, data authentication, user authentication, firewalls, and attack detection, see the HUAWEI NetEngine40EUUniversal Service Router*Feature Description - Security*.

- MPLS VPNs are created on the basis of labels and forwarding tables on network side. If an MPLS network does not connect to the Internet, internal resources on the MPLS VPN are secure. MPLS VPNs can ensure data security to some extent.

If an MPLS VPN needs to access the Internet, a channel with a firewall can be established to provide a secure connection for the VPN. The MPLS VPN is easy to manage because only one security policy is used.

An MPLS VPN is a private network that has the same security level as an FR network. Generally, user devices do not need to be configured with Internet Protocol Security (IPsec) or tunnels. On an MPLS VPN, data transmission delay is low because packets do not need to be encapsulated or encrypted. A mesh VPN is easy to create if no tunnel configuration is required.

Parent Topic: [Understanding VPN Basics](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.3 GRE Description

[Overview of GRE](#)

[Understanding GRE](#)

[Application Scenarios for GRE](#)

[Appendix](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.3.1 Overview of GRE

Definition

NOTE

If IPv4 GRE and IPv6 GRE implement a feature in the same way, details are not provided in this chapter. For details about implementation differences, see [Appendices](#).

Generic Routing Encapsulation (GRE) is a tunneling protocol that encapsulates the packets of a wide variety of network layer protocols, such as Internetwork Packet Exchange (IPX), Asynchronous Transfer Mode (ATM), IPv6, and AppleTalk, into IP tunneling packets. Then these packets can be transmitted over an IPv4 network.

GRE provides a mechanism of encapsulating packets of a protocol into packets of another protocol. This allows packets to be transmitted over heterogeneous networks. The channel for transmitting heterogeneous packets is called a tunnel.

The following types of GRE tunnels are supported on NE40E:

- GRE tunnel with the one-dimensional tunnel interface: also called distributed GRE tunnel. The tunnel interface is one-dimensional (named only by the interface number). GRE packets are encapsulated and decapsulated directly on the inbound interface board. If the multi-field classification and CAR services are configured simultaneously, bandwidth that services consume may double.
- GRE tunnel with the three-dimensional tunnel interface: also called integrated GRE tunnel. The tunnel interface is three-dimensional (named by the slot ID, subcard ID, and interface number). GRE packets are encapsulated and decapsulated directly on a service processing board.

Purpose

To allow the packets of a wide variety of network layer protocols, such as IPX, ATM, IPv6, and AppleTalk, to be transmitted over the IPv4 network, GRE is introduced. GRE solves the transmission problem faced by heterogeneous networks.

In addition, GRE serves as a Layer 3 tunneling protocol of VPNs, and provides a tunnel for transparently transmitting VPN packets. Currently, GRE is supported by IPv4 L3VPN, but not IPv6 L3VPN.

Benefits

GRE has low requirements for device performance and allows devices that do not support Multiprotocol Label Switching (MPLS) to establish tunnels.

Parent Topic: [GRE Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.3.2 Understanding GRE

[GRE Fundamentals](#)

[Keepalive Detection](#)

Parent Topic: [GRE Description](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

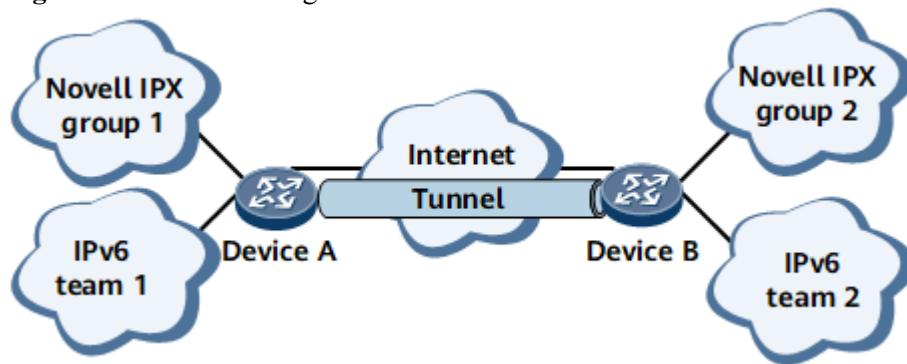
1.3.2.1 GRE Fundamentals

Background

A single network protocol, such as IPv4, is used to transmit packets on a backbone network, whereas other protocols, such as IPv6 and Internet Packet Exchange (IPX), are used to transmit packets on non-backbone networks. Because the backbone and non-backbone networks use different protocols, packets cannot be transmitted between the non-backbone networks over the backbone network. Generic Routing Encapsulation (GRE) resolves this issue by providing a mechanism of encapsulating the packets of a protocol into the packets of another protocol.

On the network shown in [Figure 1](#), groups 1 and 2 are the non-backbone networks running Novell IPX, and teams 1 and 2 are the non-backbone networks running IPv6. The backbone network is an IPv4 network. To transmit packets between groups 1 and 2 and between teams 1 and 2 over the backbone network, use GRE to establish a tunnel between Device A and Device B. When Device A receives a packet from group 1 or team 1, Device A encapsulates the packet into a GRE packet. The GRE packet is then encapsulated into an IPv4 packet for forwarding.

Figure 1 GRE networking

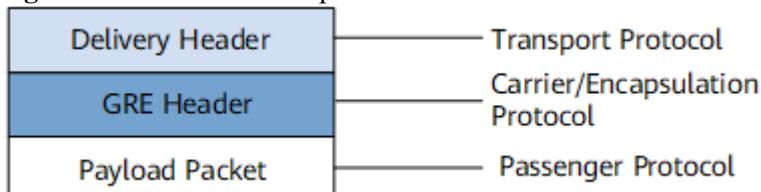


Related Concepts

- GRE packet format

After receiving a network layer protocol packet that needs to be encapsulated and routed, such as an IPX packet, the system adds a GRE header to the packet and encapsulates the packet into another protocol, such as IP. Then, the IP protocol is responsible for forwarding the packet. [Figure 2](#) shows the format of a GRE packet.

Figure 2 Format of a GRE packet

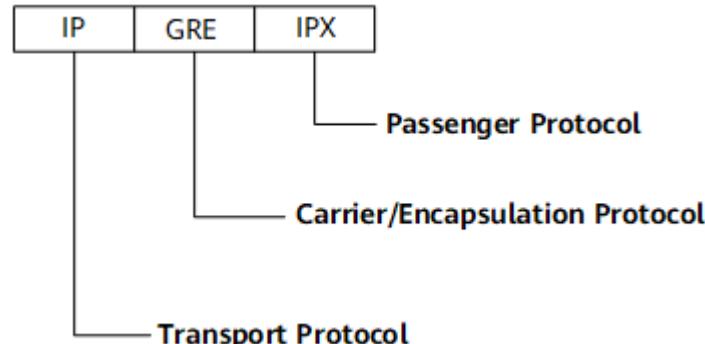


- Payload: is received by the system and needs to be encapsulated and routed.

- Passenger protocol: is used by the packet before encapsulation.
- Encapsulation protocol: is used to encapsulate passenger protocol packets. It is also called the carrier protocol.
- Transport or delivery protocol: is responsible for forwarding the encapsulated packets.

The following shows the format of an IPX packet encapsulated for transmission over an IP tunnel.

Figure 3 Format of an IPX packet transmitted over an IP tunnel



NOTE

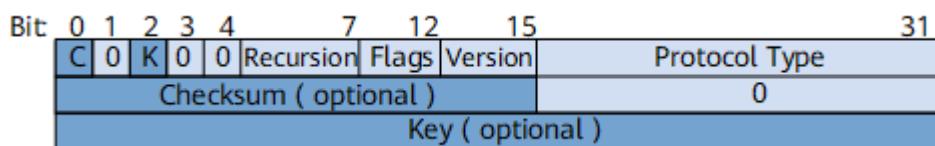
For an IPv6 GRE tunnel:

- The packet transmission protocol is IPv6.
- Only IPv4 and IPv6 packets can be encapsulated and routed.

- GRE header

[Figure 4](#) shows the format of a GRE header.

Figure 4 GRE header



The meaning of each field is as follows:

- C: indicates the Checksum bit. If it is set to 1, the Checksum field is present in the GRE header; if it is set to 0, the GRE header does not contain the Checksum field.
- K: indicates the Key bit. If it is set to 1, the Key field is present in the GRE header; if it is set to 0, the GRE header does not contain the Key field.
- Recursion: indicates the number of times that a packet is encapsulated by GRE. This field increases by one after each encapsulation. If the number of encapsulations is greater than 3, the packet is discarded. This field is used to prevent a packet from being encapsulated infinitely.

NOTE

- According to relevant standards, the default value of the Recursion field is 0.

- According to relevant standards, no errors will occur if the Recursion field value on the transmit end is different from that on the receive end. The receive end ignores this field.
- The Recursion field is only used to indicate the number of times that a packet is encapsulated by GRE. When GRE decapsulates a packet, it is unaware of this field.

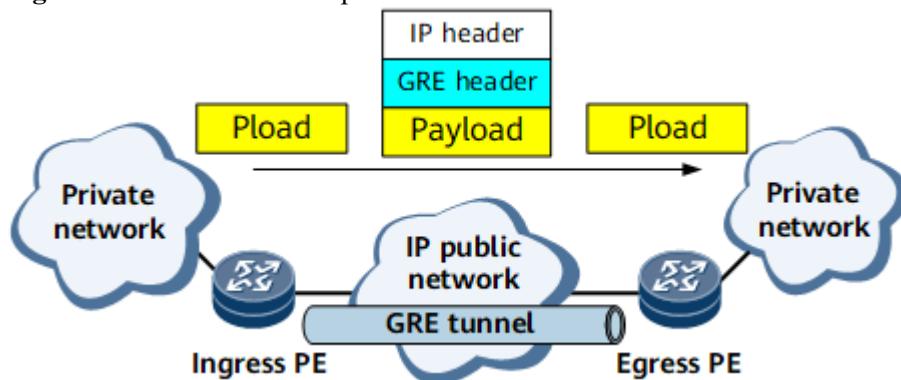
-
- Flags: indicates the reserved field. At present, it must be set to 0.
 - Version: indicates the version number. It must be set to 0. Version number 1 is used by PPTP as defined in relevant standards.
 - Protocol type: indicates the type of the passenger protocol.
 - Checksum: indicates the checksum of the GRE header and the payload.
 - Key: indicates the Key field. It is used by the receive end to authenticate the received packet.

On a device, the GRE header does not contain the Source Route field. Therefore, Bit 1, Bit 3, and Bit 4 are all set to 0.

Transmission of Packets over a GRE Tunnel

The transmission of packets over a GRE tunnel can be divided into two phases: encapsulation and decapsulation. On the network shown in [Figure 5](#), a private network packet is encapsulated on the ingress PE and decapsulated on the egress PE.

Figure 5 Interconnection of private networks over a GRE tunnel



- Encapsulation

After the ingress PE receives a private network packet, the ingress PE delivers the packet to the private network protocol module for processing.

The private network protocol module checks the destination address field in the private network packet header, searches the routing table or forwarding table of the private network for the outbound interface, and determines how to route this packet. If the outbound interface is the GRE tunnel interface, the private network protocol module sends the packet to the tunnel module.

Upon receipt of the packet, the tunnel module processes the packet as follows:

1. Adds a GRE header to the packet. Specifically, the tunnel module encapsulates the packet according to the protocol type of the Passenger packet and the Key parameter configured for the current GRE tunnel.
2. Adds a transport protocol header to the packet based on the configuration. For example, if the transport protocol is the IP protocol, the source and destination

addresses carried in the IP header are the source and destination addresses of the tunnel.

3. Delivers the packet to the IP module. Based on the destination address in the IP header, the IP module searches the public network routing table for the outbound interface and sends the packet. The encapsulated packet is then transmitted on the IP public network.

- Decapsulation

The decapsulation process is opposite to the encapsulation process. After the egress PE receives the packet, the egress PE analyzes the IP header. After determining that the destination of the packet is itself and the Protocol Type field is 47, which indicates that the protocol is GRE (see relevant standards and 2784), the egress PE delivers the packet to the GRE module for processing. The GRE module removes the IP and GRE headers and learns from the Protocol Type field in the GRE header that the Passenger protocol is the protocol running on the private network. The GRE module then delivers the packet to the module corresponding to this protocol, which forwards the packet as an ordinary packet.

Benefits

GRE offers the following benefits:

- Enables packets to be transmitted between networks running different protocols using a single network protocol.
- Enlarges the scope of route transmission.
- Connects discontinuous sub-networks for VPN establishment.

Parent Topic: [Understanding GRE](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.3.2.2 Keepalive Detection

Principles

The current GRE protocol does not have the function of link status detection. If the remote interface is unreachable, the tunnel cannot immediately close the tunnel connection. As a result, the source continuously forwards packets to the peer. The peer, however, discards all the packets because the tunnel is unreachable. A black hole is therefore generated.

The NE40E provides link status detection, also called Keepalive detection, for GRE tunnels. Keepalive detection is used to detect whether the tunnel link is in the Keepalive state at any time, specifically, whether the peer of the tunnel is reachable. If the peer is not reachable, the tunnel is disconnected to prevent data loss caused by black holes.

Implementation

After Keepalive detection is enabled, the ingress of the GRE tunnel periodically sends Keepalive detection packets to the peer. If the peer is reachable, the ingress receives a reply packet from the peer. Otherwise, the ingress cannot receive any reply packet. The details are as follows:

1. After Keepalive detection is enabled, the source of a GRE tunnel creates a timer, periodically sends the Keepalive detection packets, and counts the number of detection packets. The number increases by one each time a detection packet is sent.
2. The peer sends a reply packet to the source after receiving a detection packet.
3. If the source receives a reply packet before the counter value reaches the preset value, the source considers the peer reachable and resets the counter. If the source does not receive any reply packet before the counter reaches the preset value, specifically, the retry times, the source considers the peer unreachable.

NOTE

The endpoint of a GRE tunnel has a Keepalive detection mechanism if it has Keepalive detection configured. The peer does not need to have the Keepalive detection mechanism. After the peer receives a Keepalive detection packet, it sends a reply packet, regardless of whether it has Keepalive detection configured.

Benefits

Keepalive detection prevents data loss when the peer becomes unreachable, ensuring data transmission reliability.

Parent Topic: [Understanding GRE](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.3.2.3 Security Mechanism

GRE supports key authentication, a security mechanism used by tunnel interfaces. This security mechanism prevents tunnel interfaces from incorrectly identifying and receiving packets from other routers.

As defined in relevant standards, if the K bit in the GRE header is set to 1, the Key field is inserted to the GRE header, and both the receiver and sender perform key authentication.

The Key field contains a four-byte number, which is inserted into the GRE header during packet encapsulation. Packets of the same traffic flow have the same Key field. When decapsulating packets, a tunnel endpoint identifies packets of the same traffic flow based on the Key field.

The authentication succeeds only if the Key fields set on both endpoints of the tunnel are consistent. If they are inconsistent, the packet is discarded. "Consistent" means that the Key fields are not set on both endpoints or the same Key field is set on both endpoints.

Parent Topic: [Understanding GRE](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.
[< Previous topic](#)

1.3.3 Application Scenarios for GRE

[Enlarging the Operation Scope of the Network with Limited Hops](#)

[Connecting Discontinuous Sub-networks to Establish a VPN](#)

[CEs Connecting to the MPLS VPN over GRE Tunnels](#)

[Application of GRE on an ERSPAN Network](#)

Parent Topic: [GRE Description](#)

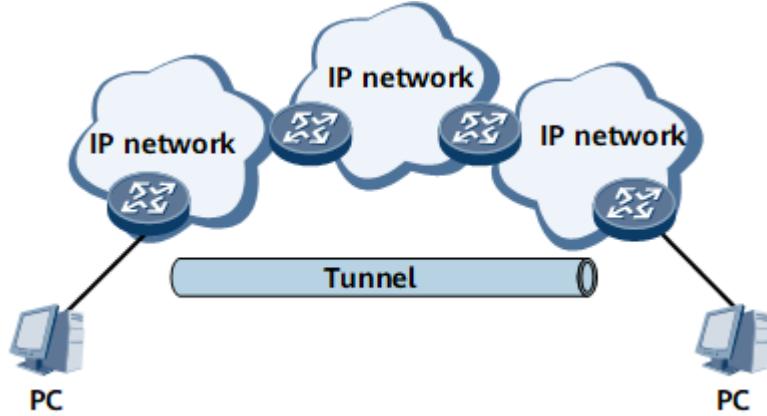
Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

< Previous topic > Next topic

1.3.3.1 Enlarging the Operation Scope of the Network with Limited Hops

On the network shown in [Figure 1](#), the IP protocol runs on the network. Assume that the IP protocol limits the hop count to 255. If the hop count between two PCs is greater than 255, the PCs cannot communicate. After a tunnel is used on the network, a few hops are hidden. This enlarges the network operation scope.

Figure 1 Enlarging the network operation scope



Parent Topic: [Application Scenarios for GRE](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.3.3.2 Connecting Discontinuous Sub-networks to Establish a VPN

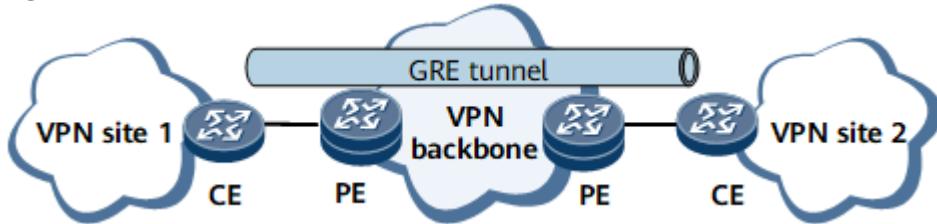
With GRE tunnels, you can connect discontinuous sub-networks to establish a VPN across a WAN.

Assume that two VPN sub-networks, Site 1 and Site 2, are deployed in two different cities. By setting up a GRE tunnel between the PEs, you can connect the two sub-networks to establish a VPN.

GRE, which applies to both L2VPNs and L3VPNs, can be used in either CPE-based VPN or network-based VPN scenarios:

- In a CPE-based VPN scenario, both ends of the GRE tunnel reside on CEs, as shown in [Figure 1](#).

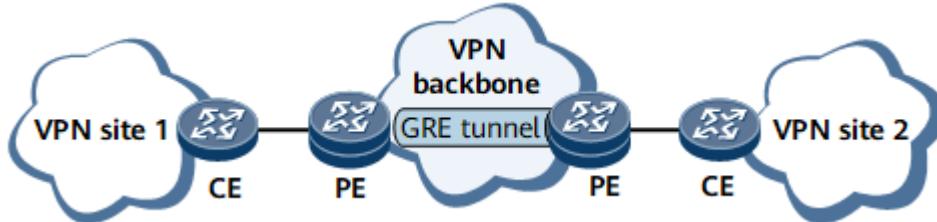
Figure 1 GRE in a CPE-based VPN scenario



In this mode, a CE refers to a CPE.

- In a network-based VPN scenario, both ends of the GRE tunnel reside on PEs, as shown in [Figure 2](#).

Figure 2 GRE in a network-based VPN scenario



Usually, the VPN backbone network uses label switched paths (LSPs) as public network tunnels. If the core devices (Ps) on the backbone network provide only the IP function whereas the PEs at the network edge provide MPLS functions, LSPs cannot be used as public network tunnels. In this situation, you can use GRE tunnels instead of LSPs in Layer 2 or Layer 3 VPN solutions. [Figure 3](#) shows the format of a private network packet transmitted on the VPN backbone network.

Figure 3 Format of a GRE packet that contains an MPLS label

Public network IP header	GRE header	MPLS label	Private network IP header	Payload
--------------------------	------------	------------	---------------------------	---------

GRE tunnels can also be used as non-MPLS VPN backbone tunnels. In this case, the private network packet cannot contain the MPLS label when being transmitted on the VPN backbone network. [Figure 4](#) shows the format of such a packet.

Figure 4 Format of a GRE packet that does not contain any MPLS label

Public network IP header	GRE header	Private network IP header	Payload
--------------------------	------------	---------------------------	---------

Parent Topic: [Application Scenarios for GRE](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

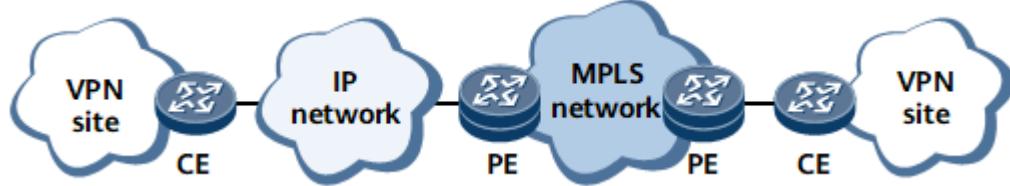
[< Previous topic](#) [Next topic >](#)

1.3.3.3 CEs Connecting to the MPLS VPN over GRE Tunnels

To connect a CE to an MPLS VPN, you must use a physical link to directly connect the CE to a PE on the MPLS backbone network. Specifically, the CE and PE must be on the same network. In this networking, you must associate the VPN with the physical interface connecting the PE to the CE.

As shown in [Figure 1](#), not all CEs and PEs can be directly connected over physical links in actual networking. For example, for multiple organizations that connect to the Internet or IP backbone network, their CEs and PEs are geographically dispersed and cannot directly access the PEs on the MPLS backbone network. These organizations cannot directly access the sites inside the MPLS VPN through the Internet or IP backbone network.

Figure 1 CEs accessing the MPLS VPN backbone network through the IP backbone network



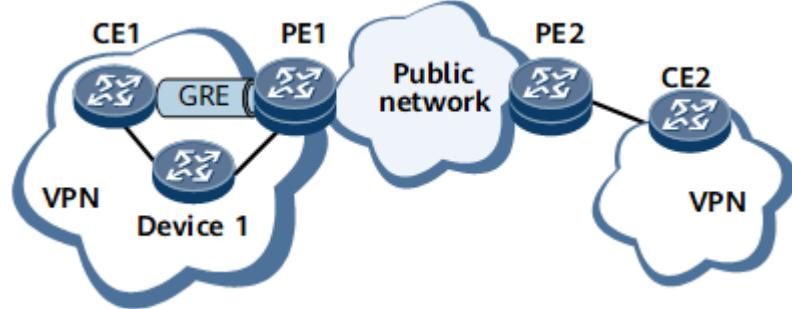
To connect a CE to the MPLS VPN and ensure data transmission security, use the public network or a private network to connect the CE to a PE on the MPLS backbone network and establish a GRE tunnel between the CE and PE. The GRE tunnel can be regarded as a physical interface. You can associate the VPN with the interface on the PE.

When a GRE tunnel is used to access an MPLS VPN, GRE can be implemented in the following modes:

- GRE of the private network: The GRE tunnel is associated with a certain VPN instance; the source and destination addresses of the GRE tunnel belong to this VPN instance.
- GRE across the public network: The GRE tunnel is associated with a certain VPN instance; the source and destination addresses of the GRE tunnel are public network addresses, which do not belong to this VPN instance.
- GRE across the VPN: The GRE tunnel is associated with a certain VPN instance, such as VPN1; the source interface of the GRE tunnel is bound to another VPN instance, such as VPN2. The GRE tunnel passes through VPN2.

GRE of the Private Network

Figure 2 GRE of the private network

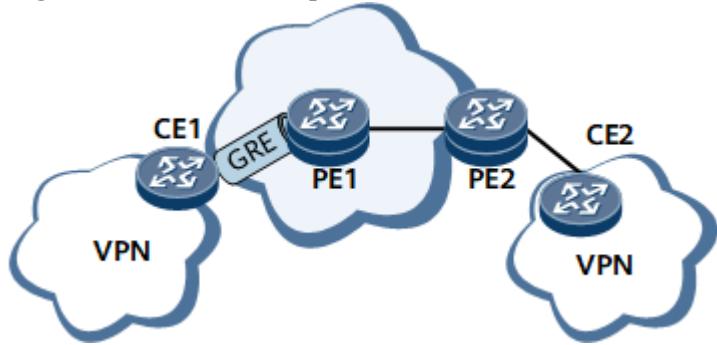


On the network shown in [Figure 2](#), the source and destination addresses of the GRE tunnel belong to the private network. Establishing another tunnel to PE1 on the VPN is not cost-effective, and therefore you are advised to use Device 1 as a CE.

GRE Across the Public Network

In this networking, a CE and a PE must have interfaces that belong to the public network. The interfaces must use public network IP addresses. The CE must have routes to PEs in its public network routing table, and the PE must also have routes to CEs in its public network routing table.

Figure 3 GRE across the public network

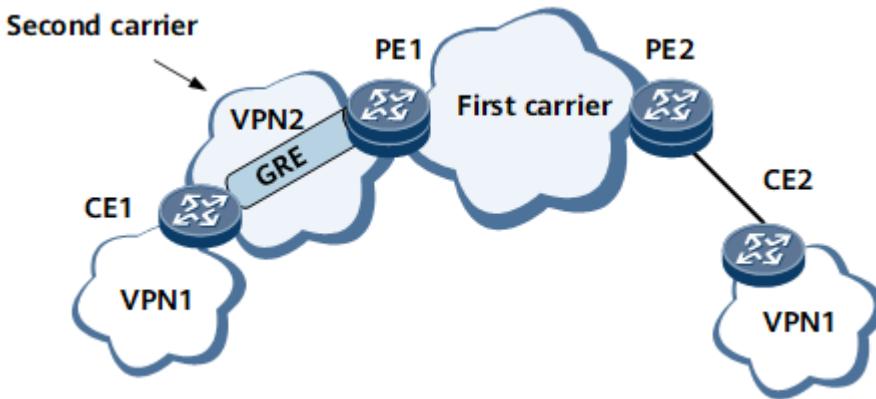


To transmit private network traffic from CEs to PEs over the tunnel, the outbound interface of the route to the remote site segment must be the GRE tunnel interface, and the next hop must be the IP address of the tunnel interface.

GRE Across the VPN

GRE across the VPN is different from GRE across the public network. In GRE across the VPN, CEs connect to PEs over a VPN such as VPN2 rather than the public network. Specifically, both the outbound interface of the private network traffic from CEs to PEs and the outbound interface of the private network traffic returned from PEs to CEs belong to VPN2.

Figure 4 GRE across the VPN



For example, in [Figure 4](#), PE1 and PE2 are the edge devices of the first carrier on the MPLS backbone network. VPN2 is a VPN of the second carrier network. CE1 and CE2 are devices of customers.

To deploy a VPN based on the MPLS network in such a networking environment, such as VPN1, you can create a GRE tunnel across VPN2 between PE1 and CE1. CE1 and PE1 are then logically directly connected.

Parent Topic: [Application Scenarios for GRE](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

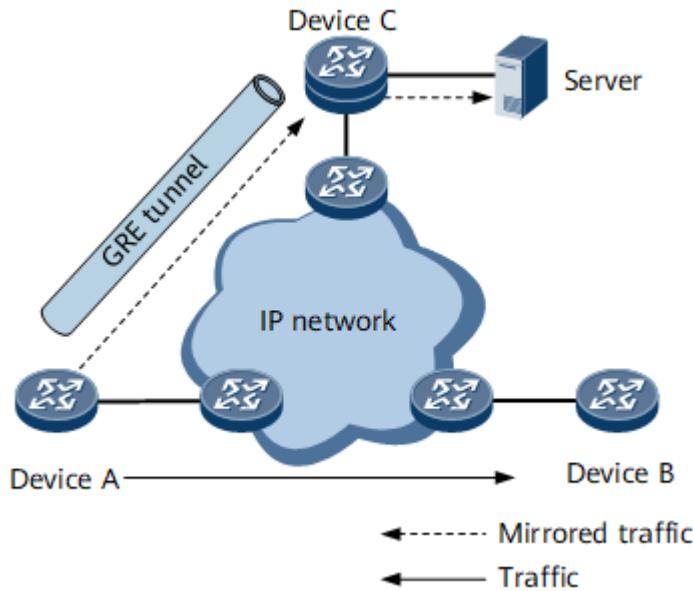
[< Previous topic](#) [Next topic >](#)

1.3.3.4 Application of GRE on an ERSPAN Network

Encapsulated remote switched port analyzer (ERSPAN) is a traffic mirroring protocol that mirrors traffic to one or more ports or virtual local area networks (VLANs). The mirrored traffic is sent to a server for monitoring.

On the network shown in [Figure 1](#), Device A sends traffic to Device B over the IP network. For traffic monitoring, Device A uses ERSPAN to mirror the traffic to a listening port of Device C over a GRE tunnel. The server connected to the listening port can then monitor the mirrored traffic on the listening port. Through the GRE tunnel, ERSPAN allows mirrored packets to traverse an IP network. ERSPAN packets are encapsulated into GRE packets, which are then transmitted over the IP network to Device C.

Figure 1 Application of GRE on an ERSPAN network



Parent Topic: [Application Scenarios for GRE](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.3.4 Appendix

Feature Name	IPv4 GRE	IPv6 GRE	Implementation Difference
Basic GRE principles	Yes	Yes	For an IPv6 GRE tunnel: <ul style="list-style-type: none"> The packet transmission protocol is IPv6. Only IPv4 and IPv6 packets can be encapsulated and routed.
Keepalive detection	Yes	No	IPv6 GRE can respond to Keepalive detection packets sent from other devices, but cannot send Keepalive detection packets.
GRE security mechanism	Yes	No	-

Parent Topic: [GRE Description](#)

Copyright © Huawei Technologies Co., Ltd.

1.4 DSVPN Description

[Overview of DSVPN](#)

[Understanding DSVPN](#)

[Application Scenarios for DSVPN](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.4.1 Overview of DSVPN

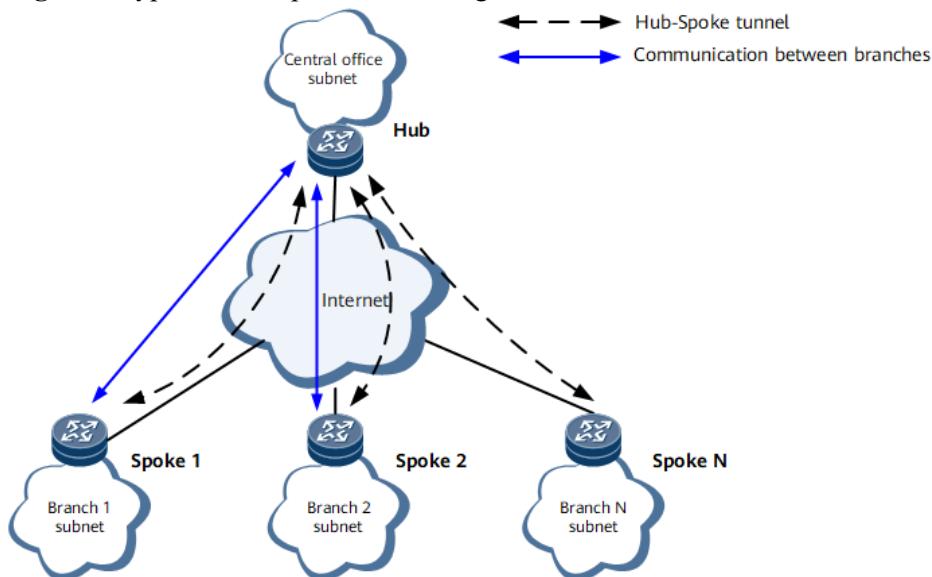
Definition

Dynamic Smart Virtual Private Network (DSVPN) establishes VPN tunnels between Spokes with dynamically variable public addresses in the Hub-Spoke model.

Purpose

More enterprises want to build the IPsec VPN in Hub-Spoke model to connect the Hub to Spokes in different geographical locations. This enhances enterprise communication security and reduces communication costs. When the Hub uses the static public address to connect to the Internet and Spokes use dynamic public addresses to connect to the Internet, Spokes cannot communicate with each other directly if traditional IPsec or GRE over IPsec is used to build the VPN. This is because Spokes cannot learn the public addresses of the remote ends in advance and tunnels cannot be set up between Spokes. In this case, communication data between Spokes must be forwarded by the Hub.

Figure 1 Typical Hub-Spoke networking without DSVPN enabled

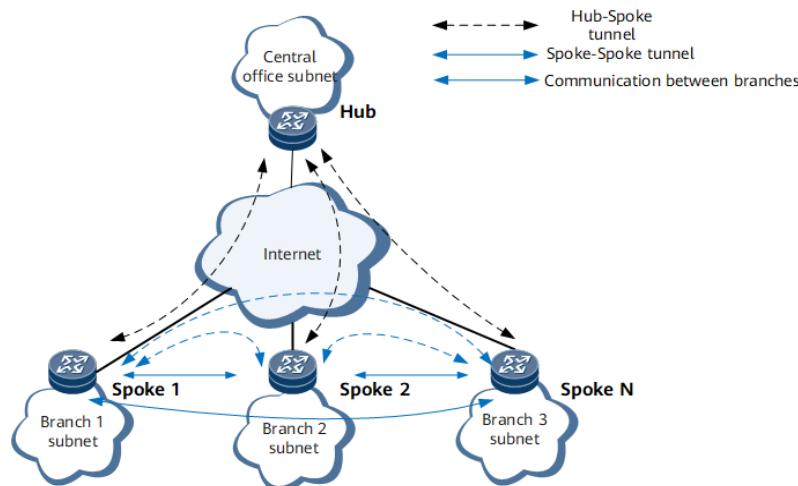


When all communication data between Spokes is forwarded by the Hub, the following problems may occur:

- Whenever a new Spoke is connected to the Hub, a VPN configuration and maintenance for the Spoke are added to the Hub. In this case, if a large number of Spokes are connected, the Hub configurations become complex. Each time when the network is adjusted, the Hub configurations have to be adjusted accordingly.
- If Spokes communicate with each other through the Hub, the transmitted data flows consume the Hub resources and result in an extra delay (especially when IPsec encryption is used). This because the Hub needs to decrypt and then encrypt data packets from the source Spoke before sending them to the destination Spoke.
- If Spokes communicate with each other directly and the Spoke egresses use dynamic IP addresses, the Spokes cannot obtain each other's IP address. As a result, a tunnel cannot be established directly between the Spokes.

To resolve this issue, DSVPN uses Next Hop Resolution Protocol (NHRP) to collect and maintain information about dynamically changing public IP addresses of the Spokes. In this manner, the Spokes can obtain each other's public IP address before establishing a tunnel with each other.

Figure 2 Typical Hub-Spoke networking without DSVPN enabled



On the network shown in [Figure 2](#), DSVPN allows the Spokes to dynamically establish a Spoke-Spoke tunnel when they use dynamic IP addresses to access the public network. This implements direct communication between the Spokes. In addition, DSVPN supports multipoint Generic Routing Encapsulation (mGRE), which allows multiple GRE tunnels to be set up on a single mGRE tunnel interface. This simplifies subnet traffic management and configurations of GRE and IPsec on devices.

Benefits

- Reduced VPN network construction costs

DSVPN implements dynamic connections between the Hub and Spokes, and between Spokes. Spokes do not need to purchase static public network addresses.

- Simplified configuration of the Hub and Spokes

The Hub and Spokes use an mGRE tunnel interface but not multiple GRE tunnel interfaces to establish tunnels. When a new Spoke is added to the network, the network administrator does not need to change configurations on the Hub or any existing Spokes. The administrator only needs to configure the new Spoke, and then the Spoke dynamically registers with the Hub.

- Reduced data transmission delay between branches

Spokes can dynamically establish tunnels to directly exchange service data, reducing the forwarding delay and improving forwarding performance and efficiency.

Parent Topic: [DSVPN Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.4.2 Understanding DSVPN

[Basic Concepts](#)

[DSVPN Fundamentals](#)

[DSVPN NAT Traversal](#)

[DSVPN IPsec Protection](#)

[Dual Hubs in Active/Standby Mode](#)

Parent Topic: [DSVPN Description](#)

Copyright © Huawei Technologies Co., Ltd.

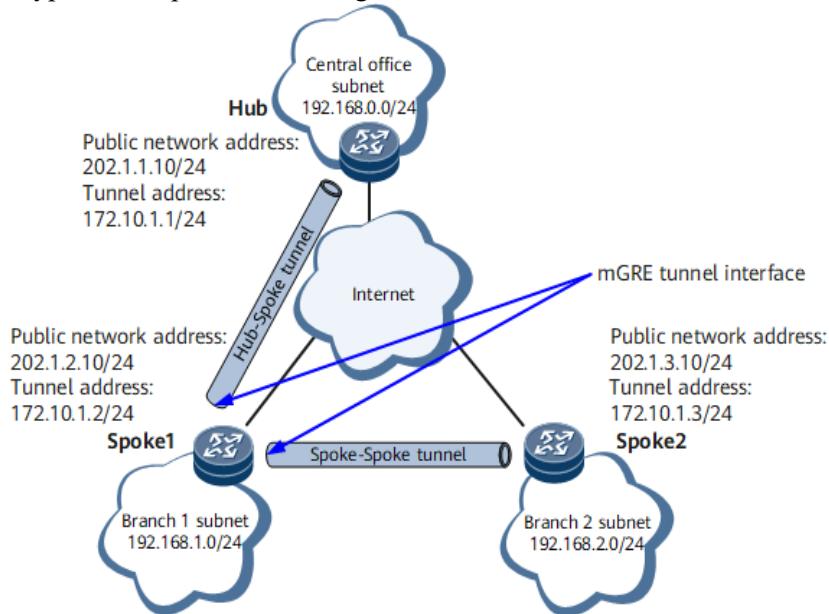
Copyright © Huawei Technologies Co., Ltd.

< Previous topic [Next topic >](#)

1.4.2.1 Basic Concepts

[Figure 1](#) shows the typical network architecture of the DSVPN solution. An enterprise connects the Hub to Spokes in different geographical locations through the public network. The Hub uses the static public address, and Spokes use dynamic public addresses.

Figure 1 Typical enterprise networking



In the figure, the public network address is considered as a Non-Broadcast Multiple Access (NBMA) address, and the tunnel address is regarded as a protocol address.

DSVPN Node

A DSVPN node is a device on which DSVPN is deployed, which can be a Spoke or Hub.

- **Spoke**

A Spoke is the network gateway of a branch. Generally, a Spoke uses a dynamic public network address.

- **Hub**

A Hub is the gateway in the headquarters and receives registration packets from Spokes. On a DSVPN network, the Hub can use a fixed public network address or a domain name.

mGRE, mGRE Tunnel Interface, and mGRE Tunnel

mGRE is a point-to-multipoint GRE technology developed based on GRE. It extends traditional P2P tunnel interfaces to P2MP mGRE tunnel interfaces. One tunnel interface can be used to establish tunnels with multiple remote devices by changing the interface type. Therefore, only one tunnel interface needs to be configured on the Hub or Spoke, reducing the GRE tunnel configuration workload.

The mGRE tunnel interface has the following attributes:

- Source tunnel address: is the source address of a GRE encapsulated packet, that is, public network address of one end in [Figure 1](#).
- Destination tunnel address: is the destination address of a GRE encapsulated packet, that is, public network address of the other end in [Figure 1](#). This address is based on NHRP, which is different from the manually specified destination address of the GRE tunnel interface.
- Tunnel interface IP address: is the tunnel address in [Figure 1](#). Similar to an IP address of a physical interface, a tunnel interface IP address is used for communication between devices, for example, routing information is obtained.

 **NOTE**

- The destination IP address of a GRE tunnel interface is manually specified. Unlike this, the destination IP address of an mGRE tunnel interface is dynamically obtained by NHRP. A single mGRE tunnel interface can establish multiple GRE tunnels with different GRE peers.
 - mGRE tunnel interfaces do not support keepalive detection of the GRE interface.
-

NHRP

NHRP allows the source Spoke to obtain the dynamic public IP address of the destination Spoke on a non-broadcast multiple access (NBMA) network over which a DSVPN is deployed. When a Spoke accesses an NBMA network, it uses the outbound interface's public IP address to send an NHRP Registration request to the Hub. The Hub creates or updates its NHRP peer entry for the Spoke node based on the received request. The Spokes create and update NHRP peer entries by exchanging NHRP Resolution Request and Reply messages with each other.

Hub-Spoke Tunnel

A Hub-Spoke tunnel is established between a Spoke and the Hub. [Figure 1](#) shows an example Hub-Spoke tunnel. Similarly, other Spokes also establish Hub-Spoke tunnels with the Hub.

On a DSVPN, Spoke information is not configured on the Hub. The Hub's public IP address and tunnel address are manually specified on the Spokes. When a Spoke accesses an NBMA network, it sends an NHRP Registration request to the Hub and notifies the Hub of its outbound interface's public IP address. After receiving the request, the Hub updates the local NHRP peer entry for the Spoke.

Spoke-Spoke Tunnel

Spoke-Spoke tunnels are established between Spokes. [Figure 1](#) shows an example Spoke-Spoke tunnel.

After the source Spoke finds the destination Spoke's next hop in the routing table, it sends an NHRP Resolution request to obtain the destination Spoke's public IP address if the public IP address corresponding to the next hop cannot be found in the local NHRP peer table. Then the Spokes dynamically establish a VPN tunnel with each other through mGRE tunnel interfaces. In this manner, the source Spoke and destination Spoke can exchange data. If no traffic is forwarded through a Spoke-Spoke tunnel within a certain period, the tunnel is automatically dismantled.

Parent Topic: [Understanding DSVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.4.2.2 DSVPN Fundamentals

Dynamic Smart VPN (DSVPN) allows VPNs to be established between branches in the following scenarios: non-shortcut scenarios on small- or medium-sized networks and shortcut scenarios on large-scale networks.

Route deployment varies according to different scenarios:

- [Non-shortcut scenario](#): route learning between branches

In a non-shortcut scenario, there are only a few branches on a small- or medium-sized network, and the branches learn routes from each other so that the next-hop address of a route from a source Spoke to a destination Spoke subnet is the tunnel address of the destination Spoke. This deployment solution applies to small- or medium-sized networks, and the number of routes dynamically learned between branches is small, which therefore does not require high performance of the Hub and Spoke nodes.

- [Shortcut scenario](#): saving routes summarized to the HQ

In a shortcut scenario, there are a large number of branches on a large-scale network, and the branches save only routes summarized to the HQ (Hub node) so that the next-hop address of a route from a source Spoke to a destination Spoke subnet is the tunnel address of the Hub node. If route learning in non-shortcut mode applies to the large-scale network, the Spoke nodes (branches) have to save network-wide routes and consume lots of CPU and memory resources to compute dynamic routes, which requires large capacity of routing tables and high performance of Spoke nodes. To compensate for this shortcoming, DSVPN is enhanced to support route learning in shortcut mode.

DSVPN Principles in Non-Shortcut Scenarios

Route Deployment

In a non-shortcut scenario, Spoke nodes establish tunnels with each other for direct communication.

The next-hop address of a route from a source Spoke to a destination Spoke subnet is the tunnel address of the destination Spoke. Spoke nodes can learn routes from each other in the following ways:

- Static routes are configured on branches.

Static routes to destination branch subnets are configured on the source branch, and the next hops of the routes are the tunnel addresses of the destination branches.

- Routes are dynamically learned between branches.

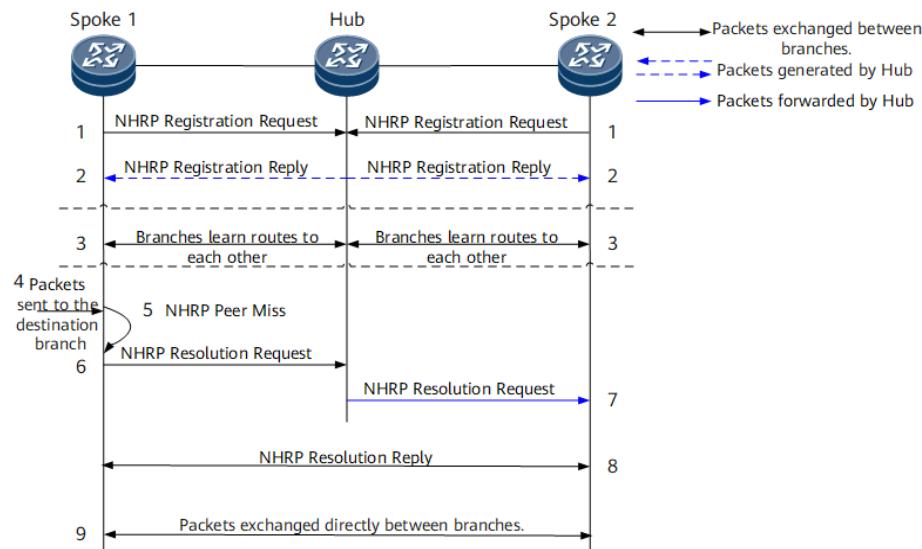
DSVPN supports OSPF and BGP, both of which can implement route learning between branch subnets and between a branch subnet and the HQ subnet. A routing protocol is configured on both the Hub node and Spoke nodes to implement dynamic route learning.

Branches learn routes from each other, and each Spoke node stores routes to all branch subnets.

Principles

DSVPN uses Next Hop Resolution Protocol (NHRP) to dynamically obtain a peer's public IP address. In non-shortcut scenarios, DSVPN working principles are as follows.

Figure 1 DSVPN principles in non-shortcut scenarios



On the network shown in [Figure 1](#):

1. The network administrator manually specifies a public IP address or tunnel address for the Hub node. All Spoke nodes on the network send registration requests to the Hub node.
2. The Hub node generates NHRP peer entries based on the received requests and sends NHRP Registration Reply messages to the Spoke nodes.
3. The Spoke nodes learn subnet routes from each other either by static route configuration or a dynamic routing protocol. The next hops of the routes are the tunnel addresses of the peer Spoke nodes.
4. When the source Spoke node forwards a data packet, it obtains the public IP address corresponding to the packet next hop (tunnel address of the destination Spoke node).
5. If the public IP address corresponding to the tunnel address of the destination Spoke does not exist, the source Spoke sends an NHRP Resolution Request message.
6. The source Spoke node constructs an NHRP Resolution Request message to request the public IP address corresponding to the tunnel address of the destination Spoke node.
7. After the NHRP Resolution Request message arrives at the Hub node, the Hub node sends the message to the destination Spoke node.
8. The destination Spoke receives the NHRP Resolution Request message and sends an NHRP Resolution Reply message to the source Spoke.
9. Then the source Spoke can directly communicate with the destination Spoke.

DSVPN Principles in Shortcut Scenarios

Route Deployment

In a shortcut scenario, the next-hop address of a route from a source Spoke to a destination Spoke subnet is the tunnel address of the Hub node. The branches are deployed to store only routes summarized to the HQ so that traffic to all destination branches is sent to the Hub node. Spoke nodes can learn routes in the following ways:

- Static routes are configured on branches.

Static routes to destination branch subnets are configured on the source branch, and the next hops of the routes are the tunnel address of the Hub node.

- Branches dynamically learn routes destined for the HQ.

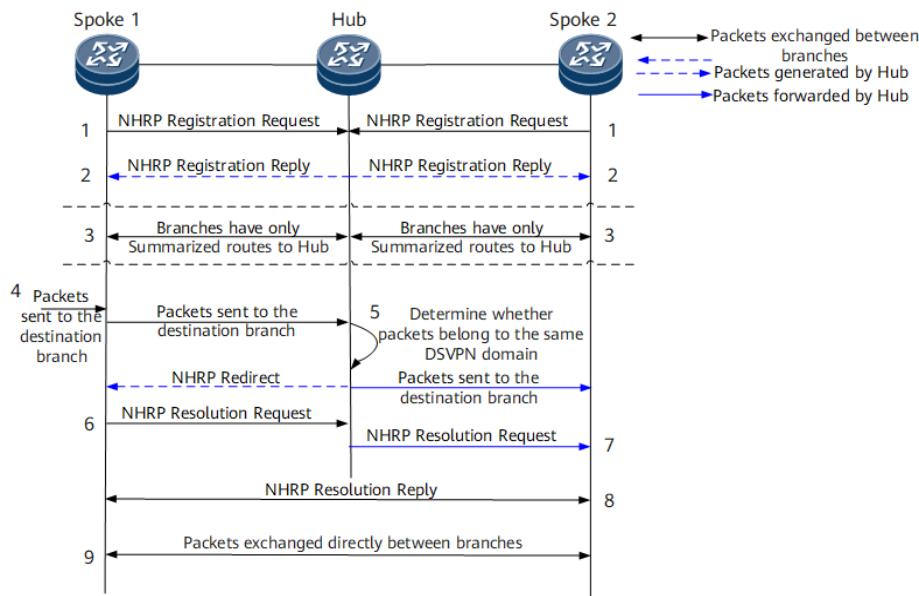
DSVPN supports OSPF and BGP. Route summarization is configured on the Hub node, and either OSPF or BGP is configured on the Spoke nodes so that the Spoke nodes store only routes summarized to the Hub node. In this manner, traffic to all destination branches is sent to the Hub node. If different routing protocols are used, the Hub and Spoke nodes must be configured separately.

In shortcut mode, the default traffic egress is the Hub node. The branches do not learn routes from each other. The Hub node aggregates the branch routes and then advertises the summarized routes. Additionally, the Hub node forwards NHRP Resolution Request messages to the destination Spoke nodes. Upon receipt, the destination Spoke nodes parse and respond to the requests.

Principles

DSVPN uses NHRP to dynamically obtain a peer's public IP address. In shortcut scenarios, DSVPN working principles are as follows.

Figure 2 DSVPN principles in shortcut scenarios



On the network shown in [Figure 2](#):

1. The network administrator manually specifies a public IP address or tunnel address for the Hub node. All Spoke nodes on the network send registration requests to the Hub node.
2. The Hub node generates NHRP peer entries based on the received requests and sends NHRP Registration Reply messages to the Spoke nodes.

3. The Spoke nodes learn routes either by static route configuration or a dynamic routing protocol and store only routes summarized to the Hub node.
4. When the source Spoke forwards a data packet, it searches for the public IP address corresponding to the packet next hop, encapsulates the data packet, and then sends the packet to the next hop. (The next hop here is the Hub node.)
5. After the data packet reaches the Hub node, the Hub node sends the packet to the destination Spoke and sends an NHRP Redirect message to the source Spoke as well.
6. Upon receipt, the source Spoke sends an NHRP Resolution Request message to the destination Spoke.
7. After the NHRP Resolution Request message arrives at the Hub node, the Hub node forwards it to the destination Spoke.
8. The destination Spoke receives the NHRP Resolution Request message and sends an NHRP Resolution Reply message to the source Spoke.
9. Then the source Spoke can directly communicate with the destination Spoke.

Parent Topic: [Understanding DSVPN](#)

Copyright © Huawei Technologies Co., Ltd.

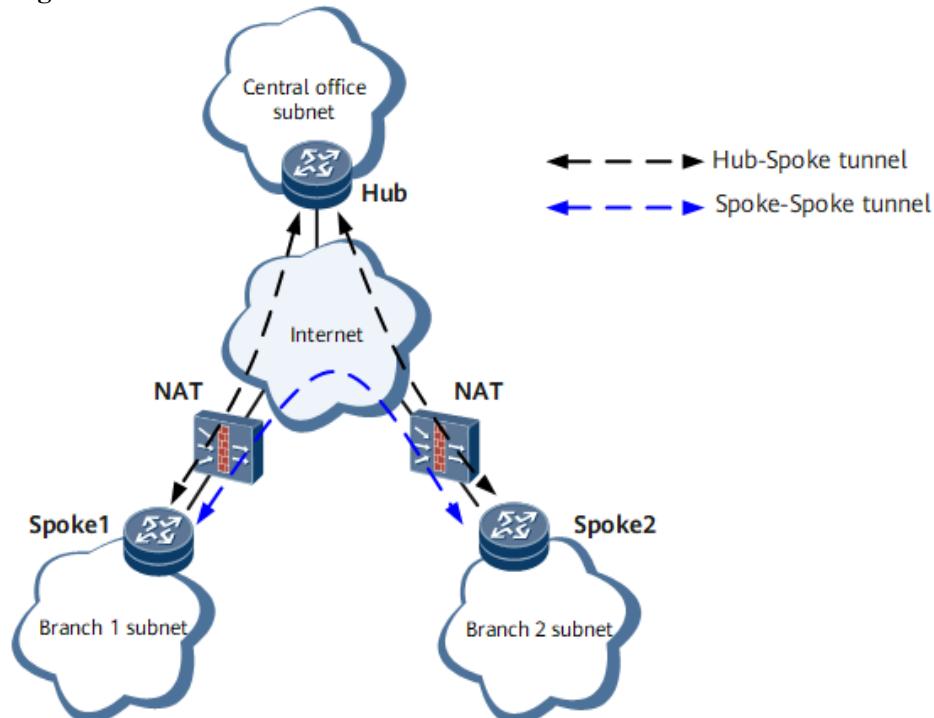
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.4.2.3 DSVPN NAT Traversal

In [Figure 1](#), when private networks of Spokes connect to the Hub through Network Address Translation (NAT), NAT traversal must be implemented to establish VPN tunnels between the Hub and Spokes, and between Spokes. DSVPN NAT traversal can be deployed so that Spokes can directly communicate across the NAT device.

Figure 1 DSVPN NAT traversal



The implementation of DSVPN NAT traversal is as follows:

1. The Spokes send NHRP Registration Request packets to the Hub. The NHRP Registration Request packets carry public network addresses of the Spokes.
2. NHRP on the Hub detects whether a NAT device exists between the Hub and Spokes. If the NAT device exists, the Hub encapsulates translated public addresses of Spokes in NAT extension fields of NHRP Registration Reply packets and sends the packets to the Spokes.
3. The source Spoke sends an NHRP Resolution Request packet to the destination Spoke. The packet carries the original private address and translated public address in NAT extension fields of the source Spoke.
4. The destination Spoke sends an NHRP Resolution Reply packet to the source Spoke. The packet carries the original private address and translated public address in NAT extension fields of the destination Spoke.
5. The source and destination Spokes learn the original private address and translated public network address of each other and establish an mGRE tunnel based on the translated public address. By doing this, Spokes can directly communicate across the NAT device.

Parent Topic: [Understanding DSVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

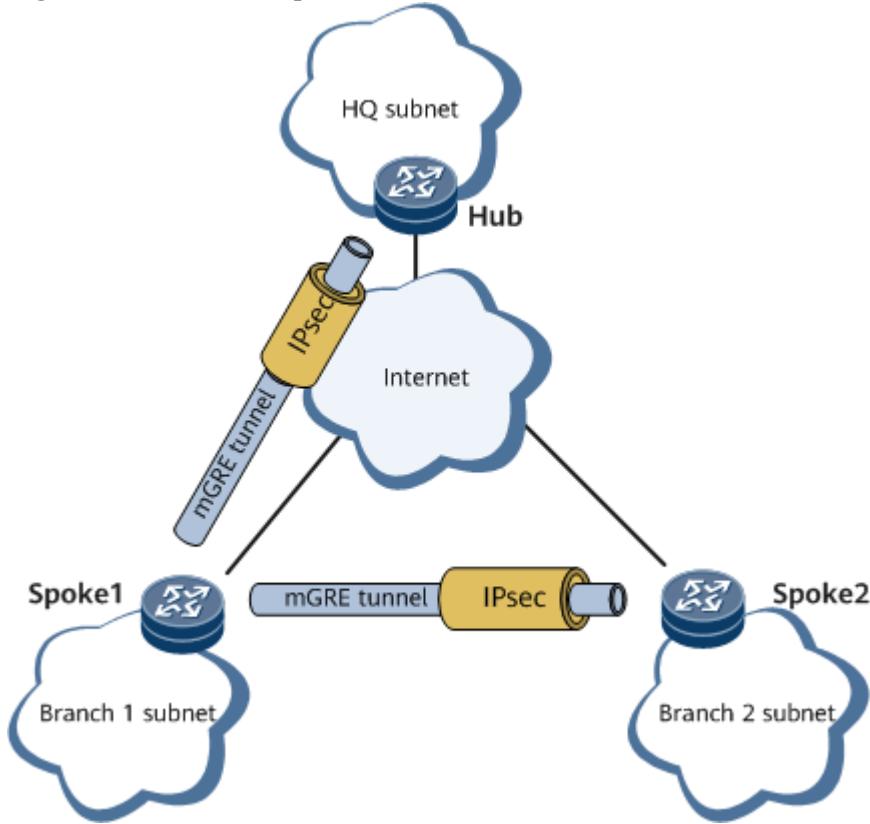
[< Previous topic](#) [Next topic >](#)

1.4.2.4 DSVPN IPsec Protection

If an enterprise needs to protect the security of data transmitted between the headquarters and branches or between branches through encryption on a DSVPN, bind an IPsec profile to the DSVPN to dynamically establish mGRE and IPsec tunnels between branches.

- The establishment of an mGRE tunnel immediately triggers the establishment of an IPsec tunnel.
- The common IPsec technology uses an ACL to identify unicast traffic to be encrypted. An IPsec security policy requires the definition of complex ACL rules, complicating implementation. If deployed with IPsec, a DSVPN using NHRP and mGRE technologies can guarantee high security for data transmission while simplifying network deployment.
- An IPsec tunnel is dynamically established between branches, preventing IPsec data exchanged between branch spokes from being decrypted and then encrypted by the hub, which in turn reducing the delay in data transmission.

Figure 1 DSVPN IPsec protection



On a DSVPN, IPsec profiles are configured on the mGRE interfaces of the hub and spokes. The DSVPN over IPsec mechanism works as follows:

1. All the spokes on the network send registration requests to the hub and report the NHRP peer information to IPsec. The Internet Key Exchange (IKE) modules of the spokes and hub negotiate IPsec tunnel parameters with each other.
2. The hub records mappings between tunnel addresses and public network addresses of the spokes and generates NHRP peer entries for the spokes based on registration requests received. The hub then sends registration replies to spokes.
3. The spokes establish an mGRE tunnel upon traffic transmission request. For details about how to establish an mGRE tunnel, see [DSVPN Principles in Non-Shortcut Scenarios](#) and [DSVPN Principles in Shortcut Scenarios](#).
4. After the spokes establish an mGRE tunnel, the IPsec module obtains NHRP peer information, adds or deletes IPsec peers based on the information, and triggers the spokes to dynamically establish an IPsec tunnel.
5. After an IPsec tunnel is established between the spokes, packets are routed based on destination IP addresses. If the outbound interface is an mGRE interface, the spoke searches the NHRP peer table for the public network address corresponding to the next hop of the route. After obtaining the public network address, the spoke searches for a matching IPsec security association (SA) to encrypt the packets.

 **NOTE**

Digital envelop authentication is not supported when NHRP and IPsec are both configured.

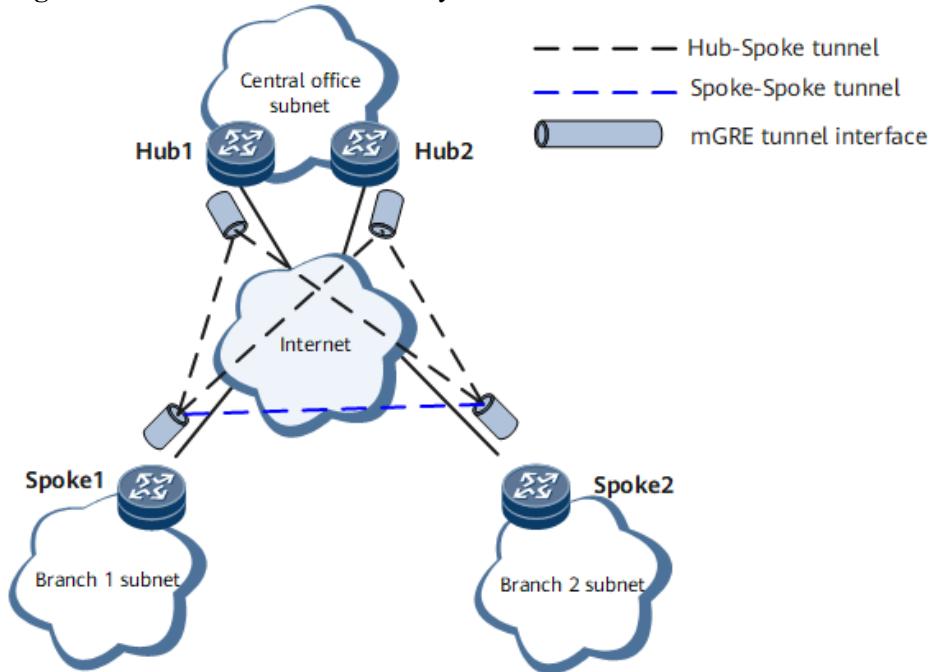
1.4.2.5 Dual Hubs in Active/Standby Mode

In basic DSVPN networking, all Spokes are connected to one Hub. Spokes cannot communicate with each other if the Hub fails. Multiple Hubs can be deployed to improve network reliability.

NOTE

Dual-hub backup is supported for DSVPN only in shortcut scenarios.

Figure 1 Dual Hubs in active/standby mode



The working mechanism is as follows:

1. All Spokes send NHRP Registration requests to Hub 1 (active) and Hub 2 (standby) at the same time. (The request messages carry the Spokes' tunnel addresses and public IP addresses.) In addition, the Spokes locally generate NHRP peer tables for the Hubs. Each Spoke records the mappings between the tunnel addresses and public IP addresses of the Hubs in the NHRP peer table.
2. Hub 1 and Hub 2 record the mapping between the tunnel addresses and public IP addresses of the Spokes based on the received requests, generate NHRP peer entries for the Spokes, and then send NHRP Registration Reply messages to the Spokes.
3. Routing policies can be deployed on the Spokes to make Hub 1's routes have higher priorities than Hub 2's routes. When the Spokes need to communicate, NHRP Resolution requests are preferentially sent to Hub 1, which forwards the messages.
4. For details about the principles of establishing tunnels between branches based on traffic, see [DSVPN Principles in Shortcut Scenarios](#).
5. If Hub 1 fails, the Spokes send NHRP Resolution requests to Hub 2, which forwards the messages. If Hub 1 recovers then, the Spokes choose Hub 1 again as the forwarder based

on the predefined routing policies.

Parent Topic: [Understanding DSVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.4.3 Application Scenarios for DSVPN

[DSVPN Deployment on a Small- or Medium-sized Network](#)

[DSVPN Deployment on a Large-sized Network](#)

[Deploying DSVPN in Hierarchical Hub Networking](#)

Parent Topic: [DSVPN Description](#)

Copyright © Huawei Technologies Co., Ltd.

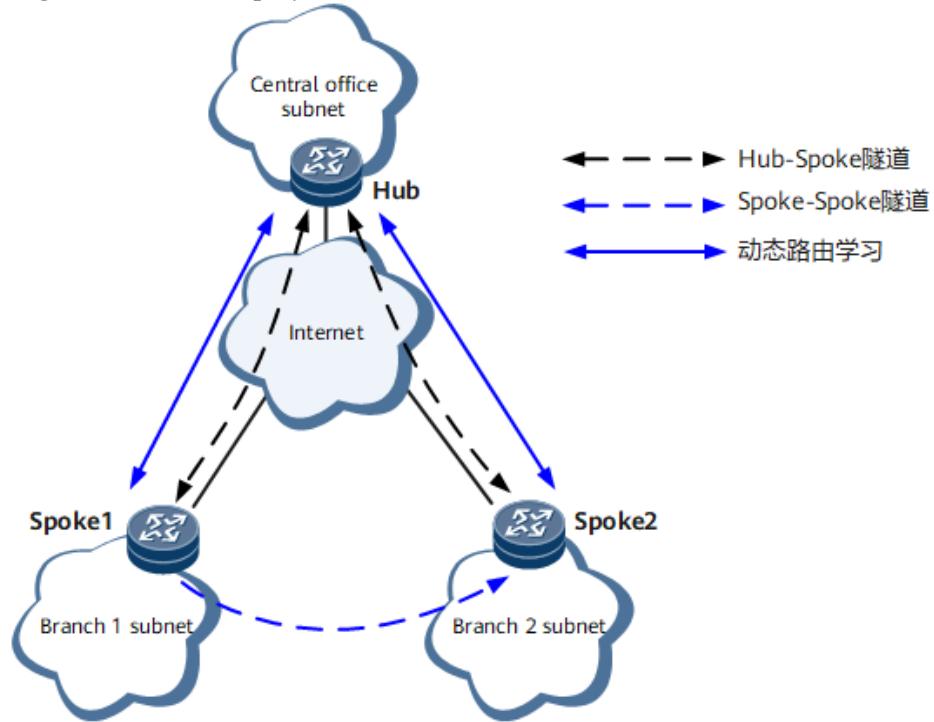
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.4.3.1 DSVPN Deployment on a Small- or Medium-sized Network

Small- and medium-sized networks have only a few branches, and the branches can dynamically establish VPNs by deploying non-shortcut scenario of DSVPN.

Figure 1 DSVPN deployment on a small- or medium-sized network



As shown in [Figure 1](#), Spoke1 and Spoke2 connect to the Hub through the public network. DSVPN is deployed to enable Spoke1 and Spoke2 to learn routes from each other. Spoke1 and Spoke2 can communicate with each other directly because they are each other's next hop.

Parent Topic: [Application Scenarios for DSVPN](#)

Copyright © Huawei Technologies Co., Ltd.

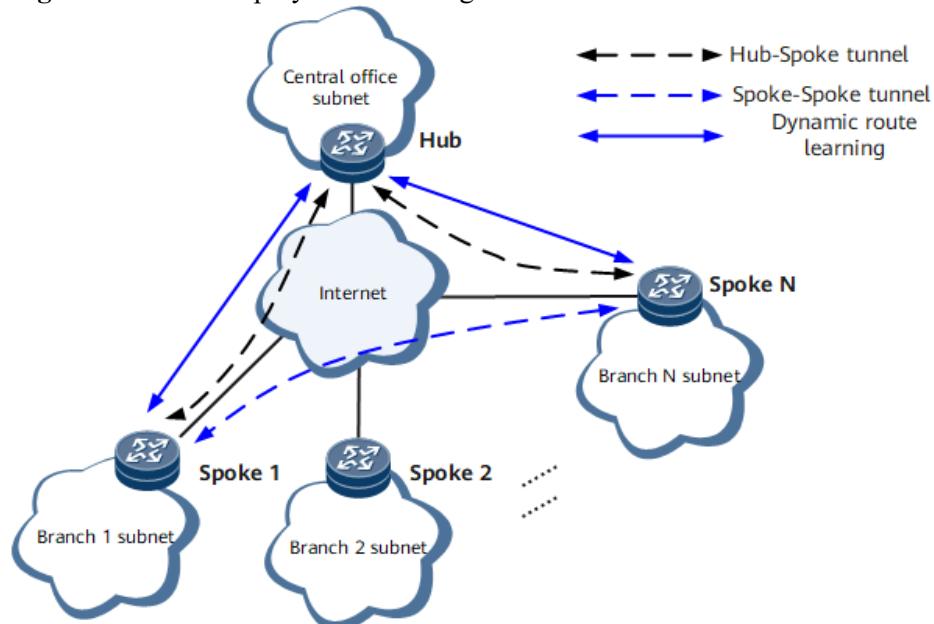
Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.4.3.2 DSVPN Deployment on a Large-sized Network

A large-sized network has a large number of branch offices. The deployment of Non-Shortcut Scenario of DSVPN requires the Spokes to have a large routing table and high forwarding performance. Shortcut Scenario of DSVPN can be deployed without upgrading the Spokes. This deployment reduces the routing entries on the Spokes, lowering the requirements on the Spokes' routing table size and forwarding performance.

Figure 1 DSVPN deployment on a large-sized network



As shown in [Figure 1](#), all the Spokes only have routes to the Hub. When two Spokes need to communicate with each other, the first packet is sent to the Hub. After that, a tunnel is established between the Spokes, and the Spokes can directly exchange data traffic.

Parent Topic: [Application Scenarios for DSVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

< Previous topic [Next topic >](#)

1.4.3.3 Deploying DSVPN in Hierarchical Hub Networking

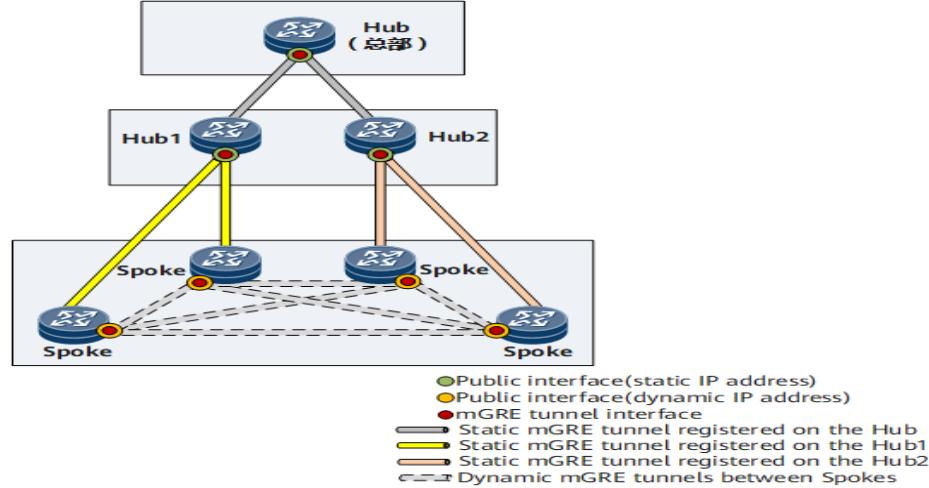
[Figure 1](#) shows a network topology of an enterprise. The organizations of the enterprise are hierarchical. DSVPN is deployed on the network. Some intermediate nodes function as both the Spokes and Hubs. For example, Hub1 and Hub2 function as Hubs for its downstream Spokes and serve as Spokes for the Hub in the headquarters.

The principles of establishing dynamic mGRE tunnels between Spokes of Hub1 and Hub2 are similar to [Principles](#). When a Spoke of Hub1 needs to establish a dynamic mGRE tunnel with a Spoke of Hub2, the source Spoke sends an NHRP Resolution Request packet to Hub1. Hub1 sends the NHRP Resolution Request packet to the Hub, the Hub sends the NHRP Resolution Request packet to Hub2, and Hub2 sends the NHRP Resolution Request packet to the destination Spoke. Finally, a dynamic mGRE tunnel is set up between Spokes in hierarchical Hub networking.

NOTE

When DSVPN is deployed in hierarchical Hub networking, branches can learn routes from each other in shortcut mode only.

Figure 1 Deploying DSVPN in hierarchical Hub networking



Parent Topic: [Application Scenarios for DSVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.5 L2TPv3 Description

[Overview of L2TPv3](#)

[Understanding L2TPv3](#)

[Application Scenarios for L2TPv3](#)

[Terminology for L2TPv3](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

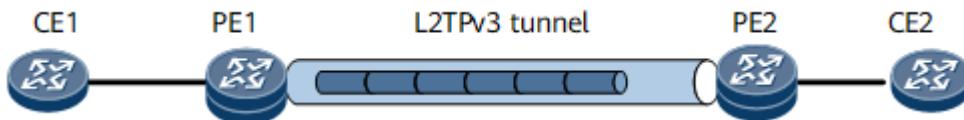
[< Previous topic](#) [Next topic >](#)

1.5.1 Overview of L2TPv3

Definition

Layer 2 Tunneling Protocol (L2TP), which integrates the advantages of Layer 2 Forwarding (L2F) and Point-to-Point Tunneling Protocol (PPTP), is an industry standard of IETF for tunneling Layer 2 circuits across a packet switched network (PSN). L2TP provides a mechanism for establishing PPP sessions over a non-P2P network. L2TPv3, the third version of L2TP, can transparently transmit Layer 2 traffic between CEs over a PSN.

Figure 1 Networking for L2TPv3



Purpose

L2TPv3 enables carriers to provide Ethernet services on public IP networks, allowing enterprises to enjoy services at lower prices. L2TPv3 does not have new requirements for the IP transmission infrastructure and therefore is easy to implement.

Benefits

L2TPv3 has a flexible identify authentication mechanism and features high security. L2TPv3 can enhance data transmission security by means of channel encryption, E2E data encryption, or application-layer data encryption.

Parent Topic: [L2TPv3 Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.5.2 Understanding L2TPv3

[L2TPv3 Basic Concepts](#)

[L2TPv3 Fundamentals](#)

Parent Topic: [L2TPv3 Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

< Previous topic [Next topic >](#)

1.5.2.1 L2TPv3 Basic Concepts

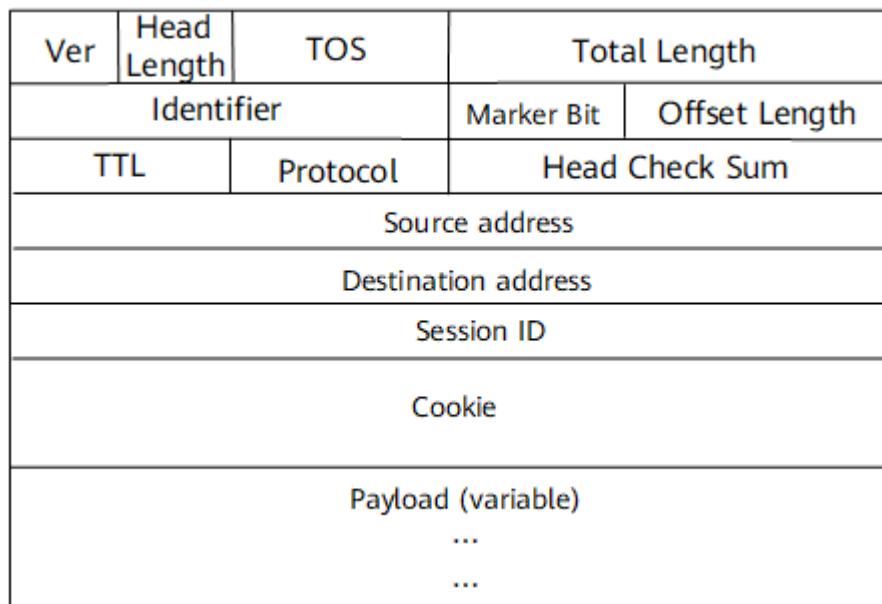
L2TPv3 over IPv4/IPv6

L2TPv3 over IPv4/IPv6 is used to establish L2TPv3 tunnels on an IPv4/IPv6 public network, so that Layer 2 user packets can be transparently transmitted across the IPv4/IPv6 public network. L2TPv3 over IPv4/IPv6 establishes tunnels based on static configurations and does not require dynamic negotiation for tunnel establishment or tear-down.

L2TPv3 uses unique source or destination IPv4/IPv6 addresses to identify tunnels, leveraging the key property that IPv6 offers, a vast number of unique IP addresses. User packets transmitted over an L2TPv3 tunnel are identified by unique source or destination IPv4/IPv6 addresses. L2TPv3 identifies Layer 2 access links by source or destination IP addresses of tunnels. In this case, processing of the

L2TPv3 session ID may be bypassed upon receipt because each tunnel has only one associated session.

Figure 1 L2TPv3 over IPv4 packet format



The following table describes the meaning of each field in the packet.

Name	Description
Ver	A 4-bit field used to indicate the version number. The value is set to 4 for IPv4.
Head Length	A 4-bit field used to indicate the packet header length.
TOS	An 8-bit field used to indicate the type of service.
Total Length	A 16-bit field used to indicate the total packet header length.
Identifier	A 16-bit field used to indicate the identifier.
Mark Bit	A 3-bit field used to indicate the flag.
Offset Length	A 13-bit field used to indicate the offset value.
Head Check Sum	A 16-bit field used to indicate the check sum of the packet header.
TTL	A 4-bit field used to indicate the time to live.
Protocol	A 4-bit field used to indicate the L2TPv3 protocol ID of 115.
Source Address	A 32-bit field used to indicate the IPv4 source address for the tunnel. The IPv4 source address is a loopback address of the local device.

Name	Description
Destination Address	A 32-bit field used to indicate the IPv4 destination address for the tunnel. The IPv4 destination address is a loopback address of the remote device.
Session ID	A 32-bit field used to indicate the session ID, which is unique globally.
Cookie	A 64-bit field. All packets must match the configured Cookie value or be discarded. This field is used in security checks performed at the endpoints of a tunnel to prevent network spoofing and attacks. The local Cookie value must match the remote one. The Cookie field can be dynamically configured.
Payload	Original Layer 2 user packet with the S-Tag or C-Tag removed. The FCS is stripped before encapsulation. A new FCS will be added at each hop when the IP packet is transmitted.

Figure 2 Packet encapsulation format

Ver	Traffic Class	Flow Label
	Payload Length	Next Header(0x73)
Source address (0:31)		
Source address (32:63)		
Source address (64:95)		
Source address (96:127)		
Destination address (0:31)		
Destination address (32:63)		
Destination address (64:95)		
Destination address (96:127)		
Session ID (32 bits)		
Cookie (0:31)		
Cookie (32:63)		
Payload (variable)		
...		
...		

The following table describes the meaning of each field in the packet.

Name	Description
Ver	A 4-bit field used to indicate the version number. The value is set to 6 for IPv6.
Traffic Class	An 8-bit field used to indicate the traffic class. This field functions in a way similar to the ToS field in IPv4.
Flow Label	A 20-bit field used to indicate the flow label. Flow labels are used to differentiate packets at the network layer.
Payload Length	A 16-bit field used to indicate the length of the packet excluding the IPv6 header, that is, the length from the session ID to the end of the packet.

Name	Description
Next Header	An 8-bit field used to identify the type of header immediately following the current header (either basic or extension header). The value is set to 0x73 to indicate that the next header is an L2TPv3 header.
Hop Limit	An 8-bit field used to indicate the hop limit. This field functions in a way similar to the TTL field in IPv4. This field is decremented by one by each node in the path to the egress router. A packet is dropped after this field is decremented to 0. The initial value is 0xFF.
Source Address	A 128-bit field used to indicate the IPv6 source address for the tunnel. The IPv6 source address is a loopback address of the local device.
Destination Address	A 128-bit field used to indicate the IPv6 destination address for the tunnel. The IPv6 destination address is a loopback address of the remote device.
Session ID	A 32-bit field used to indicate the session ID. In a static 1:1 mapping case, the IPv6 address directly resolves to an L2TPv3 session and therefore the session ID can be ignored upon receipt. For compatibility with other tunnel termination platforms, the session ID must be configurable. The session ID of 0 is reserved for use by L2TP control messages.
Cookie	A 64-bit field. All packets must match the configured Cookie value or be discarded. This field is used in security checks performed at the endpoints of a tunnel to prevent network spoofing and attacks. The local Cookie value must match the remote one. The Cookie field can be dynamically configured.
Payload	Original Layer 2 user packet with the S-tag and C-tag removed. The FCS is stripped before encapsulation. A new FCS will be added at each hop when the IP packet is transmitted.

Parent Topic: [Understanding L2TPv3](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

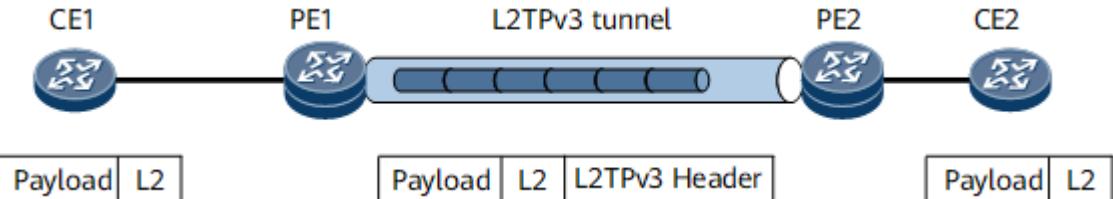
[Next topic >](#)

1.5.2.2 L2TPv3 Fundamentals

L2TPv3 Tunnel

- Access to an L2TPv3 tunnel in whole-interface mode

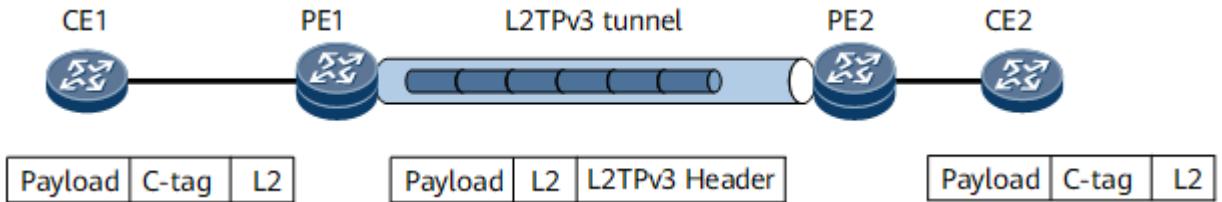
Figure 1 Networking for access to an L2TPv3 tunnel in whole-interface mode



On the network shown in [Figure 1](#), an L2TPv3 tunnel is established between PE1 and PE2. An EVC sub-interface with the default encapsulation mode serves as the L2TPv3 interface on each PE. CE1 and CE2 can exchange either tagged or untagged packets over the L2TPv3 tunnel:

- Processing on the inbound interface: Traffic from CE1 accesses PE1 through PE1's EVC sub-interface.
- Processing on the outbound interface: Traffic from PE2 arrives at CE2 through PE2's EVC sub-interface.
- Access to an L2TPv3 tunnel in C-tag termination mode

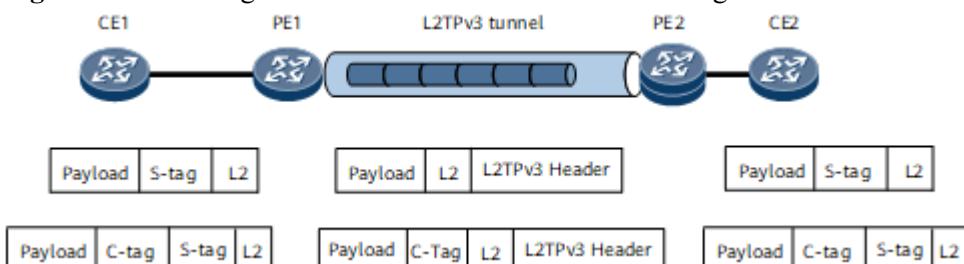
Figure 2 Networking for access to an L2TPv3 tunnel in C-tag termination mode



On the network shown in [Figure 2](#), an L2TPv3 tunnel is established between PE1 and PE2. An EVC sub-interface with the default encapsulation mode serves as the L2TPv3 interface on each PE. CE1 and CE2 can exchange packets carrying only C-tags:

- Processing on the inbound interface: Traffic from CE1 accesses PE1 through PE1's EVC sub-interface. The EVC sub-interface is configured to match user packets carrying only C-tags and strip the C-tags of these packets before forwarding these packets to the L2TPv3 tunnel.
- Processing on the outbound interface: Traffic from PE2 arrives at CE2 through PE2's EVC sub-interface. The EVC sub-interface is configured to add a C-tag to each user packet before forwarding these packets to CE2.
- Access to an L2TPv3 tunnel in S-tag termination mode

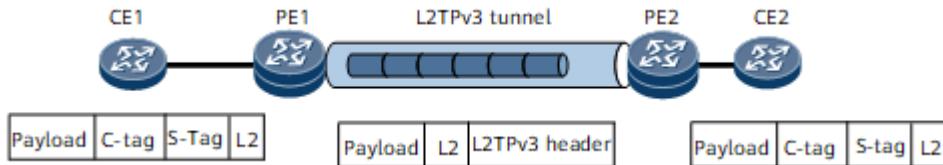
Figure 3 Networking for access to an L2TPv3 tunnel in S-tag termination mode



On the network shown in [Figure 3](#), an L2TPv3 tunnel is established between PE1 and PE2. An EVC sub-interface with the default encapsulation mode serves as the L2TPv3 interface on each PE. CE1 and CE2 can exchange packets carrying S-tags, no matter whether these packets also carry C-tags:

- Processing on the inbound interface: Traffic from CE1 accesses PE1 through PE1's EVC sub-interface. The EVC sub-interface is configured to match user packets carrying S-tags and strip the S-tags before forwarding these packets to the L2TPv3 tunnel.
- Processing on the outbound interface: Traffic from PE2 arrives at CE2 through PE2's EVC sub-interface. The EVC sub-interface is configured to add an S-tag to each user packet before forwarding these packets to CE2.
- Access to an L2TPv3 tunnel in S-tag+C-tag termination mode

Figure 4 Networking for access to an L2TPv3 tunnel in S-tag+C-tag termination mode

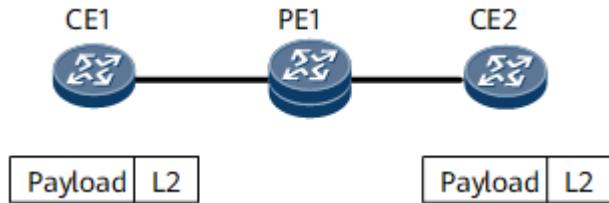


On the network shown in [Figure 4](#), an L2TPv3 tunnel is established between PE1 and PE2. An EVC sub-interface with the default encapsulation mode serves as the L2TPv3 interface on each PE. CE1 and CE2 can exchange packets carrying both C-tags and S-tags:

- Processing on the inbound interface: Traffic from CE1 accesses PE1 through PE1's EVC sub-interface. The EVC sub-interface is configured to match user packets carrying both C-tags and S-tags and strip these tags before forwarding these packets to the L2TPv3 tunnel.
- Processing on the outbound interface: Traffic from PE2 arrives at CE2 through PE2's EVC sub-interface. The EVC sub-interface is configured to add a C-tag and an S-tag to each user packet before forwarding these packets to CE2.

L2TPv3 Local Switching Connection

Figure 5 Networking for an L2TPv3 local switching connection

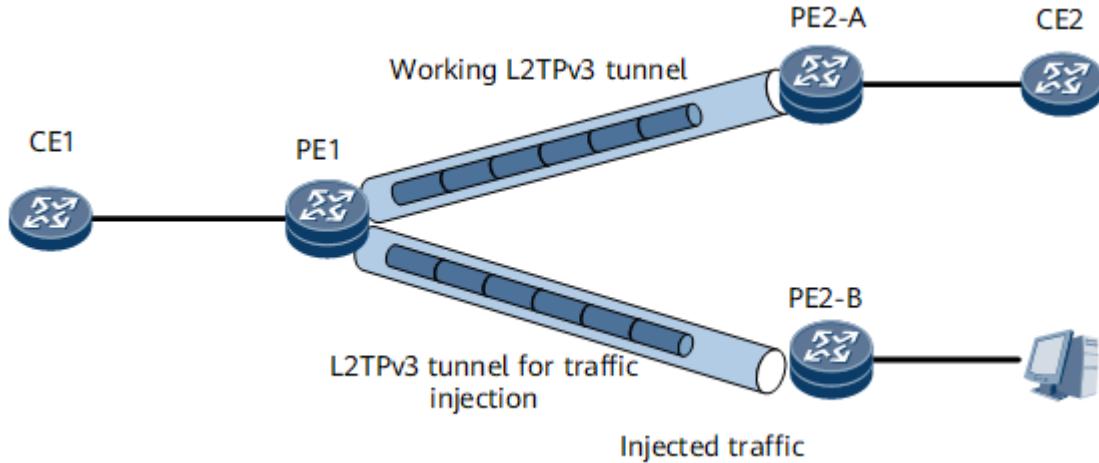


On the network shown in [Figure 5](#), CE1 and CE2 exchange packets through PE1. Two EVC sub-interfaces with the default encapsulation mode serve as L2TPv3 interfaces on PE1. PE1, which only transparently transmits service packets, allows service packets to use any encapsulation type:

- Processing on the inbound interface: Traffic from CE1 accesses PE1 through PE1's EVC sub-interface.
- Processing on the outbound interface: Traffic from PE1 arrives at CE2 through PE1's another EVC sub-interface.

Traffic Injection by an L2TPv3 Tunnel for Another L2TPv3 Tunnel

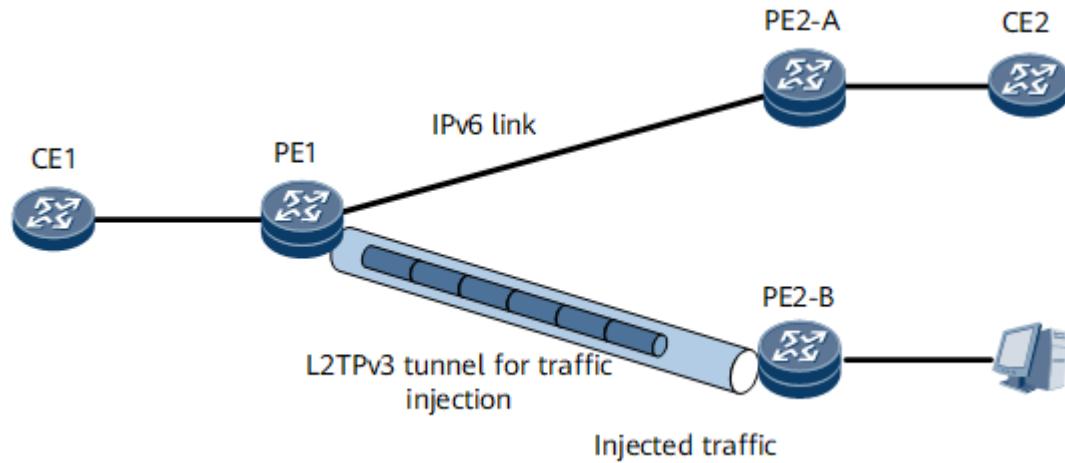
Figure 6 Networking for traffic injection by an L2TPv3 tunnel for another L2TPv3 tunnel



On the network shown in [Figure 6](#), an L2TPv3 tunnel is established between PE1 and PE2-A and between PE1 and PE2-B. CE1 and CE2 exchange packets over the L2TPv3 tunnel between PE1 and PE2-A. The L2TPv3 tunnel between PE1 and PE2-B is used to inject traffic. After PE1 receives injected packets from PE2-B, PE1 strips the L2TPv3 headers of these packets and determines whether the destination MAC addresses of these packets are the same as the MAC address of its AC interface connecting to CE1. If the destination MAC addresses of these packets are the same as the MAC address of the AC interface, PE1 sends these packets to CE1. Otherwise, PE1 sends these packets to CE2 over the L2TPv3 tunnel between itself and PE2-A to test traffic forwarding.

Traffic Injection by an L2TPv3 Tunnel for an IPv4/IPv6 link

Figure 7 Networking for traffic injection by an L2TPv3 tunnel for an IPv4/IPv6 link



On the network shown in [Figure 7](#), an IPv4/IPv6 link is established between PE1 and PE2-A for IPv4/IPv6 traffic forwarding and an L2TPv3 tunnel is established between PE1 and PE2-B to inject traffic. After packets from CE1 arrive at PE1, PE1 forwards these packets over the IPv4/IPv6 link. After PE1 receives injected packets from PE2-B, PE1 strips the L2TPv3 headers of these packets and determines whether the destination MAC addresses of these packets are the same as the MAC address of its AC interface connecting to CE1. If the destination MAC addresses of these packets are the same as the MAC address of the AC interface, PE1 sends these packets to CE1. Otherwise, PE1 sends these packets to PE2-A over the IPv4/IPv6 link between itself and PE2-A to test traffic forwarding.

Parent Topic: [Understanding L2TPv3](#)

Copyright © Huawei Technologies Co., Ltd.

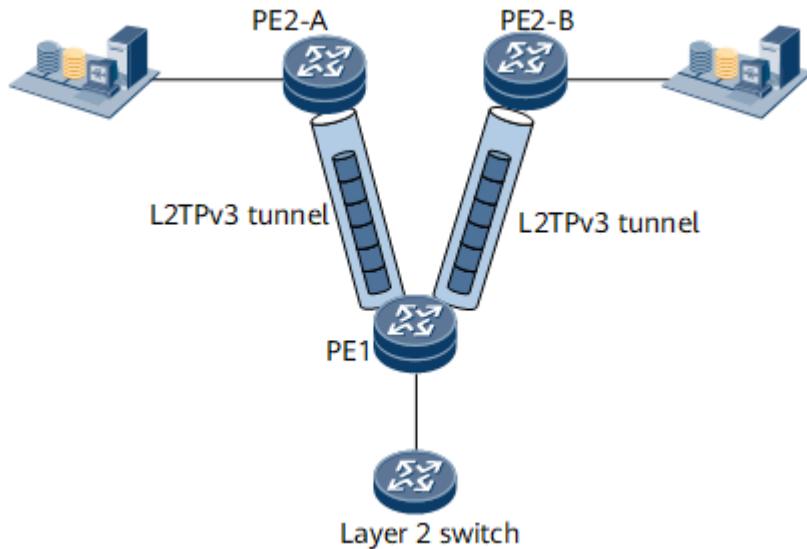
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.5.3 Application Scenarios for L2TPv3

[Figure 1](#) shows a scenario where Layer 2 Ethernet services are transmitted over L2TPv3 tunnels on a public IPv4/IPv6 network. To transmit Layer 2 Ethernet packets to the remote data center servers, an L2TPv3 tunnel is established between PE1 and PE2-A and between PE1 and PE2-B. Tags can be flexibly configured for user services accessing the L2TPv3 tunnels.

Figure 1 Typical L2TPv3 over IPv4/IPv6 scenario



An Ethernet virtual connection (EVC) sub-interface on the downstream interface of PE1 serves as the L2TPv3 interface to provide access for users. This EVC sub-interface uses the default encapsulation type. The L2TPv3 tunnels support the following access modes:

- Whole-interface mode: The EVC sub-interface transparently transmits single-tagged, double-tagged, or untagged packets. In whole-interface access mode, the downstream interface is exclusively used by the EVC sub-interface.
- C-tag termination mode: The EVC sub-interface receives only user packets carrying C-tags and strips the C-tags of these packets before forwarding these packets to an L2TPv3 tunnel for transparent transmission.
- S-tag termination mode: The EVC sub-interface receives only user packets carrying S-tags and strips the S-tags of these packets before forwarding these packets to an L2TPv3 tunnel for transparent transmission.
- S-tag+C-tag termination mode: The EVC sub-interface receives only user packets carrying both C-tags and S-tags and strips the C-tags and S-tags of these packets before forwarding these packets to an L2TPv3 tunnel for transparent transmission.

Parent Topic: [L2TPv3 Description](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.5.4 Terminology for L2TPv3

Acronyms and Abbreviations

Acronym and Abbreviation	Full Name
PW	pseudo wire
L2TPv3	Layer 2 Tunneling Protocol Version 3

Parent Topic: [L2TPv3 Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.6 Tunnel Management

[Overview of Tunnel Management](#)

[Understanding Tunnel Management](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.6.1 Overview of Tunnel Management

Definition

The tunnel management (TNLM) module is used to select a tunnel for an application according to specific configurations and notifies the application of the tunnel's status.

VPN tunnel management covers the introduction to common VPN tunnels and tunnel configuration management.

Common VPN Tunnels

Common VPN tunnels are as follows:

- LSP

LSPs are used as tunnels for VPN data forwarding over the Multiprotocol Label Switching (MPLS) backbone network. On an LSP, only provider edges (PEs) need to analyze IP packet headers. As such, the time to process VPN packets is shortened and the delay in VPN packet transmission is reduced. In addition, MPLS labels are supported by all link layer protocols.

- MPLS TE

With MPLS deployed, carriers are generally required to provide VPN users with end-to-end QoS guarantees for various services, such as the audio, video, mission-critical, and regular Internet access services. In this situation, MPLS TE tunnels can be used to optimize network resources and offer users QoS-guaranteed services.

- GRE

Generic Routing Encapsulation (GRE), which applies to both Layer 2 virtual private networks (L2VPNs) and Layer 3 virtual private networks (L3VPNs). LSPs are usually used as public network tunnels on the MPLS VPN backbone network. However, LSPs cannot be

used as public network tunnels in the scenario where MPLS is supported by PEs, but not by Ps functioning as core devices on the backbone network and providing IP functions. Instead of LSPs, you can use GRE tunnels to provide an L3VPN or L2VPN solution for the backbone network.

- **SR-MPLS TE-Policy**

Segment Routing-MPLS (SR-MPLS) TE Policy is a tunneling technology developed based on SR. An SR-MPLS TE Policy is represented by a set of candidate paths consisting of one or more segment lists, also known as segment ID (SID) lists. Each SID list identifies an end-to-end path from the source to the destination, instructing a device to forward traffic through the path, rather than the shortest path computed using an IGP. If a packet is steered into an SR-MPLS TE Policy, the ingress adds a SID list associated with that policy into the packet, so that other devices on the network can execute the instructions encapsulated into the list.

- **SRv6 TE Policy**

IPv6 Segment Routing (SRv6) TE Policy is a tunneling traffic diversion technology developed based on SRv6. An SRv6 TE Policy is a set of candidate paths consisting of one or more segment lists, that is, segment ID (SID) lists. Each SID list identifies an end-to-end path from the source to the destination, instructing a device to forward traffic through the path, rather than the shortest path computed using an IGP. The header of a packet steered into an SRv6 TE Policy is augmented with an ordered list of segments associated with that SRv6 TE Policy, so that other devices on the network can execute the instructions encapsulated into the list.

Tunnel Configuration and Management

The establishment and management of tunnels vary according to tunnel types. For example, MPLS TE tunnels, including constraint-based routed label switching paths (CR-LSPs) are established and managed using tunnel interfaces, whereas Label Distribution Protocol (LDP) LSPs are automatically created as long as corresponding protocols are configured.

This section focuses on the following aspects:

- Tunnel interface configuration: You can specify a particular tunnel type for each tunnel interface. The configurations of tunnels vary according to tunnel types.
- Tunnel management: The tunnel status is informed to the application that uses a tunnel, and a tunnel policy is provided to select a tunnel. The tunnel policy function is commonly used.

Purpose

Tunnel management allows VPNs to better use the tunneling technology to establish dedicated data transmission channels on the backbone network to transparently transmit packets.

Parent Topic: [Tunnel Management](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.6.2 Understanding Tunnel Management

[Tunnel Policy](#)

[Tunnel Policy Selector](#)

1.6.2.1 Tunnel Policy

A tunnel policy determines which tunnel can be selected for an application.

VPN service forwarding requires tunnels. By default, LSPs are selected for VPN services, multiple LDP LSPs can implement load balancing. If only BGP LSPs exist on a network, only one is selected for VPN services, and load balancing cannot be implemented. If non-LDP LSPs or non-BGP LSPs are needed for VPN service transmission or multiple BGP LSPs or TE tunnels are needed for load balancing, a tunnel policy must be applied to the VPN service.

Tunnel policies can be categorized as either tunnel type-based prioritization policies or tunnel binding policies. The two types of tunnel policies are mutually exclusive.

IPv4 Tunnel Type-based Prioritization Policy

Tunnel type-based priorities determine the sequence in which types of tunnels are selected and the maximum number of tunnels that can participate in load balancing. Tunnels that can be selected in a tunnel type-based prioritization policy include LSPs, GRE tunnels, CR-LSPs, and SR-MPLS TE Policies. Tunnels defined in a tunnel type-based prioritization policy are selected in sequence. The tunnel type specified first is selected as long as the tunnels of this type are up, regardless of whether the tunnel is selected by other services. Generally, the tunnels of a later specified type are not selected except when load balancing is required or when the preceding tunnels are all down.

For example, both the CR-LSPs and LSPs are defined in a tunnel type-based prioritization policy (with the CR-LSPs being defined first), and the maximum number of tunnels that can participate in load balancing defined in the tunnel policy is three. In this situation, the rule for selecting tunnels is as follows:

- CR-LSPs are preferred provided that they are up. If the number of CR-LSPs in the up state is greater than or equal to three, the first three CR-LSPs are selected. If the number of CR-LSPs is less than three, the system selects LSPs, in addition to existing CR-LSPs, to load-balance traffic among the three tunnels.
- Given that LSPs are available and one CR-LSP has already been selected, at most two LSPs can be selected. If no or only one LSP is available, tunnels are selected based on the default tunnel policy. If more than two LSPs are available, one CR-LSP and first two LSPs are selected.

NOTE

- If no tunnel policy is applied to an application or the tunnel policy to be applied has not been created yet, the system selects one available LSP. If no available LSP exists, a local IFNET tunnel is selected.
- If a protection group is configured for a TE tunnel (that is, CR-LSP), the protection tunnel does not participate in selection.
- CR-LSPs include RSVP-TE and SR-MPLS TE tunnels. A tunnel that goes up earlier has a higher priority.
- LSPs include LDP LSPs, BGP LSPs, and SR-LSPs, whose priorities are in descending order. Specifically, if LSPs are used, LDP LSPs are preferentially selected for load balancing. If LDP LSPs are insufficient, the system searches for available BGP LSPs. If LDP and BGP LSPs are insufficient, SR-LSPs are selected.

IPv6 Tunnel Type-based Prioritization Policy

On an IPv6 network, SRv6 TE Policies and SRv6 TE Policy groups are involved in tunnel selection. Tunnels defined in a tunnel type-based prioritization policy are selected in sequence. The tunnels with the type specified first are selected as long as the tunnels of this type are up, regardless of whether the tunnels are selected by other services. Generally, the tunnels of a later specified type are not selected except when the preceding tunnels are all down.

For example, the SRv6 TE Policy and SRv6 TE Policy group are specified in a tunnel policy in sequence. In this situation, the rules for selecting tunnels are as follows:

An available SRv6 TE Policy is preferentially selected. If the status of an SRv6 TE Policy changes to down and no SRv6 TE Policy meets the selection rules, the SRv6 TE Policy group that meets the selection rules is selected.

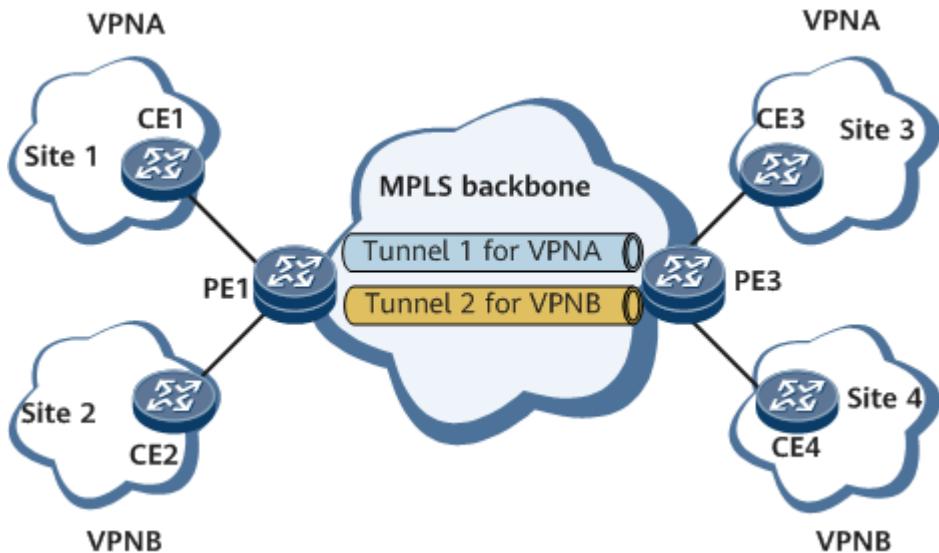
NOTE

- If no tunnel policy is applied to an application module or a nonexistent tunnel policy is applied to the application module, no tunnel is selected by default.
- SRv6 TE Policies and SRv6 TE Policy groups cannot be selected at the same time.

Tunnel Binding Policy

In a tunnel binding policy, you can bind one destination address to a tunnel. Then, VPN services applying the policy will be transmitted over the bound tunnel. The system does not check whether the bound tunnel is a TE tunnel, and the tunnel binding policy takes effect only on TE tunnels. Therefore, ensure that the tunnel binding policy is correctly configured. As shown in [Figure 1](#), two MPLS TE tunnels (tunnel 1 and tunnel 2) are set up between PE1 and PE3.

Figure 1 Application of a tunnel binding policy



If you bind VPNA to tunnel 1 and VPNB to tunnel 2, VPNA and VPNB use separate MPLS TE tunnels. This means that tunnel 1 serves only VPNA and tunnel 2 serves only VPNB. In this manner, services of VPNA and VPNB are isolated from each other and also from other services. The bandwidth for VPNA and VPNB is therefore ensured, which facilitates later QoS deployment.

In tunnel binding, you can bind one destination address to one or more TE tunnels to load-balance services. In addition, you can configure the down-switch attribute to enable other types of tunnels to be selected when the specified tunnels are unavailable, ensuring traffic continuity.

A common tunnel binding policy selects common TE tunnels based on destination addresses and tunnel interface indexes. A tunnel binding policy observes the following tunnel selection rules:

- If the tunnel binding policy does not designate any TE tunnels for a destination IP address, an available tunnel is selected based on the default tunnel policy.
- If the tunnel binding policy designates several TE tunnels for a destination IP address and more than one designated TE tunnel is available, one of the available TE tunnels is selected.
- If the tunnel binding policy designates several TE tunnels for the destination IP address but none of the designated TE tunnels is available, tunnel selection is determined by the down-switch attribute. If the down-switch attribute is not configured, no tunnels are selected. If the down-switch attribute is configured, an available tunnel is selected based on the default tunnel policy.

Comparison of Tunnel Policies

Table 1 Comparison of tunnel policies

Policy	Description
Tunnel type-based prioritization policy	Cannot ensure which tunnel is selected if there are several tunnels of the same type.
Tunnel binding policy	Accurately defines which TE tunnel can be used, ensuring QoS. This function is valid only for TE tunnels.

Parent Topic: [Understanding Tunnel Management](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.6.2.2 Tunnel Policy Selector

Principles

On a BGP/MPLS IP VPN in inter-AS VPN Option B or inter-AS VPN Option C mode, the VPN routes that an autonomous system boundary router (ASBR) or PE receives recurse only to LDP LSPs within an AS. Recursion is considered failed so long as LDP LSPs do not exist, no matter whether other types of tunnels exist. This implementation strictly confines the types of recursive tunnels, making network deployment inflexible. In addition, customers cannot use MPLS TE channels to guarantee the transmission quality by means of traffic engineering. To break the restriction of tunnel types, tunnel policy selectors are introduced.

Tunnel policy selectors achieve on-demand recursion by matching the route distinguisher (RD) and next hop of a route, facilitating tunnel selection. Tunnel policy selectors can use various tunnel policies for VPN routes to recurse to different types of tunnels, better meeting customer requirements.

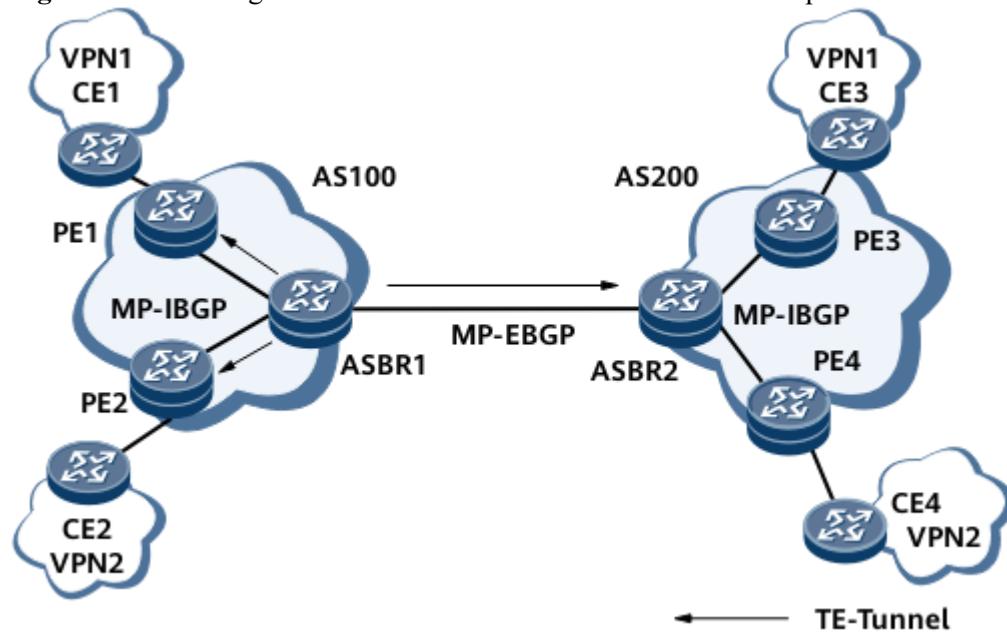
[Figure 1](#) shows the networking diagram for a BGP/MPLS IP VPN in inter-AS VPN Option B mode.

If no tunnel policy selector is configured on ASBR1, the VPN routes received by ASBR1 can only recurse to the LDP LSP between ASBR1 and PE1, PE2, or ASBR2.

After a tunnel policy selector is configured on ASBR1, the VPN routes received by ASBR1 can recurse to any type of tunnel between ASBR1 and PE1, PE2, or ASBR2. This implementation allows

flexible networking. After you configure tunnel policy selectors to select TE tunnels for route recursion, the bandwidth for data transmission can be ensured.

Figure 1 Networking for a BGP/MPLS IP VPN in inter-AS VPN Option B mode



Implementation

A tunnel policy selector consists of one or more nodes, and the relationship between these nodes is "OR". The system checks the nodes according to index numbers. If a route matches a node in the tunnel policy selector, the route stops the matching process.

Each node comprises a set of if-match and apply clauses:

- The if-match clauses define the matching rules that are used to match certain route attributes, such as the next hop and RD. The relationship between the if-match clauses of a node is "AND". A route matches a node only when the route meets all the matching rules specified by the if-match clauses of the node.
- The apply clauses specify actions. When a route matches a node, the apply clauses select a corresponding tunnel policy for the route. This tunnel policy can select other types of tunnels to carry services by means of prioritizing or tunnel binding.

The node matching modes of a tunnel policy selector are as follows:

- Permit: If a route matches all the if-match clauses of a node, the route matches the tunnel policy selector and all the actions defined by apply clauses are performed on the route. If a route does not match any if-match clauses of a node, the route continues to match the next node.
- Deny: In this mode, the apply clauses are not implemented. If a route meets all the if-match clauses of the node, the route is denied and no longer matches other nodes of the tunnel policy selector.

Usage Scenario

Tunnel policy selectors apply to BGP/MPLS IP VPNs in inter-AS VPN Option B or inter-AS VPN Option C mode.

Benefits

Tunnel policy selectors offer the following benefits:

- Routes can recurse to other types of tunnels besides LDP LSPs, allowing flexible networking.
- MPLS TE tunnels can be used for route recursion to support QoS.

Parent Topic: [Understanding Tunnel Management](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.7 BGP/MPLS IP VPN Description

[Overview of BGP/MPLS IP VPN](#)

[Understanding BGP/MPLS IP VPN](#)

[Application Scenarios for BGP/MPLS IP VPN](#)

This section describes different applications of BGP/MPLS IP VPN.

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

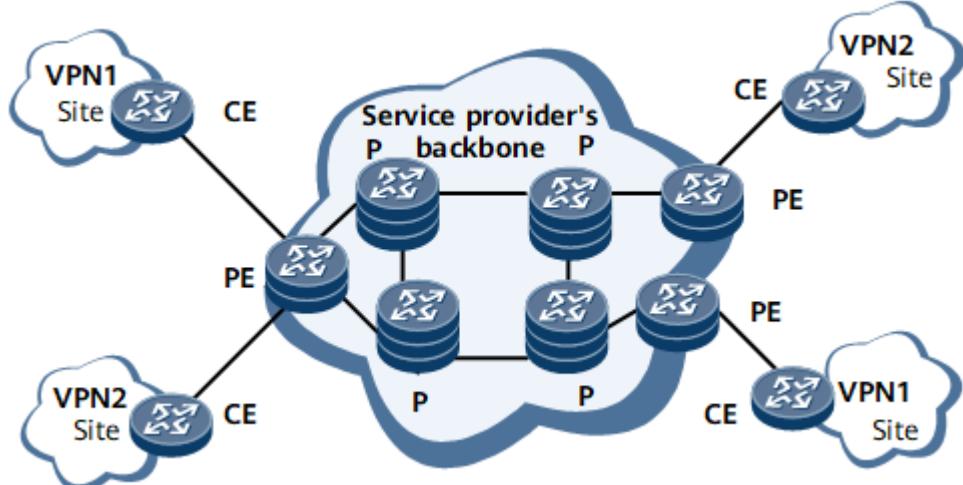
[< Previous topic](#) [Next topic >](#)

1.7.1 Overview of BGP/MPLS IP VPN

Definition

A BGP/MPLS IP VPN is a Layer 3 virtual private network (L3VPN), which uses BGP to advertise VPN routes and uses MPLS to forward VPN packets on the IP backbone networks of service providers (SPs).

Figure 1 BGP/MPLS IP VPN



As shown in [Figure 1](#), a BGP/MPLS IP VPN consists of the following roles:

- CE: An edge device on a customer network. A CE provides interfaces that are directly connected to the SP network. A CE can be a router, a switch, or a host. Usually, a CE is unaware of the VPN and does not need to support MPLS.
- PE: An edge device on an SP network. A PE is directly connected to a CE. On an MPLS network, PEs process all VPN services. The requirements on the performance of PEs are rather high.
- P: A backbone device on an SP network. A P is not directly connected to a CE. Ps only need to possess basic MPLS forwarding capabilities and do not maintain VPN information.

PEs and Ps are managed by SPs. CEs are managed by users, except that the users trust SPs with the management rights.

A PE can connect to multiple CEs. A CE can connect to multiple PEs of the same SP or of different SPs.

Purpose

- MPLS seamlessly integrates the flexibility of IP routing and simplicity of ATM label switching. A connection-oriented control plane is introduced into an MPLS IP network, which enriches the means of managing and operating the network. On IP networks, MPLS TE has become an important tool in managing network traffic, reducing network congestion, and ensuring QoS.

The VPNs using MPLS IP networks as the backbone networks are highly valued by carriers, and have become an important means of providing value-added services.

- Unlike the IGP, BGP focuses on controlling route transmission and choosing optimal routes instead of discovering and calculating routes. VPNs use public networks to transmit VPN data, and the public networks use an IGP to discover and calculate their routes. The key to constructing a VPN is to control the transmission of VPN routes and choose the optimal routes between two PEs.

BGP uses TCP (with port number 179) as the transport layer protocol, enhancing transmission reliability. VPN routes can be directly exchanged between two PEs with routers located between them.

BGP can append any information to a route as optional BGP attributes. The information is transparently forwarded by BGP devices that cannot identify those attributes. Therefore, VPN routes can be conveniently transmitted between PEs.

When routes are updated, BGP sends only updated routes rather than all routes. This implementation saves the bandwidth consumed by route transmission, making the transmission of a great number of routes over a public network possible.

As an Exterior Gateway Protocol (EGP), BGP is best suited for VPNs that cross the networks of multiple carriers.

Parent Topic: [BGP/MPLS IP VPN Description](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.7.2 Understanding BGP/MPLS IP VPN

[Basic BGP/MPLS IP VPN Fundamentals](#)

[Hub & Spoke](#)

[MCE](#)

[Inter-AS VPN](#)

[Carrier's Carrier](#)

[HVPN](#)

[BGP/MPLS IP VPN Label Allocation Modes](#)

[BGP SoQ](#)

[Route Import Between VPN and Public Network](#)

[VPN FRR](#)

[VPN GR](#)

[VPN NSR](#)

[BGP/MPLS IPv6 VPN Extension](#)

[VPN Dual-Stack Access](#)

[VPN MPLS/VPN SRv6 Dual-Stack Tunnel](#)

Parent Topic: [BGP/MPLS IP VPN Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

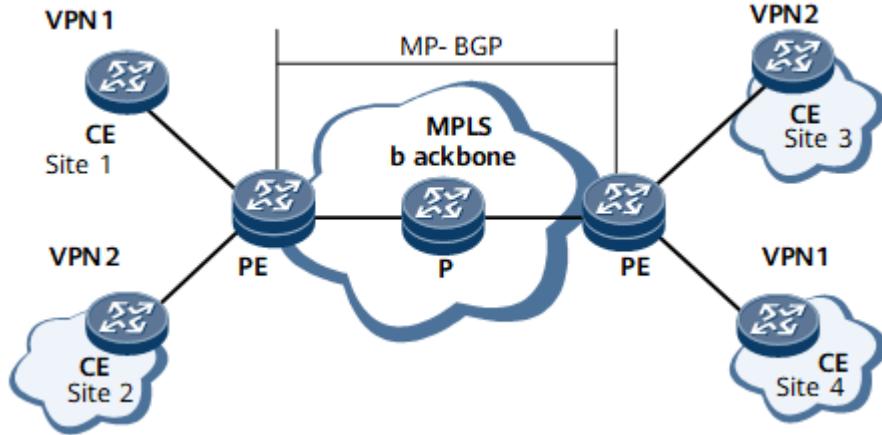
1.7.2.1 Basic BGP/MPLS IP VPN Fundamentals

Definition

As shown in [Figure 1](#), a basic BGP/MPLS VPN applies to the scenario in which there is only one carrier or the backbone networks of multiple carriers belong to the same AS. A basic BGP/MPLS IP VPN has the following characteristics:

- Transmits packets using extended BGP.
- Encapsulates and transmits VPN packets over MPLS LSPs serving as public network tunnels.
- Allows a device that can play PE, P, and CE roles to play only one role at a time.

Figure 1 Basic BGP/MPLS IP VPN networking



Related Concepts

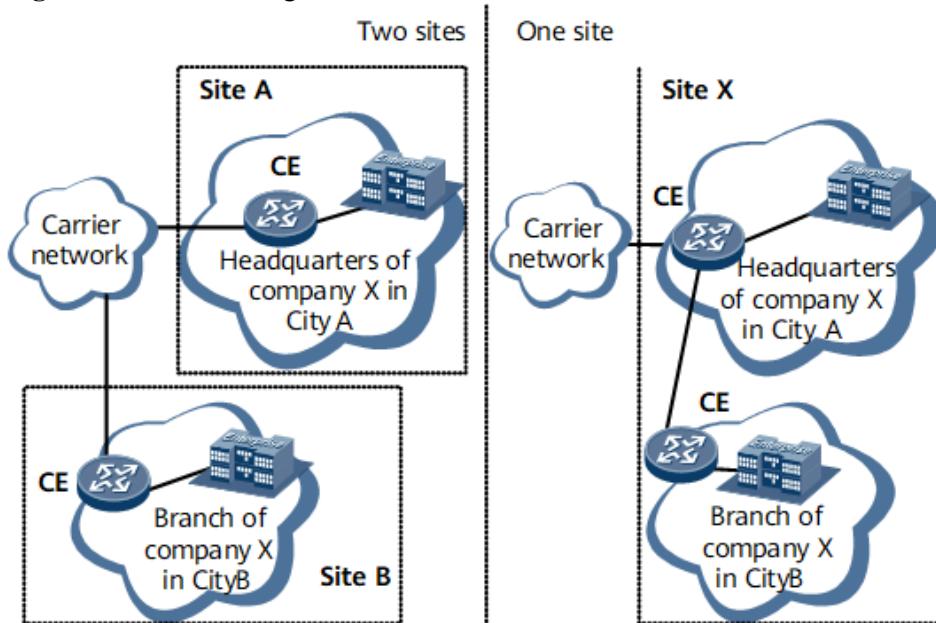
- Site

The concept of "site" is frequently mentioned in the VPN technology. The following describes a site from different aspects:

- A site is a group of IP systems that can communicate without using carrier networks.

As shown in [Figure 2](#), on the networks of the left side, the headquarters network of company X in City A is a site; the branch network of company X in City B is another site. IP devices within each site can communicate without using the SP network.

Figure 2 Schematic diagram for sites



- Sites are classified based on topological relationships between devices rather than the geographical locations of devices, even though devices in a site are geographically adjacent to each other in general. If two geographically separated IP systems are connected over a leased line instead of a carrier network, the two systems compose a site.

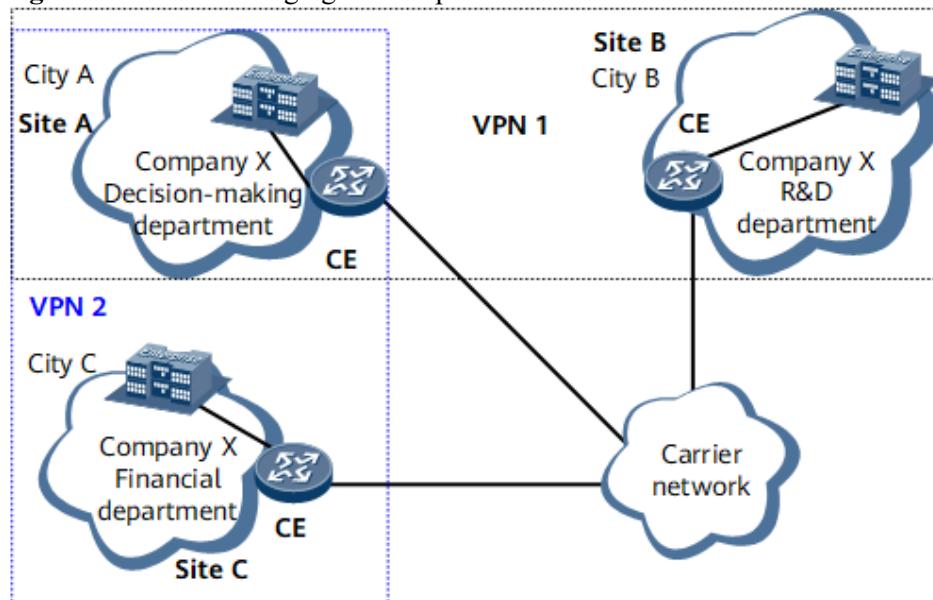
As shown in [Figure 2](#), if the branch network in City B connects to the headquarters network in City A over a leased line instead of a carrier network, the branch

network and the headquarters network compose a site.

- Devices at a site can belong to multiple VPNs. In other words, a site can belong to more than one VPN.

As shown in [Figure 3](#), the decision-making department of company X in City A (Site A) is allowed to communicate with the R&D department in City B (Site B) and the financial department in City C (Site C). Site B and Site C are not allowed to communicate with each other. In this case, VPN1 and VPN2 can be established, with Site A and Site B belonging to VPN1 and Site A and Site C belonging to VPN2. In this manner, Site A is configured to belong to multiple VPNs.

Figure 3 One site belonging to multiple VPNs



- A site connects to a carrier network through the CE and may contain more than one CE, but a CE belongs only to one site.

It is recommended that you determine the devices to be used as CEs based on the following principles:

If the site is a host, use the host as the CE.

If the site is a subnet, use switches as CEs.

If the site comprises multiple subnets, use routers as CEs.

Sites connecting to the same carrier's network can be categorized into different sets based on configured policies. Only sites that belong to the same set can access each other, and this set is a VPN.

- Address space overlapping

As a private network, a VPN independently manages an address space. Address spaces of different VPNs may overlap. For example, if both VPN1 and VPN2 use addresses on network segment 10.110.10.0/24, address space overlapping occurs.

NOTE

VPNs can use overlapped address spaces in the following situations:

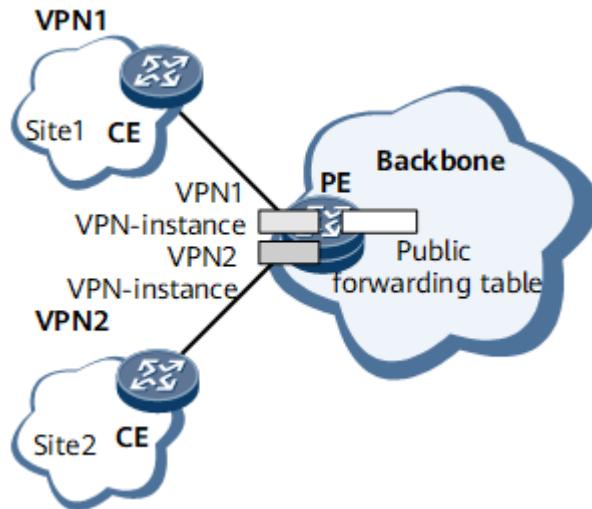
- Two VPNs do not cover the same site.
- Two VPNs cover the same site, but devices at the site and devices using addresses in overlapped address spaces in the VPNs do not access each other.

- VPN instance

CEs are user-side devices and need to send only local VPN routes to PEs, irrespective of whether the PEs connect to the public network or other VPNs. PEs are network-side devices, and a PE generally connects to multiple CEs from different VPNs. A PE may receive routes from different VPNs. Because address spaces used by different VPNs may overlap, routes sent from different VPNs may carry the same destination address. If a PE maintains only one routing and forwarding table, this table will accept only one of the routes from different VPNs but with the same destination address. To prevent this problem, the VPN technology uses VPN instances.

A VPN instance is also called a VPN routing and forwarding (VRF) table. A PE maintains multiple routing and forwarding tables, including a public routing and forwarding table and one or more VRF tables. In other words, a PE has multiple instances, including a public network instance and one or more VPN instances, as shown in [Figure 4](#). Each VPN instance maintains routes from the corresponding VPN. The public network instance maintains public network routes. This enables a PE to keep all routes from VPNs, irrespective of whether their address spaces overlap.

Figure 4 Schematic diagram for VPN instances



The differences between a public routing and forwarding table and a VRF table are as follows:

- A public routing table contains the IPv4 routes of all PEs and Ps. These IPv4 routes are static routes configured on the backbone network or are generated by routing protocols configured on the backbone network.
- A VPN routing table contains the routes of all sites that belong to the corresponding VPN instance. The routes are obtained through exchange of VPN routes between PEs or between CEs and PEs.
- Based on route management policies, a public forwarding table contains the minimum forwarding information extracted from the corresponding routing table, whereas a VPN forwarding table contains the minimum forwarding information extracted from the corresponding VPN routing table.

The VPN instances on a PE are independent of each other. They are also independent of the public routing and forwarding table.

Each VPN instance can be considered as a virtual router, which maintains an independent address space and has one or more interfaces connected to the router.

In relevant standards (BGP/MPLS IP VPNs), a VPN instance is called a per-site forwarding table. As the name suggests, one VPN instance corresponds to one site. To be specific, every connection between a CE and a PE corresponds to a VPN instance, but this is not a one-to-one mapping. The VPN instance is manually bound to the PE interface that directly connects to the CE.

A VPN instance uses an RD to identify an independent address space and uses VPN targets to manage VPN memberships and routing principles of directly connected sites and remote sites.

- Relationships between VPNs, sites, and VPN instances

The relationships between VPNs, sites, and VPN instances are as follows:

- A VPN consists of multiple sites. A site may belong to multiple VPNs.
- A site is associated with a VPN instance on a PE. A VPN instance integrates the VPN member relationships and routing rules of its associated sites. Multiple sites form a VPN based on VPN instance rules.

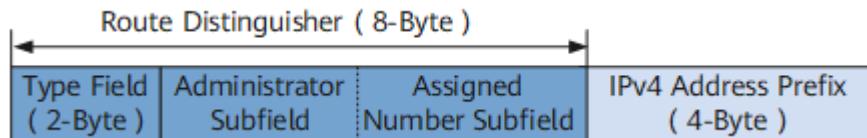
- RD and VPN-IPv4 address

Traditional BGP cannot process the routes of VPNs with overlapped address spaces. Assume that VPN1 and VPN2 use addresses on the network segment 10.110.10.0/24, and each of them advertises a route destined for this network segment. The local PE identifies the two VPN routes based on VPN instances and sends them to the remote PE. Because routes from different VPNs cannot work in load-balancing mode, the remote PE adds only one of the two routes to its VRF table based on BGP route selection rules.

This is because BGP cannot distinguish VPN routes with the same IP address prefix. To solve this problem, BGP/MPLS IP VPN uses the VPN-IPv4 address family.

A VPN-IPv4 address consists of 12 bytes. The first eight bytes represent the RD and the last four bytes represent the IPv4 address prefix, as shown in [Figure 5](#).

Figure 5 VPN-IPv4 address structure



RDs are used to distinguish IPv4 prefixes using the same address space. The format of RDs enables carriers to allocate RDs independently. An RD, however, must be unique on the entire network to ensure correct routing if CEs are dual-homed to PEs. IPv4 addresses with RDs are called VPN-IPv4 addresses. After receiving IPv4 routes from a CE, a PE converts the routes to globally unique VPN-IPv4 routes and advertises the routes on the public network.

- VPN target

The VPN target, also called the route target (RT), is a 64-bit BGP extended community attribute. BGP/MPLS IP VPN uses VPN targets to control the advertisement of VPN routing information.

A VPN instance is associated with one or more VPN targets, which are of the following types:

- Export VPN target: After learning an IPv4 route from a directly connected site, a PE converts the route to a VPN-IPv4 route and sets the export VPN target for the route. As an extended community attribute, the export VPN target is advertised with the route.

- Import VPN target: After receiving a VPN-IPv4 route advertised by another PE, the local PE checks the export VPN target of the route. If the export VPN target is identical with the import VPN target of a VPN instance on the PE, the PE adds the route to the VPN instance.

The VPN target defines the sites that can receive a VPN route, and the sites from which the PE can receive routes.

After receiving a route from a directly connected CE, a PE sets the export VPN targets of the route. The PE then uses BGP to advertise the route with export VPN targets to related PEs. After receiving the route, the related PEs compare the export VPN targets with the import VPN targets of all their VPN instances. If an export VPN target is identical with an import VPN target, the route is added to the corresponding VPN instance.

The reasons for using VPN targets instead of RDs as the extended community attributes are as follows:

- A VPN-IPv4 route has only one RD, but can be associated with multiple VPN targets. With multiple extended community attributes, BGP can greatly improve network flexibility and expansibility.
- VPN targets are used to control route advertisement between different VPNs on a PE. After being configured with matching VPN targets, different VPN instances on a PE can import routes from each other.

On a PE, different VPNs have different RDs, but the extended community attributes allowed by BGP are limited. Using RDs for route importing limits network expansibility.

On a BGP/MPLS IP VPN, VPN targets can be used to control exchange of VPN routes between sites. Export VPN targets and import VPN targets are independent of each other and can be configured with multiple values, ensuring flexible VPN access control and diversified VPN networking modes.

- Multiprotocol Border Gateway Protocol (MP-BGP)

Traditional BGP-4 standards can manage only IPv4 routing information, and cannot process VPN routes with overlapping address spaces.

To correctly process VPN routes, VPNs use MP-BGP defined in relevant standards (Multiprotocol Extensions for BGP-4). MP-BGP supports multiple network layer protocols. Network layer protocol information is contained in the Network Layer Reachability Information (NLRI) field and the Next Hop field of an MP-BGP Update message.

MP-BGP uses the address family to differentiate network layer protocols. An address family can be a traditional IPv4 address family or any other address family, such as a VPN-IPv4 address family or an IPv6 address family. For the values of address families, see relevant standards (Assigned Numbers).

Route Advertisement on a Basic BGP/MPLS IP VPN

On a basic BGP/MPLS IP VPN, CEs and PEs are responsible for advertising VPN routes, whereas Ps only need to maintain backbone network routes without knowing VPN routing information. Generally, a PE maintains the routes of VPNs that the PE accesses, rather than all VPN routes.

VPN route advertisement consists of the following phases:

- Route advertisement from the local CE to the ingress PE
- Route advertisement from the ingress PE to the egress PE
- Route advertisement from the egress PE to the remote CE

After the process of route advertisement is complete, the local and remote CEs can set up reachable routes, and VPN routing information can be advertised on the backbone network.

The following describes the three phases of route advertisement in detail:

1. Route advertisement from the local CE to the ingress PE

After the peer relationship is set up between a CE and the directly connected PE, the CE advertises the local IPv4 routes to the PE. The CE can communicate with the PE over static routes or routes established using Routing Information Protocol (RIP), Open Shortest Path First (OSPF), Intermediate System to Intermediate System (IS-IS), or BGP. Routes advertised by the CE to the PE are standard IPv4 routes, regardless of which routing protocol is used.

[VPN instances](#) on a PE are isolated from each other and independent of the public routing and forwarding table, so as to prevent problems caused by [address space overlapping](#). After learning routes from CEs, a PE decides to which routing and forwarding table the routes should be installed.

2. Route advertisement from the ingress PE to the egress PE

Route advertisement from the ingress PE to the egress PE consists of the following phases:

- After learning VPN routes from a CE, a PE stores these routes in corresponding VRFs and adds [RDs](#) to these standard IPv4 routes. The VPN-IPv4 routes are then generated.
- The ingress PE advertises VPN-IPv4 routes to the egress PE by sending [MP-BGP](#) Update messages. The MP-BGP Update messages also contain [VPN targets](#) and MPLS labels.

Before the next-hop PE receives the VPN-IPv4 routes, the routes are first filtered by BGP routing policies, including the export policy configured on the VPN instance and the peer export policy.

After these routes arrive at the egress PE, if they match the BGP peer import policy and their next hops are reachable or they can perform recursion, the egress PE performs local route leaking and filters these routes based on a VRF import policy. The egress PE then decides which routes are to be added to its VPN routing tables. Routes received from other PEs are added to a VPN routing table based on [VPN targets](#). The egress PE stores the following information for subsequent packet forwarding:

- Values of MPLS labels contained in MP-BGP Update messages
- Tunnel IDs generated after tunnel recursion

3. Route advertisement from the egress PE to the remote CE

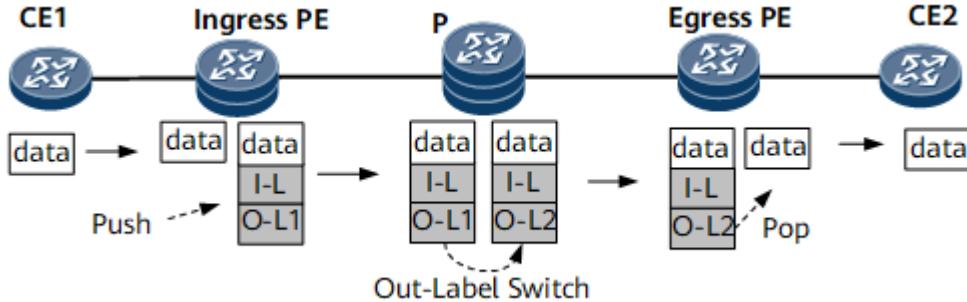
A remote CE can learn VPN routes from an egress PE over static routes or routes established using RIP, OSPF, IS-IS, or BGP. Route advertisement from the egress PE to a remote CE is similar to that from a local CE to the ingress PE. The details are not described here. Note that the routes advertised by the egress PE to the remote CE are standard IPv4 routes.

After a PE receives routes of different VPNs from a local CE, if the next hops of these routes are reachable or these routes can perform recursion, the PE matches the export VPN targets of these routes with the import VPN targets of its local VPN instances. This process is called local route leaking. During local route leaking, the PE filters these routes based on a VRF import policy and modifies the attributes of eligible routes.

Packet Forwarding on a BGP/MPLS IP VPN

On a BGP/MPLS IP VPN backbone network, a P does not know VPN routing information. VPN packets are forwarded between PEs over tunnels. [Figure 6](#) shows an example of packet forwarding on a BGP/MPLS IP VPN. A packet is transmitted from CE1 to CE2. I-L indicates an inner label, and O-L indicates an outer label. The outer label directs the packet to the BGP next hop, and the inner label identifies the outbound interface for the packet or the VPN to which the packet belongs.

Figure 6 Forwarding of a VPN packet from CE1 to CE2



The forwarding process is as follows:

1. CE1 sends a VPN packet to the ingress PE.
2. After receiving the packet from an interface bound to a VPN instance, the ingress PE performs the following steps:
 - Searches the corresponding VPN forwarding table based on the RD of the bound VPN instance.
 - Matches the destination IPv4 address with forwarding entries and searches for the corresponding tunnel ID.
 - Adds an I-L to the packet and finds the tunnel to be used based on the tunnel ID.
 - Adds an outer label to the packet and sends the packet over the tunnel. In this example, the tunnel is an LSP, and the outer label is an MPLS label (O-L1).
 - Transmits the double-tagged packet over the backbone network. Each P on the forwarding path swaps the outer label of the packet.
3. After receiving the packet with two labels, the egress PE sends the packet to MPLS for processing. MPLS removes the outer label.

NOTE

In this example, the final outer label of the packet is O-L2. If PHP is configured, O-L2 is removed on the penultimate hop, and the egress PE receives a packet with the inner label only.

4. The egress PE removes the inner label residing at the bottom of the label stack.
5. The egress PE sends the packet from the corresponding outbound interface to CE2. After its labels are removed, the packet becomes a native IP packet.

In this manner, the packet is sent from CE1 to CE2. CE2 forwards the packet to the destination in the way it sends other IP packets.

Benefits

BGP/MPLS IP VPN offers the following benefits:

- Enables users to communicate with each other over networks of geographically different regions.
- Ensures the security of VPN user data during transmission over the public network.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.7.2.2 Hub & Spoke

The Hub & Spoke networking can be used to enable an access control device on a VPN to control the mutual access of other users. The site where the access control device locates is called a Hub site, and other sites are called Spoke sites. At the Hub site, a device that accesses the VPN backbone network is called a Hub-CE; at a Spoke site, a device that accesses the VPN backbone network is called a Spoke-CE. On the VPN backbone network, a device that accesses the Hub site is called a Hub-PE; a device that accesses a Spoke site is called a Spoke-PE.

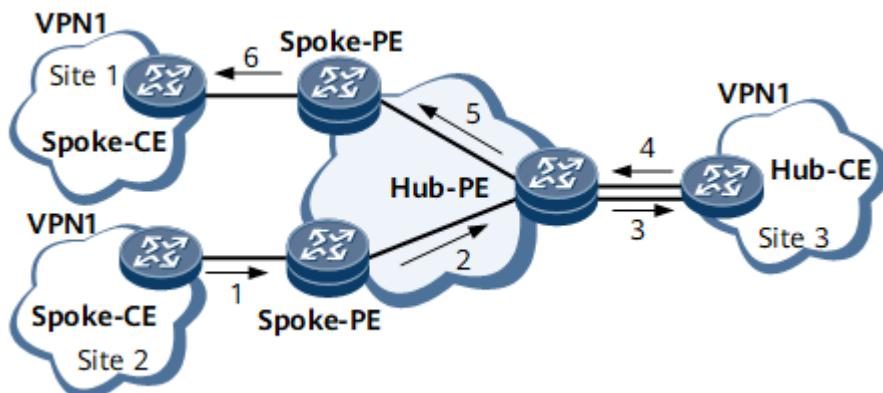
A Spoke site advertises routes to the Hub site, and the Hub site then advertises the routes to other Spoke sites. No direct route exists between the Spoke sites. The Hub site controls the communication between the Spoke sites.

In the Hub & Spoke networking model, two VPN targets are configured to stand for Hub and Spoke respectively.

The configuration of a VPN target on a PE must comply with the following rules:

- The export target and the import target of the Spoke-PE at a Spoke site are Spoke and Hub respectively. The import target of a Spoke-PE is different from the export targets of other Spoke-PEs.
- A Hub-PE requires two interfaces or sub-interfaces. One interface or sub-interface receives routes from Spoke-PEs, and the import target of the VPN instance on the interface is Spoke. The other interface or sub-interface advertises the routes to Spoke-PEs, and the export target of the VPN instance on the interface is Hub.

Figure 1 Route advertisement from Site 2 to Site 1 in Hub & Spoke networking



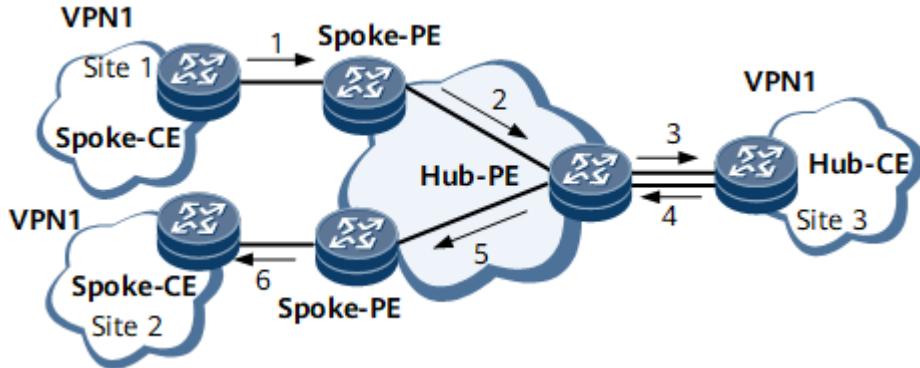
As shown in [Figure 1](#), the communication between Spoke sites is controlled by the Hub site. The lines with arrowheads show the process of advertising a route from Site 2 to Site 1.

- The Hub-PE can receive the VPN-IPv4 routes advertised by all the Spoke-PEs.

- All the Spoke-PEs can receive the VPN-IPv4 routes advertised by the Hub-PE.
- The Hub-PE advertises the routes learned from the Spoke-PEs to the Hub-CE, and advertises the routes learned from the Hub-CE to all the Spoke-PEs. The Spoke sites can access each other through the Hub site.
- The import target of a Spoke-PE is different from the export targets of other Spoke-PEs. Two Spoke-PEs cannot directly advertise VPN-IPv4 routes to each other. As a result, the Spoke sites cannot access each other.

The transmission path between Site 1 and Site 2 is shown in [Figure 2](#). The lines with arrowheads indicate the path from Site 2 to Site 1.

Figure 2 Path of data transmission from Site 1 to Site 2



Networking Description

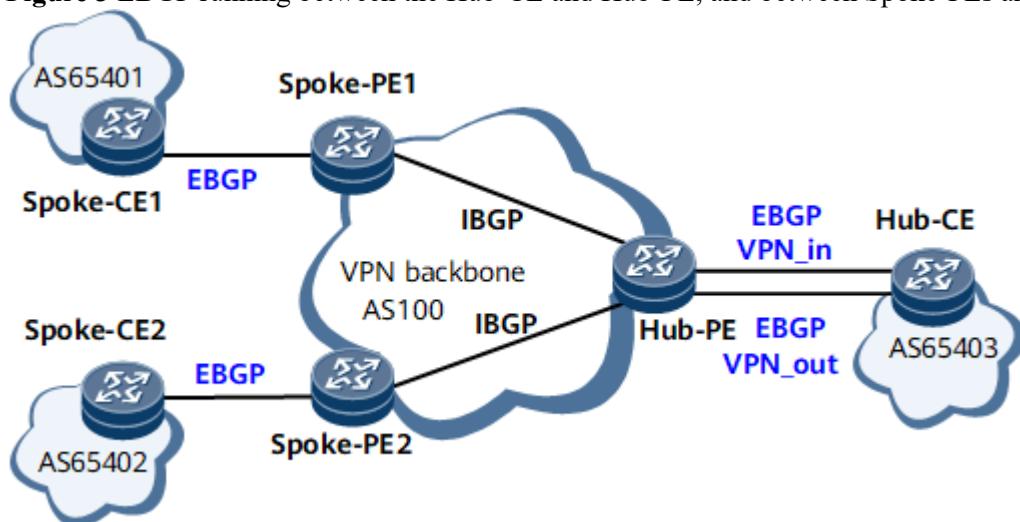
Hub & Spoke networking schemes include:

- External Border Gateway Protocol (EBGP) running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs
- IGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs
- EBGP running between the Hub-CE and Hub-PE, and IGP running between Spoke-PEs and Spoke-CEs

The following describes these networking schemes in detail:

- EBGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs

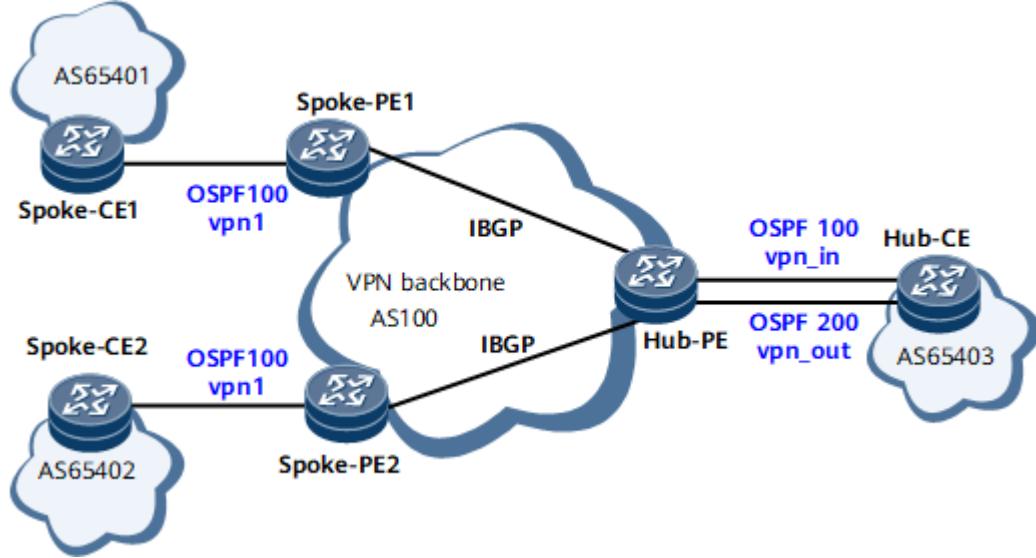
Figure 3 EBGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs



As shown in [Figure 3](#), the routing information advertised by a Spoke-CE is forwarded to the Hub-CE before being transmitted to other Spoke-PEs. If EBGP runs between the Hub-PE and Hub-CE, the Hub-PE performs the AS-Loop check on the route. If the Hub-PE detects its own AS number in the route, it discards the route. In this case, to implement the Hub & Spoke networking, the Hub-PE must be configured to permit the existence of repeated local AS numbers.

- IGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs

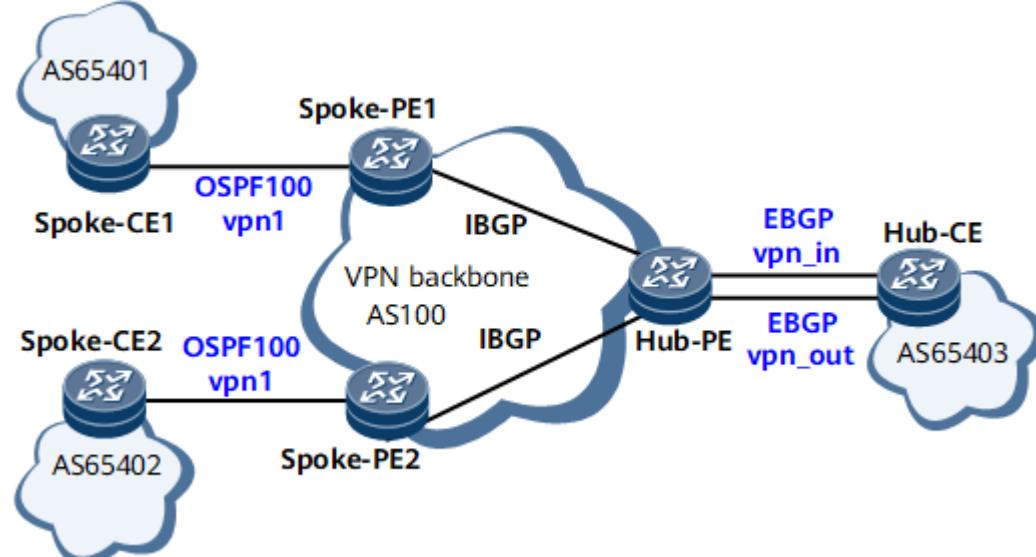
Figure 4 IGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs



Because all PEs and CEs exchange routing information through IGP and IGP routes do not contain the AS_Path attribute, the AS_Path field of BGP VPNv4 routes is null.

- EBGP running between the Hub-CE and Hub-PE, and IGP running between Spoke-PEs and Spoke-CEs

Figure 5 EBGP running between the Hub-CE and Hub-PE, and IGP running between Spoke-PEs and Spoke-CEs



The networking topology is similar to that shown in [Figure 3](#). The AS_Path attribute of the route forwarded by the Hub-CE to the Hub-PE contains the AS number of the Hub-PE. Therefore, the Hub-PE must be configured to permit the existence of repeated local AS numbers.

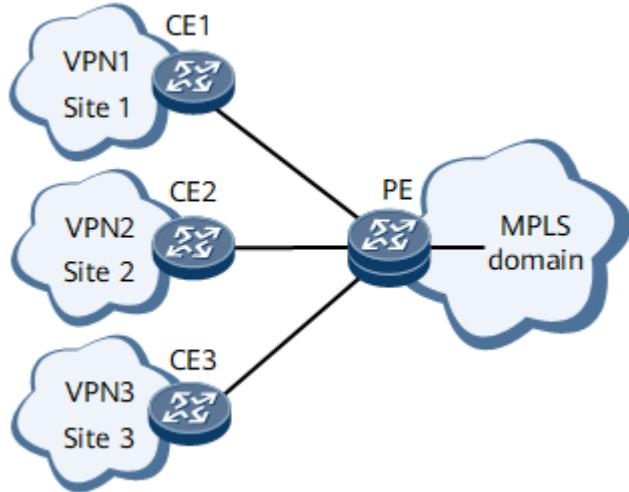
1.7.2.3 MCE

Background

The multi-VPN-instance customer edge (MCE) technology provides logically independent VPN instances and address spaces on a CE, allowing multiple VPN users to share the same CE. The MCE technology provides an economical and easy-to-use solution to solve problems concerned with VPN service isolation and security.

VPN services are becoming increasingly refined, and the demand for VPN service security is growing. Carriers must isolate different types of VPN services on networks to meet this demand. As shown in [Figure 1](#), the traditional BGP/MPLS IP VPN technology isolates VPN services by deploying one CE for each VPN, bringing in high costs and complicated network deployment. If multiple VPNs use the same CE to access upper-layer devices, these VPNs will share the same routing and forwarding table, and data security for these VPNs cannot be ensured. The MCE technology addresses conflicts between network costs and data security problems caused by multiple VPNs sharing the same CE.

Figure 1 Networking diagram for VPN service isolation using BGP/MPLS IP VPN

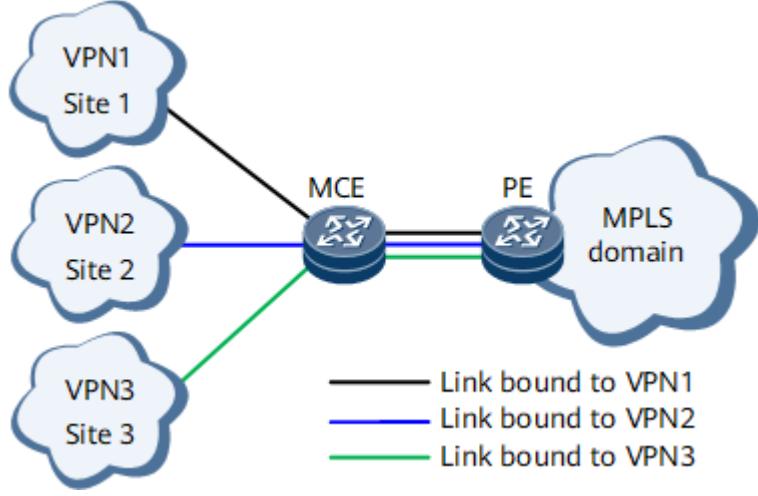


Implementation

The MCE technology creates a VPN instance for each VPN service to be isolated. Each VPN uses an independent routing protocol to communicate with the MCE to which these VPNs are connected. A VPN instance is bound to each link between the MCE and the PE to which the MCE is bound. As a result, an independent channel is established for each VPN service, and different VPN services are isolated.

As shown in [Figure 2](#), three VPN instances are configured on the MCE: VPN1, VPN2, and VPN3. To be specific, three independent VPN routing and forwarding tables are created on the MCE. VPN1 is bound to the link between the MCE and Site1 and a link between the MCE and PE, VPN2 is bound to the link between the MCE and Site2 and a link between the MCE and PE, and VPN3 is bound to the link between the MCE and Site3 and a link between the MCE and PE. These configurations allow VPN services to be isolated using only one MCE.

Figure 2 MCE networking



Benefits

The MCE technology enables CEs to provide PE functions. MCEs avoid the practice of deploying one CE for each VPN although; whereas isolating VPN services, significantly reducing maintenance costs and expenditure on devices.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.2.4 Inter-AS VPN

With the wide application of MPLS VPN solutions, different MANs of a carrier or collaborating backbone networks of different carriers frequently span multiple ASs.

Generally, an MPLS VPN architecture runs within an AS in which VPN routing information is flooded on demand. The VPN routing information within the AS cannot be flooded to the other ASs. To implement exchange of VPN routes between different ASs, the inter-AS MPLS VPN model is used. The inter-AS MPLS VPN model is an extension to the MPLS VPN framework. Through this model, route prefixes and labels can be advertised over links between different carrier networks.

The following three inter-AS VPN solutions are proposed in related standards:

- Inter-Provider Backbones Option A (inter-AS VPN Option A): VPN instances spanning multiple ASs are bound to dedicated interfaces of ASBRs to manage their own VPN routes. This solution is also called VRF-to-VRF.
- Inter-Provider Backbones Option B (inter-AS VPN Option B): ASBRs advertise labeled VPN-IPv4 routes to each other through MP-EBGP. This solution is also called EBGP redistribution of labeled VPN-IPv4 routes.
- Inter-Provider Backbones Option C: PEs advertise labeled VPN-IPv4 routes to each other through multi-hop MP-EBGP. This solution is also called multi-hop EBGP redistribution of labeled VPN-IPv4 routes.

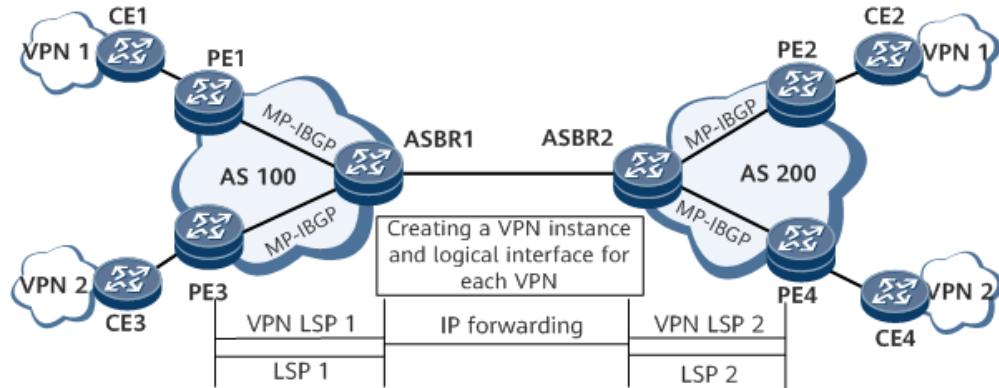
Inter-AS VPN Option A

- **Inter-AS VPN Option A overview**

As a basic BGP/MPLS IP VPN application in the inter-AS scenario, Option A does not need special configurations and MPLS does not need to run between ASBRs. In this mode, ASBRs of two ASs directly connect to each other and function as PEs in the ASs. Each ASBR views the peer ASBR as its CE, creates a VPN instance for each VPN, and advertises IPv4 routes to the peer ASBR through EBGP.

On the network shown in [Figure 1](#), for ASBR1 in AS 100, ASBR2 in AS 200 is a CE. Similarly, for ASBR2, ASBR1 is a CE. Here, a VPN LSP indicates a private network tunnel, and an LSP indicates a public network tunnel.

Figure 1 Inter-AS VPN Option A

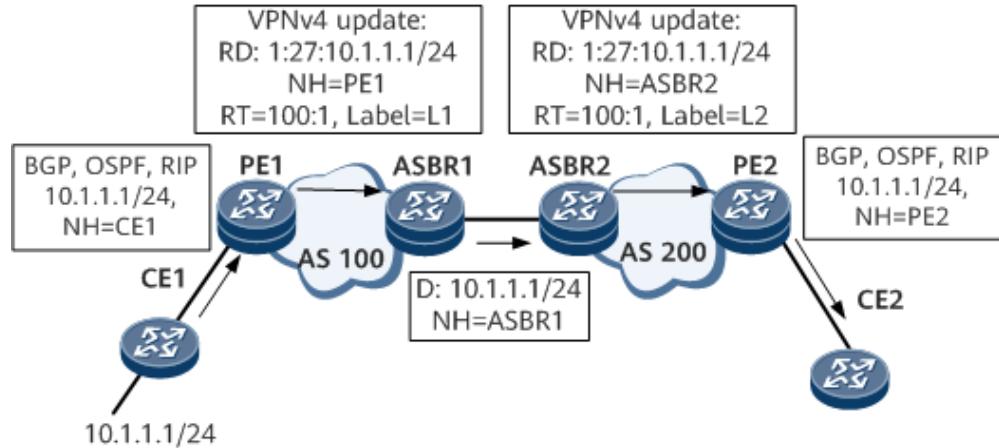


- **Route advertisement in an inter-AS VPN Option A scenario**

MP-IBGP runs between PEs and ASBRs to exchange VPN-IPv4 route information. A common PE-CE routing protocol (BGP or IGP multi-instance) or static route can be used between ASBRs for the exchange of VPN information. Because this involves interaction between different ASs, using EBGP is recommended.

For example, CE1 advertises route 10.1.1.1/24 to CE2. [Figure 2](#) shows the process. D indicates the destination address, NH the next hop, and L1 and L2 the VPN labels. This figure does not show the distribution of public network IGP routes and labels.

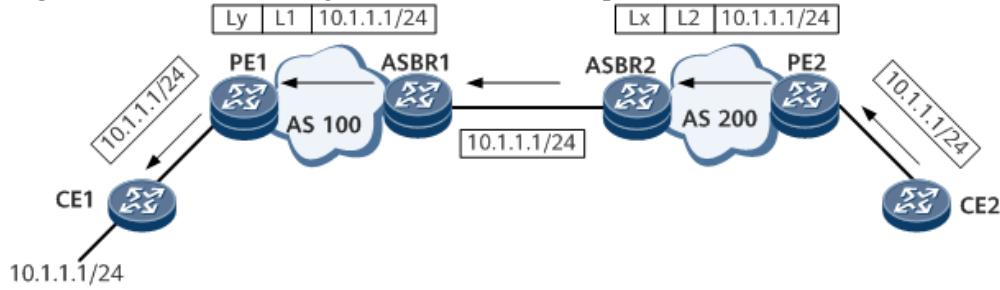
Figure 2 Route advertisement in an inter-AS VPN Option A scenario



- **Packet forwarding in an inter-AS VPN Option A scenario**

[Figure 3](#) shows the process of forwarding packets through an LSP on the public network. L1 and L2 indicate VPN labels, and Lx and Ly indicate public network labels.

Figure 3 Packet forwarding in an inter-AS VPN Option A scenario

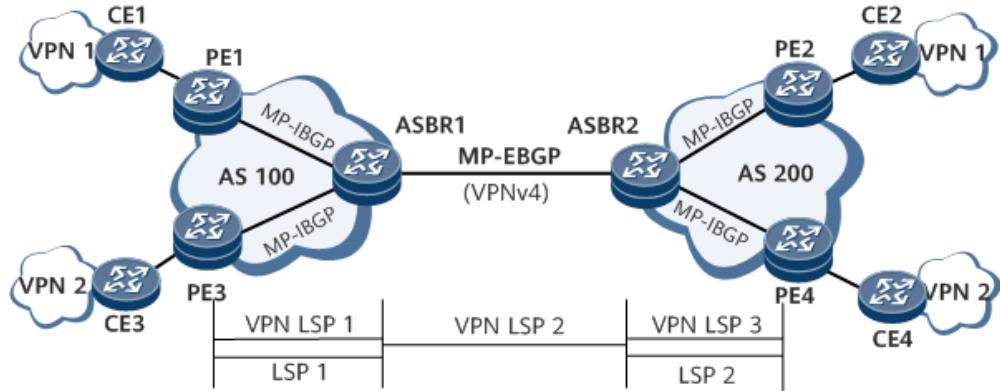


Inter-AS VPN Option B

- **Inter-AS VPN Option B overview**

On the inter-AS VPN Option B network shown in [Figure 4](#), two ASBRs use MP-EBGP to exchange labeled VPN-IPv4 routes received from local PEs in their respective ASs. A VPN LSP indicates a private network tunnel, and an LSP a public network tunnel.

Figure 4 Inter-AS VPN Option B



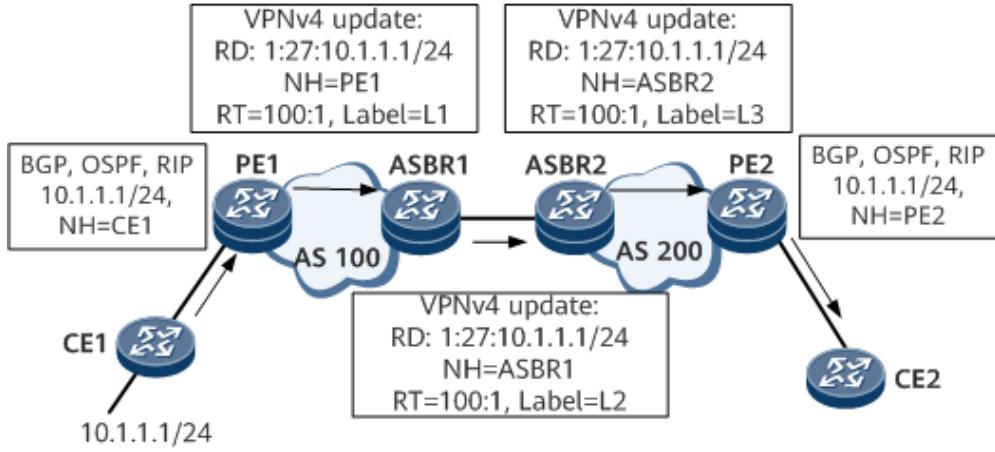
In inter-AS VPN Option B, ASBRs receive all inter-AS VPN-IPv4 routes from the local and external ASs and then advertise these routes. In basic MPLS VPN implementation, a PE stores only the VPN routes that match the VPN targets of its local VPN instances. The ASBRs are configured to store all the received VPN routes, regardless of whether these routes match the VPN targets of its local VPN instances.

The advantage of this solution is that all traffic is forwarded by ASBRs. In this way, traffic is controllable, but the loads on the ASBRs are heavy. BGP routing policies, such as VPN target-based filtering policies, can be configured on ASBRs, so that ASBRs only save some of VPN-IPv4 routes.

- **Route advertisement in an inter-AS VPN Option B scenario**

[Figure 5](#) shows a route advertisement example. In this example, CE1 advertises route 10.1.1.1/24 to CE2. NH indicates the next hop, and L1, L2, and L3 the VPN labels. This figure does not show the distribution of public network IGP routes and labels.

Figure 5 Route advertisement in an inter-AS VPN Option B scenario



The specific process is as follows:

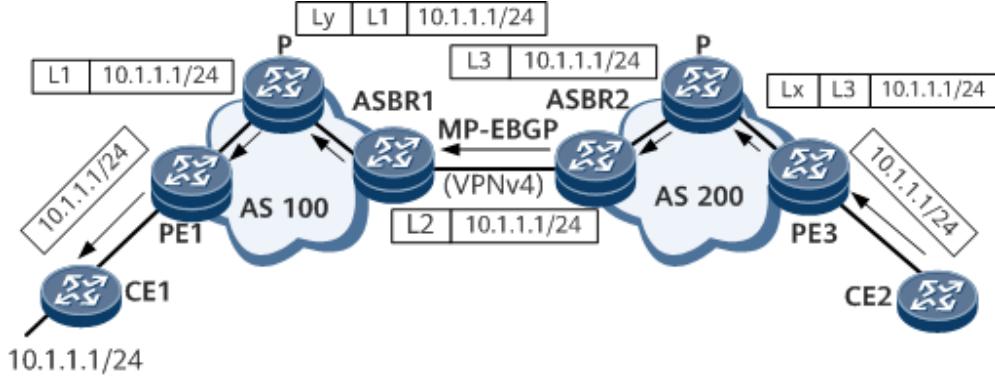
1. CE1 uses BGP, OSPF, or RIP to advertise the route to PE1 in AS 100.
2. PE1 in AS 100 uses MP-IBGP to advertise the labeled VPNv4 route to ASBR1 in AS 100. If a route reflector (RR) is deployed on the network, PE1 advertises the VPNv4 route to the RR, and the RR then reflects the route to ASBR1.
3. ASBR1 uses MP-EBGP to advertise the labeled VPNv4 route to ASBR2. Because MP-EBGP changes the next hop of a route when advertising the route, ASBR1 allocates a new label to the VPNv4 route.
4. ASBR2 uses MP-IBGP to advertise the labeled VPNv4 route to PE2 in AS 200. If an RR is deployed on the network, ASBR2 advertises the VPNv4 route to the RR, and the RR then reflects the route to PE2. When ASBR2 advertises routes to an MP-IBGP peer in the local AS, it changes the next hop of the routes to itself.
5. PE2 in AS 200 uses BGP, OSPF, or RIP to advertise the route to CE2.

ASBR1 and ASBR2 both swap the inner labels of VPNv4 routes and use BGP to transmit inter-AS label information. Therefore, LDP does not need to run between ASBRs.

• Packet forwarding in an inter-AS VPN Option B scenario

In an inter-AS VPN Option B scenario, the two ASBRs need to swap VPN LSPs once during packet forwarding. [Figure 6](#) shows the process of forwarding packets through an LSP on the public network. Here, L1, L2, and L3 indicate VPN labels, and Lx and Ly indicate public network labels (outer tunnel labels).

Figure 6 Packet forwarding in an inter-AS VPN Option B scenario



Inter-AS VPN Option C

- **Inter-AS VPN Option C overview**

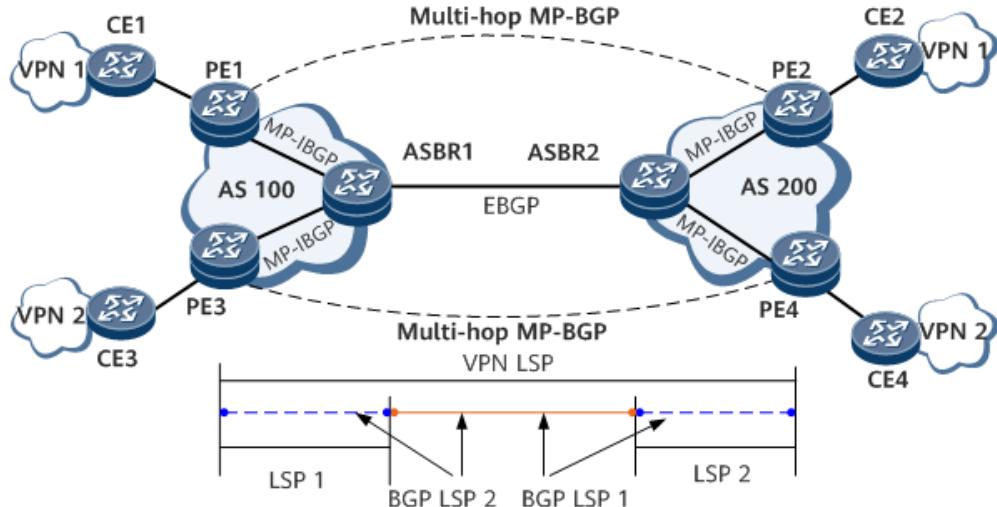
In the preceding two inter-AS VPN modes, ASBRs need to maintain and distribute VPN-IPv4 routes. When each AS needs to exchange a large number of VPN routes, ASBRs may hinder network expansion.

One solution to the problem is that PEs directly exchange VPN-IPv4 routes with each other and ASBRs do not maintain or advertise such routes.

- ASBRs use MP-IBGP to advertise labeled IPv4 routes to PEs in their respective ASs, and advertise labeled IPv4 routes received by PEs in their respective ASs to the peer ASBRs in other ASs. ASBRs in the intermediate AS also advertise labeled IPv4 routes. Therefore, a VPN LSP needs to be established between the ingress and egress PEs.
- The PEs in different ASs establish multi-hop EBGP connections with each other to exchange VPN-IPv4 routes.
- The ASBRs neither store VPN-IPv4 routes nor advertise VPN-IPv4 routes to each other.

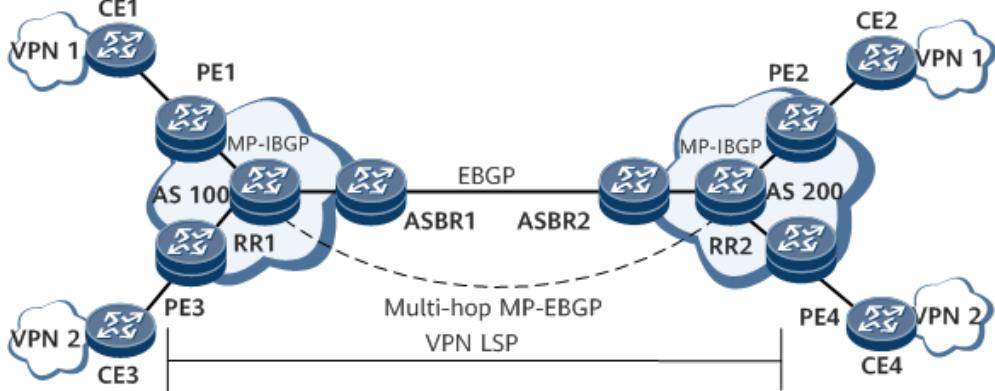
[Figure 7](#) shows the networking of inter-AS VPN Option C. In the figure, VPN LSPs are private network tunnels, and LSPs are public network tunnels. A BGP LSP enables two PEs to exchange loopback route information. A BGP LSP consists of two unidirectional BGP LSP. For example, BGP LSP1 is established from PE1 to PE2, and BGP LSP2 is established from PE2 to PE1.

Figure 7 Inter-AS VPN Option C



To improve scalability, you can specify an RR in each AS. The RR stores all VPN-IPv4 routes and exchanges VPN-IPv4 routes with PEs in the same AS. The RRs in two ASs establish MP-EBGP connections with each other to advertise VPN-IPv4 routes.

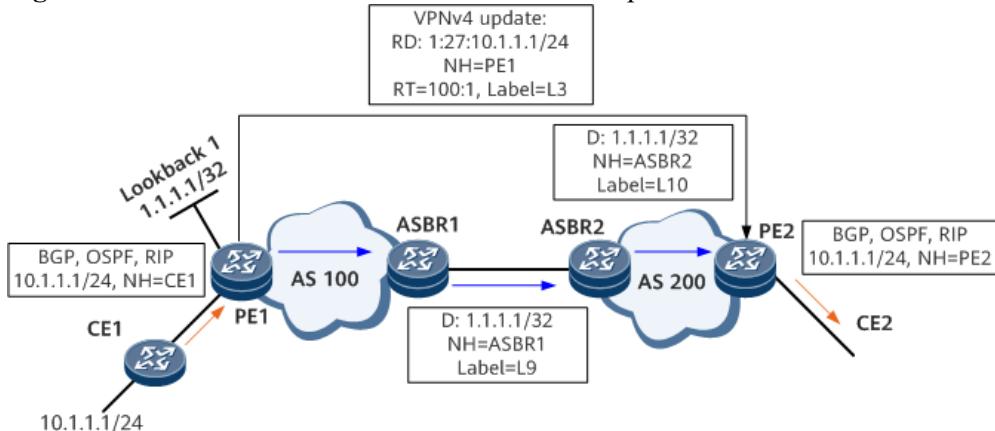
Figure 8 Inter-AS VPN Option C with RRs deployed



- **Route advertisement in an inter-AS VPN Option C scenario**

The key to implementing inter-AS VPN Option C is to establish inter-AS public network tunnels. For example, [Figure 9](#) shows how CE1 advertises route 10.1.1.1/24. D indicates the destination address, NH the next hop, L3 the VPN label, and L9 and L10 the BGP LSP labels. This figure does not show the distribution of public network IGP routes and labels.

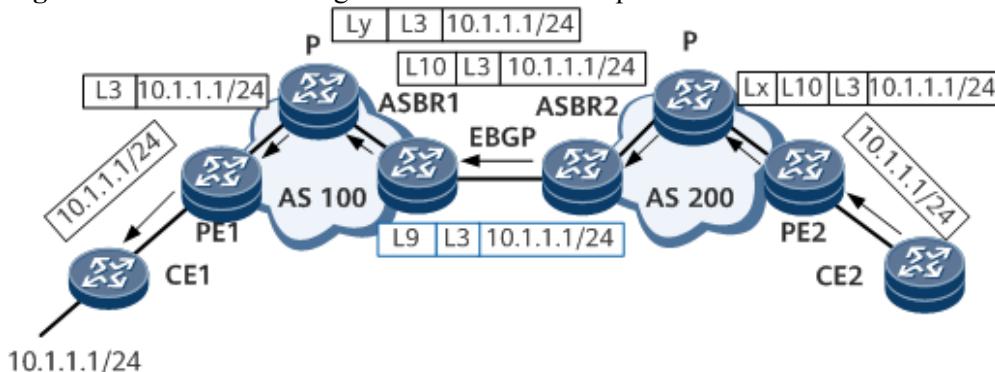
Figure 9 Route advertisement in an inter-AS VPN Option C scenario



- **Packet forwarding in an inter-AS VPN Option C scenario**

[Figure 10](#) shows the process of forwarding packets through an LSP on the public network. L3 indicates the VPN label, L10 and L9 BGP LSP labels, and Lx and Ly public network labels (outer tunnel labels).

Figure 10 Packet forwarding in an inter-AS VPN Option C scenario



When PE2 forwards a packet to PE1, PE2 needs to add three labels to the packet: a VPN label, a BGP LSP label, and a public network label. When the packet reaches ASBR2, only the VPN label and BGP LSP label remain. After the packet reaches ASBR1, ASBR1 removes the BGP LSP label and forwards the packet as a common MPLS VPN packet.

Comparison of the Three Inter-AS VPN Modes

Table 1 Comparison of the three inter-AS VPN modes

Inter-AS VPN	Description
Option A	<p>Easy configuration: MPLS is not required between ASBRs, and no special configuration is required for inter-AS connections.</p> <p>Poor scalability: ASBRs need to manage all VPN routes, and a VPN instance needs to be configured for each VPN. This results in numerous VPN-IPv4 routes on the ASBRs. In addition, because common IP forwarding is implemented between ASBRs, each inter-AS VPN requires a different interface, which can be a sub-interface, physical interface, or bundled logical interface. This poses high requirements for ASBRs. If a VPN spans multiple ASs, the intermediate ASs must support VPN services. This requires complex configurations and greatly affects the intermediate ASs. If only a few inter-AS VPN instances are used, Option A is recommended.</p>
Option B	<p>Unlike Option A, Option B is not restricted by the number of links between ASBRs. VPN route information is stored on and forwarded by ASBRs. If a large number of VPN routes exist, the overloaded ASBRs tend to become faulty points. Therefore, in scenarios where MP-EBGP is used, ASBRs that maintain VPN route information generally do not perform IP forwarding on the public network.</p>
Option C	<p>VPN routes are directly exchanged between the ingress and egress PEs. The routes do not need to be stored or forwarded by intermediate devices.</p> <p>Only PEs maintain VPN route information, and Ps and ASBRs are only responsible for packet forwarding. This means that the intermediate devices only need to support MPLS forwarding instead of MPLS VPN services. ASBRs are no longer bottlenecks. Option C, therefore, is suitable for VPNs spanning multiple ASs.</p> <p>MPLS VPN load balancing is easier to implement in Option C mode.</p> <p>The disadvantage of this mode is that it costs too much to manage an E2E BGP LSP between PEs.</p>

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

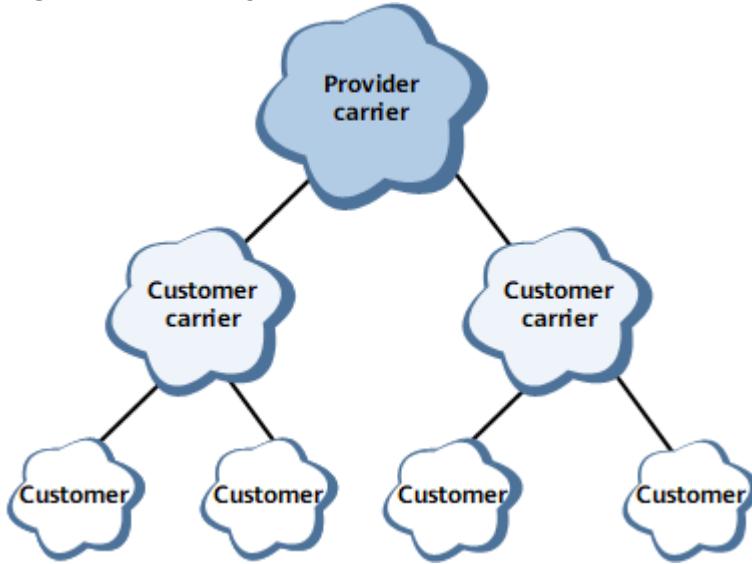
[< Previous topic](#) [Next topic >](#)

1.7.2.5 Carrier's Carrier

Background

A customer of an SP providing BGP/MPLS IP VPN services may also be an SP. In this case, the SP providing the BGP/MPLS VPN service is called the provider carrier or Level 1 carrier and the customer is called the customer carrier or Level 2 carrier, as shown in [Figure 1](#). This networking model is called carrier's carrier. In this model, the Level 2 carrier is a VPN user of the Level 1 carrier.

Figure 1 Networking of carrier's carrier



Related Concepts

To ensure good expansibility, the Level 2 carrier uses an operation mode similar to that of a stub VPN. In other words, the Level 1 carrier CE advertises only Level 2 carrier's internal routes, instead of the Level 2 carrier routes, to the Level 1 carrier PE. In this section, the internal and external routes of the Level 2 carrier are called internal and external routes for short, respectively.

The differences between internal and external routes are as follows:

- Routes to Level 2 carrier SP sites are called internal routes. The routes to VPNs of the Level 2 carrier are called external routes.
- Level 1 carrier PEs exchange internal routes using BGP. The external routes are exchanged using BGP between Level 2 carrier PEs, but are not advertised to Level 1 carrier PEs.
- The VPN-IPv4 routes of the Level 2 carrier are considered as external routes. The Level 2 carrier PEs import only internal routes and not external routes to their VRFs, reducing the number of routes that need to be maintained on the Level 1 carrier network. The Level 2 carrier network has to maintain both internal and external routes.

NOTE

A Level 1 carrier CE is a device through which the Level 2 carrier network accesses the Level 1 carrier network. The Level 1 carrier CE means a CE for a Level 1 carrier network and a PE for a Level 2 carrier network. The device through which users access the Level 2 carrier network is called a user CE.

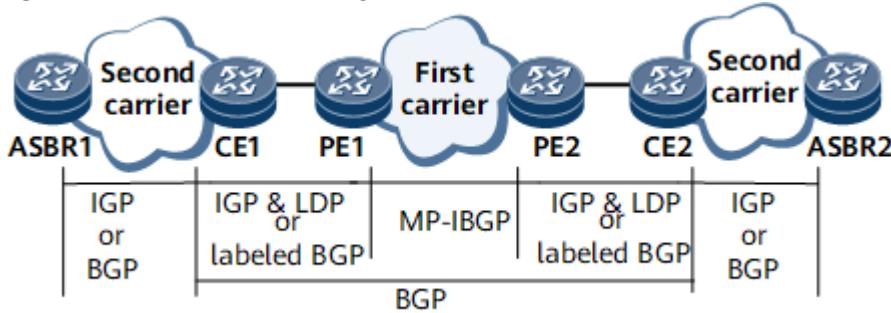
Scenario Categories

The Level 2 carrier can be a common SP or a BGP/MPLS IP VPN SP.

If a Level 2 carrier is a common SP, MPLS does not need to be configured on Level 2 carrier PEs. Level 2 carrier PEs communicate with Level 1 carrier PEs using an IGP. Level 2 carrier PEs exchange external routes with each other over BGP sessions, as shown in [Figure 2](#).

In this scenario, BGP needs to run between CE1 and CE2 to transmit the internal routes of the Level 2 carrier. If CE1 and CE2 are in the same AS, establish an IBGP peer relationship between CE1 and CE2 and configure CE1 and CE2 as RRs. If CE1 and CE2 are in different ASs, establish an EBGP peer relationship between CE1 and CE2.

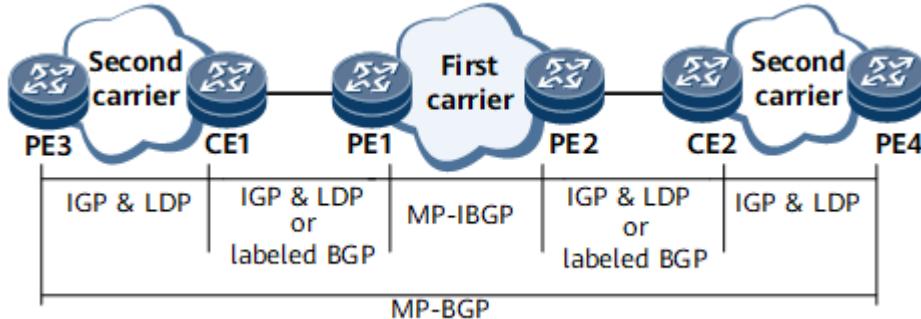
Figure 2 Level 2 carrier serving as a common SP



If a Level 2 carrier is a BGP/MPLS IP VPN SP, Level 2 carrier PEs must be configured with MPLS. Level 2 carrier PEs communicate with Level 1 carrier CEs using an IGP and LDP. Level 2 carrier PEs exchange external routes between each other using MP-BGP, as shown in [Figure 3](#).

Because PE3 and PE4 need to provide VPN services, an LSP needs to be established between PE3 and PE4. Generally, LDP runs between the Level 1 carrier and Level 2 carrier. Two solutions are available for connecting the Level 1 carrier CE to the Level 1 carrier PE: using the LDP multi-instance to establish an LDP LSP, or using BGP labeled routes to establish a BGP LSP.

Figure 3 Level 2 carrier serving as a BGP/MPLS IP VPN SP



Classification of Implementation Solutions

When the Level 2 carrier is a common SP, the IP network and the BGP/MPLS IP VPN are co-constructed. The key point is that a BGP peer relationship is re-established after Level 1 carrier CEs can communicate with each other.

When the Level 2 carrier is a BGP/MPLS IP VPN SP, the LDP multi-instance solution is called carrier's carrier solution 1 and the BGP label routing solution is called carrier's carrier solution 2 according to the method that the Level 2 carrier CE uses to access the Level 2 carrier PE.

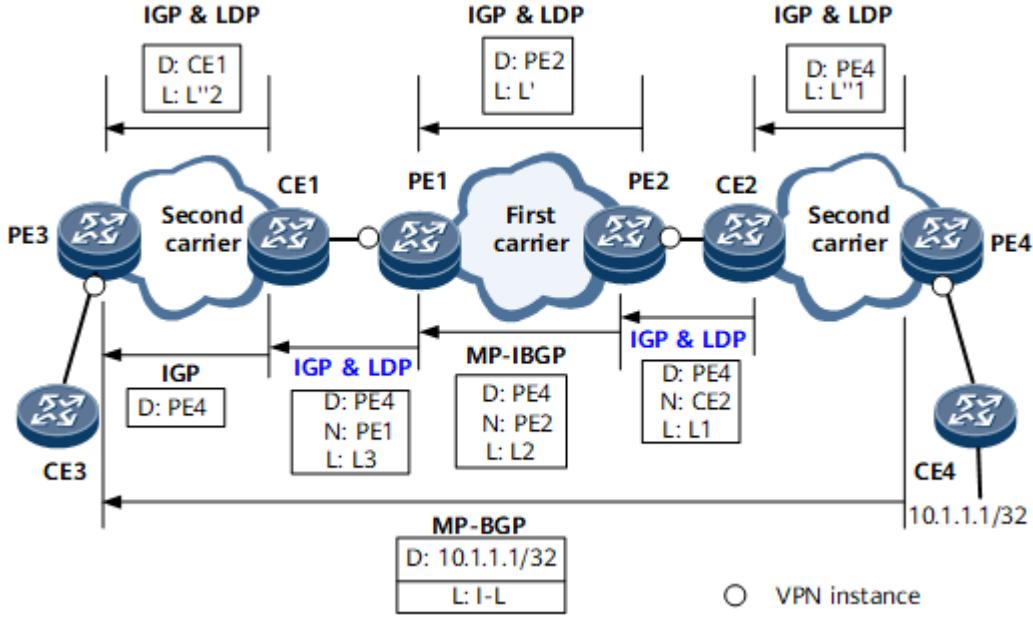
NOTE

The scenario where the Level 2 carrier is a BGP/MPLS IP VPN SP is more commonly used. The scenario where the Level 2 carrier is a common SP is seldom used and the configuration scheme is simple. Therefore, the following description focuses on the scenario where the Level 2 carrier is a BGP/MPLS IP VPN SP.

Carrier's Carrier Solution 1 (LDP Multi-Instance)

When the Level 1 carrier CE uses the LDP multi-instance to set up an LDP LSP to access the Level 1 carrier PE, the routing information exchange process is shown in [Figure 4](#). D represents the destination address of a route, N the next hop, and L the label.

Figure 4 Route information exchange process in carrier's carrier solution 1



The following uses the advertisement of a VPN route destined for 10.1.1.1/32 advertised by PE4 to PE3 as an example to describe VPN route exchange inside the Level 2 carrier network.

1. PE4 advertises a route destined for itself to CE2 using an IGP running on the Level 2 carrier network. Meanwhile, PE4 assigns label L^{"1} to the IGP next hop and establishes a public network LSP to CE2.
2. CE2 advertises the route destined for PE4 to PE2 using an IGP running between CE2 and PE2. In addition, LDP is used to allocate label L1 to the route. (LDP multi-instance needs to be configured on PE2's interface connected to CE2.)
3. PE2 assigns label L2 to the route destined for PE4 and advertises the route to PE1 using MP-IBGP. Previously, PE2 has advertised its routes to PE1 using an IGP running on the Level 1 carrier backbone network and assigned label L['] to the routes destined for itself. A public network LSP has been established between PE2 and PE1.
4. PE1 assigns label L3 to the route destined for PE4 based on the LDP multi-instance peer relationship with CE1, and advertises the route carrying label L3 to CE1.
5. CE1 uses an IGP to advertise the route to PE4 to PE3.

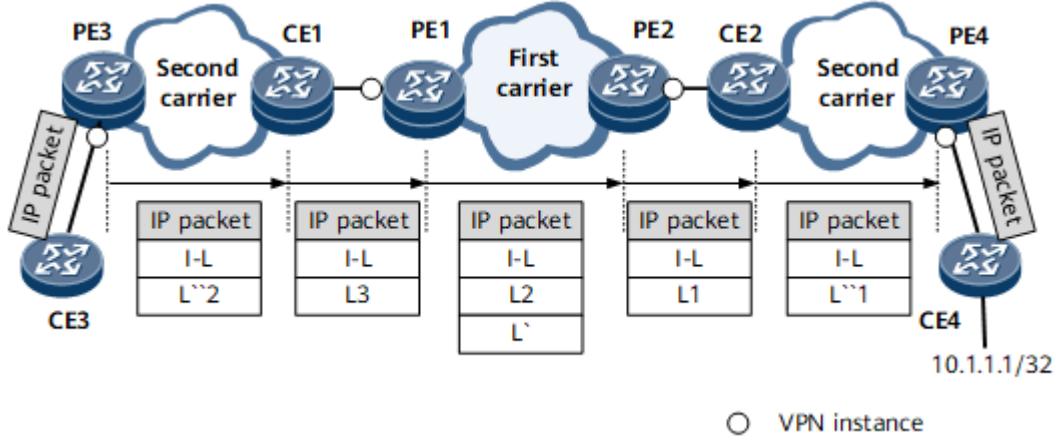
Previously, CE1 has advertised its routes to PE3 using an IGP running on the Level 2 carrier backbone network and assigned label L^{"2} to the routes destined for itself. A Level 2 public network LSP has been established between CE1 and PE3.

6. After the route destined for PE3 is advertised to PE4, an MP-IBGP connection is established between PE3 and PE4.
7. PE4 assigns VPN label I-L to the VPN route destined for 10.1.1.1/32 and advertises the route to PE3 using MP-IBGP.

The advertisement of a VPN route from PE3 to PE4 is similar to that from PE4 to PE3 and therefore is not described here.

[Figure 5](#) shows the transmission of VPN packets on the carrier network. I-L indicates a VPN label assigned by MP-BGP. L' indicates the public network label used on the Level 1 carrier network. L^{"1} and L^{"2} stand for public network labels used on the Level 2 carrier network. L1, L2, and L3 represent labels assigned to packets destined for PE4.

Figure 5 Packet forwarding process in carrier's carrier solution 1



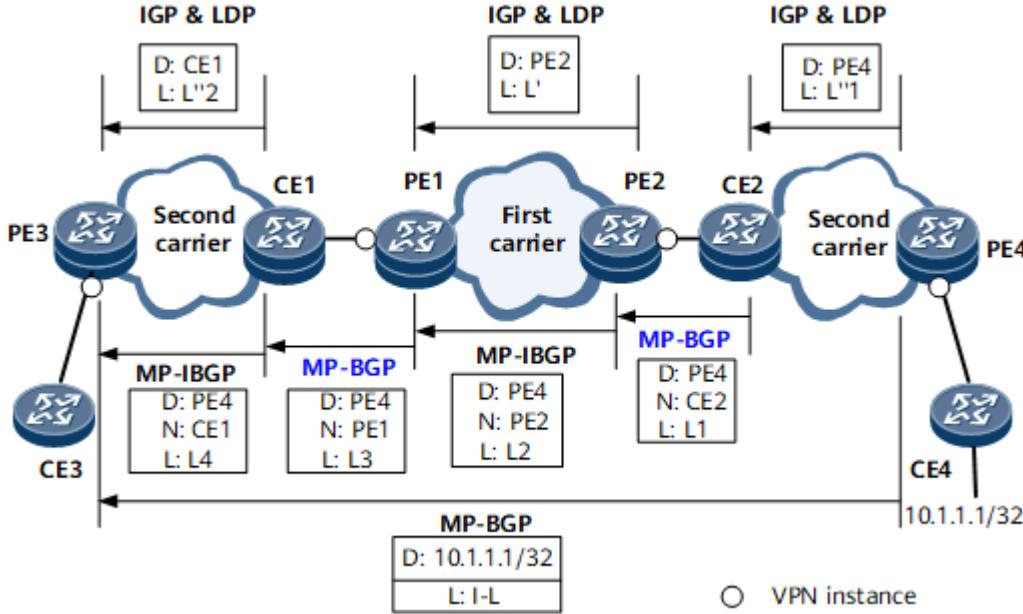
The following uses forwarding of the VPN packet destined for 10.1.1.1/32 from PE3 to CE4 as an example to describe VPN packet forwarding over carrier networks.

1. After receiving a VPN packet destined for 10.1.1.1/32, PE3 adds the VPN label I-L to this packet and transparently transmits the packet to CE1 over the public network LSP on the Level 2 carrier network.
Before the packet arrives at CE1, the penultimate LSR removes the outer public network label of the packet.
2. CE1 adds label L3 to the packet and forwards this packet to PE1.
3. PE1 replaces label L3 with label L2 and adds label L' to the packet. PE1 then forwards the packet to PE2 over the public network LSP. Before the packet arrives at PE2, the penultimate LSR removes label L'.
Before the packet arrives at PE4, the penultimate LSR removes label L''1.
4. PE2 replaces label L2 with label L1 and forwards the packet to CE2.
5. CE2 removes label L1, adds label L''1, and transparently forwards the packet to PE4 over the public network LSP on the Level 2 carrier network.
Before the packet arrives at PE4, the penultimate LSR removes label L''1.
6. PE4 removes label I-L and forwards the packet to CE4 based on label I-L.

Carrier's Carrier Solution 2 (BGP Labeled Route)

[Figure 6](#) shows the route exchange process when the Level 1 carrier CE uses the BGP labeled route solution to establish a BGP LSP and accesses the Level 1 carrier PE. D represents the destination address of a route, N the next hop, and L the label.

Figure 6 Route information exchange process in carrier's carrier solution 2



The following uses the advertisement of a VPN route destined for 10.1.1.1/32 from PE4 to PE3 as an example to describe VPN route exchange inside the Level 2 carrier network.

1. PE4 advertises a route destined for itself to CE2 using an IGP running on the Level 2 carrier network. Meanwhile, PE4 assigns label L"1 to the IGP next hop and establishes a public network LSP to CE2.
2. CE2 assigns label L1 to the route to PE4 based on the MP-BGP peer relationship with PE2, and advertises the labeled route to PE2.
3. PE2 assigns label L2 to the route and advertises the route destined for PE4 to PE1 using MP-IBGP.

Previously, PE2 has advertised its routes to PE1 using an IGP running on the Level 2 carrier's backbone network and assigned label L' to the routes destined for itself. A public network LSP has been established between PE2 and PE1.

4. PE1 assigns label L3 to the route destined for PE4 based on the MP-BGP peer relationship with CE1, and advertises the route carrying label L3 to CE1.
5. CE1 assigns label L4 to the route destined for PE4, and advertises the route carrying label L4 to PE3 through the MP-IBGP peer relationship between CE1 and PE3.

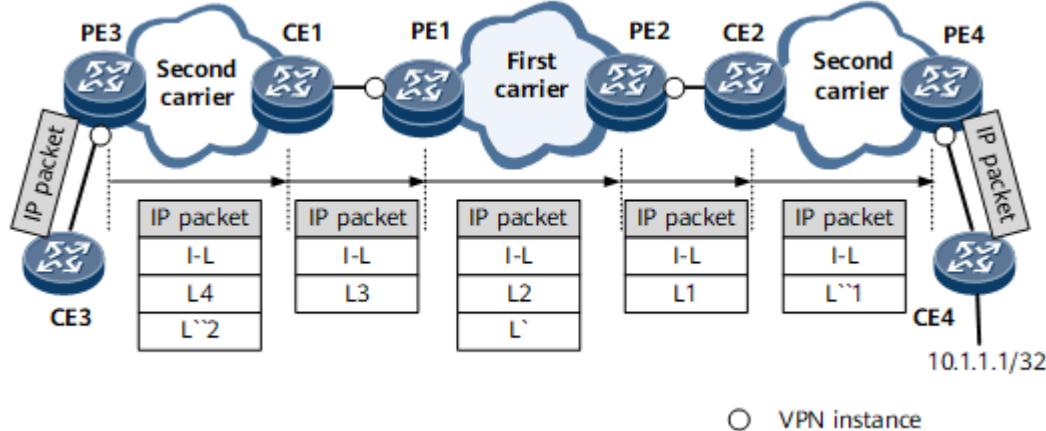
Previously, CE1 has advertised its route to PE3 using an IGP running on the Level 2 carrier backbone network and assigned label L"2 to the route destined for itself. A Level 2 public network LSP has been established between CE1 and PE3.

6. The route destined for PE4 and label assigned to the route are advertised to PE3. A BGP LSP is established between CE2 and PE3.
- After the route destined for PE3 is advertised to PE4, an MP-EBGP connection is successfully established between PE3 and PE4.
7. PE4 assigns VPN label I-L to the VPN route destined for 10.1.1.1/32 and advertises the route to PE3 using MP-EBGP.

The advertisement of the VPN route from PE3 to PE4 is similar to that from PE4 to PE3 and therefore is not described here.

[Figure 7](#) shows the transmission of VPN packets on the carrier network. I-L represents the VPN label assigned using MP-BGP. L' indicates the public network label used on the Level 1 carrier network. L"1 and L"2 stand for public network labels used on the Level 2 carrier network. L1, L2, L3, and L4 represent labels assigned to packets destined for PE4.

Figure 7 Packet forwarding process in carrier's carrier solution 2



The following uses forwarding of the VPN packet destined for 10.1.1.1/32 from PE3 to CE4 as an example to describe VPN packet forwarding over carrier networks.

1. After receiving the VPN packet destined for 10.1.1.1/32, PE3 adds the VPN label I-L and BGP LSP label L4 to this packet and transparently forwards the packet to CE1 over the public network LSP on the Level 2 carrier network.
Before the packet arrives at CE1, the penultimate LSR removes the outer public network label of the packet.
2. CE1 replaces L4 with L3 and forwards the packet to PE1.
3. PE1 replaces label L3 with label L2, adds label L', and forwards the packet to PE2 over the public network LSP. Before the packet arrives at PE2, the penultimate LSR removes label L'.
Before the packet arrives at PE4, the penultimate LSR removes label L"1.
4. PE2 replaces label L2 with label L1 and forwards the packet to CE2.
5. CE2 removes label L1, adds label L"1, and transparently forwards the packet to PE4 over the public network LSP on the Level 2 carrier network.
Before the packet arrives at PE4, the penultimate LSR removes label L"1.
6. PE4 removes label I-L and forwards the packet to CE4 based on label I-L.

Benefits

The carrier's carrier model has the following advantages:

- Part of the configuration, management, and maintenance work used to be carried out by the Level 2 carrier can be undertaken by the Level 1 carrier.
- The Level 2 carrier can flexibly plan addresses, as its addresses are independent of those of the customers and the Level 1 carrier.
- The Level 1 carrier can provide VPN services for multiple Level 2 carriers over a backbone network, and can provide Internet services at the same time. This increases the profits of the Level 2 carrier.
- The Level 1 carrier manages and maintains VPN services of each Level 2 carrier in the same manner instead of maintaining individual backbone networks for Level 2 carriers. This

simplifies the operation of the Level 1 carrier.

The carrier's carrier model has the following disadvantages: As a strict symmetrical networking mode, only VPN users at the same network level can communicate with each other.

VPN users at the same network level need to directly exchange VPN routing information between each other. Therefore, these user devices must be routable. The user devices at the same network level must maintain all routing information of this network level. The PEs at the same network level need to directly exchange VPNV4 routes between each other.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

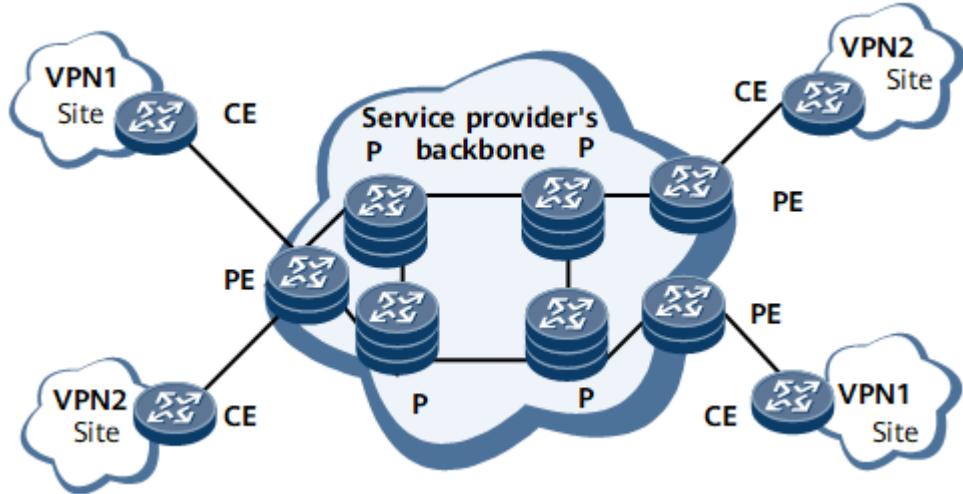
1.7.2.6 HVPN

Background

Currently, hierarchical architectures are used in most networking designs. For example, metropolitan area networks (MANs) typically use a three-layer architecture consisting of an access layer, an aggregation layer, and a core layer. On the network shown in [Figure 1](#), all PEs reside on the same plane and must provide the following functions:

- Provide access services for users. This function requires each PE to provide a large number of interfaces.
- Manage and advertise VPN routes and process user packets. This function requires each PE to have a high-capacity memory and strong forwarding capabilities.

Figure 1 Basic architecture of a BGP/MPLS IP VPN



To deploy VPN functions in a hierarchical architecture, a BGP/MPLS IP VPN must use a hierarchical model instead of a plane model. As a result, the concept of HVPN is introduced.

Related Concepts

[Figure 2](#) shows a basic HVPN architecture consisting of mainly UPEs, SPEs, and NPEs:

- UPE: a type of PE directly connected to CEs and provides access services for users.

- SPE: a type of PE connected to UPEs and located on the core of a network. SPEs manage and advertise VPN routes.
- NPE: a type of PE connected to SPEs and located on the network side.

A UPE and an SPE are connected by only one link and exchange packets based on labels. An SPE does not need to provide a large number of interfaces for access users. UPEs and SPEs can be connected by physical interfaces with physical links, by sub-interfaces with VLANs or PVCs, or by tunnel interfaces with LSPs. If an IP or MPLS network resides between a UPE and an SPE, the UPE and SPE can be connected by tunnel interfaces to exchange labeled packets over a tunnel.

The capabilities of SPEs and UPEs differ according to the roles they play on a network. SPEs require large-capacity routing tables and high forwarding performance, but few interface resources. UPEs, on the other hand, require only low-capacity routing tables and low forwarding performance, but high access capabilities.

NOTE

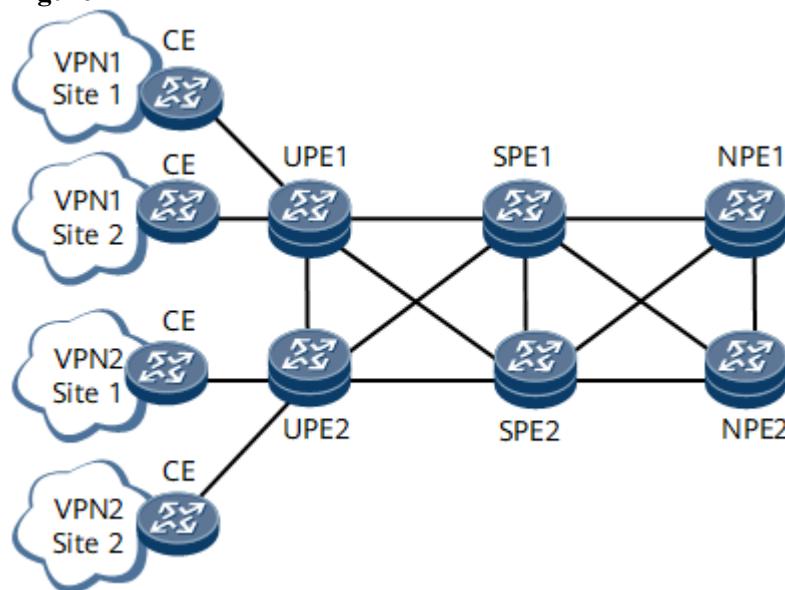
The roles of UPEs and SPEs are relative. On an HVPN, a superstratum PE is the SPE of an understratum PE, and an understratum PE is the UPE of a superstratum PE.

An HoPE is compatible with common PEs on an MPLS network.

If a UPE and an SPE belong to the same AS, they use MP-IBGP. If they belong to different ASs, they use MP-EBGP.

If MP-IBGP is used, an SPE can function as the RR for multiple UPEs to advertise routes between IBGP peers. To reduce the number of routes on UPEs, ensure that an SPE that is already acting as the RR for UPEs is not used as the RR for other PEs.

Figure 2 HVPN architecture



HVPN can be classified into HoVPN and H-VPN.

Table 1 Comparison of HoVPN and H-VPN

HVPN Mode	Characteristics
-----------	-----------------

HVPN Mode	Characteristics
HoVPN	An SPE advertises only default or aggregated routes to UPEs. <ul style="list-style-type: none"> An export policy must be configured on an SPE so that the SPE only advertises specific routes, such as the default routes, to UPEs. VPN instances must be configured on an SPE for the SPE to import default routes locally or aggregate routes received from remote SPEs or NPEs, so that the SPE advertises only default routes or aggregated routes to UPEs.
H-VPN	An SPE advertises all VPN routes to UPEs. <ul style="list-style-type: none"> VPN instances do not need to be configured on SPEs. MP-BGP peer relationships must be configured between SPEs and NPEs and between SPEs and UPEs. The NPEs and UPEs must be configured as the clients of SPEs that function as RRs, and the SPEs must be configured to set the next hops of routes to themselves when advertising these routes to NPEs and UPEs.

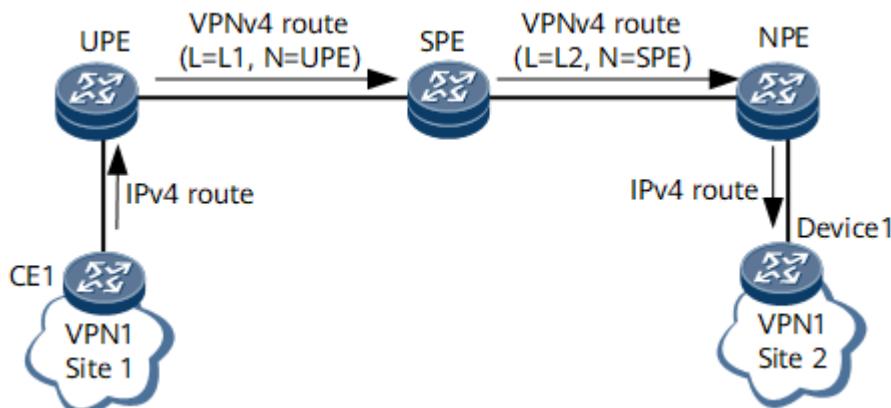
The following describes the route exchanging and packet forwarding processes on an HoVPN and an H-VPN. In the following figures, N indicates a next hop, and L indicates a label.

Route Advertisement from CE1 to Device1 on an HoVPN or H-VPN

[Figure 3](#) shows route advertisement from CE1 to Device1 on an HoVPN or H-VPN.

1. CE1 advertises IPv4 routes to the UPE using the IP protocol.
2. The UPE applies for label L1 for the received IPv4 routes and converts these routes to VPNV4 routes. Then, the UPE sets itself as the next hops of these routes and advertises them to the SPE.
3. After receiving the VPNV4 routes, the SPE saves label L1 locally and applies for label L2 for these VPNV4 routes. Then, the SPE sets itself as the next hops of these routes and advertises them to the NPE.
4. After receiving the VPNV4 routes, the NPE converts these routes to IPv4 routes and imports routes with reachable next hops to its VPN IPv4 routing table. The NPE retains label L2 and recursive tunnel ID information of these routes for later packet forwarding.
5. The NPE advertises the IPv4 routes to Device1 using the IP protocol.

Figure 3 Route advertisement from CE1 to Device1 on an HoVPN or H-VPN

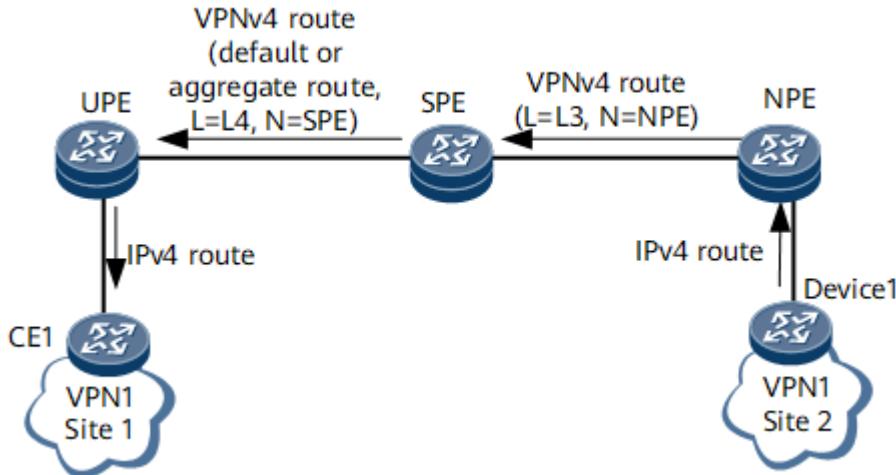


Route Advertisement from Device1 to CE1 on an HoVPN

[Figure 4](#) shows route advertisement from Device1 to CE1 on an HoVPN.

1. Device1 advertises IPv4 routes to the NPE using the IP protocol.
2. The NPE applies for label L3 for the received IPv4 routes and converts these routes to VPNV4 routes. Then, the NPE sets itself as the next hops of these routes and advertises them to the SPE.
3. After receiving the VPNV4 routes, the SPE saves label L3 locally and converts these routes to IPv4 routes and imports routes with reachable next hops to its VPN IPv4 routing table.
4. The SPE imports a default route to its VPN IPv4 routing table or generates an aggregated VPN route based on the received IPv4 routes in its VPN IPv4 routing table and applies for label L4 for the default route or aggregated VPN route. Then, the SPE converts the default route or aggregated VPN route to a VPNV4 route, sets itself as the next hop of the VPNV4 route, and advertises the route to the UPE.
5. After receiving the VPNV4 route, the UPE converts the route to an IPv4 route and imports the route to its VPN IPv4 routing table if the next hop of the route is reachable.
6. The UPE advertises the IPv4 route to CE1 using the IP protocol.

Figure 4 Route advertisement from Device1 to CE1 on an HoVPN

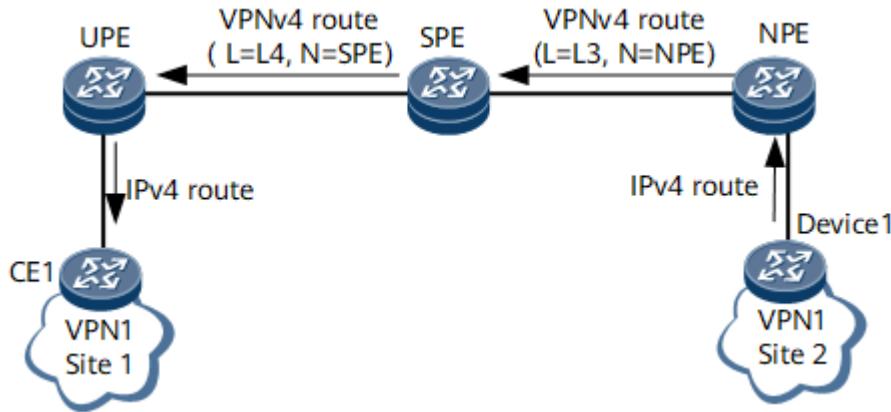


Route Advertisement from Device1 to CE1 on an H-VPN

[Figure 5](#) shows route advertisement from Device1 to CE1 on an H-VPN.

1. Device1 advertises IPv4 routes to the NPE using the IP protocol.
2. The NPE applies for label L3 for the received IPv4 routes and converts these routes to VPNV4 routes. Then, the NPE sets itself as the next hops of these routes and advertises them to the SPE.
3. After receiving the VPNV4 routes, the SPE saves label L3 locally and applies for label L4 for these VPNV4 routes. Then, the SPE sets itself as the next hops of these routes and advertises them to the UPE.
4. After receiving the VPNV4 routes, the UPE converts these routes to IPv4 routes and imports routes with reachable next hops to its VPN IPv4 routing table.
5. The UPE advertises the IPv4 routes to CE1 using the IP protocol.

Figure 5 Route advertisement from Device1 to CE1 on an H-VPN

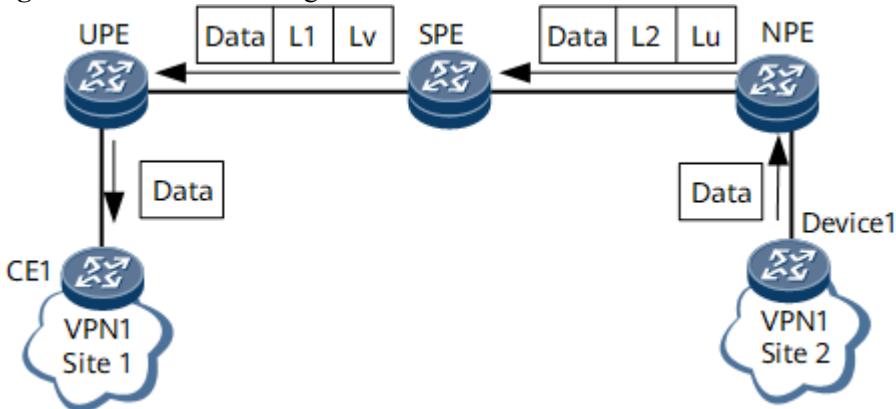


Packet Forwarding from Device1 to CE1 on an HoVPN or H-VPN

[Figure 6](#) shows packet forwarding from Device1 to CE1 on an HoVPN or H-VPN.

1. Device1 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 an inner label L2 and an outer label Lu to the packet and sends the packet to the SPE over the found tunnel.
3. After receiving the packet, the SPE replaces the outer label Lu with Lv and the inner label L2 with L1. Then, the SPE sends the packet to the UPE over the same tunnel.
4. After receiving the packet, the UPE removes the outer label Lv, searches for a VPN instance corresponding to the packet based on the inner label L1, and removes the inner label L1 after the VPN instance is found. Then, the UPE searches the VPN forwarding table of this VPN instance for the outbound interface of the packet based on the destination address of the packet and sends the packet through this outbound interface to CE1. The packet sent by the UPE is a pure IP packet with no label.

Figure 6 Packet forwarding from Device1 to CE1 on an HoVPN or H-VPN



Packet Forwarding from CE1 to Device1 on an HoVPN

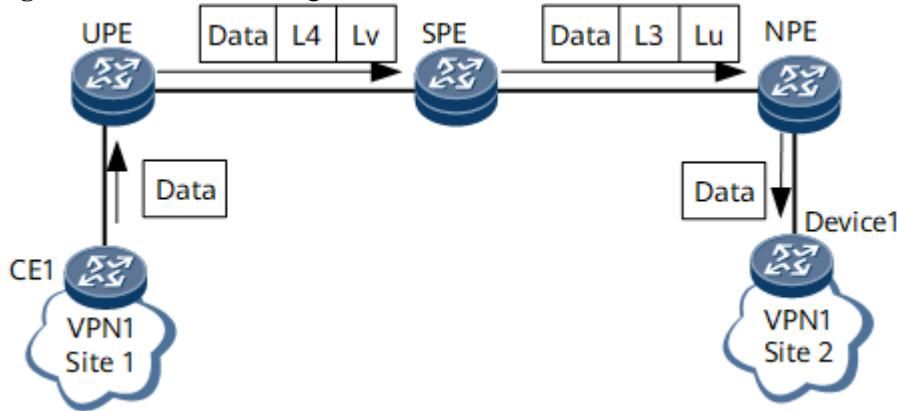
[Figure 7](#) shows packet forwarding from CE1 to Device1 on an HoVPN.

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 aggregated route). Then, the UPE adds an inner label L4 and an outer label Lv to the packet and sends the packet to the SPE over the found tunnel.

3. After receiving the packet, the SPE removes the outer label Lv and searches for the VPN instance corresponding to the packet based on the inner label L4. Then, the SPE removes the inner label L4 and searches the VPN forwarding table of the found VPN instance for a tunnel to forward the packet based on the destination address of the packet. Finally, the UPE adds an inner label L3 and an outer label Lu to the packet and sends the packet to the NPE over the found tunnel.
4. After receiving the packet, the NPE removes the outer label Lu, searches for a VPN instance corresponding to the packet based on the inner label L3, and removes the inner label L3 after the VPN instance is found. Then, the NPE searches the VPN forwarding table of this VPN instance for the outbound interface of the packet based on the destination address of the packet and sends the packet through this outbound interface to Device1. The packet sent by the NPE is a pure IP packet with no label.

Figure 7 Packet forwarding from CE1 to Device1 on an HoVPN

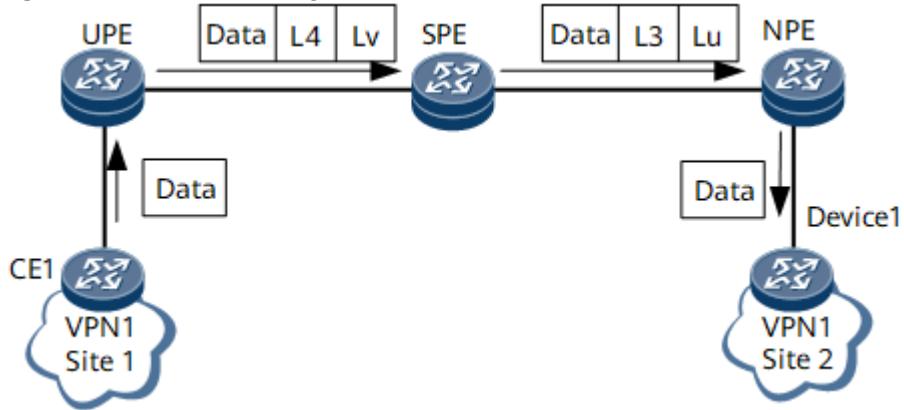


Packet Forwarding from CE1 to Device1 on an H-VPN

[Figure 8](#) shows packet forwarding from CE1 to Device1 on an H-VPN.

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 entries for specific routes received from the SPE). Then, the UPE adds an inner label L4 and an outer label Lv to the packet and sends the packet to the SPE over the found tunnel.
3. After receiving the packet, the SPE replaces the outer label Lv with Lu and the inner label L2 with L3. Then, the SPE sends the packet to the NPE over the same tunnel.
4. After receiving the packet, the NPE removes the outer label Lu, searches for a VPN instance corresponding to the packet based on the inner label L3, and removes the inner label L3 after the VPN instance is found. Then, the NPE searches the VPN forwarding table of this VPN instance for the outbound interface of the packet based on the destination address of the packet and sends the packet through this outbound interface to Device1. The packet sent by the NPE is a pure IP packet with no label.

Figure 8 Packet forwarding from CE1 to Device1 on an H-VPN



Related Functions

H-VPN supports HoPE embedding:

- You can connect a new SPE to an existing SPE and configure the existing SPE to be the UPE of the new SPE.
- You can connect a new UPE to an existing UPE and configure the existing UPE to be the SPE of the new UPE.
- HoPEs can be embedded repeatedly in the preceding two methods.

HoPE embedding can infinitely expand a VPN in theory.

[Figure 9](#) shows a three-layer H-VPN, and the PEs in the middle are referred to as middle-level PEs (MPEs). MP-BGP runs between the SPE and MPEs, and between the MPEs and UPEs.

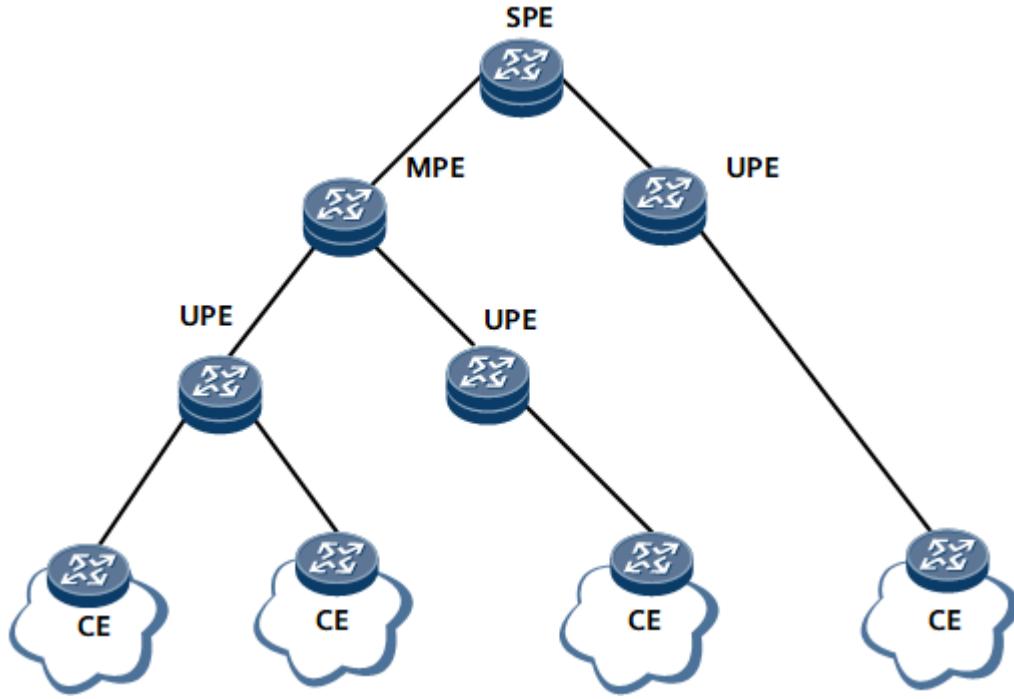
NOTE

The MPE concept is introduced solely for descriptive purposes and does not actually exist in an H-VPN model.

MP-BGP advertises all the VPN routes of UPEs to the SPE, but advertises only the default routes of the VPN instances of the SPE to UPEs.

An SPE maintains the routes of all VPN sites connected to its understratum PEs, whereas a UPE maintains only the routes of its directly connected VPN sites. The numbers of routes maintained by an SPE, an MPE, and a UPE are in descending order.

Figure 9 HoPE embedding



Benefits

HVPN networking offers the following benefits:

- Flexible expansibility

If the performance of a UPE is insufficient, you can add an SPE for the UPE to access. If the access capabilities of an SPE are insufficient, add more UPEs to the SPE.

- Reduced interface resource requirements

Since a UPE and an SPE exchange packets based on labels, they only need to be connected over a single link.

- Reduced burdens on UPEs

A UPE needs to maintain only local VPN routes. The remote VPN routes are represented by a default or aggregated route, lightening the burdens on UPEs.

- Simpler configuration

SPEs and UPEs use MP-BGP, a dynamic routing protocol, to exchange routes and advertise labels. Each UPE only needs to establish a single MP-BGP peer relationship with an SPE.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.2.7 BGP/MPLS IP VPN Label Allocation Modes

Background

In BGP/MPLS IP VPN networking, a device assigns a VPN label (MPLS label) to each VPN instance by default. On a network with a large number of VPN routes, the one-label-per-instance mode helps conserve label resources and reduce PE capacity requirements. [Table 1](#) describes three label allocation modes that are supported.

Table 1 Comparison of label allocation modes

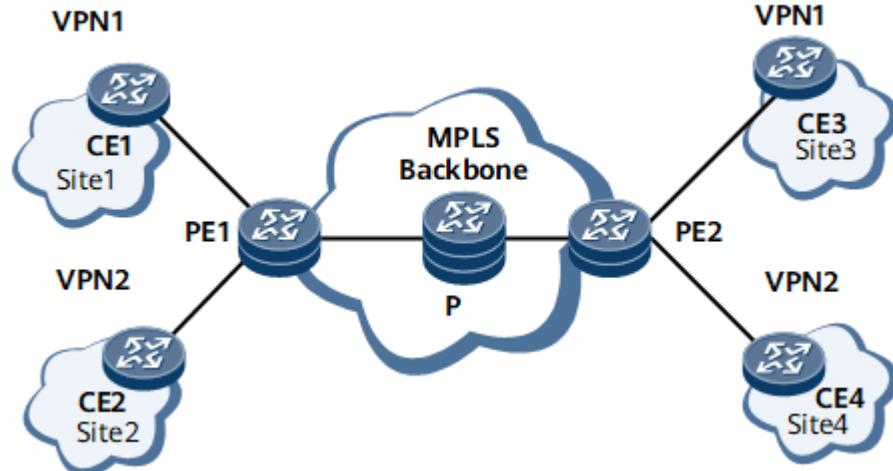
Mode	Description	Applicable Networking	Where to Configure
One-label-per-instance	All VPN routes from a VPN instance are assigned the same VPN label.	All types of BGP/MPLS IP VPNs	Devices on which VPN instances are configured
One-label-per-route	A label is assigned to each VPN route.	All types of BGP/MPLS IP VPNs	Devices on which VPN instances are configured
One-label-per-next-hop	All routes with the same next hop and VPN label are assigned the same label.	Method 1: The one-label-per-next-hop mode configured in the BGP VPN instance IPv4 or IPv6 address family is mainly applicable to inter-AS VPN Option B networking. Method 2: The one-label-per-next-hop mode configured in the VPN instance view is applicable to all BGP/MPLS IP VPN networking.	Method 1: ASBRs in inter-AS VPN Option B networking Method 2: Devices on which VPN instances are configured

Implementation

One-label-per-instance

After one-label-per-instance label distribution is configured in a VPN instance IPv4 or IPv6 address family, all VPN routes from such an address family share the same VPN label. On the network shown in [Figure 1](#), PE1 has two VPN instances: VPN1 and VPN2. If PE1 receives 10,000 VPN routes from sites of VPN1 and VPN2, respectively, by default, PE1 assigns only one label to the 10,000 VPN routes received from each of VPN1 and VPN2.

Figure 1 Networking diagram for the one-label-per-instance mode



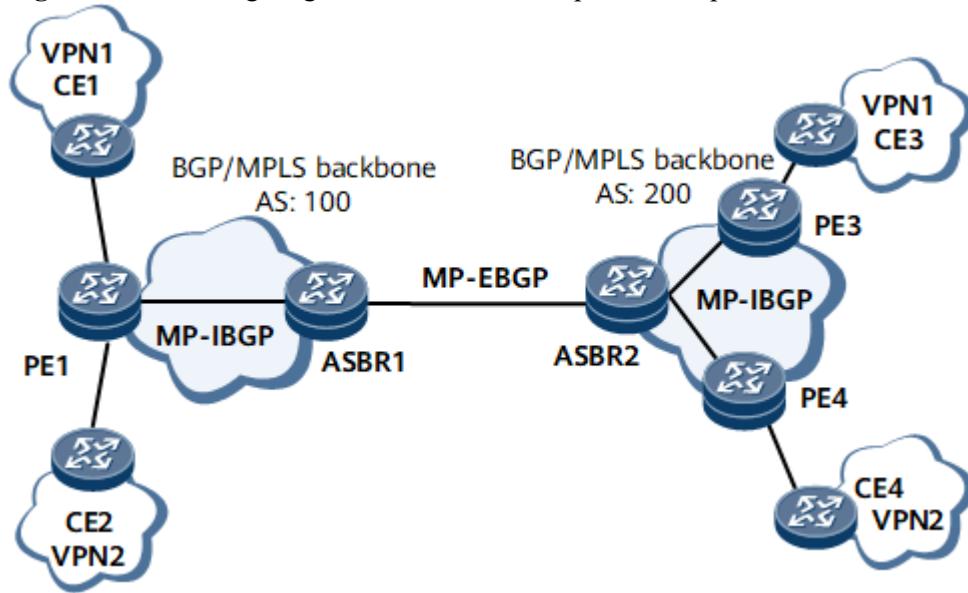
One-label-per-route

After the one-label-per-route function is configured in the VPN instance IPv4 or IPv6 address family view, each VPN route is assigned a label. During the forwarding of private network packets, the packets are forwarded directly to the next hop whose information is carried in a label, and the forwarding speed is fast.

One-label-per-nexthop

By default, an ASBR or PE assigns a label to each VPN route. Alternatively, you can enable one-label-per-next-hop label allocation. After one-label-per-next-hop label allocation is enabled on an ASBR or PE, the ASBR or PE re-advertises an MP-BGP Update message to its peers. Each MP-BGP Update message carries a VPNv4 route and a label re-allocated based on the next hop. After a peer receives the MP-BGP Update message, the peer updates its local label forwarding table and re-establishes an LSP. After the label forwarding tables of the ASBR or PE and its peers are updated, service traffic is forwarded according to the new label forwarding table.

Figure 2 Networking diagram for the one-label-per-next-hop mode



One-label-per-next-hop label allocation is applicable to VPN routes learned by PE1 and peer routes learned by ASBRs:

- Method 1: On the network shown in [Figure 2](#), two VPN instances named VPN1 and VPN2 are configured on PE1 in the inter-AS VPN Option B scenario, and the label distribution mode is one-label-per-route. If 10,000 VPN routes are imported to CE1 and CE2 belonging to VPN1 and VPN2, respectively, 20,000 labels are consumed when ASBR1 advertises 20,000 routes learned from PE1 to ASBR2. After one-label-per-next-hop label allocation is enabled on ASBR1, ASBR1 allocates only one label to the VPN routes with the same next hop and outgoing label. In this case, ASBR1 only needs to allocate two labels to the 20,000 routes.
- Method 2: On the network shown in [Figure 2](#), two VPN instances named VPN1 and VPN2 are configured on PE1 in the inter-AS VPN Option B scenario. Each of CE1 and CE2 belonging to VPN1 and VPN2, respectively, sends 10,000 VPN routes to PE1. If the label allocation mode is one-label-per-route, 20,000 labels are consumed when PE1 advertises 20,000 routes to ASBR1. After one-label-per-next-hop label allocation is enabled on PE1, PE1 allocates only one label to the VPN routes with the same next hop and outgoing label. In this case, PE1 only needs to allocate two labels to the 20,000 routes.



The one-label-per-route mode and one-label-per-next-hop mode can be flexibly switched to each other. During label allocation mode switching, service packets are lost for a short period due to the update of label forwarding tables.

In the inter-AS VPN Option B scenario, one-label-per-instance label allocation must be configured on PEs if one-label-per-next-hop label allocation is configured on ASBRs.

Benefits

Using an appropriate label distribute mode conserves label resources.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

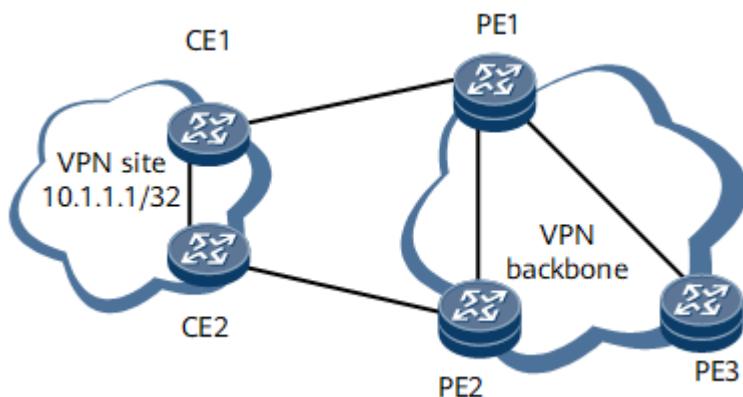
1.7.2.8 BGP SoO

If multiple CEs in a VPN site access different PEs and BGP peer relationships are established between PEs and CEs, VPN routes sent from CEs to PEs may return to this VPN site after traveling across the backbone network. This may cause routing loops in the VPN site.

After the SoO attribute is configured on a PE, the PE adds the SoO attribute to the route sent from a CE and then advertises the route to other PE peers. Before advertising the VPN route to the connected CE, the PE peers check the SoO attribute carried in the VPN route. If the PE peers find that this SoO attribute is the same as the locally configured SoO attribute, the PE peers do not advertise this VPN route to the connected CE.

On the network shown in [Figure 1](#), CE1 and CE2 belong to the same VPN site and can advertise routes to each other. CE1 advertises the route destined for 10.1.1.1/32 in the VPN site to PE1, and PE1 advertises the route to PE2 by using Multiprotocol Internal Border Gateway Protocol (MP-IBGP). PE2 then advertises the route to CE2 by using BGP. As a result, the route returns to the original VPN site from which the route is advertised, which may cause a routing loop in the VPN site.

Figure 1 Networking diagram for BGP SoO application



To avoid routing loops in a VPN site, you can configure an SoO attribute on PE1 for CE1. The SoO attribute identifies the site where the CE1 resides. The routes advertised by CE1 to PE1 then carry this SoO attribute, and PE1 advertises the routes with the SoO attribute to PE2 across the backbone network. Before advertising the received routes to its peer CE2, PE2 checks whether the routes carry the SoO attribute specified for the site where CE2 resides. If a route carries this SoO attribute, this route is advertised from the site where CE2 resides. PE2 then refuses to advertise such a route to CE2, avoiding routing loops in the site.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

1.7.2.9 Route Import Between VPN and Public Network

Background

In BGP/MPLS IP VPN networking, the users of a VPN can communicate with the users of another VPN provided that the two VPNs have matching VPN targets, but cannot communicate with public network users. To enable VPN users and public network users to communicate, configure route import between VPN and public network.

Implementation

After route import between VPN and public network is configured, the VPN and public network will be able to import protocol-specific routes from each other. The imported routes retain their route attributes and recursion information. The VPN or public network instance compares each imported route with local routes that have the same prefix as the imported route and then delivers the optimal route to the IP routing table to guide traffic forwarding.

The VPN and public network can import the following types of routes from each other:

- Static routes
- Direct routes
- OSPF routes
- IS-IS routes
- BGP routes (including active BGP routes preferentially selected in the IP routing table and valid BGP routes with reachable next hops)
- Vlink direct routes

NOTE

Traffic forwarding relies on direct routes (Vlink direct routes) generated based on user entries. When QinQ or Dot1q VLAN tag termination sub-interfaces are used for route import between VPN and public network, Vlink direct routes cannot be imported. As a result, traffic forwarding is interrupted. To solve this problem, route import between VPN and public network newly supports import of Vlink direct routes.

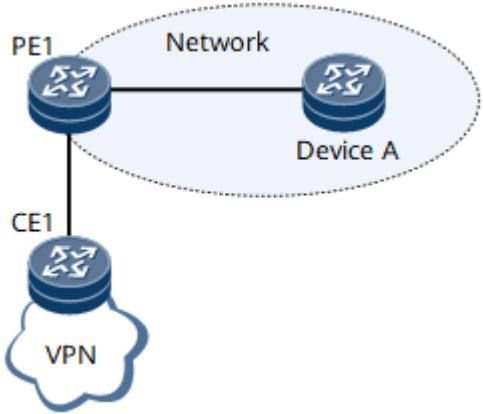
Usage Scenario

Route import between VPN and public network applies to scenarios where VPN users need to communicate with public network users in BGP/MPLS IP VPN networking. On the network shown in [Figure 1](#), CE1 resides on a VPN and Device A resides on the public network. To enable VPN users to communicate with public network users, specifically, to enable CE1 to communicate with Device A, configure route import between VPN and public network on PE1.

After PE1 receives a BGP route from Device A, PE1 imports the route to its VPN instance. After PE1 determines based on a preconfigured routing policy that the newly imported route is an optimal route, PE1 adds the route to its VPN IP forwarding table and advertises the route to CE1, its VPN BGP peer. After PE1 receives a route from CE1, PE1 imports the route to its public network instance. After PE1

determines based on a preconfigured routing policy that the newly imported route is an optimal route, PE1 adds the route to its public IP forwarding table and advertises the route to Device A. CE1 and Device A can then communicate.

Figure 1 Route import between VPN and public network



Benefits

Route import between VPN and public network allows VPN users to communicate with public network users.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.2.10 VPN FRR

Background

As networks develop rapidly, the time used for end-to-end service convergence if a fault occurs on a carrier's network has been used as an indicator to measure bearer network performance. MPLS TE FRR is one of the commonly used fast switching technologies. The solution is to create an end-to-end TE tunnel between two PEs and a backup label switched path (LSP) that protects a primary LSP. When either of the PEs detects that the primary LSP is unavailable because of a node or link failure, the PE switches the traffic to the backup LSP.

MPLS TE FRR protects services in case a link or node fails between two PEs at both ends of a TE tunnel. MPLS TE FRR, however, cannot protect services against endpoint PE faults. If a PE fault occurs, services can only be restored through end-to-end route convergence and LSP convergence. The service convergence time is closely related to the number of routes inside an MPLS VPN and the number of hops on the bearer network. The more VPN routes, the longer the service convergence time.

VPN FRR sets in advance on a remote PE forwarding entries pointing to the active and standby PEs, respectively. In collaboration with fast PE fault detection, VPN FRR can reduce end-to-end service convergence time if a fault occurs on an MPLS VPN where a CE is dual-homed to two PEs. In VPN FRR, service convergence time depends on only the time required to detect remote PE faults and change tunnel status. VPN FRR enables the service convergence time to be irrelevant to the number of VPN routes on the bearer network.

Implementation

As shown in [Figure 1](#), normally, CE1 accesses CE2 over Link A. If PE2 is Down, CE1 accesses CE2 over Link B.

Based on the traditional BGP/MPLS IP VPN technology, both PE2 and PE3 advertise routes destined for CE2 to PE1, and assign VPN labels to these routes. PE1 then selects a preferred VPNv4 route based on the routing policy. In this example, the preferred route is the one advertised by PE2, and only the routing information, including the forwarding prefix, inner label, selected LSP, advertised by PE2 is filled in the forwarding entry of the forwarding engine to guide packet forwarding.

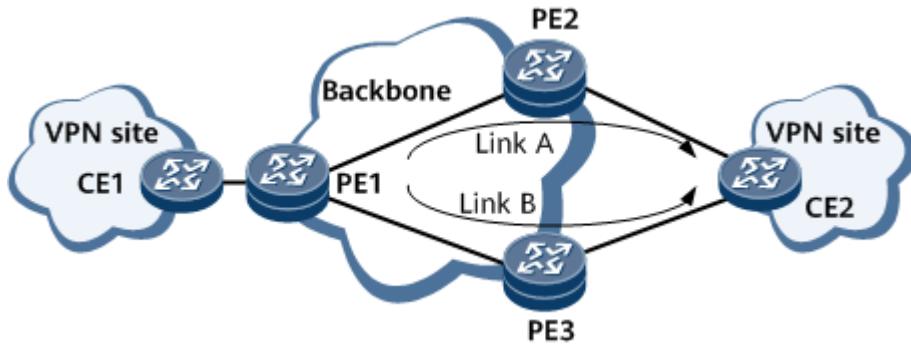
If PE2 fails, PE1 detects the fault on PE2 (the BGP peer goes Down or the MPLS LSP is unavailable), re-selects the route advertised by PE3, and updates the forwarding entry to complete E2E service convergence. Before PE1 re-delivers the forwarding entry for the route advertised by PE3, CE1 cannot access CE2 for a certain period. This is because PE2 is the end point of the MPLS LSP to which the forwarding entry refers and fails. As a result, E2E services are interrupted.

VPN FRR is an improvement on the traditional reliability technology. VPN FRR enables PE1 to add the optimal route advertised by PE2 and the second optimal route advertised by PE3 to a forwarding entry. The optimal route is used for traffic forwarding, and the second optimal route is used as a backup route.

If a fault occurs on PE2, the MPLS LSP between PE1 and PE2 becomes unavailable. After detecting the fault by means of techniques such as BFD, PE1 marks the corresponding entry in the LSP status table as unavailable, and delivers the setting to the forwarding table. After selecting a forwarding entry, the forwarding engine examines the status of the LSP corresponding to the forwarding entry. If the LSP is unavailable, the forwarding engine uses the second optimal route carried in the forwarding entry to forward packets. After being tagged with the inner labels assigned by PE3, packets are transmitted to PE3 over the LSP between PE1 and PE3 and then forwarded to CE2. In this manner, fast end-to-end service convergence is implemented and traffic from CE1 to CE2 is restored.

If both EVPN L3VPN over SRv6 and L3VPN over SRv6 are deployed on the network, PE2 and PE3 advertise four routes destined for CE2 to PE1. To prevent routes from the same device (PE2 or PE3) from being selected as the optimal route and sub-optimal route, configure an export routing policy to change the local preference of routes. This policy ensures that the route with the highest preference is preferred, and the selected optimal route and sub-optimal route correspond to link A and link B, respectively.

Figure 1 VPN FRR networking



Other Functions

VPN FRR is a fast switching technique based on inner labels. The outer tunnels can be LDP LSPs, RSVP-TE tunnels. When the forwarding engine detects that the outer tunnel is unavailable during packet forwarding, fast switching based on inner labels can be implemented.

VPNv6 FRR implements fast switching of IPv6 VPN routes on an IPv6 VPN where a CE is dual-homed to two PEs. The working principle of VPNv6 FRR is similar to that of VPN FRR.

Usage Scenario

On a VPN where a CE is dual-homed to two PEs, after a PE fails, VPN FRR ensures that the VPN services from the CE to the PE can be rapidly switched to the standby PE for transmission.

Benefits

On a VPN where a CE is dual-homed to two PEs, VPN FRR speeds up service convergence and enhances network availability in the case of PE failures.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.2.11 VPN GR

Graceful restart (GR) is a type of high availability (HA) technology, which comprises a comprehensive set of technologies such as fault-tolerant redundancy, link protection, faulty node recovery, and traffic engineering. As a fault-tolerant redundancy technology, GR ensures normal forwarding of data when the routing protocol restarts to prevent interruption of key services. Currently, GR has been widely applied to active/standby switchovers and system upgrade.

GR is usually used when the active route processor (RP) fails due to a software or hardware error, or used when an administrator performs a master/slave main control board switchover.

Prerequisite for GR Implementation

On a traditional router, a processor performs both control and forwarding. The processor finds routes based on routing protocols and maintains the routing and forwarding tables of a device. High- and medium-end devices generally use the multi-RP structure to improve forwarding performance and reliability. The processor responsible for routing protocols is mostly located on the main control board, whereas the processor responsible for data forwarding is located on the interface board. This design helps to ensure the continuity of packet forwarding on the interface board during the restart of the main processor. The technology that separates control from forwarding satisfies the prerequisite for GR implementation.

At present, a GR-capable device must have two main control boards. In addition, the interface board must have an independent processor and memory.

Related Concepts

GR involves the following concepts:

- GR restarter: A GR-capable router that performs a master/slave main control board switchover upon the occurrence of a failure or under the instructions of an administrator.
- GR helper: Neighbor of a GR restarter. A GR helper must support GR.
- GR session: A session over which a GR restarter and a GR helper can negotiate GR capabilities.
- GR time: Time when the GR helper keeps the topology information or routes obtained from the GR restarter after detecting that the GR restarter is Down.



Currently, the HUAWEI NetEngine40E can only function as a GR helper.

VPN GR Overview

VPN GR is the application of the GR technology on a VPN. VPN GR ensures that VPN traffic is not interrupted when a master/slave main control board switchover is performed on a router that transmits VPN services. VPN GR offers the following benefits:

- Reduces the impact of VPNV4 route or BGP label route flapping on an entire network during the route processor switchover.
- Decreases the packet loss ratio to almost 0%.
- Reduces the impact on important VPN services.
- Reduces PE or CE single-point failures to improve the reliability of an entire VPN.

To support VPN GR, a BGP/MPLS IP VPN must support IGP GR and BGP GR. When using an MPLS LDP LSP as a tunnel, the BGP/MPLS IP VPN must support MPLS LDP GR. If traffic engineering is used, the BGP/MPLS IP VPN must also support RSVP GR. After the master/slave main control board switchover is performed on a PE or CE, the router and its connected PE can keep the forwarding information of all VPN routes for a certain period to ensure that VPN traffic is not interrupted. In addition, the CE that connects to the PE on which the master/slave main control board switchover is performed also needs to keep the forwarding information of all VPN routes for a certain period.

On a common L3VPN, the master/slave main control board switchover can be performed on the router that functions as a PE, CE, or P.

Master/Slave Main Control Board Switchover of a PE

The master/slave main control board switchover of a PE consists of three phases:

1. Before the switchover

The PE negotiates the IGP GR and MPLS LDP GR capabilities with a P, and negotiates the IGP GR or BGP GR capabilities with the connected CE. The PE also negotiates BGP GR capabilities with the peer PE and sends the Open message containing the GR capability field of <AFI=Unicast, SAFI=VPNV4>.

2. During the switchover

The PE keeps the status of forwarding VPNV4 routes and the following procedures are involved:

- MPLS LDP GR

If a neighbor detects that the corresponding TCP session enters the Down state, the neighbor backs up all LSPs on the slave board and marks these LSPs as invalid.

- BGP GR

BGP session messages are lost during the switchover. Then, the PE does not keep any routing information but the forwarding information. GR-aware BGP peers mark all the routes related to the GR routers as Stale. The BGP peers, however, still forward packets based on these routes within the GR time.

3. After the switchover

The PE instructs all the IGP neighbors, BGP IPv4 peers, and private network IGP neighbors between the PE and CE to reestablish connections. The following procedures are involved:

- IGP convergence

To resynchronize the link state database (LSDB) of OSPF or IS-IS with the neighboring P, the PE sends a signal to each neighboring P and reestablishes the neighbor relationship list after receiving a response. If IS-IS or OSPF multi-instances are run between the PE and CE, the PE also needs to resynchronize the LSDB with the CE. The PE obtains the topology or routing information by establishing sessions with all the neighbors. After obtaining the topology and routing information, the PE recalculates the routing table and deletes the routes in the Stale state to complete IGP convergence.

- BGP convergence

The PE also exchanges routing information with BGP peers, including public network BGP peers, MP-BGP peers, and private network BGP peers. The PE then updates the routing table and the forwarding table according to the new routing information and replaces the invalid routing information to complete BGP convergence.

After receiving the End-of-Rib message from a BGP peer on a public or private network, the PE notifies the routing management (RM) module. The End-of-Rib message is used to notify the peer that the first routing information update after a BGP session is established has been completed.

Before all routing protocols complete the GR, only FIB information on the main control board is updated, and the FIB information on the interface board is not updated.

After all routing protocols complete the GR, the RM module sends a message to notify each protocol that the GR is complete, and then updates the FIB information on the interface board. BGP sends BGP public network IPv4 routes, private network IPv4 routes, and VPNv4 routes to each peer. After sending the routes, BGP sends End-of-Rib messages.

The processing on routers connected to the PE is as follows:

- After detecting the restart of the PE, the CE connected to this PE uses the same processing flow as that of the GR helper in the common IGP GR or BGP GR and keeps information about all IPv4 routes for a certain period.
- After the P connected to this PE detects the restart of the PE, either of the following situations occurs:
 - If BGP is not configured, the P uses the same processing flow as that of the GR helper in the common IGP GR or MPLS LDP GR.
 - If BGP is configured, the BGP processing flow is the same as that of the GR helper in the common BGP GR except that the BGP processing flow includes additional IGP GR processing and MPLS LDP GR processing, and the P then keeps information about all the public IPv4 routes for a certain period.
- After detecting the restart of the PE, the RRs reflecting VPNv4 routes and the other PEs (including ASBRs) connected to this PE use the same processing flow as that of the GR helper in the BGP GR. They then keep information about all the public IPv4 routes and VPNv4 routes for a certain period.

Master/Slave Main Control Board Switchover of a P

The processing flow of a P is the same as that of the GR restarter in common IGP GR, MPLS LDP GR, or BGP GR.

After detecting the restart of a P, other Ps and PEs that connect to the P use the same processing flow as that of the GR helper in common IGP GR or BGP GR. That means that they keep information about all the public network IPv4 routes for a certain period.

Master/Slave Main Control Board Switchover of a CE

The processing flow of a CE is the same as that of the GR restarter in common IGP GR or BGP GR.

After detecting the restart of a CE, the PEs that connect to the CE use the same processing flow as that of the GR helper in common IGP GR or BGP GR. That means that they keep information about all the private network IPv4 routes for a certain period.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.2.12 VPN NSR

VPN non-stop routing (NSR) is a technique which ensures that if an active main board (AMB) failure causes the control plane of a specific node to fail, the node retains its control plane connections to neighboring nodes and ensures uninterrupted traffic transmission on the forwarding plane.

Background

Carriers have increasing demands for IP network reliability. Conventional non-stop forwarding (NSF) and GR techniques cannot prevent traffic interruptions if a neighboring node does not support GR or multiple nodes fail simultaneously during a GR process. Interruptions occur because a GR-enabled node cannot obtain routing information from or establish control plane connections to neighboring nodes during the GR process. As a result, traffic is interrupted temporarily before the GR process is complete.

NSR can be used to ensure uninterrupted traffic transmission and retain control plane connections if a software or hardware fault occurs on the control plane. The control planes of neighboring nodes are unaware of the fault during the NSR process.

Related Concepts

- HA: supports a backup channel between the AMB and standby main board (SMB).
- NSR: allows a standby control plane to take over traffic from an active control plane if the active control plane fails, although; whereas preventing the control planes of neighbor nodes from detecting the fault.
- NSF: enables a node to use the GR mechanism to ensure uninterrupted transmission during an AMB/SMB switchover.
- AMB and SMB: implement control plane processes.

Implementation

The AMB backs up VPN data to the SMB on a specific node to implement NSR. The following key VPN data is synchronized between the AMB and SMB:

- VPN routes, including:
 - Routes imported by running the **import-route** or **network** command in the BGP-VPN instance IPv4 address family view or BGP-VPN instance IPv6 address family view
 - Locally crossed routes
 - Remotely crossed routes
- Attributes carried by routes
- VPN labels
- Next hop information about routes

NOTE

Active and standby control planes must be able to run on different boards of an NSR-capable device.

The NSR process is as follows:

1. Batch backup

The AMB backs up VPN data in batches to the SMB immediately after the SMB starts.

2. Real-time backup

The AMB backs up VPN data in real time to the SMB. Both the AMB and SMB receive packets to implement real-time data synchronization between them.

3. Switchover

If the AMB fails, the SMB takes over services. The SMB retains uninterrupted operation of both the control and forwarding planes because the SMB has the same data as the AMB.

NOTE

For details about NSR, see chapter "Uninterruptible Service Technology" in *HUAWEI NetEngine40EUniversal Service RouterFeature Description - Reliability*.

Other Usage

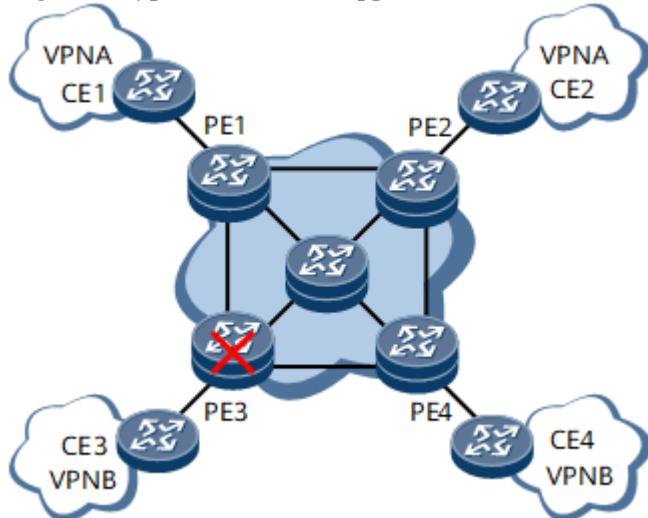
An NSR-enabled device can function as a GR helper. The GR helper communicates with NSR-disabled devices and responds to neighboring nodes' GR requests during an AMB/SMB switchover.

Usage Scenario

NSR is enabled on a node (for example, PE3 shown in [Figure 1](#)) when the node has multiple links that load-balance traffic. NSR helps prevent traffic interruptions that result from a single point of failure.

[Figure 1](#) shows a typical NSR application.

Figure 1 Typical VPN NSR application



NSR minimizes the impact of control plane faults and prevents route flapping on heavily loaded networks.

Benefits

NSR offers the following benefits:

- NSR ensures the uninterrupted operation of the forwarding plane and allows a standby control plane to retain connections between a local node and its neighboring nodes if the VPN control plane of the local node fails.
- A VPN node can use NSR to work independently without the assistance of neighboring nodes. NSR can help multiple VPN nodes implement AMB/SMB switchovers if control planes of these nodes over an MPLS network fail.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.2.13 BGP/MPLS IPv6 VPN Extension

On a BGP/MPLS IP VPN network, IPv4 routing protocols, such as BGP, OSPF, and IS-IS, run between PEs, and between PEs and CEs. After a VPN customer network transits from IPv4 to IPv6, the preceding routing protocols cannot be used between PEs and CEs. IPv6 VPN packets are transmitted on the backbone network. With BGP/MPLS IPv6 VPN extension, the backbone network can provide IPv6 VPN services for customers without having to be upgraded to an IPv6 network.

NOTE

BGP/MPLS IPv6 VPN networking solutions include:

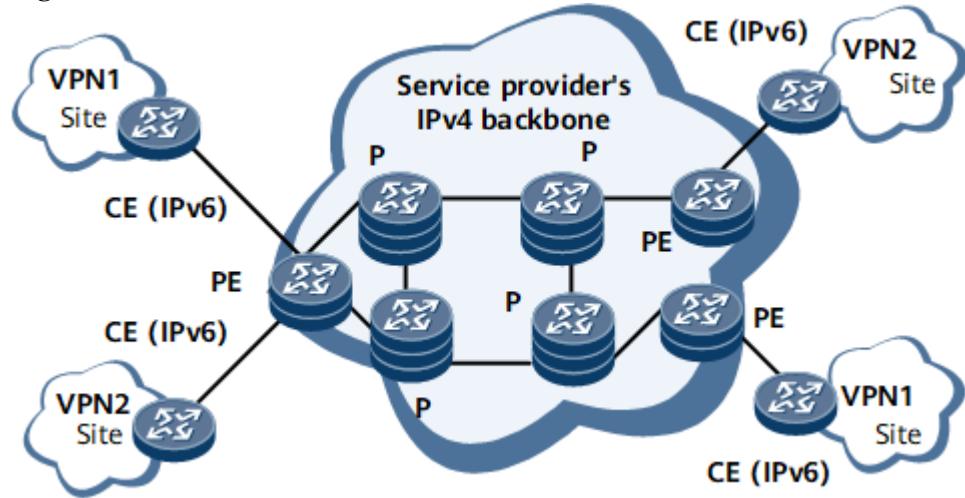
- Solution using carriers' IPv4 backbone networks to carry IPv6 VPN services (also called the 6VPE solution)
- Solution using carriers' IPv6 backbone networks to carry IPv6 VPN services

Currently, only the 6VPE solution is supported.

[Figure 1](#) is the BGP/MPLS IPv6 VPN extension model. After BGP/MPLS IPv6 VPN extension is configured, IPv6 routing protocols run between PEs and CEs, and the following IPv6 routing protocols can be used to provide IPv6 VPN services:

- BGP4+
- Static IPv6 routes
- OSPFv3
- IS-IS IPv6
- Routing Information Protocol next generation (RIPng)

Figure 1 BGP/MPLS IPv6 VPN extension model



IPv4 protocols can still run on the carriers' backbone networks that provide IPv6 VPN services between PEs. In this manner, the carriers' networks can smoothly transit from IPv4 to IPv6.

If the backbone network is an IPv4 network, IPv4 addresses are used to establish VPNV6 peers between PEs to transmit IPv6 VPN routes. IPv6 VPN routes can be carried over IPv4 tunnels on the backbone network to transmit IPv6 VPN services.

BGP/MPLS IPv6 VPN has the same principles and functions as BGP/MPLS IPv4 VPN, except that the routing protocols running between PEs and CEs are different.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

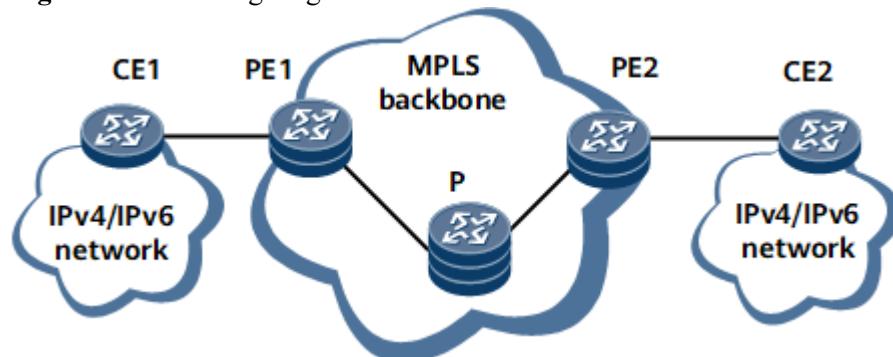
1.7.2.14 VPN Dual-Stack Access

IPv4 address exhaustion forces carriers to put IPv6 network planning issues high on their agendas. For the BGP/MPLS IPv4 VPN services that have been widely deployed, supporting VPN dual-stack access is a practical and readily available solution to the transition from IPv4 to IPv6.

Originally, interfaces in a VPN support only a single type of protocol stack. After VPN dual-stack access is configured, both IPv4 and IPv6 address families can be configured in a VPN so that the interfaces bound to this VPN can support not only IPv4 VPN access but also IPv6 VPN access. This implementation greatly improves the feasibility of the transition from IPv4 to IPv6.

As shown in [Figure 1](#), IPv4/IPv6 VPN dual-stack access allows VPN sites to support both IPv4 and IPv6 networks and to be connected to a PE through the same interface.

Figure 1 Networking diagram for VPN dual-stack access



Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.2.15 VPN MPLS/VPN SRv6 Dual-Stack Tunnel

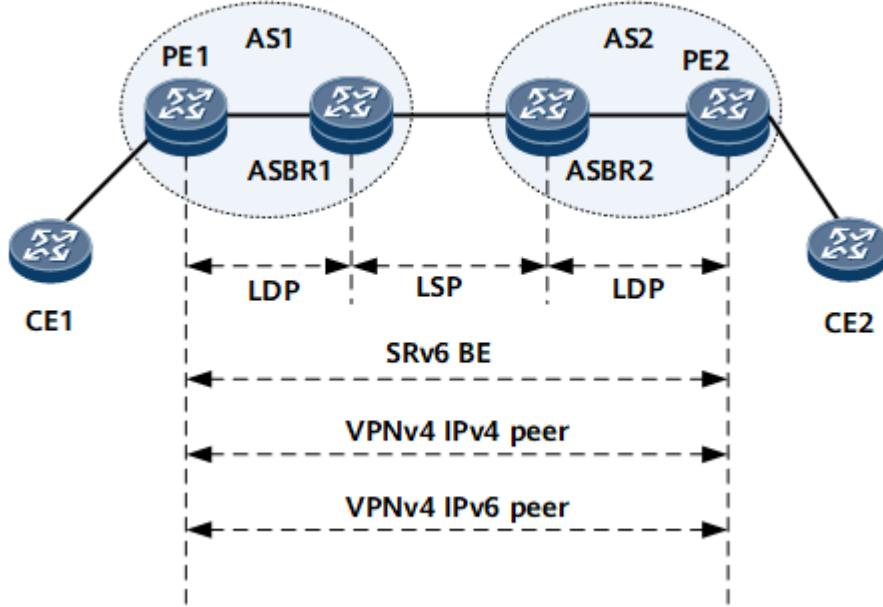
Typical Application Scenario of VPN MPLS/VPN SRv6 Dual-Stack Tunnels

When an MPLS backbone network that carries VPN routes spans multiple ASs, inter-AS VPN technology is used to deploy L3VPN over MPLS services. As IPv4 addresses gradually run out, IPv6 networks will be increasingly deployed to solve this issue. However, such an evolution cannot take place overnight, causing IPv4 and IPv6 services to coexist.

To prevent existing services from being compromised during the upgrade and evolution of existing networks, VPN MPLS/VPN SRv6 dual-stack tunnels can be used. Such tunnels not only prevent traffic interruption when IPv4 and IPv6 services coexist, but also make it much more feasible to transition from IPv4 to IPv6.

On the network shown in [Figure 1](#), an end-to-end VPNV4 IPv4 BGP peer relationship, a VPNV4 IPv6 BGP peer relationship, and an SRv6 BE tunnel are established between PE1 and PE2.

Figure 1 Typical application scenario of VPN MPLS/VPN SRv6 dual-stack tunnels



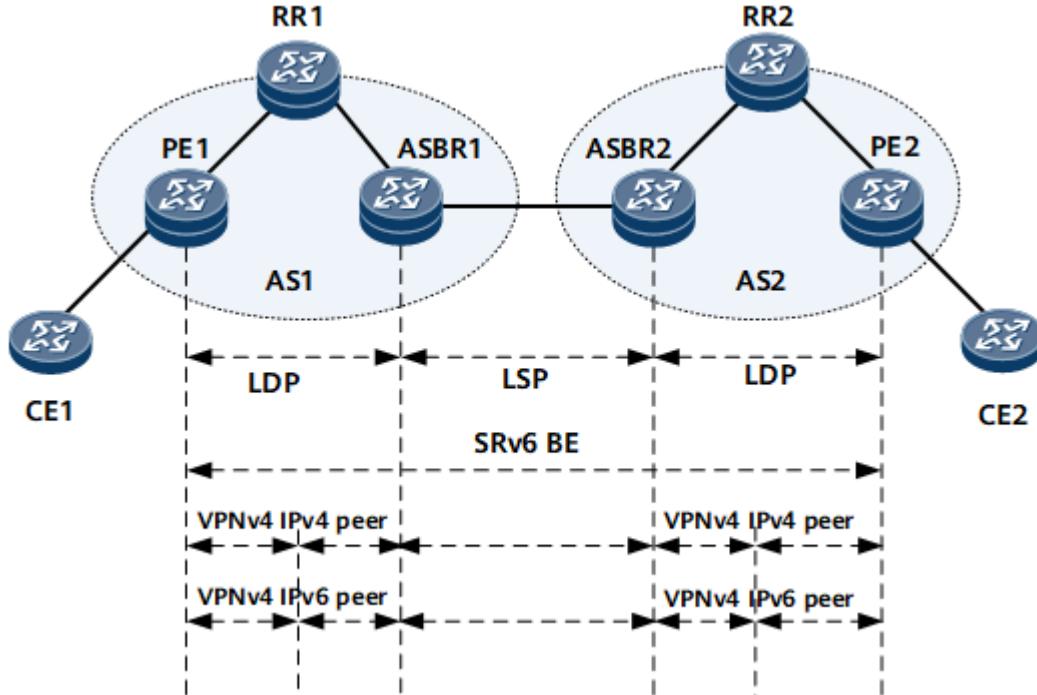
Services are deployed in the following process:

1. An end-to-end VPNv4 IPv4 BGP peer relationship is established between PE1 and PE2. The route to CE1 is advertised to PE2 through the VPNv4 IPv4 BGP peer relationship. After receiving the route, PE2 adds the route to its VPN instance, which then sends the route to CE2.
2. PE1 and PE2 are configured to preferentially select routes with IPv4 next hops based on their high priority using the **peer ipv4-address high-priority** command.
3. SRv6 BE services are deployed between PE1 and PE2.
4. A VPNv4 IPv6 BGP peer relationship is established between PE1 and PE2, and the devices are enabled to exchange IPv4 prefix SID information with each other using the **peer ipv6-address prefix-sid** command.
5. PE1 is enabled to add SIDs to VPN routes using the **segment-routing ipv6 locator locator-name** command in PE1's VPN instance that connects to CE1.
6. After PE1 receives a VPN route from its VPN instance, PE1 advertises a copy of this route to the VPNv4 IPv4 BGP peer and applies for an MPLS label. PE1 then advertises another copy to the VPNv4 IPv6 BGP peer, with the route carrying a SID. A route without a SID cannot be advertised to the VPNv4 IPv6 BGP peer.
7. PE2 receives two VPN routes with the same prefix, but one with an IPv4 next hop and the other with an IPv6 next hop. PE2 then adds the two routes to its VPN instance that connects to CE2.
8. After CE2 receives the two routes, the route with the IPv4 next hop recurses to the MPLS tunnel, and the route with the IPv6 next hop recurses to the SRv6 tunnel.
9. After the route with the IPv6 next hop becomes stable, the **peer ipv6-address high-priority** and **undo peer ipv4-address high-priority** commands are run on PE2 to trigger a change in route preference. In this way, the preferred route switches from that with the IPv4 next hop to that with the IPv6 next hop, and user traffic switches from the traditional MPLS tunnel to the SRv6 tunnel accordingly.

Application Scenario of VPN MPLS/VPN SRv6 Dual-Stack Tunnels (with RRs)

On the network shown in [Figure 2](#), an end-to-end VPNV4 IPv4 BGP peer relationship, a VPNV4 IPv6 BGP peer relationship, and an SRv6 BE tunnel are established between PE1 and PE2.

Figure 2 Application scenario of VPN MPLS/VPN SRv6 dual-stack tunnels (with RRs)



Services are deployed in the following process:

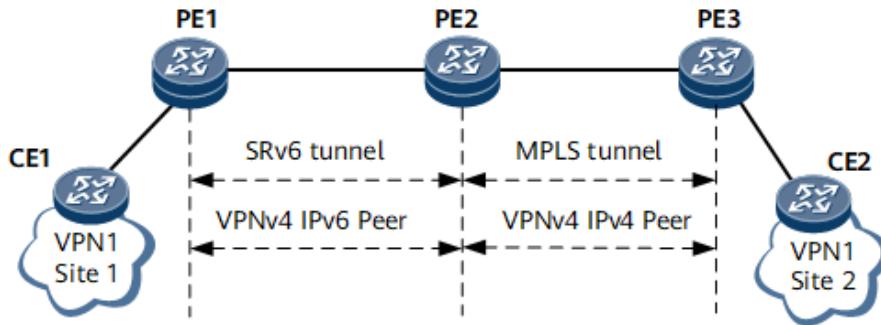
1. In AS1, VPNV4 IPv4 BGP peer relationships are established between PE1 and RR1 and between RR1 and ASBR1. The next hop of a VPNV4 route is not changed when the route is being reflected by RR1.
2. In AS2, VPNV4 IPv4 BGP peer relationships are established between PE2 and RR2 and between RR2 and ASBR2. The next hop of a VPNV4 route is not changed when the route is being reflected by RR2.
3. PE1, ASBR1, ASBR2, and PE2 are configured to preferentially select routes with IPv4 next hops based on their high priority using the **peer ipv4-address high-priority** command.
4. SRv6 BE services are deployed between PE1 and PE2.
5. VPNV4 IPv6 BGP peer relationships are established between PE1 and RR1, between RR1 and ASBR1, between ASBR2 and RR2, and between RR2 and PE2. The devices are enabled to exchange IPv4 prefix SID information with specified IPv6 peers using the **peer ipv6-address prefix-sid** command.
6. CE1 advertises VPN route A to PE1, and PE1 sends the route to the VPNV4 address family. After finding that route A is a locally leaked VPN route, the VPNV4 address family advertises route A to RR1 through the VPNV4 IPv4 and IPv6 peer relationships separately.
7. RR1 receives two routes with the same prefix but different next hops from its IPv4 and IPv6 BGP peers. According to the BGP route advertisement policy, the RR advertises only the optimal route to ASBR1. This may lead to just one link being available to the same destination address during data transmission. To address this issue, you can deploy the BGP Add-Path feature on RR1 so that the router can advertise multiple routes with the same prefix to its BGP peers. Specifically, RR1 reflects the VPNV4 route with the IPv4 next hop to the IPv4 BGP peer, and reflects the VPNV4 route with the IPv6 next hop to the IPv6 BGP peer. In this way, multiple links to the same destination address are formed.

8. ASBR1 receives two VPN routes with the same prefix, but one with an IPv4 next hop and the other with an IPv6 next hop. ASBR1 adds both routes to its VPN instance. After the VPN instance receives the two remotely leaked routes, the route with the IPv4 next hop recurses to the MPLS tunnel, and the route with the IPv6 next hop recurses to the SRv6 tunnel.
9. CE2 switches traffic to the SRv6 BE tunnel, through which traffic is forwarded to ASBR1. Then ASBR1 also switches traffic to the SRv6 BE tunnel.

Route Regeneration Scenario with VPN MPLS/VPN SRv6 Dual-Stack Tunnels

On the network shown in [Figure 3](#), a VPNV4 IPv6 BGP peer relationship and an SRv6 tunnel are established between PE1 and PE2. A VPNV4 IPv4 BGP peer relationship and an MPLS tunnel are established between PE2 and PE3.

Figure 3 Route regeneration scenario with VPN MPLS/VPN SRv6 dual-stack tunnels



Services are deployed in the following process:

1. PE3 advertises a route to PE2 through the VPNV4 IPv4 BGP peer relationship, and then PE2 advertises the route to PE1 through the VPNV4 IPv6 BGP peer relationship. Route advertisement from PE1 to PE3 follows the same process but in reverse.
2. PE2 advertises the route (received from PE3) with the IPv4 next hop to PE1, and advertises the route (received from PE1) with the IPv6 next hop to PE3.
3. Route regeneration is enabled on PE2. Specifically, after the VPNV4 routes received from PE3 and PE1 are added to PE2's VPN instance, the VPN instance is configured to advertise regenerated routes to the VPNV4 address family using the **advertise route-reoriginate** command.
4. PE2 is enabled to advertise the routes regenerated in the VPNV4 address family to the VPNV4 IPv4 and IPv6 BGP peers using the **peer ipv4-address advertise route-reoriginated vpnv4** and **peer ipv6-address advertise route-reoriginated vpnv4** commands, respectively.

Parent Topic: [Understanding BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.7.3 Application Scenarios for BGP/MPLS IP VPN

This section describes different applications of BGP/MPLS IP VPN.

[Application of MCEs on a Campus Network](#)

[Application of MCEs on a Data Center Network](#)

[Application of HVPN on an IP RAN](#)

[Application of Route Import Between VPN and Public Network in the Traffic Cleaning Networking](#)

This section describes how to apply the function of route import between VPN and public network to the traffic cleaning networking.

Parent Topic: [BGP/MPLS IP VPN Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

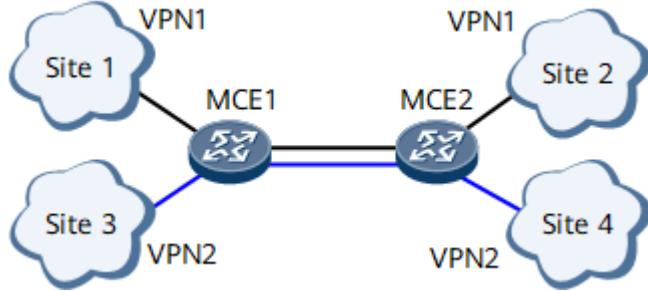
[< Previous topic](#)

1.7.3.1 Application of MCEs on a Campus Network

Networking Description

BGP/MPLS IP VPN, which can isolate users on a campus network, is complicated to deploy and expensive to maintain. Since the number of users to be isolated on a campus network is small, you can use the MCE multi-hop technology instead.

Figure 1 Networking diagram for MCEs on a campus network



As shown in [Figure 1](#), devices at the access layer of the campus network are configured to function as MCEs. These MCEs connect to multiple VPNs and carry VPN routes. VPN routes can be exchanged between local and remote MCEs without using the carrier network. This networking mode offers the following benefits:

- Simplified configuration
MPLS or MP-BGP does not need to be configured.
- Reduced costs
PEs and Ps do not need to be deployed.

Feature Deployment

Routing protocols are deployed on the MCE multi-hop network as follows:

1. The link between MCE1 and MCE2 is bound to a VPN and a routing protocol is configured on both ends of the link.

2. Each link between an MCE and a site is bound to a VPN and a routing protocol is configured on both ends of these links.
3. The route import function is configured on MCEs, if the routing protocol running between MCEs and sites is different than that running between the MCEs.

Route Transmission

Routes are transmitted as follows on the MCE multi-hop network for Site 1 and Site 2 to learn the routes of each other:

1. Site 1 sends its VPN routes to MCE1. Upon receipt, MCE1 adds these routes to its VPN routing and forwarding table.
2. MCE1 sends received routes to MCE2.
3. Upon receipt, MCE2 determines whether the export RTs of these routes match its import RTs. If yes, MCE2 adds these routes to its VPN routing and forwarding table and sends them to Site 2.

Packet Forwarding

Packets are transmitted as follows on the MCE multi-hop network:

1. Site 2 sends a packet destined for Site 1 to MCE2 after searching its local routing and forwarding table.
2. Upon receipt, MCE2 searches the corresponding VPN routing and forwarding table based on the inbound interface of the packet and forwards the packet to MCE1.
3. Upon receipt, MCE1 searches the corresponding VPN routing and forwarding table based on the inbound interface of the packet and forwards the packet to Site 1.
4. Upon receipt, Site 1 checks the destination address of the packet and processes this packet normally after finding that the destination address is itself.

Parent Topic: [Application Scenarios for BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.7.3.2 Application of MCEs on a Data Center Network

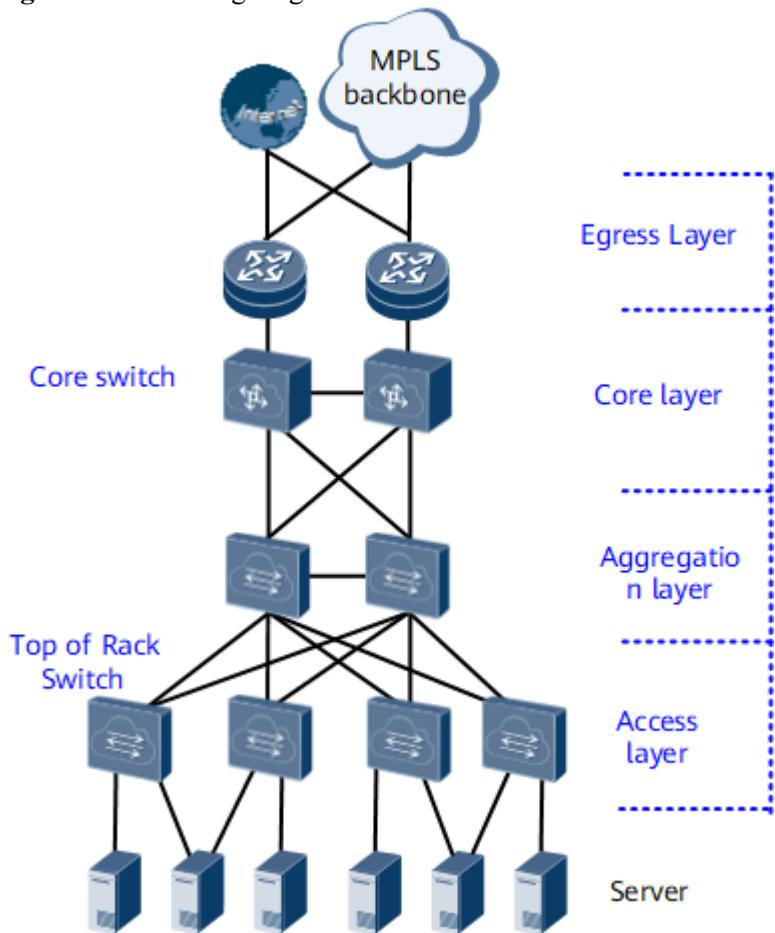
Networking Description

As shown in [Figure 1](#), a data center network consists of the following layers:

- Gateway layer: provides high-speed forwarding for inbound and outbound data of the data center.
- Core layer: provides high-speed forwarding for data from the aggregation layer.
- Aggregation layer: provides functions such as gateway redundancy for servers, load balancing, and firewall.

- Access layer: provides high-density network interfaces for data center servers.

Figure 1 Networking diagram for a data center



VPN services are becoming increasingly refined, and the demand for VPN service security is growing. Carriers must isolate different types of VPN services on data center networks to meet this demand, by configuring devices at the core layer to function as MCEs. The interfaces of an MCE can be bound to different VPNs based on service types, so that the MCE can create and maintain an independent routing and forwarding table for each VPN to completely isolate VPN services.

Feature Deployment

The MCE function is configured on the data center network shown in [Figure 1](#) as follows:

1. Different VPN instances are configured for different VPN services on core-layer devices.
2. Each interface of a core-layer device is bound to the corresponding VPN instance.
3. Gateway-layer devices that function as PEs are configured to connect to core-layer devices. The interfaces of gateway-layer devices connecting to core-layer devices are bound to the corresponding VPN instances.
4. A routing protocol is configured on devices at the gateway and core layers for these devices to communicate.

Parent Topic: [Application Scenarios for BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.3.3 Application of HVPN on an IP RAN

Service Overview

As 3G and Long Term Evolution (LTE) services develop, mobile operators keep building and expanding RANs. This situation imposes high requirements on the bandwidth, scalability, and configuration flexibility of the IP RAN between base transceiver stations (BTSs)/NodeBs/eNodeBs and base station controllers (BSCs)/radio network controllers (RNCs)/mobility management entities (MMEs). IP datacom networks, as the mainstream of datacom networks, are large in scale and provide a variety of access modes. To maximize carriers' return on investment, reduce network construction costs, and evolve the existing network smoothly to an LTE network, an IP RAN solution is introduced.

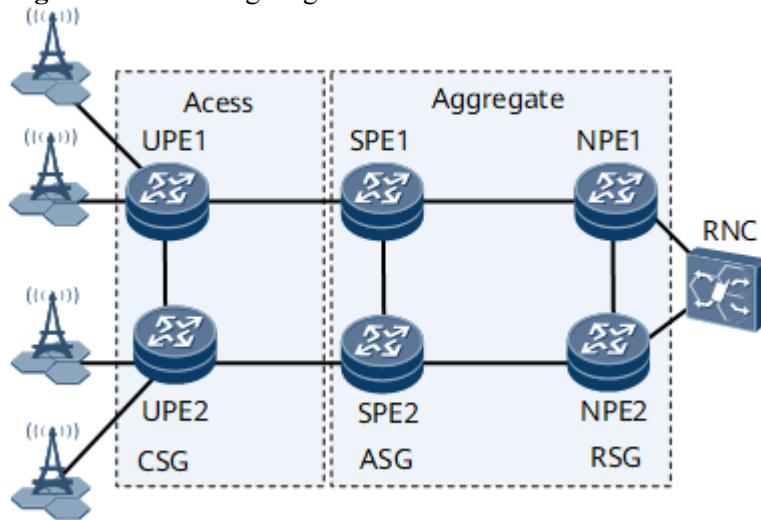
In the HVPN bearer solution, the RAN provides excellent fixed-mobile convergence (FMC) capabilities and has simple and flexible networking. The hierarchical network between CSGs and radio service gateways (RSGs) can bear various types of services.

Networking Description

As shown in [Figure 1](#), the HVPN bearer solution involves three types of devices:

- CSG: On an HVPN, CSGs function as UPEs to provide access services for BTSs/NodeBs/eNodeBs.
- Access service gateway (ASG): On an HVPN, ASGs function as SPEs to provide access services for UPEs.
- RSG: On an HVPN, RSGs function as NPEs to connect to BSCs/RNCs/MMEs.

Figure 1 Networking diagram for the HVPN bearer solution



Feature Deployment

The HVPN bearer solution applies to large-scale networks with dynamic routing capabilities. [Table 1](#) uses an HoVPN bearer solution as an example to describe feature deployment for E2E Ethernet service bearer and protection.

Table 1 Feature deployment in an HoVPN bearer solution

Feature	Description
---------	-------------

Feature	Description
Service	<ul style="list-style-type: none"> LTE S1/X2 3G Eth
IGP	<p>IGP multi-processes must be deployed between CSGs and RSGs to ensure proper data forwarding. A routing policy needs to be configured on ASGs to aggregate and filter routes. This configuration reduces bandwidth requirements for updating and maintaining routes.</p> <p>Recommended routing protocol: IS-IS multi-processes</p>
MPLS tunnel	<p>MPLS tunnels must be established between CSGs and ASGs, and between ASGs and RSGs.</p> <p>Recommended tunnel protocol: MPLS TE (tunnel selectors must be configured)</p>
VPN	Hierarchical L3VPN must be configured between CSGs and ASGs, and between ASGs and RSGs to isolate services.
Protection switching	<p>Protection switching must be configured for nodes and links to provide high availability:</p> <ul style="list-style-type: none"> BGP NSR and BGP GR must be enabled network-wide to ensure stability of neighbor relationships during protection switching. BGP tracking must be enabled network-wide to speed up IBGP route convergence. The ConnectRetry interval for re-establishing neighbor relationships must be prolonged between a CSG and the master ASG. VPN FRR must be configured on CSGs, ASGs, and RSGs. MPLS TE hot standby and BFD for TE CR-LSP must be configured to protect links between devices.
QoS	<p>E2E QoS must be configured between CSGs and RSGs to ensure service quality.</p> <p>Recommended QoS solutions: DiffServ</p>
Clock synchronization	<p>E2E clock synchronization must be configured between CSGs and RSGs to ensure real-time data transmission.</p> <p>Recommended clock synchronization technologies: synchronous Ethernet and 1588v2</p>

Parent Topic: [Application Scenarios for BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.7.3.4 Application of Route Import Between VPN and Public Network in the Traffic Cleaning Networking

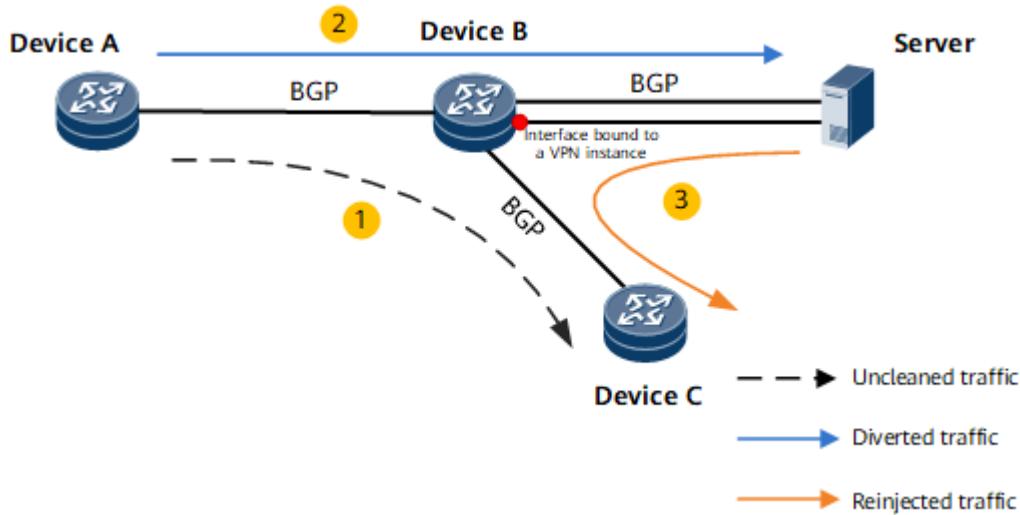
This section describes how to apply the function of route import between VPN and public network to the traffic cleaning networking.

In BGP/MPLS IP VPN networking, the users of a VPN can communicate with the users of another VPN if the two VPNs have matching VPN targets, but cannot communicate with public network users. [Figure 1](#) shows a traffic cleaning scenario. If attack traffic is detected, it is imported to the traffic cleaning server for cleaning. The cleaned traffic is injected back to the network through Device B. In this case, the public network routes destined for Device C need to be imported to the VPN routing table of Device B to forward the cleaned traffic to Device C. In addition, the public network routes sent by the cleaning server should not be imported to the VPN routing table. This prevents the reinjected traffic from being sent back to the cleaning server after reaching Device B, thereby preventing loops. To implement the preceding process, configure route import between VPN and public network on Device B, and configure a route-policy on the Device B sub-interface bound to a VPN instance.

On the network shown in [Figure 1](#), public BGP peer relationships are established between Device A and Device B and between Device B and Device C. The following uses the process of forwarding the Device A -> Device C traffic as an example:

1. If no attack traffic is detected, Device A learns public network routes from Device C, and traffic is forwarded along the expected path Device A -> Device B -> Device C.
2. If attack traffic is detected, the server advertises a public network route with a 32-bit mask as a traffic diversion route to Device B, which then diverts the Device A -> Device C traffic to the server for cleaning.
3. The server sends the cleaned traffic to Device B through a sub-interface. A VPN instance is bound to a sub-interface of Device B and is used to establish a VPN BGP peer relationship with the server. This allows Device B to import the public network service routes sent from Device C into its VPN routing table. After receiving the reinjected traffic through the VPN sub-interface, Device B searches its VPN routing table and forwards the traffic to Device C.

Figure 1 Application of route import between VPN and public network in the traffic cleaning networking



Parent Topic: [Application Scenarios for BGP/MPLS IP VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.8 VPWS Description

[Overview of VPWS](#)

[Understanding VPWS](#)

[Application Scenarios for VPWS](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.1 Overview of VPWS

Definition

VPWS is an L2VPN technology that transmits Layer 2 services by simulating the basic behaviors and features of services, such as ATM, FR, Ethernet, low-speed time division multiplexing (TDM) circuits, and synchronous optical network (SONET)/synchronous digital hierarchy (SDH) on a packet switched network (PSN). VPWS emulates the traditional leased line on an IP network and provides asymmetric and low-cost digital data network (DDN) services. For users at both ends of a VPWS connection, the connection is similar to the traditional leased line. VPWS functions as a point-to-point virtual private wire technology that can support almost all the link layer protocols.

Purpose

As IP networks develop, the expansibility, upgradability, and compatibility of IP networks have been greatly enhanced. Nevertheless, the expansibility, upgradability, and interworking capability of traditional communications networks are relatively poor. Confined by transmission modes and service types, resource sharing among newly constructed communications networks is also poor, complicating interworking management. In the process of upgrading and expanding a traditional communications network, you must determine whether to achieve your goal by constructing more traditional communications networks, or by utilizing existing network resources at your disposal, or by utilizing public network resources. VPWS is a solution that enables traditional communications networks to interwork with existing PSNs.

Benefits

VPWS offers the following benefits:

- Extended network functions and service capabilities

Carriers can provide MPLS L2VPN services by using only one network. In addition, carriers can use enhanced MPLS-related technologies, such as traffic engineering (TE) and quality of service (QoS), to provide users with different classes of services to meet their requirements.

- Higher scalability

On an ATM or FRnetwork where MPLS is not enabled, virtual circuits (VCs) are used to provide the L2VPN service. For each VC, both the PE and provider (P) devices on the network need to maintain complete VC information. That means that when the PEs of a carrier connect to multiple CEs, multiple VCs are required, and information about multiple VCs must be maintained on both PEs and Ps. VPWS uses label stacking to multiplex multiple VCs on a label switched path (LSP). As a result, Ps only need to maintain information about one LSP. This improves the scalability of a system.

- Well-defined administration roles

On a VPWS network, a carrier provides only Layer 2 connectivity and users are responsible for Layer 3 connectivity, such as routing. This implementation prevents route flapping caused by incorrect configurations from affecting the stability of the carrier's network.

- Support for multiple protocols

Carriers provide only Layer 2 connections, whereas users can use any Layer 3 protocol, such as IPv4 and IPv6.

- Smooth network upgrade

The VPWS network is transparent to users. When a carrier upgrades the network from a traditional L2VPN, such as an ATM and FRnetwork, to a VPWS network, users do not need to perform any configuration. The network upgrade does not affect user services, except for data loss for a short period during the network cutover.

Parent Topic: [VPWS Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.8.2 Understanding VPWS

[VPWS Basic Functions](#)

[VPWS in CCC Mode](#)

[LDP VPWS](#)

[VPWS in SVC Mode](#)

[VPWS in BGP Mode](#)

[Heterogeneous VPWS](#)

[ATM Cell Relay](#)

[VCCV](#)

[PW Redundancy](#)

[PW APS](#)

[Comparison of VPWS Implementation Modes](#)

[Comparison of LDP VPWS and BGP/MPLS IP VPN](#)

[Inter-AS VPWS](#)

[Flow-Label-based Load Balancing](#)

[Mutual Protection Between an LDP VC and a CCC VC](#)

[Multi-Segment PW Redundancy](#)

Parent Topic: [VPWS Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

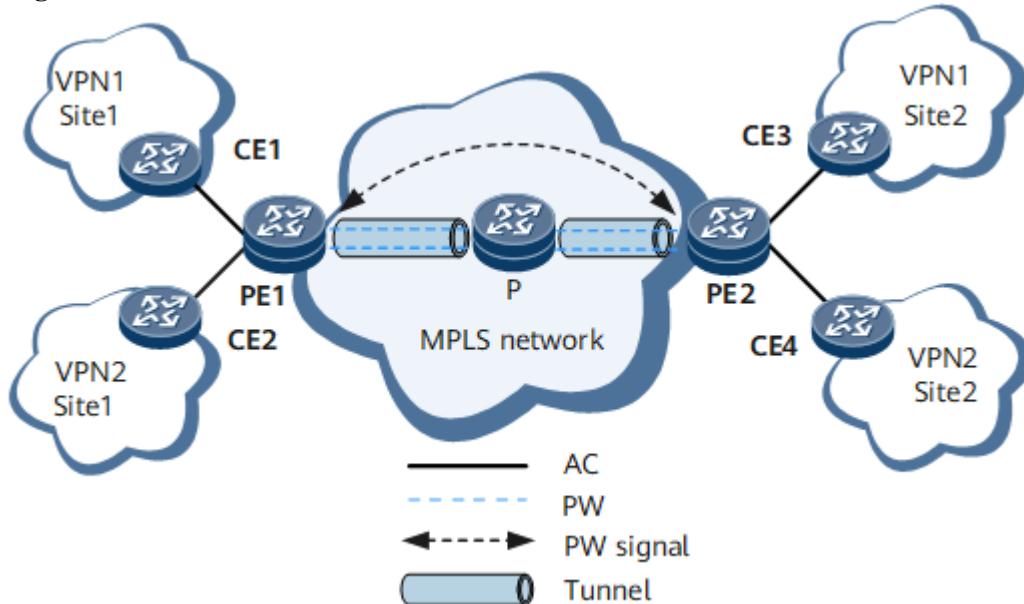
< Previous topic [Next topic >](#)

1.8.2.1 VPWS Basic Functions

Basic VPWS Architecture

As shown in [Figure 1](#), the VPWS architecture consists of ACs, PWs, and tunnels.

Figure 1 Basic VPWS architecture



Functional Modules

VPWS has the following functional modules:

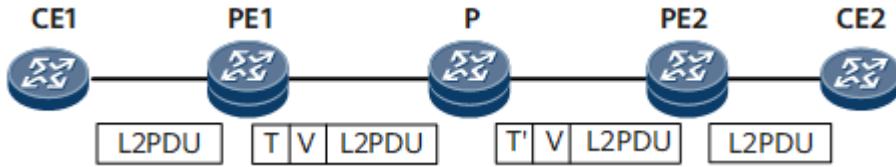
- AC: An independent physical or virtual circuit connecting a CE and a PE. An AC interface can be a physical interface or a virtual interface. The AC attributes include the encapsulation type, maximum transmission unit (MTU), and interface parameters of the specified link type.
- PW: A virtual link or path between two nodes on a network. In this document, it is a virtual connection between two PEs.
- Tunnel: A virtual link used to transparently transmit service data.
- PW signaling: A type of signaling for PW negotiation.

[Figure 1](#) uses the flow direction of VPN1 packets from CE1 to CE3 as an example to show data transmission.

- CE1 sends user packets to PE1 over an AC.
- Upon receipt, PE1 selects a PW to forward the packets.
- PE1 generates double MPLS labels (one VPN label and one public network label) according to the PW forwarding entry. The VPN label identifies a PW, and the public network label identifies a public network tunnel.
- After user packets travel along the public network tunnel to PE2, PE2 removes the VPN label. The public network label is removed by means of penultimate hop popping (PHP) on the P.
- PE2 selects an AC and forwards these packets to CE3.

[Figure 2](#) shows label changes in packet forwarding over a VPWS network.

Figure 2 VPWS label processing



In [Figure 2](#):

- L2PDU: Layer 2 protocol data unit, a type of link-layer packet.
- T: a tunnel label.
- V: a virtual circuit (VC) label.
- T': a substitute tunnel label during packet forwarding.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.8.2.2 VPWS in CCC Mode

CCC is short for circuit cross connect. A CCC is manually configured to implement L2VPN.

A CCC must be configured by network administrators and is best suited for small MPLS networks with simple topologies. CCC VPWS does not require signaling negotiation or exchange of control packets. Compared with other types of Layer 2 connections, CCC VPWS consumes relatively few resources and are easy to configure. However, CCC VPWS has poor scalability and is inconvenient to maintain.

Network Topology of VPWS in Local CCC Mode

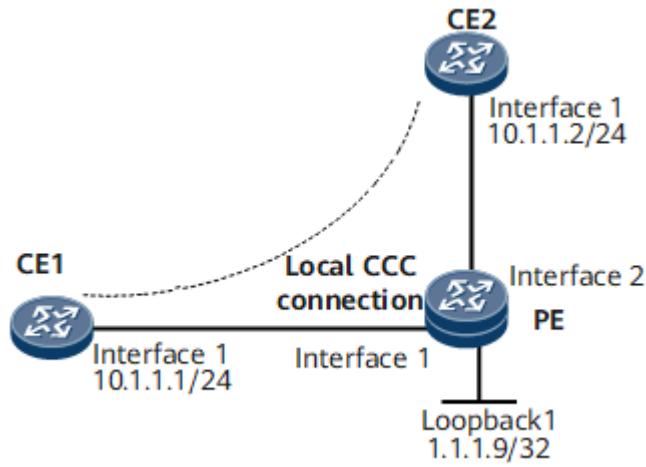
A local CCC is a connection between two CEs connected to the same PE. Similar to a Layer 2 switch, a PE can directly complete switching without the need to configure a label switched path (LSP).

[Figure 1](#) shows the topology in local CCC mode.

Figure 1 Topology of local CCC VPWS



- Interface 1 and interface 2 in this example represent GE 1/0/0 and GE 1/0/0, respectively.
-



As shown in [Figure 1](#), the CEs and PE are connected through GE interfaces. A local CCC connection is set up between CE1 and CE2. PE to which CE1 and CE2 are connected functions as a Layer 2 switch. Data of different link encapsulation types, such as VLAN, PPP, HDLC, and Ethernet, can be directly exchanged between the CEs.

The advantage of this mode is that no label signaling is required to transmit Layer 2 VPN information, and the ISP network merely needs to support MPLS forwarding.

Network Topology of VPWS in Remote CCC Mode

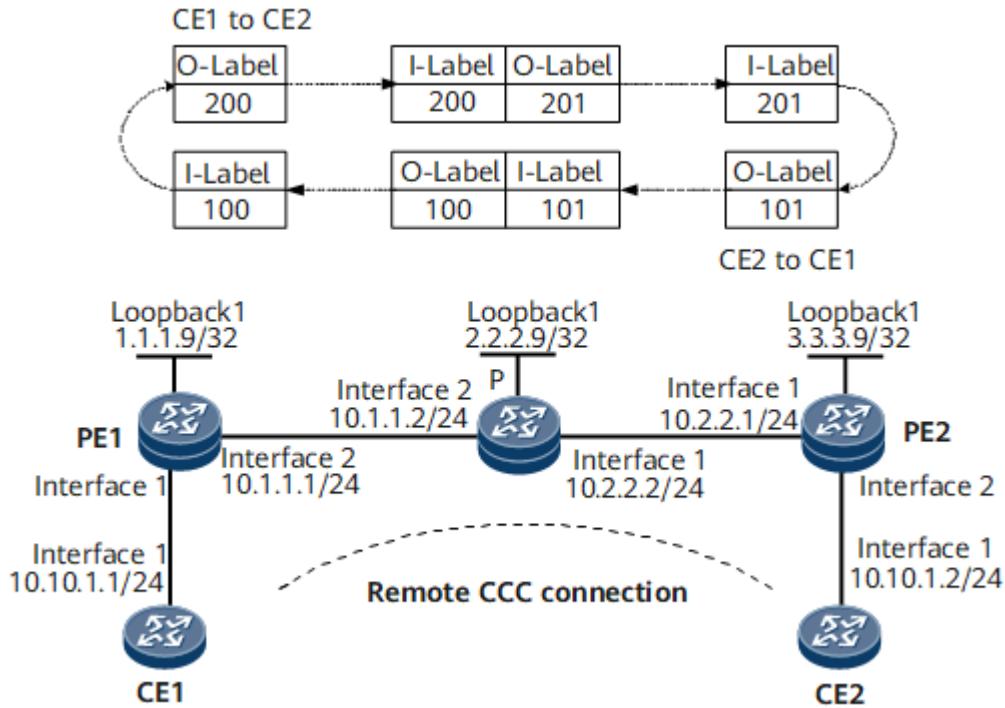
A remote CCC connection is set up between a local CE and a remote CE. Two CEs are connected to different PEs. Static CR-LSPs need to be configured to transmit packets from one PE to the other PE. A static CR-LSP must be configured on each PE to map to CCC connections.

[Figure 2](#) shows the topology in remote CCC mode.

Figure 2 VPWS Topology in remote CCC mode

NOTE

- Interfaces 1 and 2 in this example represent GE 1/0/0 and GE 1/0/0, respectively.
-



As shown in [Figure 2](#), CE1 and CE2 are connected to different PEs. To allow the two CEs to communicate, create a remote CCC connection; configure two static CR-LSPs on the P to transmit packets in both directions.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.3 LDP VPWS

Description

LDP VPWS is an MPLS L2VPN technology that establishes point-to-point links to implement L2VPN and uses the LDP signaling to transmit VC information.

LDP VPWS uses double labels for traffic transmission. The inner label is exchanged using extended LDP, and the outer label is a tunnel label.

On an LDP VPWS network, multiple VCs can be established over one LSP between two PEs. In addition, PEs store only a small amount of L2VPN information, such as mappings between VC labels and LSPs. Ps do not store any L2VPN information. Therefore, LDP VPWS has excellent scalability. To add a VC, you only need to configure a unidirectional VC on each endpoint PE. This operation does not affect network operations.

LDP VPWS uses the LDP signaling protocol, being independent from the periodic refresh mechanism. Therefore, LDP VPWS supports fast fault detection.

Basic Concepts

In LDP VPWS, the VC type and VC ID together uniquely identify a VC between CEs.

- The VC type indicates the encapsulation type of a VC. The VC type can be PPP, High-Level Data Link Control (HDLC), Ethernet, or VLAN.
- The VC ID identifies a VC. VCs of the same type must have unique IDs on a PE.

The endpoint PEs exchange VC labels using LDP, and each CE is bound to the peer CE according to the VC ID. A VC can be successfully established if the following conditions are all met:

- The physical status of AC interfaces is up.
- The tunnel between the PEs has been established.
- Labels have been exchanged between two PEs and the CE binding have been completed.

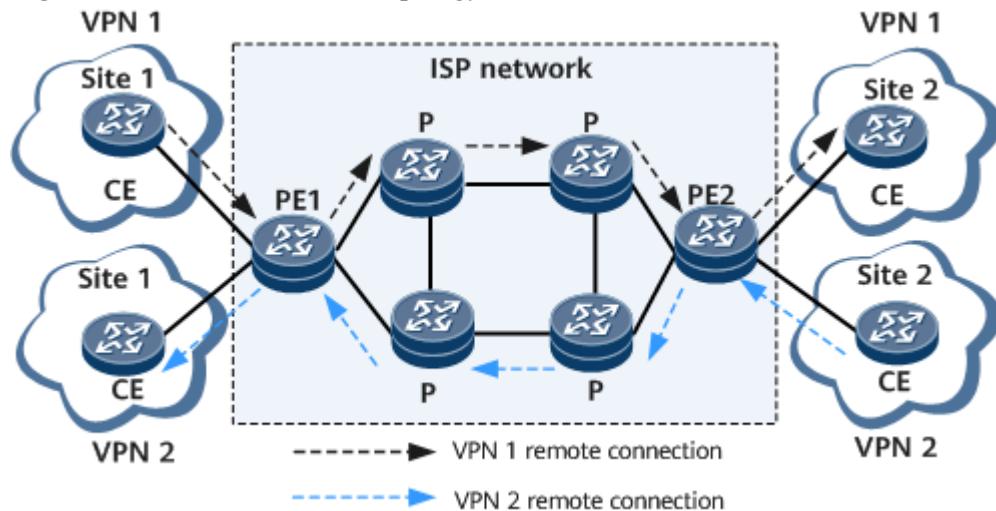
In LDP VPWS, the outer label is used to transmit the data of each VC over an ISP network, and the inner VC label is used to identify the type of service data. In light of this, an LSP on the ISP network can be shared by multiple VCs.

To support LDP VPWS, an ISP network must be able to automatically establish LSPs. This means that the ISP network must support MPLS forwarding and MPLS LDP.

LDP VPWS Network Topology

[Figure 1](#) shows the LDP VPWS network topology.

Figure 1 LDP VPWS network topology



As shown in [Figure 1](#), site 1 and site 2 of VPN 1 are connected over a remote LDP connection (as indicated by the black dashed line). Site 1 and site 2 of VPN 2 are also connected over a remote LDP connection (as indicated by the blue dashed line). Either one or two LSPs can be established within the ISP network for communication between VPN 1 and VPN 2.

PWE3 VPWS

LDP VPWS is classified into PWE3-compatible VPWS and PWE3 VPWS.

- PWE3-compatible VPWS: does not use Notification messages.
- PWE3 VPWS: uses Notification messages.

PWE3 simulates the basic behaviors and characteristics of services, such as ATM, FR, Ethernet, low-speed TDM circuit, and SONET/SDH, on a PSN to transmit Layer 2 traffic.

PWE3 is a type of VLL implementation and an extension to the LDP protocol. By extending the LDP signaling, reducing the signaling cost, and defining the multi-segment negotiation mode, PWE3

improves networking flexibility. Compared with LDP VPWS, PWE3 VPWS exchanges fewer packets when the network is unstable, preventing repeated PW establishment and deletion.

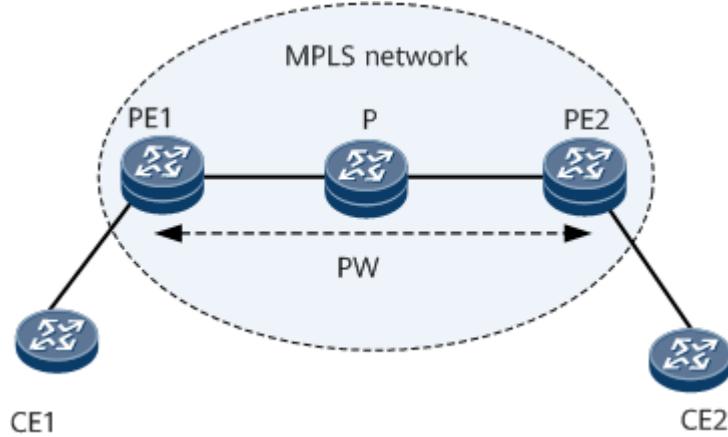
- PWE3 networking modes
 - Single-segment PWE3 networking

Single-segment PWE3 means that only one PW exists between two PEs, and no inner label swapping is needed. Because the PW uses LDP as a signaling protocol to transmit VC information, an LDP session must be established between the PEs:

- If Ps exist between the PEs, a remote LDP session must be created.
- If the PEs are directly connected, a local LDP session can be created.

[Figure 2](#) shows the typical single-segment PWE3 networking with a PW established using LDP signaling.

Figure 2 Single-segment PWE3 topology



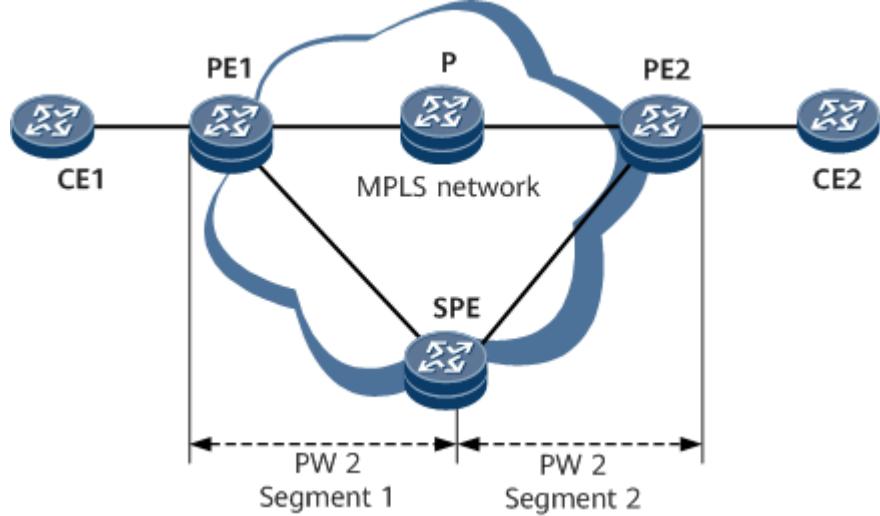
- Multi-segment PWE3 networking

An MS-PW is a set of two or more PW segments between two PEs. The MS-PW forwarding mechanism is the same as the SS-PW forwarding mechanism on PEs. The only difference is that PW labels are swapped on switching PEs (SPEs) for MS-PWs. [Figure 3](#) shows the typical multi-segment PWE3 networking.

If two PEs cannot establish a connection using signaling or cannot establish a direct tunnel, configure an MS-PW between the two PEs. Supporting PWE3 MS-PWs improves networking flexibility.

Besides being classified into SS-PWs and MS-PWs, PWs can also be classified into static and dynamic PWs. The two types of PWs can be used together. For example, an MS-PW can be a set of static and dynamic PW segments.

Figure 3 Multi-segment PWE3 networking



- Dynamic PW

Dynamic PWs are established using signaling protocols. UPFs swap VC labels using LDP and bind the corresponding CEs to AC interfaces based on VC IDs. A VC is established when all the following conditions are satisfied: two PEs have established a tunnel; the two PEs have exchanged VC labels and bound the VC ID to corresponding CEs; the ACs of the two PEs are up.

Messages used by a dynamic PW include:

- Request: requests a peer PE to allocate labels.
- Mapping: notifies a peer PE of a label allocated by a local PE. Whether the Label Mapping message carries the Status field depends on the default signaling. By default, LDP VPWS does not support the Status field.
- Notification: advertises and negotiates the PW status, reducing the number of messages to be exchanged.
- Withdraw: carries label and status information to instruct the peer PE to withdraw labels.
- Release: responds to a Label Withdraw message, and notifies the peer that sends the Label Withdraw message of the label release event.

- Establishment, Maintenance, and Deletion of Dynamic PWs

Dynamic PWs are established using LDP, whose TLV is extended to carry VC information. Before a dynamic PW is established between two PEs, an LDP session must be established between the two PEs. During the establishment of a dynamic PW, the label distribution control mode is downstream unsolicited (DU) and the label retention mode is liberal.

NOTE

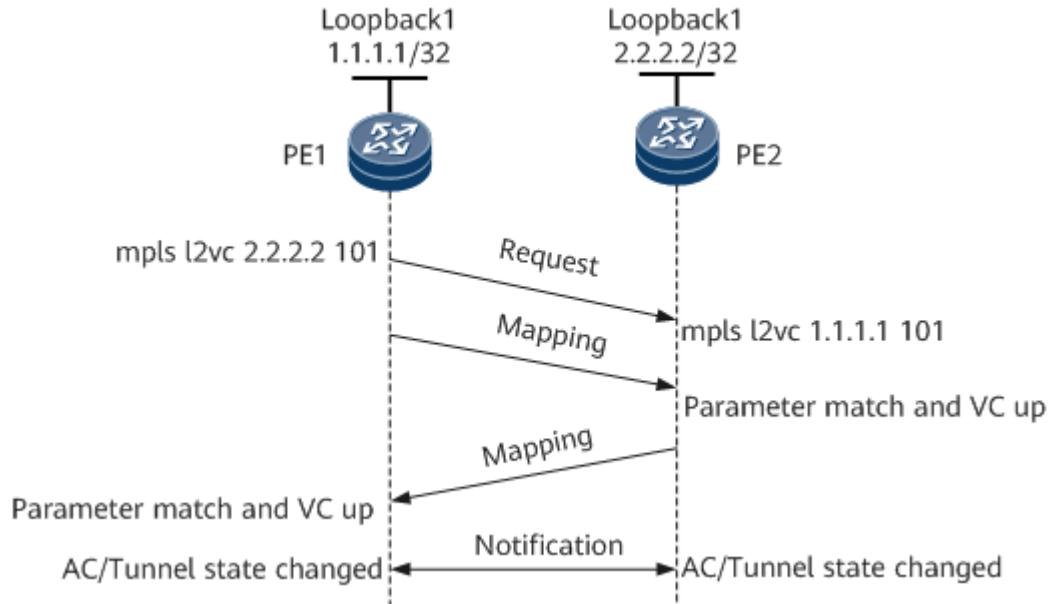
If Ps exist between the two PEs, a remote LDP session can be established. If the two PEs are directly connected, a local LDP session can be established.

After PWE3 is configured on the two PEs and an LDP session is established between the two PEs, the dynamic PW starts to be established. [Figure 4](#) shows the process of establishing a dynamic PW.

1. PE1 sends a Label Request message and a Label Mapping message to PE2.

2. After receiving the Label Request message from PE1, PE2 sends a Label Mapping message to PE1.
3. After receiving the Label Mapping message from PE1, PE2 determines whether the configuration of a PW is consistent with that on PE1. If its PW configurations such as the VC ID, VC type, MTU, and control word (CW) enabling status are consistent with those on PE1 after negotiation, PE2 sets the PW status to up.
4. After receiving the Label Mapping message from PE2, PE1 determines whether its PW configurations are consistent with those on PE2 after negotiation. If the parameters are consistent, PE1 sets the PW status to up. After that, a dynamic PW is established between PE1 and PE2.
5. After the dynamic PW is established, PE1 and PE2 learn the status of each other by exchanging Notification messages.

Figure 4 Process of establishing and maintaining an SS-PW

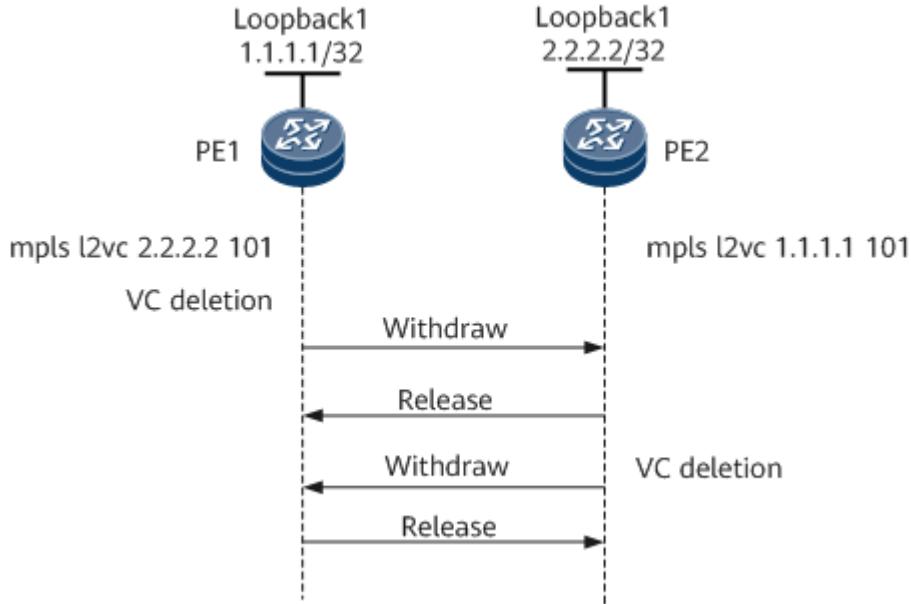


If the AC interface of a PW is down or the corresponding tunnel is down, PWE3-compatible VPWS and PWE3 VPWS use different processing mechanisms:

- In PWE3-compatible mode, the local PE sends a Label Withdraw message to its peer to tear down the PW. After the AC interface or tunnel goes up, another round of negotiation is required for the PEs to establish a PW.
- In PWE3 mode, the local PE sends a Notification message to notify its peer that data packets cannot be forwarded, but the PW is not torn down. After the AC interface or tunnel goes up, the local PE sends a Notification message to notify its peer that data packets can be forwarded.

A PW is torn down only when PW configurations are deleted from the PEs or when the LDP session is interrupted. Using Notification messages prevents repeated PW establishment and deletion caused by link flapping.

Figure 5 Process of tearing down an SS-PW



[Figure 5](#) shows the process of tearing down a PW.

1. When PW configurations are deleted from PE1, PE1 withdraws its VC label and sends Label Withdraw and Label Release messages to PE2 in succession.
2. After receiving the Label Withdraw message from PE1, PE2 withdraws its VC label and sends a Label Release message to PE1.
3. After PE1 and PE2 receive the Label Release message from each other, PE1 and PE2 have deleted the PW.

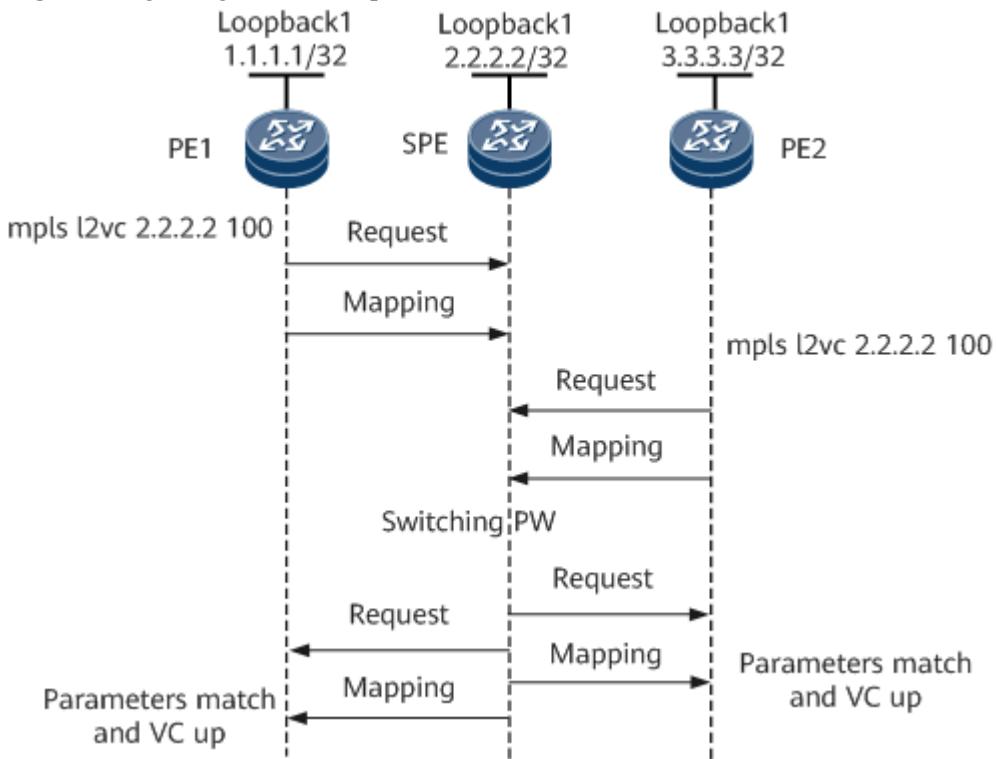
NOTE

The Label Withdraw message instructs a PE to withdraw the PW label. The Label Release message is a response to the Label Withdraw message to inform the PE sending the Label Withdraw message that the PW label has been withdrawn on the peer. To tear down the PW more quickly, PE1 can send the Label Withdraw and Label Release messages in succession.

The difference between an SS-PW and an MS-PW is that one or more SPEs exist between the endpoint PEs of an MS-PW. [Figure 6](#) shows an example of the signaling interaction process for an MS-PW between PE1 and PE2. The SPE connects the two PW segments that are established to PE1 and PE2.

During signaling negotiation, the SPE forwards to PE2 the parameters carried in the Label Mapping message sent by PE1 and forwards to PE1 the parameters carried in the Label Mapping message sent by PE2. If these parameters are consistent through negotiation, the PW status becomes up. Similar to Label mapping messages, Label Release, Withdraw, and Notification messages are forwarded segment by segment.

Figure 6 Signaling interaction process for an MS-PW



- Extensions on the PWE3 control plane

- Signaling extension

The means of sending a Notification message is added to the LDP signaling, which is merely to advertise the status but will not tear down a signaling connection. A signaling disconnection occurs only when the PW configurations are deleted or the signaling is interrupted. This signaling extension allows for fewer control packet exchanges, reduces the signaling cost, and is compatible with the original LDP mode.

- Other extensions

Other extensions on the control plane are as follows:

- Mechanism for negotiating fragmentation
 - PW connectivity detection, such as virtual circuit connectivity verification (VCCV), is added, improving the fast network convergence capability and network reliability.

- Extensions on the PWE3 data plane

- Real-time information extension
 - Bandwidth, jitter, and delay assurance of electrical signals
 - Retransmission of disordered packets

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.4 VPWS in SVC Mode

Definition

SVC VPWS is an L2VPN technology that uses manually configured VC labels for data transmission. Unlike LDP VPWS, which uses LDP to exchange VC labels, SVC VPWS uses VC labels manually assigned on the PE according to the VC ID. The SVC mode can be regarded as the simplified LDP mode.

A VC label of a VPWS in SVC mode is statically configured and does not require VC label mapping, avoiding the need of LDP to transmit VC labels.

Network Topology of VPWS in SVC Mode

Specifying the outer label (which identifies a public tunnel) in SVC mode is similar to that in LDP mode. The inner label is specified manually during the VC configuration. PEs do not need the signaling protocol to transmit VC labels. The network topology and packet exchange process in SVC mode are similar to those in LDP mode.

When creating a static Layer 2 VC connection in SVC mode, you can specify a tunnel type, such as LDP or TE. In addition, you can configure a tunnel policy for tunnels to work in load balancing mode.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

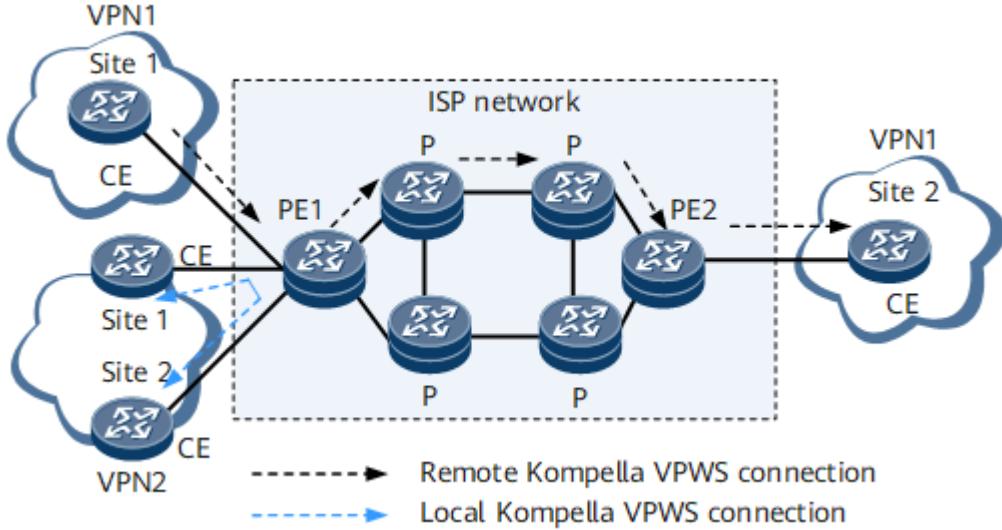
1.8.2.5 VPWS in BGP Mode

Definition

BGP VPWS uses BGP as the signaling protocol to transmit Layer 2 information and VC labels between PEs.

BGP VPWS uses VPN targets to identify different VPNs, creating greater flexibility for VPN networking. BGP VPWS assigns VC labels from label blocks. A label block is allocated to each CE in advance. The size of the label block allocated to a CE determines the number of connections that the CE can establish with other CEs. BGP VPWS allows allocation of additional labels to a CE for future VPN capacity expansion. PEs figure out the inner labels of packets based on these label blocks and then transmit packets based on inner labels. BGP VPWS has good scalability and supports both local and remote connections.

Figure 1 BGP VPWS network



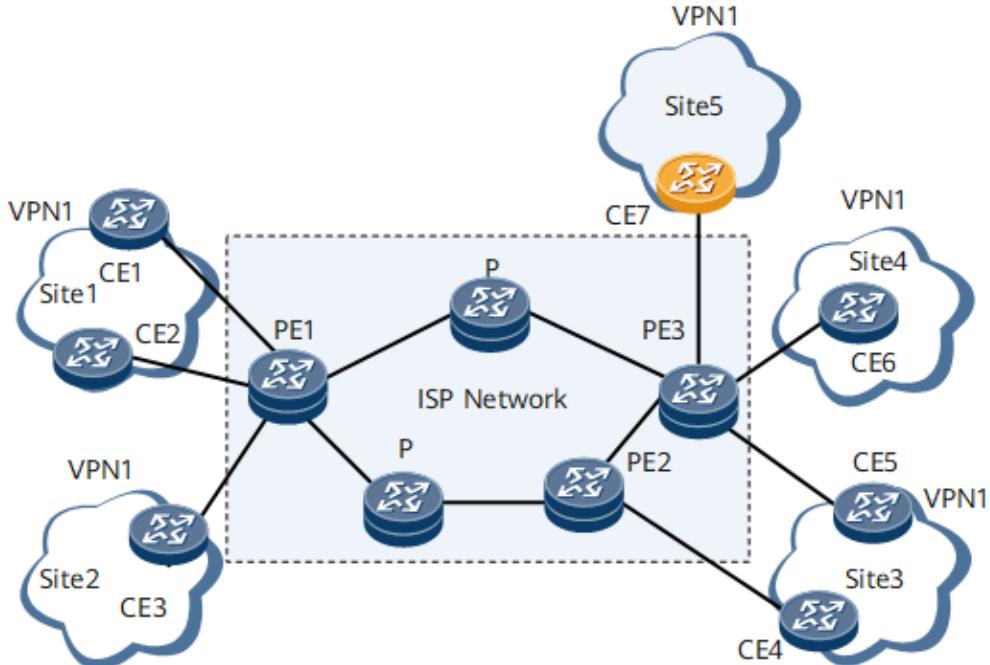
Basic Concepts

BGP VPWS can use label blocks to allocate labels to multiple connections at the same time. The CE range specified for a CE indicates the number of connections that can be established between this CE and other CEs. Only one label block can be allocated to a CE at one time. The label block size equals the CE range. The practice of additional label allocation may waste labels in the short term, but will reduce configuration workload during future VPN capacity expansion. Assume that an enterprise VPN has 10 CEs and the number may increase to 20 due to service expansion in the future. In this situation, you can set the CE range for each CE to 20 to reserve labels for the 10 CEs to be added later.

Implementation

Packet forwarding by BGP VPWS is similar to that by LDP VPWS. The two types of VPWS both use standard Layer 2 labels, but use different signaling protocols to exchange these labels: LDP VPWS uses extended LDP, whereas BGP VPWS uses MP-BGP.

Figure 2 Packet forwarding by BGP VPWS



On the network shown in [Figure 2](#), six CEs ranging from CE1 to CE6 access VPN1. To enable these CEs to communicate, full-mesh connections must be established. In other words, a CE must establish a VC with each of the other CEs. To establish these connections, perform the following configurations on PE1, PE2, and PE3:

1. Configure VPN1 on each PE and bind local CEs to VPN1. For example, bind CE1, CE2, and CE3 to VPN1 on PE1.
2. Allocate label blocks to CEs. Here, each CE connects to five CEs. Therefore, a label block containing at least five labels must be allocated to each CE.
3. On each PE, specify peer CE IDs and PE interfaces connecting to local CEs.

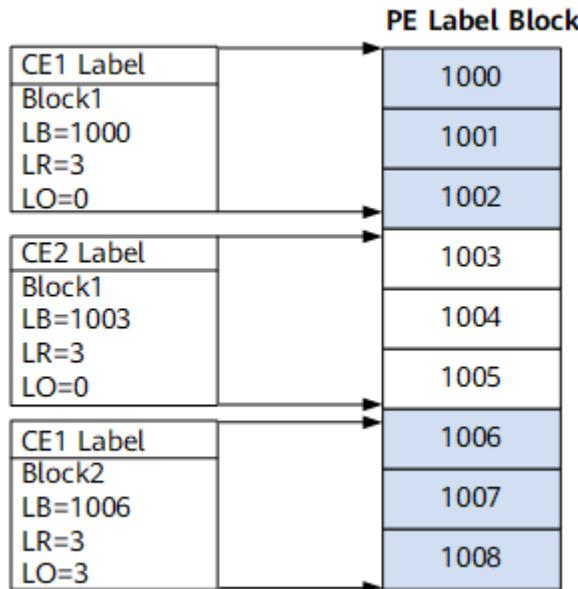
Like CCC VPWS, BGP VPWS also supports local connections. It is easy to use BGP VPWS to establish full-mesh connections.

VC Label Calculation

A label block is a consecutive range of labels. BGP VPWS uses MP-BGP as the signaling protocol to transmit label block information.

For a clear description of the label block, several parameters are defined: Label Base (LB), initial label of the label block, Label Range (LR), and Label-Block Offset (LO), as shown in [Figure 3](#).

Figure 3 Calculating VC labels



When adding CE information to a PE, you must specify the LR. The LB is automatically allocated by the PE. This label block is used as a network layer reachable information (NLRI) entry and transmitted to other PEs through BGP. When CE information is deleted or the connection between the PE and CE becomes invalid, the label block is deleted. Meanwhile, BGP sends a Withdraw message for notification. Assume that when BGP VPWS is deployed, CE1 needs to establish two VCs with other remote CEs; then the size of the label block cannot be smaller than 2. Considering future capacity expansion, you can define the range as 10.

The labels may be insufficient as VCs increase, regardless of the range. If this situation occurs, redefine the range for a larger label space. A problem then arises. As mentioned earlier, the data of the label block is transmitted through the BGP NLRI, and this label block is used to calculate VC labels and forward data. To protect the original VC connection, you can allocate a new label block to this CE and advertise the label block as a new NLRI through BGP. In other words, the label space of a CE may consist of multiple label blocks. The LO defines the relationship between multiple labels. The

LO of a label block identifies the total size of the label blocks preceding this label block. For example, if the LR of the first label block is 100 and the LO is 0, and the LR of the second label block is 50, then the LO of the second label block is 100. Then the LO of the third label block if any is 150. The LO is used in calculating VC labels. Therefore, a label block can be defined by three parameters: LB, LR, and LO.

CE IDs are used as follows:

- A CE ID uniquely identifies a CE in a VPN. In a VPN, CE IDs must be unique. The CE ID is carried in each NLRI; different label blocks can therefore be associated with their corresponding CEs.
- CE IDs can also be used to calculate VC labels. For this reason, they cannot be chosen at random. The condition $x > y$ must be met if the range of the local CE is x and the local CE connects to a peer CE with the CE ID being y . Otherwise, the value of x must be increased to meet the preceding condition.

Figure 4 Calculating label blocks

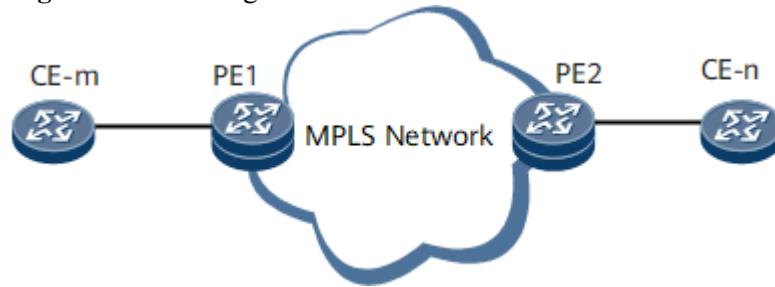


Table 1 Calculating label blocks

Item	Description	Item	Definition
Label block allocated by PE-1 to CE-m	Lm	Label block allocated by PE-2 to CE-n	Ln
LO of Lm	LOm	LO of Ln	LOn
LB of Lm	LBm	LB of Ln	LBn
LR of Lm	LRm	LR of Ln	LRn

Assume that PE1 and PE2 establish a VC between CE-m and CE-n that belong to the same VPN.

PE1 receives a label block LBn/LRn/LOn from PE2.

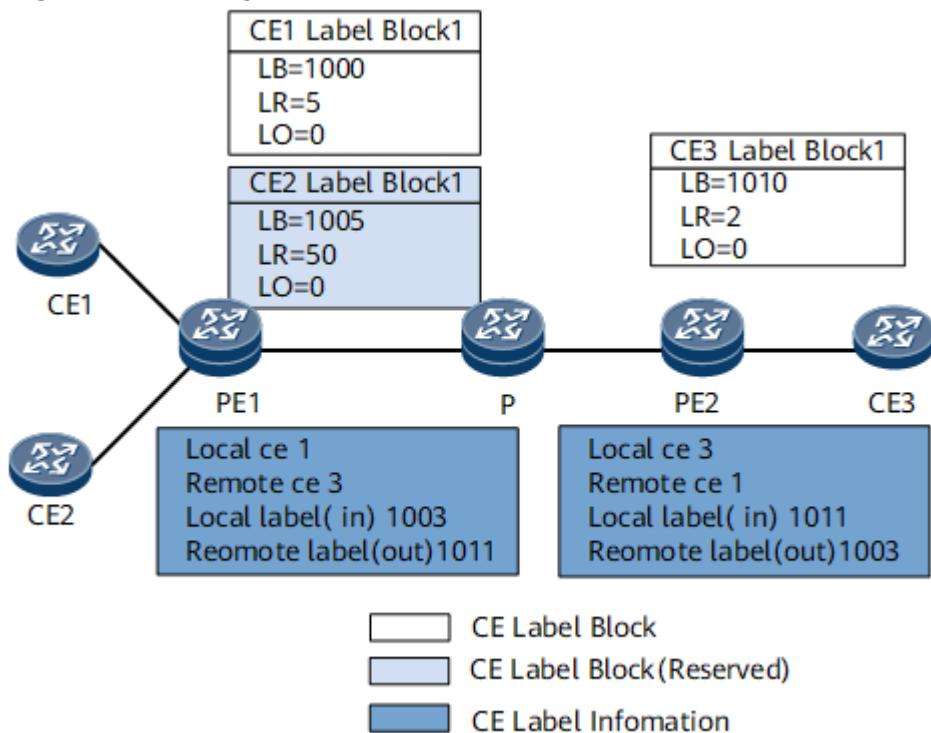
1. PE1 checks whether the encapsulation type of CE-n received from PE2 is the same as that of CE-m. If not, PE1 stops processing.
2. PE1 checks the CE ID to see whether $m = n$. If $m = n$, PE1 reports an error and stops processing.
3. If CE-m has multiple label blocks, PE1 checks whether these label blocks meet the condition $LOm \leq n < LOm + LRm$. If the condition is not met, PE1 reports an error and stops processing.
4. PE1 checks whether all the label blocks related to CE-n meet the condition $LOn \leq m < LOn + LRn$. If the condition is not met, PE1 reports an error and stops processing.

5. PE1 checks whether the outer tunnel between PE-m and PE-n is established normally. If not, PE1 stops processing. The outer tunnel is assumed as an LSP tunnel with the label as Z.
6. PE1 allocates an inner label ($LBn + m - LOn$), the outgoing label of the VC, to CE-n; PE1 allocates an inner label ($LBm + n - LOm$), the incoming label of the VC, to CE-m.
7. The label of the outer tunnel from PE2 to PE1 is Z.
8. After the inner and outer labels are calculated and the VC is Up, Layer 2 packets can be transmitted.

The following example describes the process of allocating CE label blocks phase by phase.

Assume that all PEs exchange label block information through BGP. All the public network LSP tunnels are Up. Only VC labels need to be calculated. The ID of CE1 is 1; the ID of CE2 is 2; and the rest can be deduced by analogy.

Figure 5 Calculating VC labels



On the network shown in [Figure 5](#), label blocks are allocated to CE1 and CE3 as follows:

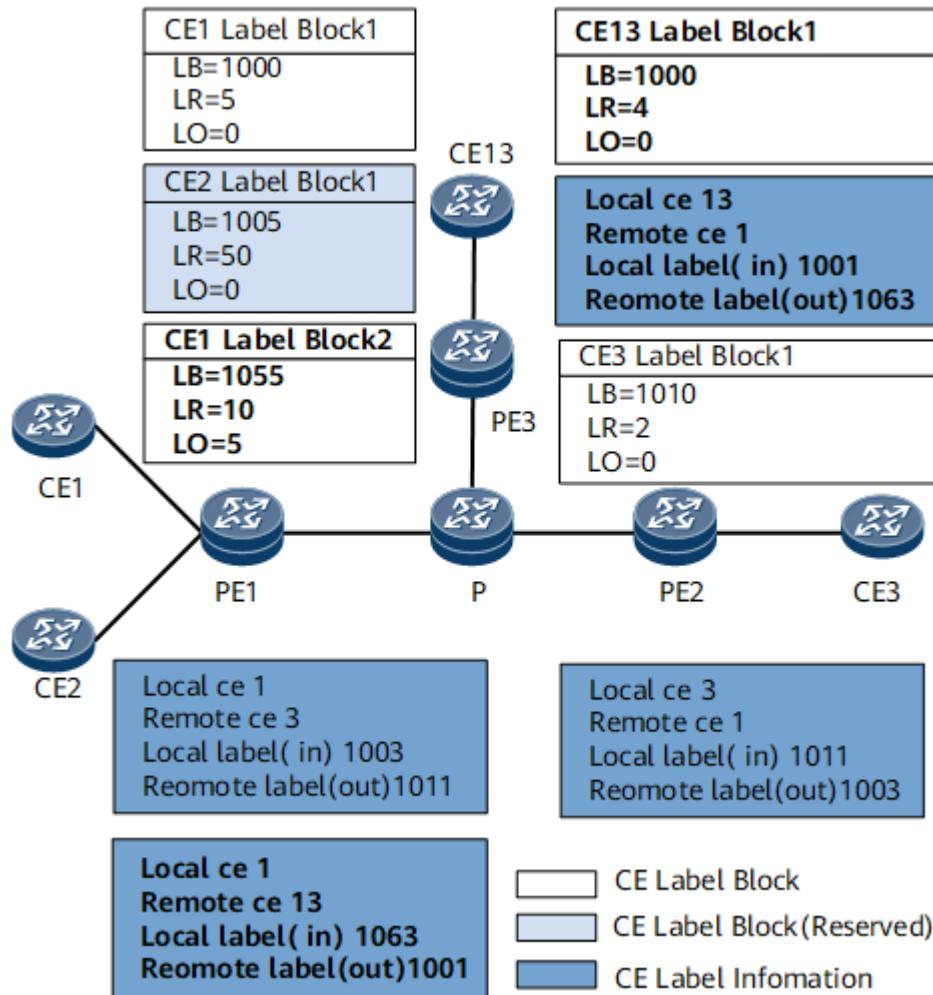
1. PE1 allocates a label block LB/LR/LO = 1000/5/0 to CE1 and receives the label block LB/LR/LO = 1010/2/0 allocated to CE3 from PE2. According to the preceding calculation rule, the incoming and outgoing labels of the VC can be calculated.
2. PE1 also connects to CE2. Following the allocation of the label block to CE1, PE1 allocates a label block LB/LR/LO = 1005/50/0 to CE2. At the moment, CE2 does not establish any connection with other CEs. The label block is allocated for future capacity expansion. Therefore, PE1 does not calculate labels for CE2.
3. PE1 determines whether the label block allocated to CE1 meets the requirement $LOm \leq n < LOm + LRm$. Because, LOm is 0, n is 3, and LRm is 5, this requirement is met. PE1 also re-determines whether this label block meets the requirement $LOn \leq m < LOn + LRn$. Because LOn is 0, m is 1, and LRn is 2, this requirement is also met.
4. PE1 calculates the incoming and outgoing VC labels. The outgoing VC label (inner label of CE3) is $LBn + m - LOn = 1010 + 1 - 0 = 1011$. The incoming VC label (inner label of CE1) is $LBm + n - LOm = 1005 + 1 - 0 = 1006$.

is $LBm + n - LOm = 1000 + 3 - 0 = 1003$.

5. PE2 determines whether the label block allocated by PE1 to CE1 meets the requirement $LOm <= n < LOm + LRm$. Because LOm is 0, n is 1, and LRm is 2, this requirement is met. PE1 also re-determines whether this label block meets the requirement $LOn <= m < LOn + LRn$. Because LOn is 0, m is 3, and LRn is 5, this requirement is also met.
6. PE2 calculates the incoming and outgoing VC labels. The outgoing VC label (inner label of CE1) is $LBn + m - LOn = 1000 + 3 - 0 = 1003$. The incoming VC label (inner label of CE3) is $LBm + n - LOm = 1010 + 1 - 0 = 1011$.

As shown in [Figure 6](#), if CE13 is added to the VPN, a VC must be established between CE13 and CE1. PE3 allocates a label block LB/LR/LO = 1000/4/0 to CE13. A new VC needs to be established between CE1 and CE13.

Figure 6 Calculating VC labels after a CE is added



Whether the ID of CE13 is appropriate is judged in a similar way. PE1 determines whether the label block allocated to CE1 meets the requirement $LOm <= n < LOm + LRm$. Because LOm is 0, n is 13, and LRm is 5, this requirement is not met. PE1 determines whether the label block allocated to CE1 meets the requirement $LOn <= m < LOn + LRn$. Because, LOn is 0, m is 1, and LRn is 4, this requirement is met.

The CE ID 13 is greater than the $LO + LR$ of CE1; therefore, the outgoing VC label (inner label of CE1) cannot be calculated. You need to modify the label range of CE1 on PE1 to 15 in this example, and allocate another label block to CE1 with the label range as 10. Following the allocation of the label block to CE1, PE1 allocates a second label block LB/LR/LO = 1055/10/5 to CE1. PE1 re-determines whether this label block meets the requirement $LOm <= n < LOm + LRm$. Because, LOm is 5, n is 13, and LRm is 10, this

requirement is also met. PE1 also re-determines whether this label block meets the requirement $LOn \leq m < LOn + LRn$. Because LOn is 0, m is 1, and LRn is 4, this requirement is also met.

Similarly, PE3 receives the two label blocks $LB/LR/LO = 1000/5/0$ and $LB/LR/LO = 1055/10/5$ allocated by PE1 to CE1. Only the label block $LB/LR/LO = 1055/10/5$ meets the requirement.

PE1 determines that labels must be allocated from the second label block of CE1. The outgoing VC label calculated by PE1 is $LBn + m - LOn = 1000 + 1 - 0 = 1001$. The incoming VC label calculated by PE1 is $LBm + n - LOm = 1055 + 13 - 5 = 1063$.

According to the formula, PE3 calculates the VC labels. The outgoing VC label calculated by PE3 is $LBm + n - LOm = 1055 + 13 - 5 = 1063$; the incoming label calculated by PE3 is $LBn + m - LOn = 1000 + 1 - 0 = 1001$.

The preceding example shows that though the label allocation method used by BGP VPWS consumes a large number of labels, the number of VCs established on a PE is limited. Therefore, the label consumption can be ignored.

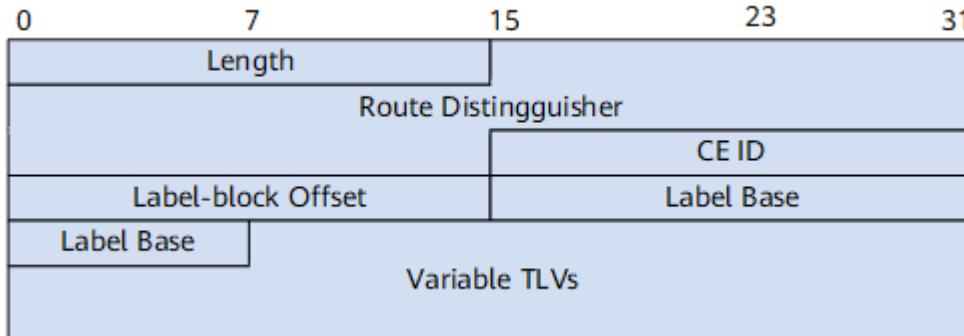
In practical network deployment, the network administrator is used to identifying the location of a CE by CE ID. If large CE IDs are used, a lot of label spaces will be consumed. In extreme cases, the valid label range will be insufficient. To solve this problem, use CE names to describe CE locations and create a table to record actual CE IDs.

Signaling for Transmitting VC Labels

BGP VPWS extends the NLRI of MP-BGP to transmit VC information. Like L3VPN, BGP VPWS also uses the route distinguisher (RD) and VPN target. Because VPWS is a P2P technology, a CE has to use multiple interfaces or sub-interfaces to establish VCs with multiple other CEs. Even if in the same VPN, two CEs must connect over a VC.

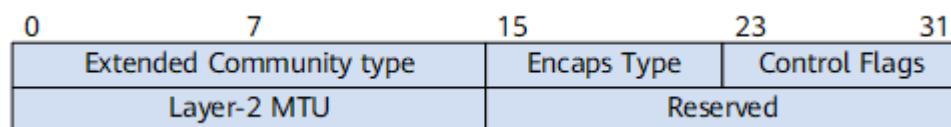
[Figure 7](#) describes label block information in the NLRI. The circuit status vector (CSV) in the TLV with variable length is used to describe the LR and tunnel status of the label block.

Figure 7 MP-BGP extension



An extended community attribute is defined to carry more L2VPN information, as shown in [Figure 8](#).

Figure 8 Extended community attribute for Layer 2 information



[Table 2](#) describes each field shown in [Figure 8](#).

Table 2 Description of each field in the extended community attribute

Field	Description	Number of Bits	Description
-------	-------------	----------------	-------------

Field	Description	Number of Bits	Description
Extended Community Type	Extended information type	16	Extended community type
Encaps Type	Encapsulation type	8	Layer 2 encapsulation type
Control Flags	Control word	8	Control word
Layer-2 MTU	Layer 2 MTU	16	-
Reserved	Reserved	16	Reserved

Usage Scenario

BGP VPWS applies to networks with dense Layer 2 connections, such as a network with the mesh topology.

Benefits

BGP VPWS does not directly perform operations on connections between CEs. It partitions the entire ISP network into different VPNs and numbers each CE in each VPN. Similar to BGP/MPLS VPN, BGP VPWS uses the VPN target to control the sending and receiving of VPN routes, creating greater networking flexibility.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.6 Heterogeneous VPWS

On a heterogeneous VPWS network, a VC is established between CEs of different link types, and AC interfaces are enabled with IP interworking to transparently transmit Layer 3 data, that is, IP packets, over an MPLS network.

If the link types of CEs on both ends of an L2VPN link are different, the heterogeneous VPWS feature is required.

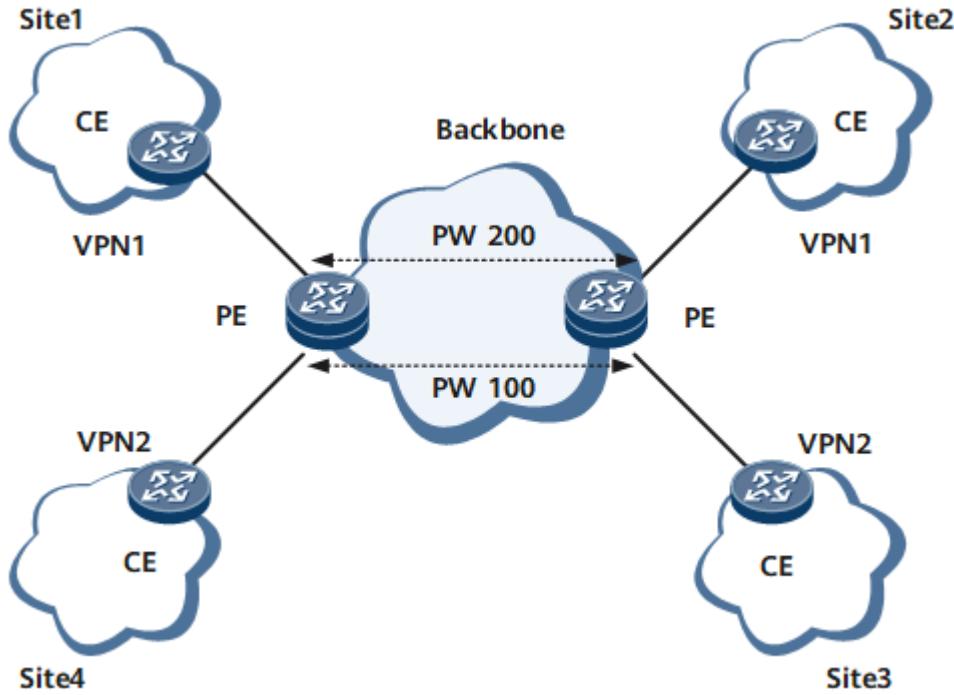
Introduction

Heterogeneous VPWS is used when the link types of CEs on both ends of an L2VPN link are different. After a PE on a heterogeneous VPWS network receives a frame from a CE, the PE removes the frame header and transparently transmits the IP packet over an MPLS network to the peer PE. The peer PE re-encapsulates the IP packet according to its link-layer protocol and transmits the packet to the connected CE. The link-layer control packet sent by a CE is processed by the corresponding PE and is not transmitted over the MPLS network. All non-IP packets, such as MPLS and Internet Packet Exchange (IPX) packets, are discarded.

Heterogeneous VPWS Network Topology

If heterogeneous sites access an L2VPN backbone network, heterogeneous VPWS must be configured. On the network shown in [Figure 1](#), Site 3 and Site 4 access the L2VPN backbone network by the same means, but Site 1 and Site 2 access the L2VPN backbone network by different means.

Figure 1 Heterogeneous VPWS network topology



Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.7 ATM Cell Relay

ATM cell relay is a technique that transmits ATM cells over PWE3 VCs.

Background

ATM is a traditional multi-service bearer technology used on backbone networks. ATM networks can carry services such as IP, FR, voice, teleconference, and ISDN/DSL and provide well-designed quality of service (QoS) mechanisms for these services. ATM networks have been used to carry important services.

IP networks have developed rapidly in recent years, owing to their advantages in upgradability, expansibility, and interoperability. The traditional ATM networks, however, are less compatible with newly deployed networks due to limitations on transmission modes and service types. An urgent demand is to upgrade traditional ATM networks and integrate them with existing PSNs, so that existing network resources can be fully utilized to meet expanded service demand.

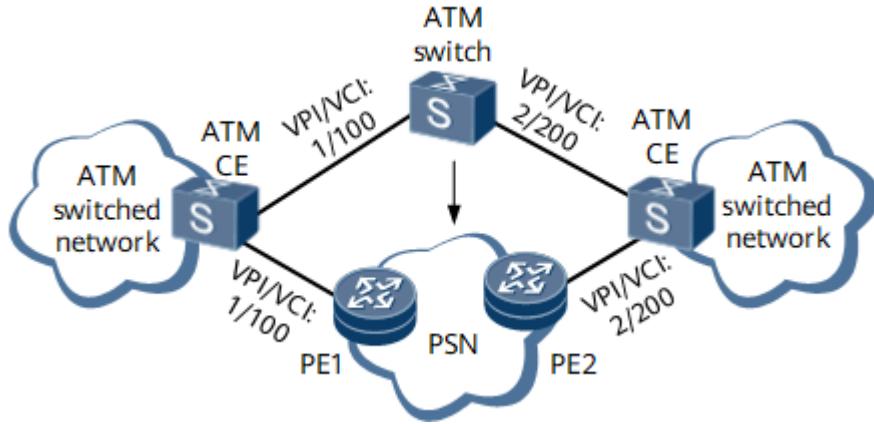
By interconnecting ATM networks over a PSN, ATM cell relay emulates traditional ATM services when they are transmitted over the PSN. This allows end users to be unaware of network differences and protects carriers' investment during network convergence and construction.

Related Concepts

- ATM cell: A cell is the basic ATM transmission unit. An ATM cell consists of 53 bytes, comprising a 5-byte header and a 48-byte payload. Each ATM cell is transmitted independently with a short transmission delay.

- VC: ATM is a VC-based and connection-oriented switching technology. Each VC is identified by a virtual path identifier (VPI) and a virtual channel identifier (VCI). A VPI/VCI pair is valid for only a link between ATM devices.
- Permanent virtual circuit (PVC): A PVC is a type of ATM connection configured by a network administrator. The establishment of a PVC does not require signaling.
- Switched virtual circuit (SVC): An SVC is a type of ATM connection dynamically established using signaling.
- Virtual circuit connection (VCC): A VCC is a type of ATM connection established based on VCI switching.
- Virtual path connection (VPC): A VPC is a type of ATM connection established based on VPI switching.
- ATM adaptation layer (AAL): The ATM adaptation layer is similar to the data link layer of the OSI reference model and is integrated with the ATM layer. The AAL is responsible for separating the upper layer from the ATM layer. The AAL prepares for conversion between service data and ATM cells by fragmenting service data into 48-byte payloads for ATM cells.
- VPI/VCI mapping: As shown in [Figure 1](#), a PW is used to emulate an ATM Switch. To retain configurations on ATM Switch, VPI/VCI pairs 1/100 and 2/200 must be mapped to each other on PE1 and PE2. In this manner, VPI/VCI pairs for CEs of a VC are mapped. If the PW emulates only one VPC or VCC, the PW functions as an ATM switch and mapping between VPI/VCI pairs does not need to be configured on PE1 and PE2. If the PW emulates two or more VPCs or VCCs, mapping between VPI/VCI pairs need to be configured on PE1 and PE2.

Figure 1 Networking for ATM cell relay over a P2P tunnel on a PSN

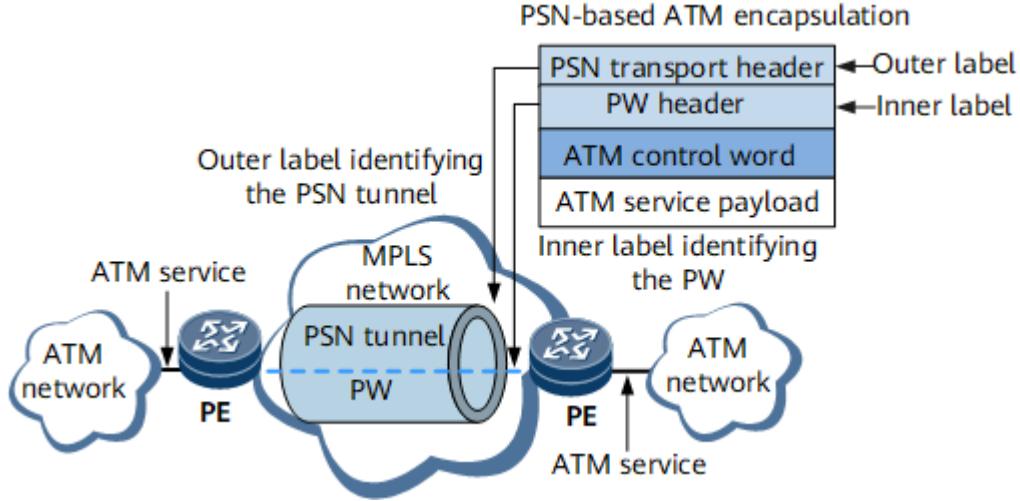


Implementation

ATM cell relay interconnects traditional ATM networks and carries ATM cells over a point-to-point PW on a PSN.

[Figure 2](#) shows the label encapsulation mode for ATM cell relay over a PSN. The outer label is the MPLS tunnel label and the inner label is the VC label used to identify the PW.

Figure 2 Networking for ATM cell relay over a PSN



A VPI/VCI pair is used to identify an ATM VC. Based on PW emulation types and comparison of ATM cell relay and AAL5 SDU relay, the following ATM cell relay modes are defined:

- One-to-one (1-to-1): One PW emulates one VCC or VPC to carry ATM cells.
- N-to-one (N-to-1): One PW emulates two or more VCCs or VPCs to carry ATM cells.

[Table 1](#) lists the characteristics of different ATM cell relay modes.

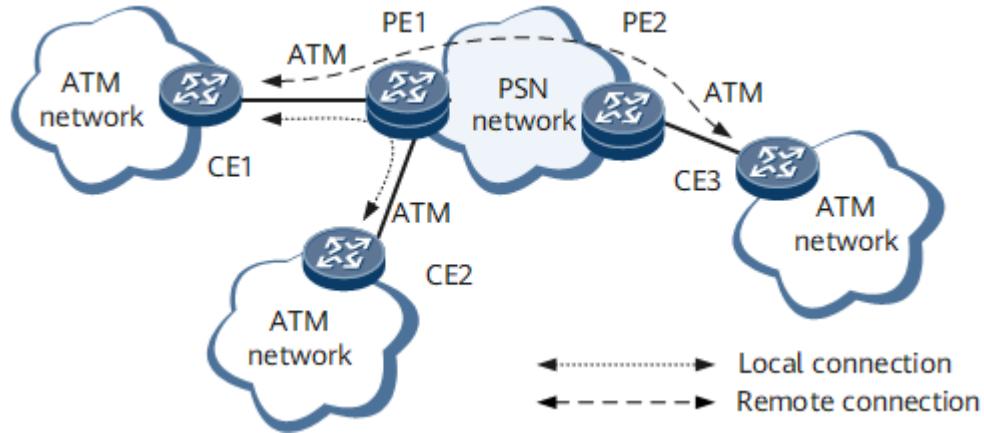
Table 1 Characteristics of different ATM cell relay modes

Encapsulation Mode	AAL Type	Connection Type	Encapsulation Method
N-to-1 VCC cell relay	All AAL types	VC	The VPI/VCI pair is encapsulated into the ATM cell. The control word is optional for the PW. This PW encapsulation mode supports VPI/VCI switching.
1-to-1 VCC cell relay	All AAL types	VC	The VPI/VCI pair is not encapsulated into the ATM cell. The control word is required for the PW. This PW encapsulation mode supports VPI/VCI switching.
N-to-1 VPC cell relay	All AAL types	VP	The VPI/VCI pair is encapsulated into the ATM cell. The control word is optional for the PW.
1-to-1 VPC cell relay	All AAL types	VP	The VCI but not the VPI is encapsulated into the ATM cell. The control word is required for the PW.
AAL5-SDU	AAL5	VC	AAL5 SDUs are directly mapped to PWs for transmission, fully utilizing PSN bandwidth. The control word is optional for the PW.

As shown in [Figure 3](#), ATM cell relay is classified into the following modes based on PWE3 networking modes:

- Remote ATM cell relay: CEs are connected to two different PEs on the PSN, and ATM cells need to be transparently transmitted over the PSN.
- Local ATM cell relay: CEs are connected to the same PE on the PSN. ATM cells are directly forwarded by the PE, instead of being transparently transmitted over the PSN.

Figure 3 Networking diagram for local and remote ATM cell relay



Usage Scenario

The following describes usage scenarios for different ATM cell relay modes.

ATM VCC Cell Relay

[Figure 4](#) shows an example of a VCC. A VCC is the basic transmit unit of an ATM network. VCCs can carry various ATM services.

Figure 4 ATM VCC cell relay



ATM VPC Cell Relay

[Figure 5](#) shows an example of a VPC. A VPC is a set of VCCs with the same destination. VPCs can carry various ATM services. ATM VPC cell relay applies to the scenario in which packets from multiple users are bound to the same destination. ATM VPC cell relay features rapid transmission, easy management, and convenient configuration.

Figure 5 ATM VPC cell relay



Benefits

By interconnecting traditional ATM network resources over a PSN, ATM cell relay emulates traditional ATM services when they are being transmitted over the PSN. This allows end users to be unaware of network differences and protects carriers' investment during network convergence and construction.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.8 VCCV

Virtual circuit connectivity verification (VCCV) is an end-to-end fault detection and diagnostic mechanism for a PW. VCCV is, in its simplest description, a control channel between a PW's ingress and egress points over which connectivity verification messages can be sent.

VCCV can detect and diagnose the connectivity of the forwarding paths on a PW.

A Label Mapping packet uses the BFD CV Type field to indicate whether a PW supports BFD fault detection. The BFD CV Type can be 0x04, 0x08, or 0x10. By default, the BFD CV Type is 0x08. If a remote peer does not support the 0x08 BFD CV Type, change the BFD CV Type in the Label Mapping packet.

- 0x04: supports BFD IP/UDP encapsulation, used for PW fault detection only.
- 0x08: supports BFD IP/UDP encapsulation, used in PW fault detection and AC/PW fault status signaling.
- 0x10: supports BFD PW-ACH encapsulation, used in PW fault detection only.

VCCV ping, an extension to LSP ping, is a tool used to manually test the connectivity of a VC. VCCV defines a series of messages exchanged between PEs for PW connectivity verification. To ensure that VCCV packets and PW data packets are transmitted along the same path, VCCV packets must be encapsulated in the same way and transmitted over the same tunnel as PW data packets.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.9 PW Redundancy

PW Redundancy Signaling

In conventional Pseudowire Emulation Edge-to-Edge (PWE3), one-to-one mapping is implemented between ACs and PWs. To ensure the same forwarding capability, the PW redundancy mechanism to be used must allow only a single PW in a PW group to forward traffic.

Relevant standards specify the PW Status TLV to transmit the PW forwarding status. The PW Status TLV is transported to the remote PW end using a Label Mapping or LDP Notification message. The PW Status TLV is 32 bits long. Each bit can be set individually to indicate a PW forwarding state. PW redundancy introduces a new PW status code of 0x00000020 indicating "PW forwarding standby."



PW redundancy is supported only in PWE3 VPWS.

Primary/Secondary and Active/Inactive PWs

PW redundancy involves the following terms:

- Primary/Secondary: indicates the forwarding priority of a PW and can be specified.
The primary PW is preferentially used to forward traffic, and the secondary PW is used to protect the primary PW. The primary PW forwards traffic when the primary and secondary PWs work in the active state. Currently, only a single secondary PW can be configured for a primary PW. Note that a bypass PW can be considered as a secondary PW.
- Active/Inactive: indicates the PW forwarding status, as known as the PW running status. The forwarding status cannot be configured.
Traffic can only be forwarded along active PWs. The signaling status and configured forwarding priority (primary/secondary) determine the PW forwarding status (active/inactive). Only the PW with the optimal signalling status and the higher priority is selected as an active PW to forward traffic. Other PWs are in the inactive state. Inactive PWs cannot forward traffic. If VLL PWs are used, inactive VLL PWs can be configured to receive traffic.

PW Redundancy Modes

PW redundancy modes are specified on PEs where primary and secondary PWs are configured. If no working mode is specified, PWE3 FRR is used.

NOTE

In PWE3 FRR, a PE locally determines the primary/secondary PW status and does not notify a remote PE of such status. Therefore, the remote PE is unaware of the primary and secondary PWs. PWE3 FRR is implemented on Huawei devices only and is not recommended.

Master/Slave mode:

In this mode, the local PE determines the primary/secondary PW status and uses a signaling protocol to notify the remote PE of such status. The remote PE then can obtain the primary/secondary PW status. The PW-side and AC-side primary/secondary status do not affect each other, and therefore, PW- and AC-side faults are isolated.

As shown in [Figure 1](#), in a 3PE PW redundancy scenario, the master/slave mode is used, which means that the local end determines the primary/secondary PW status. If the master/slave mode is deployed on PE1, PE1 selects a PW based on the configured priority and the PW up/down status. If the primary PW is up, PE1 always selects the configured primary PW to forward traffic. In master/slave PW redundancy mode, PE1 notifies the remote PEs (PE2 and PE3) of the selection result. PE2 and PE3 determine whether to forward traffic over this PW based on the PW status notified by PE1.

Independent mode:

In this mode, the primary/secondary status of the local PWs is determined by the negotiation result on the remote end (PE2 and PE3), and the remote end notifies the local end of the primary/secondary status. If a fault occurs on the AC side, it affects both the AC and PW sides and triggers protection switching on both sides, indicating a fault isolation failure.

As shown in [Figure 1](#), in a 3PE PW redundancy scenario, the independent mode is used, which means that the remote end determines the primary/secondary PW status. If the independent mode is deployed on PE1, PE1 does not concern the locally configured primary/secondary priority, but uses the status notified by the remote PEs (PE2 and PE3). If PE2 notifies its active PW state to PE1, PE1 selects PW1 connected to PE2 to forward traffic. If PE3 notifies its active PW state to PE1, PE1 selects PW2

connected to PE3. If both remote PEs notify PE1 of their active PW states, PE2 selects the PW whose status is notified later.

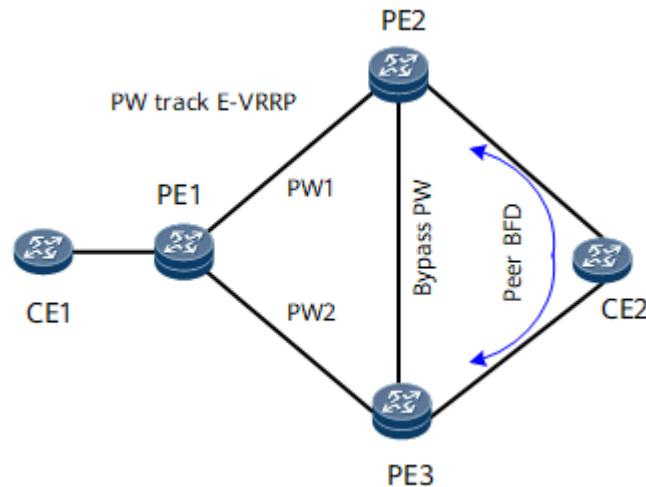
NOTE

In PWE3 scenarios, the independent mode is usually used to improve switching performance.

PW Redundancy Usage Scenarios

3PE PW redundancy

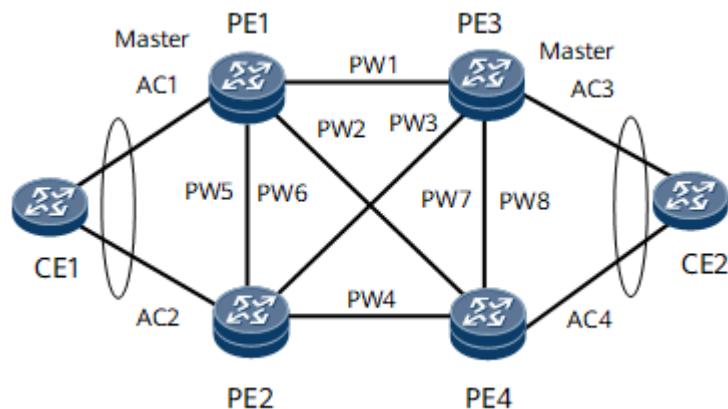
Figure 1 3PE PW redundancy



- MC-LMSP 1:1 protection or E-Trunk 1:1 protection is used.
- The master/slave mode is supported, and in this mode, bypass PWs can be configured.
- The independent mode is supported, and in this mode, bypass PWs can be configured.

4PE PW redundancy

Figure 2 4PE PW redundancy



- VRRP or E-Trunk 1:1 protection is supported.
- The master/slave mode is not supported.
- The independent mode is supported.
- Dual bypass PWs can be configured only for Ethernet services.

- Dual bypass PWs do not support heterogeneous interworking.
- Bypass PWs do not support VPN QoS.
- The bypass PW must be fault-free.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.10 PW APS

Definition

Two PW reliability mechanisms are currently available: PW redundancy and PW automatic protection switching (APS).

APS instructs the source and destination ends to implement protection switching in the same manner to achieve traffic switching, delayed switching, and wait-to-restore. APS protocol packets are always transmitted along the backup channel. Both the transmit and receive ends know that they receive APS protocol packets through each other's backup channel. This implementation helps determine whether both ends are configured with the same primary and backup channels.

PW APS is an application of APS on PWs. PW APS uses PW OAM to monitor the PW status. If a PE detects that the primary PW fails, PW APS is triggered, and traffic is switched to the secondary PW, implementing service protection.

Purpose

PWs are generally used to transmit 2G services between base transceiver stations (BTSs) and base station controllers (BSCs), 3G services between NodeBs and RNCs, and long term evolution (LTE) services between eNodeBs and mobility management entities (MMEs)/serving gateways (S-GWs). PWs meet requirements for bandwidth, expansion, and flexible configuration of these services. The bearer network solution includes:

- Static solution: uses static routes, static LSPs, and static PWs.
- Dynamic solution: uses dynamic routes, dynamic LSPs/TE tunnels, and dynamic PWs.

As static PWs do not use signaling, the primary and secondary PW status negotiation, PW switchover, and PW switchback cannot be implemented using signaling. PW redundancy currently supported addresses only PWE3 reliability, but not reliability for PWs in SVC or LDP mode. SVC PWs are static PWs. PW APS can provide reliability for PWs in SVC, LDP, or PWE3 mode.

- PW APS uses PW OAM (MPLS OAM or MPLS-TP OAM) to rapidly monitor PW status and notifies APS of the status.

NOTE

The primary and secondary PWs must use OAM of the same type. To configure OAM for a bypass PW, it must also use the same type of OAM as the primary and secondary PWs.

- The primary/secondary PW protection group is associated with APS instances. APS instructs the source and destination ends to implement bidirectional PW protection switching in the

same manner, as defined in G.8131.

PW APS applies to SVC, LDP, or PWE3 PWs.

NOTE

Using PW APS or PW redundancy solely on the entire network is recommended. PW APS and PW redundancy are both reliability mechanisms but are implemented differently. Implementing both mechanisms on a network increases the difficulties for network maintenance.

Basic Concepts

Protection Type

PW APS can work in 1:1 or 1+1 mode, in which primary and secondary PWs back up each other. In PW APS 1:1 mode, traffic is transmitted and received over a single link. In PW APS 1+1 mode, traffic is transmitted and received over double links but accepted over only one link.

Switching Type

- Dual-ended switching

Dual-ended switching refers to bidirectional switching, that is, if a fault is detected on the forward or reverse working PW, APS allows bidirectional services on the working PW to be switched to the protection PW.

- Single-ended switching

Single-ended switching refers to unidirectional switching, that is, if a fault is detected on the forward or reverse working PW, the service only in the fault direction of the working PW is switched to the protection PW. Single-end switching is also called pseudo wire fast protection switching (PW FPS).

Revertive Mode

The PW APS mode can be either revertive or non-revertive. In non-revertive mode, traffic will not be switched back from the protection PW to the working PW even if the working PW recovers. In revertive mode, traffic will return to the working PW after the wait-to-restore (WTR) timer configured for the working PW expires.

WTR Time

The WTR time is counted from the time when the primary PW recovers to the time when traffic is switched back from the secondary PW. Setting a WTR time prevents frequent traffic switching.

Delayed Switching Time

The delayed switching time is the time after which a protection switching is triggered if a signal fail (SF) is still detected on a PW. Setting a delayed switching timer prevents switching from immediately occurring after an SF is detected.

Dual-Homing Protection

Dual-homing protection is implemented by connecting two PEs to a CE through respective ACs. This protects PE services on the bearer network.

PW APS Bundling

The device usually needs to undergo a great deal of PW APS protection switching. If PW APS enables a state machine for each protection switching, the device will not be able to implement all protection switching due to limited resources and capabilities. Configuring an APS state machine to

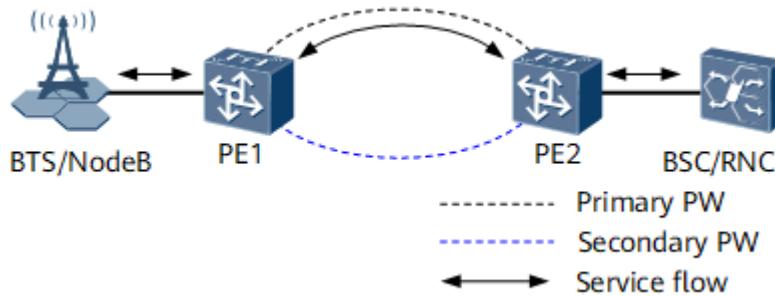
process a great deal of PW APS protection switching decreases resource consumption. This APS state machine is shared by multiple PWs, which is called PW APS bundling.

Switching Mechanism

PW APS uses PW OAM to monitor the primary and secondary PW status. PW OAM sends detection packets from the ingress to the egress periodically. If the egress fails to receive any detection packets in a certain period, it considers that an SF occurs and notifies the remote APS module of the fault. This implements service switching and protection.

As shown in [Figure 1](#), PW APS is configured on PE1 and PE2. Normally, upstream traffic from a BTS/NodeB is transmitted along the path PE1 -> primary PW -> PE2 on the PSN. PE2 forwards the traffic to a BSC/RNC. Downstream traffic from a BSC/RNC is transmitted along the path PE2 -> primary PW -> PE1 on the PSN. PE1 forwards the traffic to a BTS/NodeB.

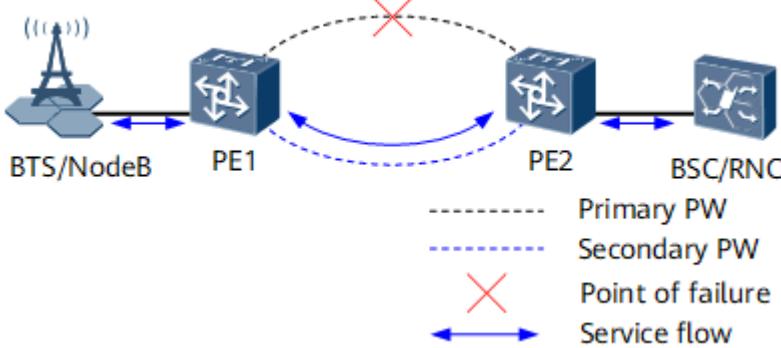
Figure 1 PW APS deployment



As shown in [Figure 2](#), if the primary PW fails, PW OAM on PE1 and PE2 detects the failure and triggers APS. Both upstream and downstream traffic are switched to the secondary PW.

The delayed revertive operation mode is used for PW APS by default. After the primary PW recovers, PW OAM on PE1 and PE2 detects the recovery but waits a delayed switching time before triggering an APS revertive operation. Both upstream and downstream traffic are then switched back to the primary PW.

Figure 2 PW APS implementation



Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.11 Comparison of VPWS Implementation Modes

[Table 1](#) describes the differences between VPWS implementation modes.

Table 1 Comparison of VPWS implementation modes

Implementation Mode	Signaling Protocol	Tunnel	Usage Scenario	Scalability	Support for Local Connections
Circuit cross connect (CCC)	None	Local CCC (public network tunnels not needed)	N/A	Comparatively poor	Yes
LDP	LDP	Needs a shared LSP tunnel.	Sparse mode	Poor	No
PWE3	LDP	Needs a shared LSP tunnel.	Sparse mode	Poor	No
Static virtual circuit (SVC)	None	Needs a shared LSP tunnel.	N/A	Poor	No

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.12 Comparison of LDP VPWS and BGP/MPLS IP VPN

[Table 1](#) describes the differences between LDP VPWS and BGP/MPLS IP VPN.

Table 1 Comparison of LDP VPWS and BGP/MPLS IP VPN

Item	LDP VPWS	BGP/MPLS IP VPN
Cost of PEs	The memory cost is low; the consumption of interface resources is high; the signaling cost is high.	The memory cost is high; the consumption of interface resources is low; the signaling cost is low.
Flooding mode of the VPN topology	Manual configuration.	BGP automatic discovery.
Flooding mode of VPN routes	The VPN routes are flooded directly between CEs and converge rapidly.	The VPN routes are flooded by PEs and converge slowly.
Access mode of CEs	CEs with different link encapsulation types can interwork over a heterogeneous LDP VPWS network.	Different sites in the same VPN can have different access modes.
VPN nesting	Not supported.	Supported.
Support for multicast	The protocol cost is low; the forwarding cost is high.	The protocol cost is high; the forwarding cost is low.
Protocol independence	Over any Layer 3 protocol.	Over only IP.

Item	LDP VPWS	BGP/MPLS IP VPN
Variety of tunnels	LSPs tunnels are supported.	LSPs and IPsec tunnels are supported.
Inheritance from the traditional VPN	Inherits and improves the traditional L2VPN.	Inherits and improves the traditional L2VPN.
Maturity	Immature.	Mature.
Easy-of-use	Complex.	Simple.
Manageability	Outsourced topology and centralized management.	Outsourced route and role-based management.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.13 Inter-AS VPWS

Definition

Inter-AS VPWS is an L2VPN technology that enables users to communicate across multiple autonomous systems (ASs).

Unlike inter-AS VLL, which provides inter-AS communication by setting a static LSP between autonomous system boundary routers (ASBRs), inter-AS VPWS is more like inter-AS L3VPN in terms of implementation.

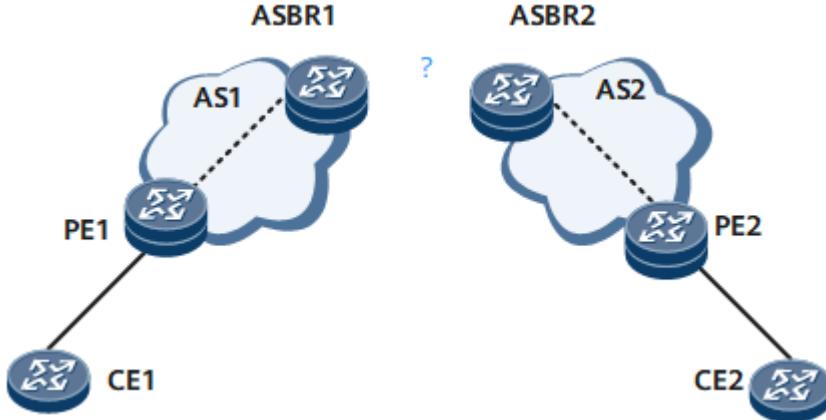
In inter-AS VPWS Option A, the link type between ASBRs must be the same as the VC type. The disadvantage of inter-AS VPWS Option A is that each ASBR must reserve a sub-interface for each inter-AS VC. Inter-AS VPWS Option A applies to scenarios in which the number of inter-AS VCs is small. Compared with inter-AS L3VPN Option A, inter-AS VPWS Option A consumes more resources and requires more configuration workload.

In inter-AS VPWS Option C, the devices on a service provider network only need to set up an outer tunnel on PEs in different ASs; the ASBRs do not need to maintain information about inter-AS VPWS or reserve sub-interfaces for inter-AS VCs; L2VPN information is exchanged only between PEs. Compared with inter-AS VPWS Option A, inter-AS VPWS Option C consumes fewer resources and requires less configuration workload.

Purpose

With the popularity of MPLS VPN, the requirements for communication between the MANs of different carriers or between different backbone networks become common. As a result, inter-AS VPWS is introduced to solve the inter-AS communication problem.

Figure 1 Origin of inter-AS VPN

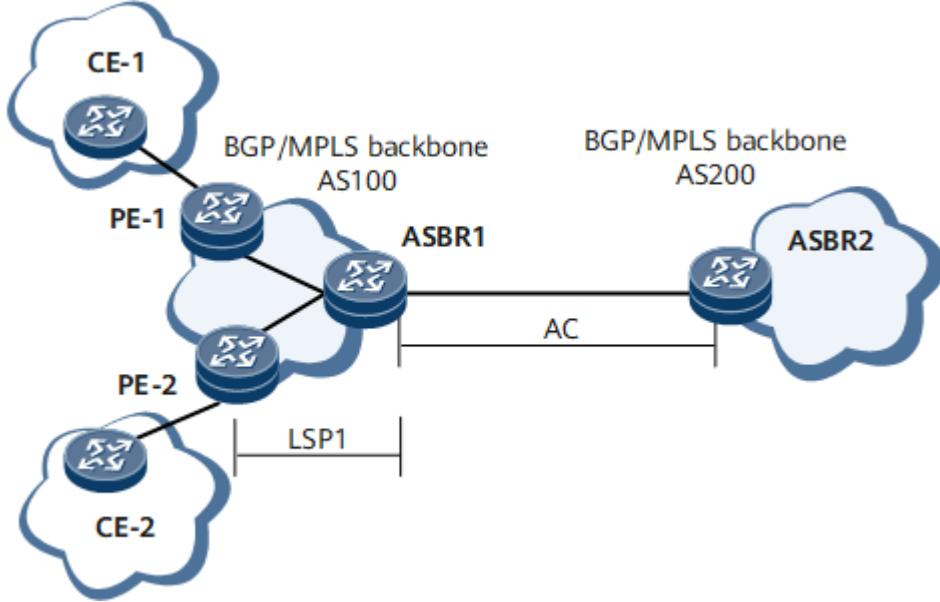


On the L2VPN shown in [Figure 1](#), some users belong to AS1, and some to AS2. If MPLS forwarding is not implemented, L2VPN users in different ASs cannot communicate. [Figure 1](#) shows a scenario in which L2VPN users of only two ASs need to communicate. In reality, L2VPN users of more ASs may need to communicate.

Inter-AS VPWS Option A

In inter-AS VPWS Option A, ASBRs in two ASs are directly connected. The two ASBRs function as PEs in their respective ASs, but regard each other as a CE.

Figure 2 Networking diagram of the inter-AS L2VPN Option A



On the network shown in [Figure 2](#), ASBR2 is a CE for ASBR1. Similarly, ASBR1 is a CE for ASBR2. The characteristics of inter-AS VPWS Option A are as follows:

- Easy to implement

MPLS forwarding is not required between the PEs functioning as ASBRs, because common IP forwarding can be adopted, and no special inter-AS configuration is required.

- Poor scalability

- The PEs functioning as ASBRs need to manage information about all L2VPNs, which leads to a great deal of L2VPN information on the PEs and therefore burdens the PEs.

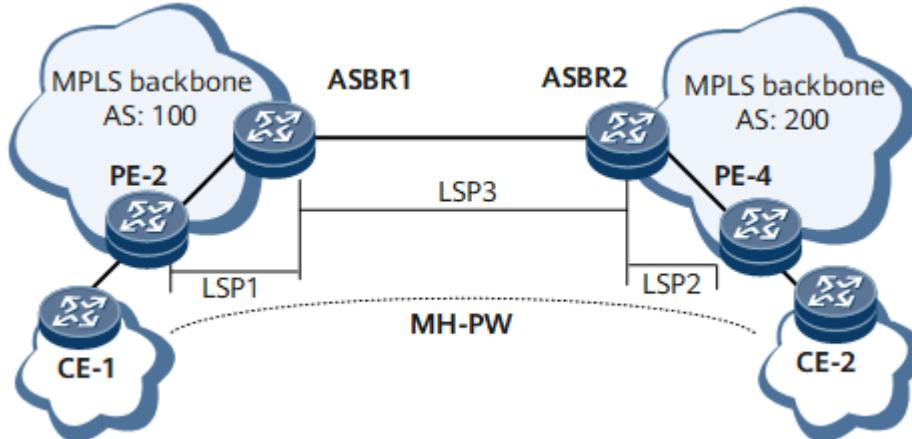
- An AC interface must be reserved for each PW, because the PE functions as an ASBR in the local AS.
- If users need to communicate across multiple ASs, intermediate ASs must support L2VPN. This causes heavy configuration and affects the intermediate ASs.

Inter-AS VPWS Option A applies to scenarios in which a VPWS network crosses only a few ASs.

Inter-AS Multi-segment PWE3

In inter-AS multi-segment PWE3, multi-segment PWs are set up. As shown in [Figure 3](#), PW switching needs to be performed on two ASBRs, and an LDP session and a tunnel must be set up between the two ASBRs.

Figure 3 Networking diagram of inter-AS multi-segment PWE3



Compared with inter-AS VPWS Option A, inter-AS multi-segment PWE3 provides better scalability and has no limit to the number of links between ASBRs, because ASBRs exchange PW information over an LDP session rather than a private link.

Inter-AS multi-segment PWE3, however, has the following limitations:

- The PEs functioning as ASBRs need to manage information about all L2VPNs.
- If users need to communicate across multiple ASs, intermediate ASs must support L2VPN.
- LDP sessions and LSPs need to be set up between ASBRs.

Inter-AS VPWS Option C

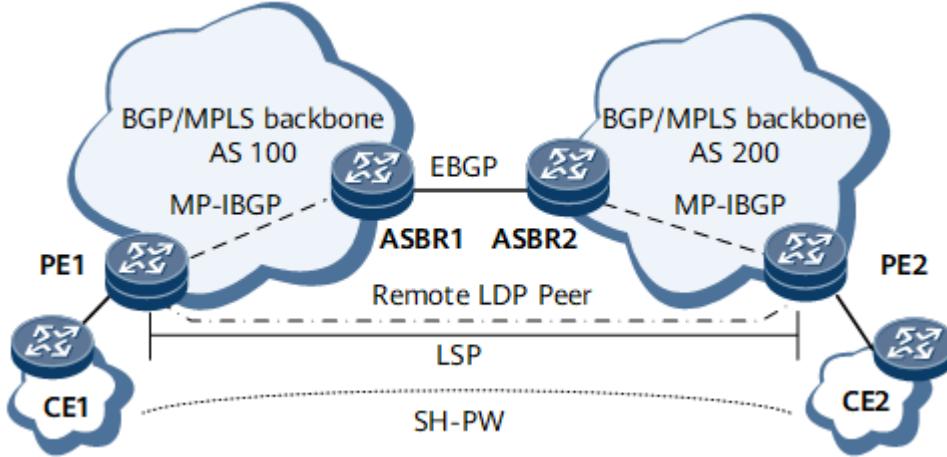
The preceding two schemes require ASBRs to participate in the distribution and maintenance of PW labels. When multiple inter-AS PWs exist in each AS, ASBRs may be a bottleneck in network expansion.

Inter-AS VPWS Option C avoids this problem by freeing ASBRs from setting up and maintaining PWs. PW labels are directly switched between PEs, as shown in [Figure 4](#).

In inter-AS VPWS Option C:

- ASBRs advertise labeled IPv4 routes to PEs in their respective ASs through Multiprotocol Interior Border Gateway Protocol (MP-IBGP), and advertise labeled IPv4 routes received by PEs in their respective ASs to the ASBR peers in other ASs. ASBRs in the intermediate AS also advertise labeled IPv4 routes. As a result, an BGP LSP is set up between the ingress PE and the egress PE.
- PEs in different ASs set up remote MPLS LDP sessions to exchange PW information.

Figure 4 Networking diagram of inter-AS VPWS Option C



Inter-AS VPWS Option C has the following advantages:

- Similar to the network on which L2VPN users belong to the same AS, intermediate devices do not need to store L2VPN information.
- Only PEs need to store L2VPN information. The devices in intermediate ASs only need to function as ordinary ASBRs that support IP forwarding and do not need to support L2VPN. Inter-AS VPWS Option C is preferred when users need to communicate across a large number of ASs.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.14 Flow-Label-based Load Balancing

Background

Packets of multiple data flows on the same L2VPN carry the same VC labels, which are encapsulated on a PE. When these packets reach a P device, they can be forwarded only over one path.

To load-balance different data flows, configure flow-label-based load balancing on the PE. After you have completed the configuration, the PE encapsulates a data packet and adds a flow label following the PW label. The P device load-balances different data flows based on flow labels.

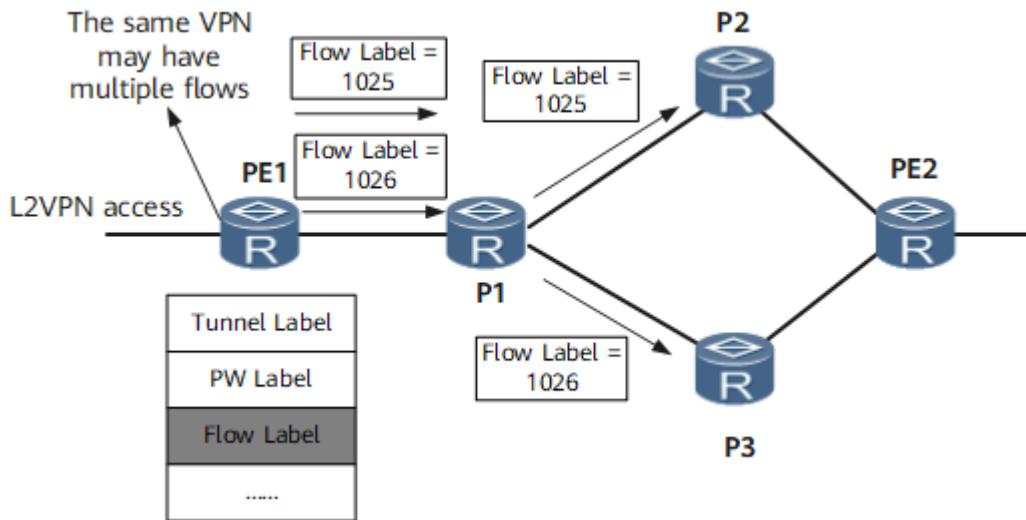
Implementation

On the L2VPN shown in [Figure 1](#) where two data flows exist:

1. PE1 calculates flow labels based on the source and destination IP addresses of the two data flows. In this scenario, the flow labels are calculated as 1025 and 1026.
2. PE1 adds the flow labels following the PW labels of the packets in the two data flows.
3. When the two data flows reach P1, P1 performs a hash calculation based on the flow labels, and the two data flows are mapped onto different paths. In this example, the next hop of the data flow with the flow label 1025 is P2, and the next hop of the data flow with the flow label 1026 is P3.

- When the two data flows reach PE2, the PW and flow labels are sequentially removed. PE2 then forwards the two data flows to their destination CEs based on their PW labels.

Figure 1 Flow-label-based load balancing



Usage Scenario

Flow-label-based load balancing applies to an L2VPN on which multiple links exist between P devices.

Benefits

Flow-label-based load balancing allows data flows on the same VPN to be load-balanced along different paths based on flow labels, improving resource usage.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.15 Mutual Protection Between an LDP VC and a CCC VC

The use of Layer 2 virtual private network (L2VPN) technologies increases reliability requirements for L2VPNs. This is especially true of L2VPNs that provide real-time services such as VoIP and Internet Protocol television (IPTV).

Configuring mutual protection between a Label Distribution Protocol (LDP) virtual channel (VC) and a circuit cross connect (CCC) VC can meet these requirements. A CCC VC and an LDP VC work in the active/standby mode to protect traffic over each other. If an active path goes Down, traffic is switched to the standby path, therefore improving CCC VC's and LDP VC's reliability.

Single-Homing 2PE Scenario

[Figure 1](#) shows the following:

- CE1 is single-homed to PE1 through AC1 and preferentially accesses CE2 through AC2.

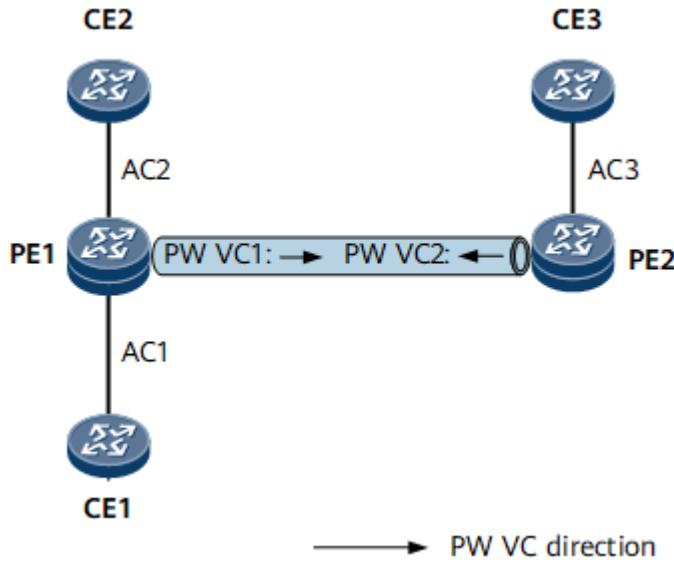
- A PW is established between PE1 and PE2, with PW VC1 protecting traffic transmitted through AC2.

If AC2 fails, CE1 accesses CE3 through the path CE1->AC1->PW VC1->AC3->CE3.

Because PE1 and PE2 are connected to different CEs, they cannot detect the active/standby status of the CEs. In this case, they have to use the active/standby mode to determine the CEs' status. The configuration roadmap is as follows:

- For AC1->AC2 traffic, the PW functions as the standby path of AC2.
- The path AC2->AC1 does not have a standby path.
- The path AC3->PW VC1 does not have a standby path.

Figure 1 Mutual protection between a CCC VC and an LDP VC (single-homing 2PE scenario)



CE2 and CE3 are the active and standby devices respectively in the preceding illustration. In that case, an LDP VC (PW VC1) protects traffic transmitted through a CCC VC (AC2). In real-time deployment, CE3 and CE2 may be the active and standby devices, respectively. In this case, a CCC VC (AC2) protects traffic transmitted through an LDP VC (PW VC1). The configuration roadmap is the same in both cases, so the roadmap for configuring a CCC VC to protect an LDP VC is not described here.

Table 1 Scenarios where typical faults trigger switchovers

Fault Point	Switchover
AC2 or CE2 failure when: <ul style="list-style-type: none"> • CE2 is in the active state. • CE3 is in the standby state. In this case, an LDP VC protects a CCC VC.	<ol style="list-style-type: none"> When detecting AC2 or CE2 failure, PE1 switches traffic to the PW. Then traffic between CE1 and CE3 is transmitted through the path CE1->AC1->PW->AC3->CE3. After AC2 or CE2 recovers, PE1 switches traffic back to AC2 based on a configured revertive switching policy.

Fault Point	Switchover
<p>PE2, AC3, CE3, or PW failure when:</p> <ul style="list-style-type: none"> • CE3 is in the active state. • CE2 is in the standby state. <p>In this case, a CCC VC protects an LDP VC.</p>	<ol style="list-style-type: none"> 1. When detecting the failure, PE1 switches traffic to AC2. Then traffic between CE1 and CE2 is transmitted through the path CE1<->AC1<->AC2<->CE2. 2. After PE2, AC3, CE3, or the PW recovers, PE1 switches traffic back to the PW based on the configured revertive switching policy.

Dual-Homing 2PE Scenario

[Figure 2](#) shows the following:

- CE1 is dual-homed to PE1 and PE2 through an Eth-Trunk, with AC1 serving as the active path.
- CE2 is dual-homed to PE1 and PE2 through another Eth-Trunk, with AC3 serving as the active path.
- PW1 VC1 is deployed between PE1 and PE2 to protect AC3. If AC3 fails, CE1 accesses CE2 through the PW1 VC1->AC4 path.
- PW2 VC1 is deployed between PE1 and PE2 to protect AC1. If AC1 fails, CE2 accesses CE1 through the PW2 VC1->AC2 path.

Based on the preceding, CE1 and CE2 select AC1 and AC3 as the active paths, respectively. They select AC2 and AC4 as the standby paths, respectively. If the active paths fail, the standby paths take over service traffic.

Because CE1 and CE2 both use Eth-Trunks to access the PEs, the active/standby Eth-Trunk status determines the active/standby status of the PWs connected to PE1 and PE2. The roadmap for configuring an LDP VC to protect a CCC VC is as follows:

- For AC1->AC3 traffic, configure PW1 VC1 on the AC1 interface to protect AC3.
- For AC2->PW2 VC2 traffic, configure PW2 VC2 on the AC2 interface to protect AC4.
- For AC3->AC1 traffic, configure PW2 VC1 on the AC3 interface to protect AC1.
- For AC4->PW1 VC2 traffic, configure PW1 VC2 on the AC4 interface to protect AC2.

Figure 2 LDP VC protecting CCC VC (dual-homing 2PE scenario)

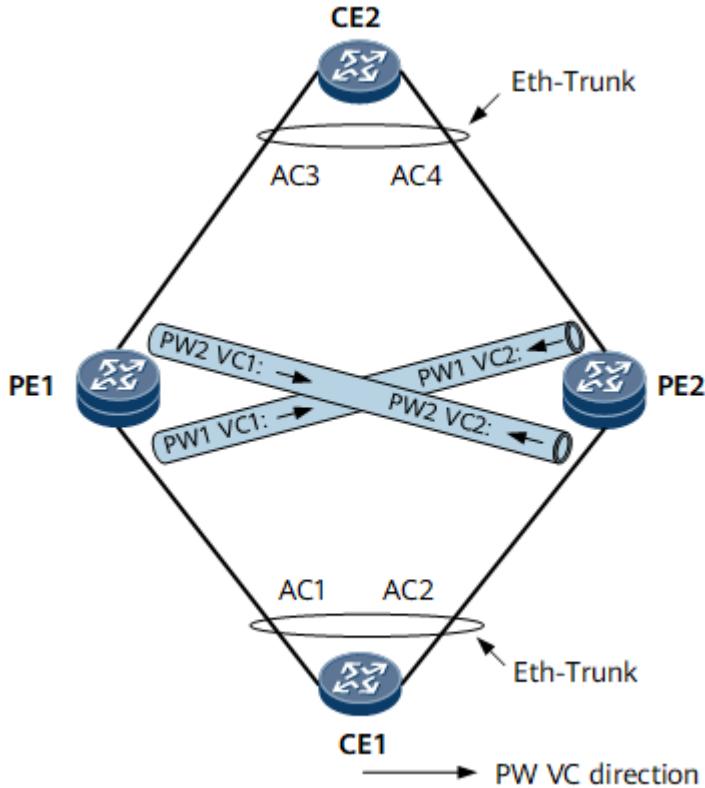


Table 2 Scenarios where typical faults trigger switchovers

Fault Point	Switchover
AC1 fails.	<ol style="list-style-type: none"> After detecting AC1 failure, the Eth-Trunk on CE1 switches traffic from AC1 to AC2. Because AC4 is blocked, PW2 VC2 functions as the active path, and the traffic between CE1 and CE2 is transmitted through the path CE1<->AC2<->PW2<->AC3<->CE2. After AC1 recovers, the Eth-Trunk on CE1 switches traffic back to AC1. Then traffic between CE1 and CE2 is transmitted through the path CE1<->AC1<->AC3<->CE2.
AC3 fails.	<ol style="list-style-type: none"> After detecting AC3 failure, the Eth-Trunk on CE2 switches traffic from AC3 to AC4. Because AC2 is blocked, PW1 VC2 functions as the active path, and the traffic between CE2 and CE1 is transmitted through the path CE2<->AC4<->PW1<->AC1<->CE1. After AC3 recovers, the Eth-Trunk on CE2 switches traffic back to AC3. Then traffic between CE2 and CE1 is transmitted through the path CE2<->AC3<->AC1<->CE1.

Fault Point	Switchover
PE1 fails.	<ol style="list-style-type: none"> After detecting PE1 failure, the Eth-Trunks on CE1 and CE2 switch traffic from AC1 to AC2 and from AC3 to AC4, respectively. Then traffic between CE1 and CE2 is transmitted through the path CE1<->AC2<->AC4<->CE2. After PE1 recovers, the Eth-Trunks on CE1 and CE2 switch traffic back to AC1 and AC3. Then traffic between CE1 and CE2 is transmitted through the path CE1<->AC1<->AC3<->CE2.

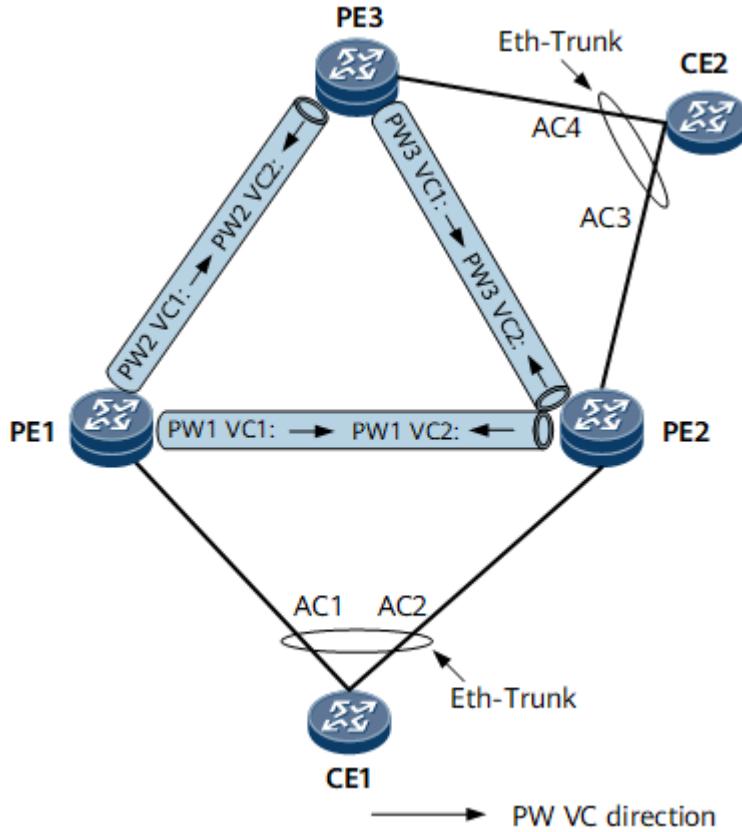
Dual-Homing 3PE Scenario

[Figure 3](#) shows the following:

- CE1 is dual-homed to PE1 and PE2 through an Eth-Trunk. AC1 serves as the active path, and AC2 serves as the standby path.
- CE2 is dual-homed to PE2 and PE3 through another Eth-Trunk. AC3 serves as the active path, and AC4 serves as the standby path.
- PW redundancy in Independent mode is configured on PE1, with PW1 VC1 being the active PW and PW2 VC1 being the standby PW.
- PW redundancy in Independent mode is configured on PE3, with PW2 VC2 being the active PW and PW3 VC1 being the standby PW.

The PW1 VC1 (active) and PW2 VC1 (standby) paths can be used to transmit traffic from AC1 to CE2. The AC3 (active) and PW3 VC2 (standby) paths can be used to transmit traffic from AC2 to CE2.

Figure 3 LDP VC protecting CCC VC (dual-homing 3PE scenario)



An Eth-Trunk priority determines the active/standby status of the ACs on PE1, PE2, and PE3. There are four combinations of AC status in this scenario. The following example uses two combinations of AC status:

- AC1 and AC3 serve as the active paths, and AC2 and AC4 serve as the standby paths:
 - For AC1->PW1 VC1 traffic, configure PW redundancy on the AC1 interface, with PW1 VC1 being the active path and PW2 VC1 being the standby path.
 - For AC2->AC3 traffic, configure PW3 VC2 on the AC2 interface to protect AC3.
 - For AC3->PW1 VC2 traffic, configure PW1 VC2 on the AC3 interface to protect AC2.
 - For AC4->PW2 VC2 traffic, configure PW redundancy on the AC4 interface, with PW2 VC2 being the active path and PW3 VC1 being the standby path.
- AC2 and AC3 serve as the active paths, and AC1 and AC4 serve as the standby paths:
 - For AC1->PW1 VC1 traffic, configure PW redundancy on the AC1 interface, with PW1 VC1 being the active path and PW2 VC1 being the standby path.
 - For AC2->AC3 traffic, configure PW3 VC2 on the AC2 interface to protect AC3.
 - For AC3->AC2 traffic, configure PW1 VC2 on the AC3 interface to protect AC2.
 - For AC4->PW3 VC1 traffic, configure PW redundancy on the AC4 interface, with PW2 VC2 being the active path and PW3 VC1 being the standby path.

Table 3 Scenarios where typical faults trigger switchovers

Fault Point	Switchover
-------------	------------

Fault Point	Switchover
<p>AC3 failure when:</p> <ul style="list-style-type: none"> • AC1 and AC3 serve as the active paths. • AC2 and AC4 serve as the standby paths. 	<ol style="list-style-type: none"> 1. After detecting AC3 failure, the Eth-Trunk on PE3 switches traffic from AC3 to AC4, and AC4 serves as the active path. 2. PE3 instructs PW2 VC2 to switch to the active state, and PE2 instructs PW1 VC2 to switch to the standby state. Then traffic between CE1 and CE2 is transmitted through the path CE1<->AC1<->PW2<->AC4<->CE2. <p>NOTE:</p> <p>It takes some time for PE3 to detect the status change of AC4 and to notify PE1 that PW2 VC2 has switched to the active state. The amount of time that elapses depends on the signaling convergence speed. At this time, PW2 VC2 remains in the standby state, and traffic may be discarded if transmitted through the path CE2->AC4>PW2 VC2->AC1->CE1. To resolve this problem, enable PE1 to accept traffic sent from a standby PW.</p> <ol style="list-style-type: none"> 3. After AC3 recovers, PE2 switches to the active state, and PE3 switches to the standby state. After signaling convergence is complete, traffic between CE1 and CE2 is transmitted through the path CE1<->AC1<->PW1<->AC3<->CE2. 4. If AC1 also fails after AC3 fails, the Eth-Trunk on PE2 switches traffic from AC1 to AC2, and AC2 serves as the active path. Because PW3 VC2 switches to the active state when AC3 fails, traffic between CE1 and CE2 is transmitted through the path CE1<->AC2<->PW3<->AC4<->CE2.
<p>AC1 failure when:</p> <ul style="list-style-type: none"> • AC1 and AC3 serve as the active paths. • AC2 and AC4 serve as the standby paths. 	<ol style="list-style-type: none"> 1. After detecting AC1 failure, the Eth-Trunk on PE2 switches traffic from AC1 to AC2, and AC2 serves as the active path. Then traffic between CE1 and CE2 is transmitted through the path CE1<->AC2<->AC3<->CE2. 2. After AC1 recovers, PW1 switches to the active state, and AC2 switches to the standby state. Then traffic between CE1 and CE2 is transmitted through the path CE1<->AC1<->PW1<->AC3<->CE2. 3. If AC3 also fails after AC1 fails, the Eth-Trunk on PE3 switches traffic from AC3 to AC4, and PW3 serves as the active PW. Then traffic between CE1 and CE2 is transmitted through the path CE1<->AC2<->PW3<->AC4<->CE2.

Fault Point	Switchover
AC3 failure when: <ul style="list-style-type: none"> • AC2 and AC3 serve as the active paths. • AC1 and AC4 serve as the standby paths. 	<ol style="list-style-type: none"> 1. After detecting AC3 failure, the Eth-Trunk on PE3 switches traffic from AC3 to AC4. At this time, AC4 serves as the active path, and PW3 serves as the active PW. Traffic between CE1 and CE2 is transmitted through the path CE1<->AC2<->PW3<->AC4<->CE2. 2. After AC3 recovers, traffic between CE1 and CE2 is transmitted through the path CE1<->AC2<->AC3<->CE2.
PE2 failure when: <ul style="list-style-type: none"> • AC2 and AC3 serve as the active paths. • AC1 and AC4 serve as the standby paths. 	<ol style="list-style-type: none"> 1. After detecting PE2 failure, the Eth-Trunks on PE1 and PE3 switch AC1 and AC4 to the active state, respectively. 2. After detecting the AC status change, PE1 and PE3 switch traffic to PW2. Traffic between CE1 and CE2 is transmitted through the path CE1<->AC1<->PW2<->AC4<->CE2. 3. After PE2 recovers, traffic between CE1 and CE2 is transmitted through the path CE1<->AC2<->AC3->CE2.

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.8.2.16 Multi-Segment PW Redundancy

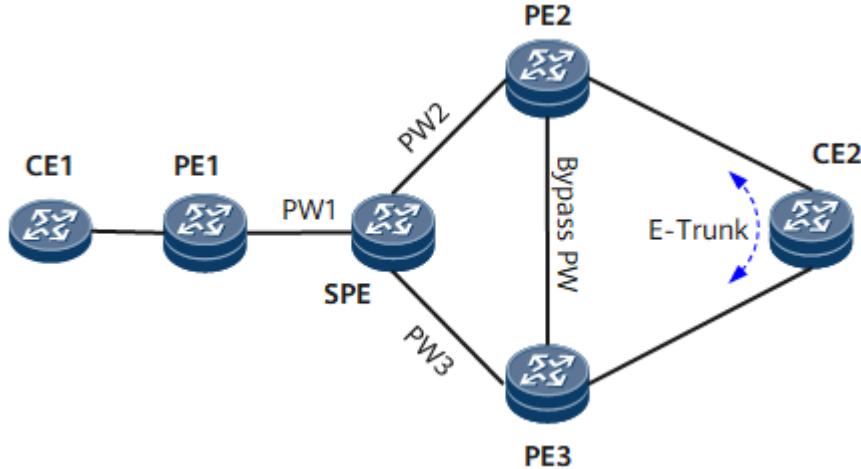
Multi-segment PW (MS-PW) redundancy applies to the following scenarios:

- A PE is single-homed to an SPE.
- A PE is dual-homed to two SPEs.

MS-PW Redundancy When a PE Is Single-Homed to an SPE

On the network shown in [Figure 1](#), when primary and secondary PWs cannot be configured on PE1, you can configure primary and secondary PWs on the SPE to implement protection. The L2VPN service between PE1 and the SPE is in PWE3 mode. A common PW is configured on PE1. The primary and secondary PWs with label switching are configured on the SPE. A bypass PW is configured between PE2 and PE3.

Figure 1 MS-PW redundancy when a PE is single-homed to an SPE



Normal Scenario

- Uplink: After the packets sent by CE1 reach PE1, they pass through PW1 to the SPE and then through PW2 to PE2. Finally, the packets are forwarded to CE2 through the AC interface on PE2.
- Downlink: After the packets sent by CE2 reach PE2, PE2 forwards the packets to the SPE through PW2. Then, the SPE forwards the packets to PE1 through PW1. Finally, PE1 forwards the packets to CE1.

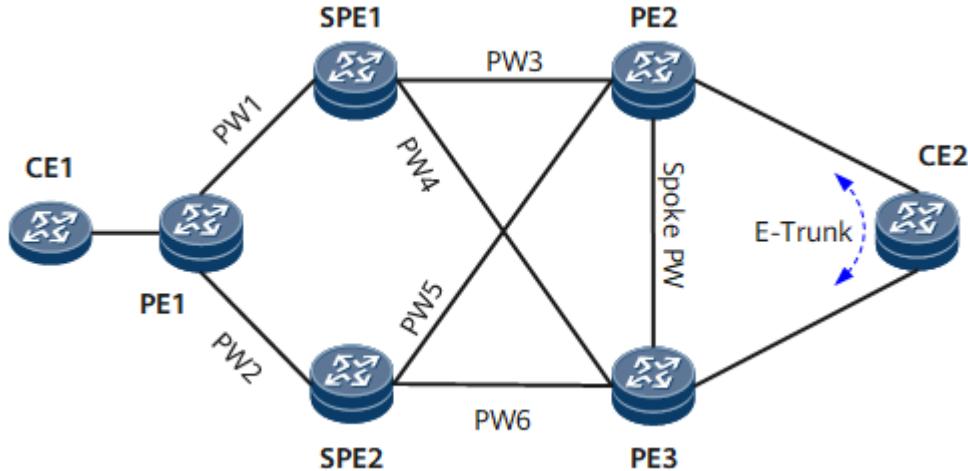
Fault Scenario

- If the link between the SPE and PE2 fails, packets are transmitted along the following paths:
 - Uplink: CE1 -> PE1 -> SPE -> PE3 -> PE2-> CE2
 - Downlink: CE2 -> PE2 -> PE3 -> SPE -> PE1 -> CE1
- If PE2 fails, packets are transmitted along the following paths:
 - Uplink: CE1 -> PE1 -> SPE -> PE3 -> CE2
 - Downlink: CE2 -> PE3 -> SPE -> PE1 -> CE1
- If the link between PE2 and CE2 fails, packets are transmitted along the following paths:
 - Uplink: CE1 -> PE1 -> SPE -> PE2 -> PE3 -> CE2
 - Downlink: CE2 -> PE3 -> PE2 -> SPE -> PE1 -> CE1

MS-PW Redundancy When a PE Is Dual-Homed to SPEs

On the network shown in [Figure 2](#), PE1 is dual-homed to SPE1 and SPE2, which are in turn dual-homed to PE2 and PE3. A Spoken PW is deployed between PE2 and PE3. Dynamic SS-PWs are configured on PE1, and PW redundancy is in master/slave mode. Dual receive mode is not configured for the PWs. Dynamic MS-PWs are configured on SPE1 and SPE2. SPE1 switches PW1 to PW3 and PW4, whereas SPE2 switches PW2 to PW5 and PW6. The PW redundancy protection group is in master/slave mode. Dual receive mode is not configured for the PWs.

Figure 2 MS-PW redundancy when a PE is dual-homed to SPEs



Normal Scenario

- Uplink: After the packets sent by CE1 reach PE1, PE1 forwards the packets to SPE1 through PW1. Then, SPE1 forwards the packets to PE2 through PW3. Finally, the packets are forwarded to CE2 through the AC interface on PE2.
- Downlink: After the packets sent by CE2 reach PE2, PE2 forwards the packets to SPE1 through PW3. Then, SPE1 forwards the packets to PE1 through PW1. Finally, PE1 forwards the packets to CE1.

Fault Scenario

- If the link between PE1 and SPE1 fails, packets are transmitted along the following paths:
 - Uplink: CE1 -> PE1 -> SPE2 -> PE2 -> CE2
 - Downlink: CE2 -> PE2 -> SPE2 -> PE1 -> CE1
- If SPE1 fails, packets are transmitted along the following paths:
 - Uplink: CE1 -> PE1 -> SPE2 -> PE2 -> CE2
 - Downlink: CE2 -> PE2 -> SPE2 -> PE1 -> CE1
- If the link between SPE1 and PE2 fails, packets are transmitted along the following paths:
 - Uplink: CE1 -> PE1 -> SPE1 -> PE3 -> PE2 -> CE2
 - Downlink: CE2 -> PE2 -> PE3 -> SPE1 -> PE1 -> CE1
- If PE2 fails, packets are transmitted along the following paths:
 - Uplink: CE1 -> PE1 -> SPE1 -> PE3 -> CE2
 - Downlink: CE2 -> PE3 -> SPE1 -> PE1 -> CE1
- If the link between CE2 and PE2 fails, packets are transmitted along the following paths:
 - Uplink: CE1 -> PE1 -> SPE1 -> PE2 -> PE3 -> CE2
 - Downlink: CE2 -> PE3 -> PE2 -> SPE1 -> PE1 -> CE1

Parent Topic: [Understanding VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

1.8.3 Application Scenarios for VPWS

[Enterprise Leased Line Service Bearer Using PWE3](#)

[HSI Service Bearer Using PWE3](#)

[PW APS Application](#)

Parent Topic: [VPWS Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.8.3.1 Enterprise Leased Line Service Bearer Using PWE3

Service Overview

As globalization gains momentum, more and more enterprises set up branches in foreign countries and requirements for office flexibility are increasing. An urgent demand for carriers is to provide Layer 2 links for enterprises to set up their own enterprise networks, so that enterprise employees can conveniently visit enterprise intranets from outside their offices.

Various types of backbone networks have been constructed by carriers, using different technologies. For example, carriers construct backbone public switched telephone networks (PSTNs) to carry voice services, HDLC backbone networks to carry HDLC data, and PPP (Point-to-Point Protocol) backbone networks to transmit PPP data. With the exponential growth of IP services, carriers have also constructed IP backbone networks. In addition to the backbone networks, diverse access networks are constructed, which are of different types and difficult to achieve interoperability. Therefore, it is a crucial task for carriers to find a method to effectively integrate these networks, enhance network utilization, and provide more types of services to users.

VPWS is a key technology to the MAN. VPWS enables the original access mode to be well integrated with existing IP backbone networks. This implementation reduces repetitious network construction and operational costs. With VPWS, the IP backbone network joins diverse access networks, achieving extension to and enhancement of the conventional data network. After the MPLS backbone network is constructed, the conventional data communications networks, such as HDLC and PPP networks, can function as access networks. Users, however, are not aware of such changes in the network architecture. In addition, VPWS enables access networks using different protocols to interwork with each other. For example, HDLC users and PPP users can communicate with each other.

Usage Scenario

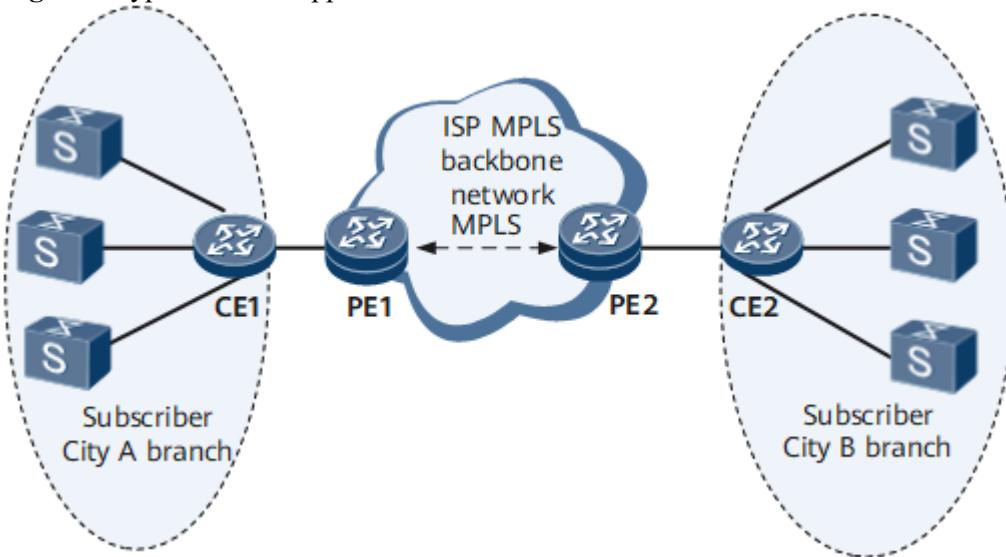
[Figure 1](#) shows a typical single-segment VPWS application, with the IP network being the backbone network. Different LANs access the backbone network.

A carrier constructs a national backbone network that provides VPWS services to a client with two branches, one in City A, and the other in City B. The branch in City A accesses the backbone network using PPP, whereas the branch in City B accesses the backbone network using HDLC or PPP. In this

situation, the carrier can establish a VPWS connection between the two access points, that is, PE1 in City A and PE2 in City B.

In so doing, the carrier can provide the user point-to-point VPN services that traverse the Wide Area Network (WAN). Besides establishing a VPWS connection, no other special measures have to be taken by the carrier. With the simple and convenient VPWS solution, the user can enjoy point-to-point VPN services without having to modify the original intranet, and the carrier can implement smooth transition to the IP backbone network without having to modify existing access modes.

Figure 1 Typical VPWS application



Feature Deployment

1. IP addresses and IGPs are configured on the carrier's MPLS backbone network for communication between PEs.
2. MPLS is enabled on the carrier's backbone network. TE tunnels are configured between PE1 and PE2. Usually, two TE tunnels are configured between PE1 and PE2, one as the primary tunnel and the other as the backup tunnel.
3. MPLS L2VPN is enabled on PE1 and PE2 and a remote MPLS LDP session is configured between PE1 and PE2.
4. PWE3 is configured on the AC interfaces of PE1 and PE2, so that PE1 and PE2 can communicate over MPLS L2VCs.

Parent Topic: [Application Scenarios for VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.8.3.2 HSI Service Bearer Using PWE3

Service Overview

High speed Internet (HSI) services are provided over IP networks. As shown in [Figure 1](#), users access the broadband remote access server (BRAS) by means of Ethernet access. Usually:

- DSLAMs are far away from a BRAS, and an intermediate network is required to connect them.
- Large numbers of DSLAMs are deployed on a network, but the number of DSLAMs for which each BRAS can provide access services is limited and a BRAS is expensive.

To resolve the preceding issues, establish a PWE3 network between the DSLAMs and BRAS. Multiple DSLAMs are connected to the BRAS in Layer 2 aggregation mode, improving BRAS usage efficiency and reducing costs.

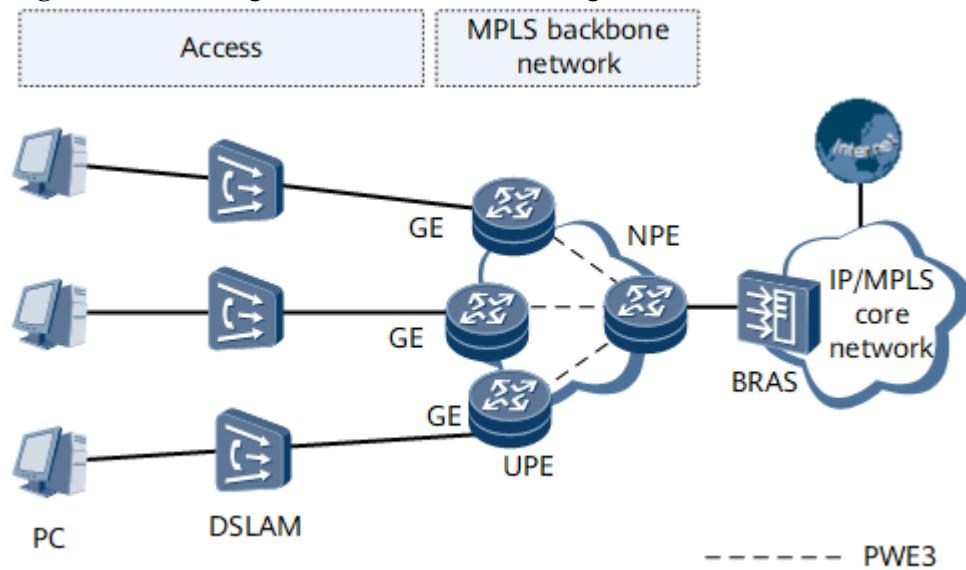
HSI services are data services that do not require low delays. Access and bearer networks can meet the following requirements:

- Strong expansibility
- Clear management responsibilities
- Effective control on Layer 2 broadcast domains
- Guaranteed user data security
- Support for multiple protocols and smooth network upgrade
- Convenient configuration

Networking Description

In [Figure 1](#), digital subscriber line access multiplexers (DSLAMs) converge HSI service packets to UPEs through VLANs. The UPEs transmit the packets to the NPE through PWs. After receiving the packets, the NPE removes PW labels, adds VLAN tags to the packets, and transmits the packets to the BRAS. The BRAS removes the VLAN tags. HSI service users dynamically access the BRAS by means of Ethernet access to obtain IP addresses.

Figure 1 Networking for HSI service bearer using PWE3



Feature Deployment

1. IP addresses and IGPs are configured on the carrier's MPLS backbone network for communication between PEs.
2. MPLS is enabled on the carrier's backbone network. TE tunnels are configured between UPEs and NPEs. Usually, two TE tunnels are configured between each UPE and NPE, one

as the primary tunnel and the other as the backup tunnel.

3. MPLS L2VPN is enabled on each UPE and NPE. A remote MPLS LDP session is configured between each UPE and NPE.
4. PWE3 is configured on AC interfaces of each UPE and NPE, so that UPEs and NPEs can communicate over MPLS L2VCs.

Parent Topic: [Application Scenarios for VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

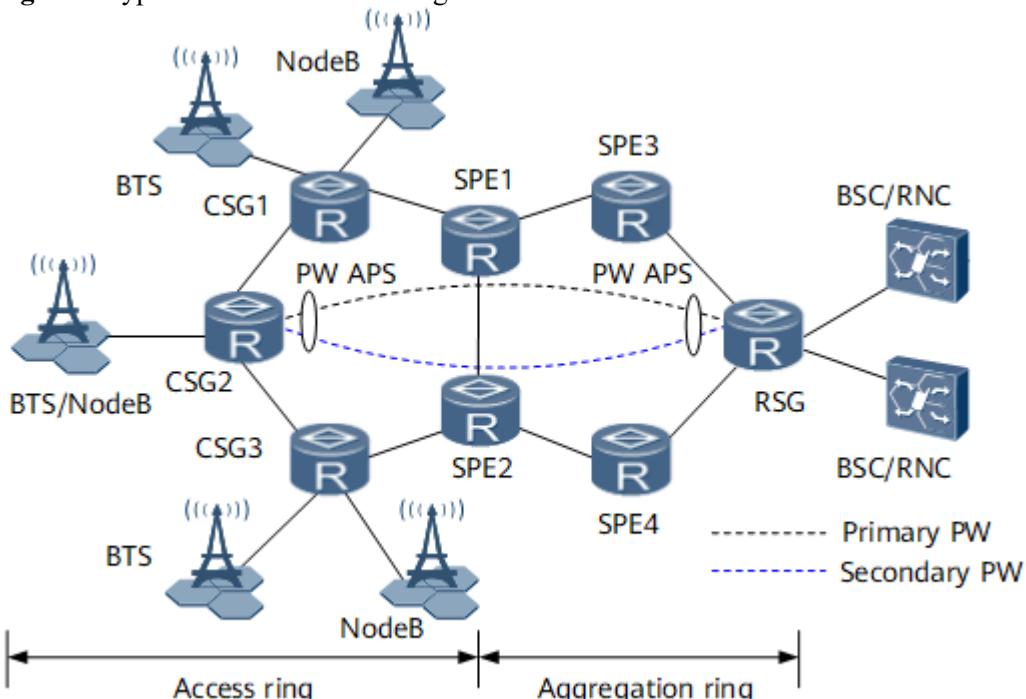
[< Previous topic](#) [Next topic >](#)

1.8.3.3 PW APS Application

[Figure 1](#) shows typical pseudo wire (PW) automatic protection switching (APS) networking. The network comprises an access ring and an aggregation ring. A BTS/NodeB is connected to a cell site gateway (CSG). A BSC/RNC is connected to an RSG. Primary and secondary PWs are established between a CSG and an RSG. The PWs can be either single-segment PWs (SS-PWs) or multi-segment PWs (MS-PWs). A BTS/NodeB communicates with a BSC/RNC through a mobile broadband (MBB) network.

PW APS is deployed on the bearer network to improve reliability. APS instances are configured on CSGs and RSGs, and the primary/secondary pseudo wire (PW) protection group is associated with each APS instance. APS instructs the source and destination ends to implement bidirectional protection switching in the same manner to achieve delayed switching and wait-to-restore (WTR) for PW protection.

Figure 1 Typical PW-APS networking



Parent Topic: [Application Scenarios for VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.9 IP Hard Pipe Description

[Overview of IP Hard Pipe](#)

[Understanding IP Hard Pipe](#)

This section describes the implementation principles of IP hard pipe.

[Application Scenarios for IP Hard Pipe](#)

[Terminology for IP Hard Pipe](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

< Previous topic > Next topic

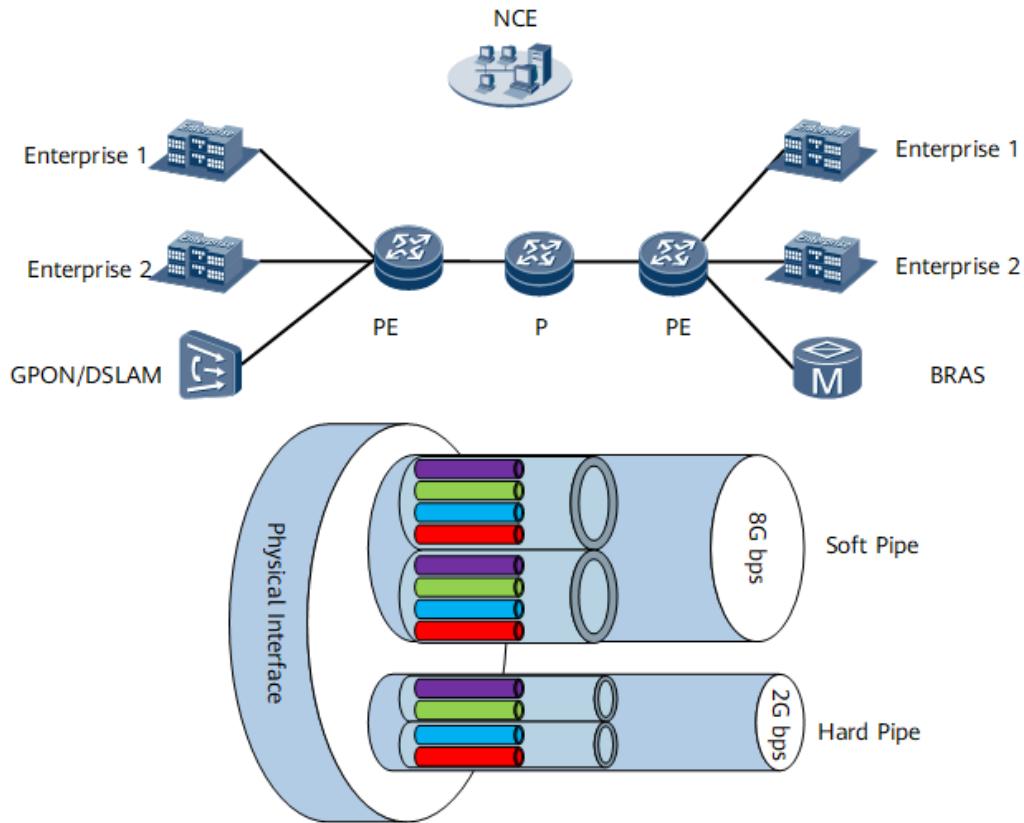
1.9.1 Overview of IP Hard Pipe

Definition

IP hard pipe is an IP-network-based access technology that strictly isolates soft and hard pipes by reserving hardware on routers. The hard pipe can preempt the bandwidth resources of the soft pipe while not being affected by soft pipe congestion. The hard pipe can provide guaranteed bandwidth and low delay for the leased line services of high-value customers.

In the IP hard pipe solution, NCE is used to manage bandwidth resources network-wide. The physical interface bandwidth on the public network is divided and allocated to hard and soft pipes. For example, on a 10G Ethernet interface, 2 Gbit/s bandwidth is allocated to the hard pipe, and the remaining 8 Gbit/s is allocated to the soft pipe. The hard and soft pipes are isolated. The hard pipe can preempt the bandwidth resources of the soft pipe, but the soft pipe cannot preempt the bandwidth of the hard pipe.

Figure 1 IP hard pipe networking



Purpose

Customers that have high requirements on bandwidth, low delay, and high security prefer synchronous digital hierarchy (SDH) networks. To retain these customers, carriers must keep both IP and SDH networks, which costs a lot in maintenance. Therefore, carriers expect to migrate the customer network to the IP network to reduce maintenance costs and facilitate user management.

IP hard pipe has been developed to meet up with this expectation. It provides bandwidth guarantee and low delay on IP networks, allowing the IP networks to provide access services with SDH service quality. It also provides service-specific granular OAM and SLA monitoring, which will accelerate the migration of SDH networks to IP leased line networks.

Benefits

IP hard pipe offers the following benefits to carriers:

- Deployment of high-quality leased lines for VIP customers on newly deployed or existing routers, reducing SDH network construction and costs for maintaining both SDH and IP networks
- Rapid service protection, ensuring high-reliability service quality
- Granular service quality measurement using IP FPM, providing flexible and effective maintenance and management means to leased lines of VIP customers

Parent Topic: [IP Hard Pipe Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.9.2 Understanding IP Hard Pipe

This section describes the implementation principles of IP hard pipe.

[Centralized Management of IP Hard-Pipe-based Leased Line Services on the NMS](#)

[Interface-based Hard Pipe Bandwidth Reservation](#)

[AC Interface Service Bandwidth Limitation](#)

[Hard-Pipe-based TE LSP](#)

[Hard Pipe-based VPWS/VPLS](#)

[Hard Pipe Reliability](#)

[Hard Pipe Service Quality Monitoring](#)

Parent Topic: [IP Hard Pipe Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

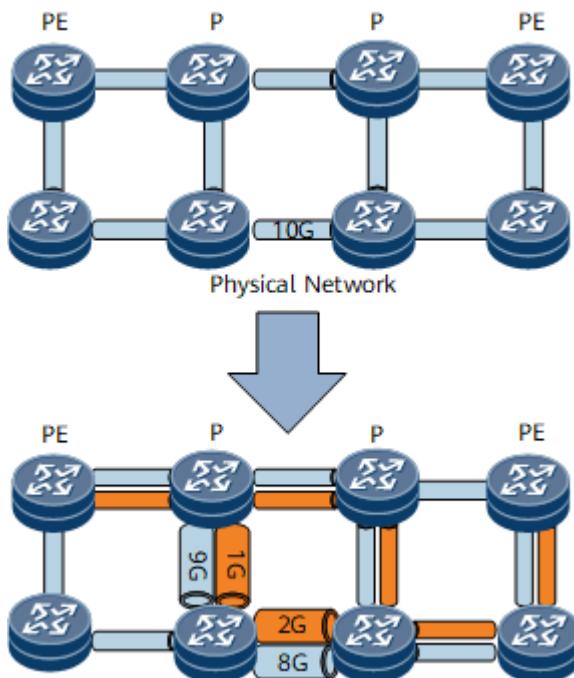
< Previous topic > Next topic

1.9.2.1 Centralized Management of IP Hard-Pipe-based Leased Line Services on the NMS

In the IP hard pipe solution, the NMS centrally manages bandwidth resources and implements service provisioning. Hard pipe service provisioning involves two steps:

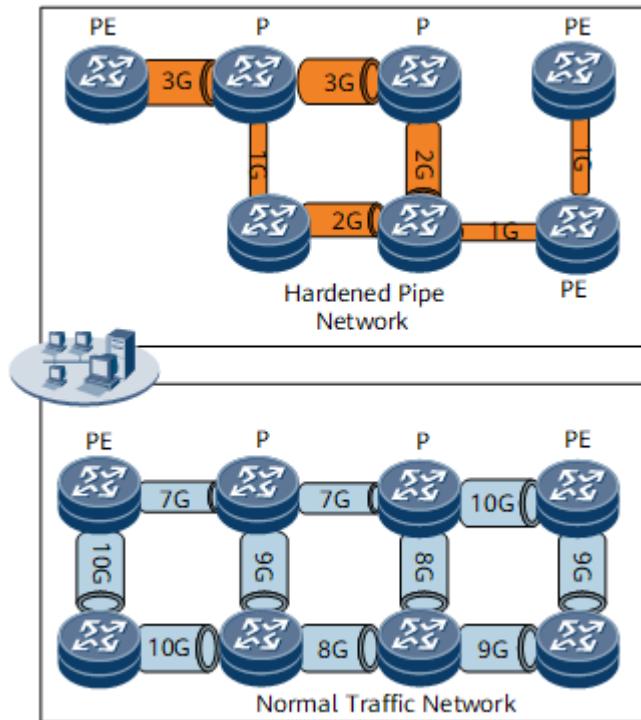
1. Establish a hard pipe plane.

Figure 1 IP hard pipe topology establishment 1



In the physical network topology, select the public network links that require hard pipe deployment and set the hard pipe bandwidth for each link. The hard pipe topology is then established. On the network shown in [Figure 2](#), after hard pipes are classified on the entire network, the original network is divided into two logical networks: a hard pipe network and a normal service network (called a soft pipe network).

Figure 2 IP hard pipe topology establishment 2



2. Service provisioning

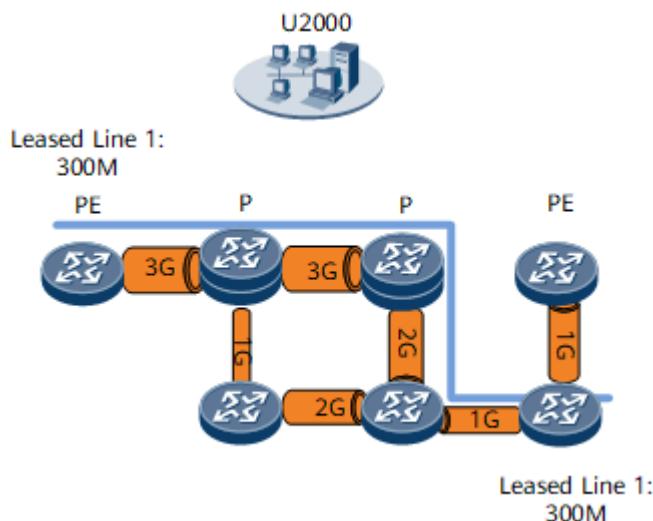
The service bandwidth, source and destination devices, and service IDs are manually configured for VIP customers. The intermediate path can be configured or automatically calculated by the NMS.

The NMS checks the hard pipe bandwidth on each node to see if the bandwidth is adequate for service provisioning. If not, the NMS stops service provisioning and informs users of it.

The NMS delivers configurations to devices.

After service provisioning succeeds, the NMS updates the bandwidth resource database.

Figure 3 IP hard pipe service provisioning



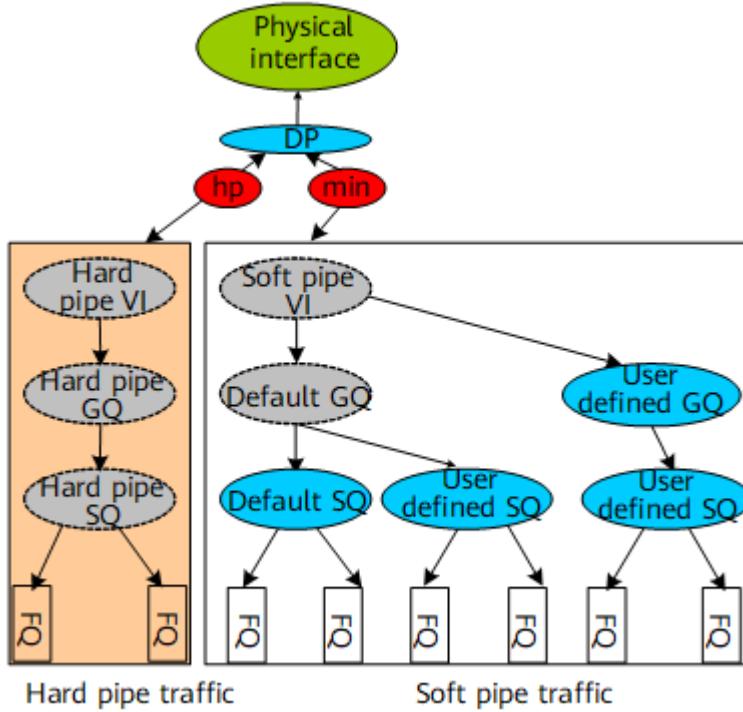
1.9.2.2 Interface-based Hard Pipe Bandwidth Reservation

During network planning, traffic is classified as hard pipe traffic and soft pipe traffic based on interfaces. Hard pipe traffic must have low delay and no packet loss and not be affected by soft pipe traffic. Therefore, fixed bandwidth must be allocated to hard pipe traffic on interfaces. This bandwidth is exclusive to hard pipe traffic.

On devices, hard pipe and soft pipe services on interfaces are assigned forwarding paths with different priorities. When both hard and soft pipes have traffic, the hard pipe traffic is preferentially forwarded, guaranteeing low delay. The maximum bandwidth is also configured for the forwarding paths to ensure that the sum of hard and soft pipe bandwidth does not exceed the interface bandwidth. Subsequently, the hard and soft pipe services do not affect each other.

On the network shown in [Figure 1](#), bandwidth limitation is applied to the hard pipe VI and soft pipe VI. The hard pipe VI traffic is preferentially scheduled on the physical interface, ensuring guaranteed bandwidth and low delay.

Figure 1 Hard and soft pipe bandwidth reservation



End-to-end hard pipe service deployment can be implemented only through the NMS. The NMS delivers hard pipe VLL/PWE3 or TE LSP configurations based on the bearer capabilities of the hard pipe. The device establishes VLL/PWE3 PWs and TE LSPs based on the delivered data and transmits VLL/PWE3 or TE services through the hard pipe.

The NMS supports alarm thresholds for services exceeding the hard pipe's processing capabilities, ensuring that services transmitted over the hard pipe do not exceed the hard pipe's processing capabilities.

Parent Topic: [Understanding IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.9.2.3 AC Interface Service Bandwidth Limitation

User-specific bandwidth limitation is implemented using hierarchical QoS (HQoS) Interface-based HQoS schedules user services based on priorities. If the access user service bandwidth reaches the maximum leased line bandwidth, high-priority user packets are preferentially forwarded. This ensures service quality for high-priority user packets.

When you configure HQoS on AC interfaces to limit user bandwidth, configuring a flow-mapping template is recommended. In this template, map all eight priorities of incoming traffic to CS7 (highest priority) to ensure that hard pipe services are scheduled with the highest priority.

Parent Topic: [Understanding IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.9.2.4 Hard-Pipe-based TE LSP

Principles

After the hard pipe bandwidth is reserved on a physical interface on a carrier network, the logical hard network is partitioned. Then path planning is required for services provisioned to users. P2P leased line services are carried over static bidirectional co-routed TE LSPs between two PEs.

After a carrier determines the PE for user access on the NMS, the network transmission paths can be manually specified or automatically generated. When the NMS automatically plans paths, hard pipe bandwidth is reserved hop by hop on the transmission path based on the user access bandwidth. If the hard pipe bandwidth of all links on the transmission path meets the user access service requirements, a hard-pipe-based TE LSP is established between the PEs. The NMS then updates the bandwidth resource database. This implements hard pipe services over the TE LSP.

DiffServ Model

TE LSPs use the DiffServ pipe mode. When the hard pipe attribute is configured, DiffServ is switched to enhanced pipe mode automatically. The commands for configuring hard pipe attributes and Pipe/Short-pipe in diffserv mode are mutually exclusive. Therefore, you cannot manually change the DiffServ mode for TE LSPs.

Parent Topic: [Understanding IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.9.2.5 Hard Pipe-based VPWS/VPLS

Principles

A logical hard pipe network can be defined if you specify hard-pipe attributes and reserve bandwidth resources on the network-side interfaces of PEs and the P. To provide users with IP hard-pipe leased line services, configure hard-pipe attributes for static PWs of a VPWS or VPLS.

- To implement a P2P IP hard pipe, enable the IP hard pipe function for the static PW of a VPWS.
- To implement an MP2MP IP hard pipe, enable the IP hard pipe function for static PWs of a VPLS.

A hard pipe-based VPWS and a hard pipe-based VPLS use the same IP hard pipe scheduling model and tunnel type. After a carrier determines the PEs for user access on the NMS and plans a transmission path, a hard-pipe TE tunnel can be established between the PEs. The carrier can use the NMS to reserve a section of hard-pipe bandwidth for each node along the TE tunnel based on committed user access bandwidth and establish PWs to be bound to the TE tunnel.

A hard pipe-based VPWS and a hard pipe-based VPLS require different configurations of IP hard pipe bandwidth. For a hard pipe-based VPLS, establish PWs to connect different access sites on multiple PEs and restrict access bandwidth for the PWs. For a hard pipe-based VPWS, establish a PW to connect different access sites on two PEs and restrict access bandwidth for the PW. In this way, both a hard pipe-based VPWS and a hard pipe-based VPLS ensure sufficient bandwidth and low delay for leased line services.

DiffServ Model

IP hard pipe VPWS services use the pipe enhanced model. When hard pipe attributes are configured, VPWS services are switched to the pipe enhanced model. Therefore, you do not need to configure the DiffServ command separately. The commands for configuring hard pipe attributes and a DiffServ mode are mutually exclusive. Therefore, you cannot manually change the DiffServ mode for TE LSPs.

The uniform mode is configured on the PE's user-to-network inbound interface to identify user data, differentiating priorities and mapping VC label priorities. The pipe mode is used on the PE's network-to-user outbound interface to process user data, without changing the user packet priorities.

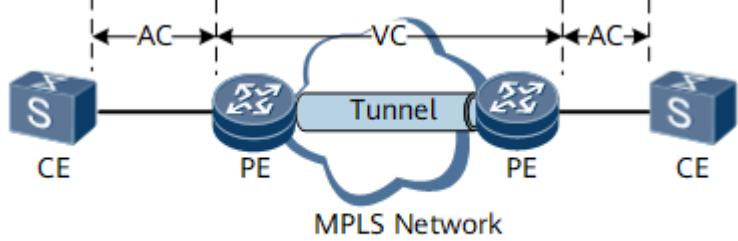
Bandwidth Expansion for Hard Pipe-based VPWS

In compliant with the VPWS/PWE3 service packet encapsulation standards, the outbound interface on the public network encapsulates a public network header to access user packets before sending them out. Therefore, the packet length is increased, and the access user bandwidth must be increased on the public network side.

The NMS must reserve bandwidth for the hard pipe on the public network side based on the expanded access user bandwidth.

On the network shown in [Figure 1](#):

Figure 1 MPLS L2VPN networking



In this example, the length of packets received on a PE from a CE is L1 (CRC length included). The length of the packets sent by the PE to the public network interface is L2. The public network interface is an Ethernet interface that sends double-tagged packets. The L2 is calculated as follows:

$L2 = L1 + \text{Public network header length}$ (length of the destination MAC address, source MAC address, Eth_Type, outer VLAN tag, inner VLAN tag, TE label, and VC label)

$$L2 = L1 + 30$$

The calculation shows that the public network packet length is determined by the following factors:

- User packet length
- Public network link type

The user data packet length varies. Even in the data flow from the same access user, packets are of varied lengths. Therefore, there is no fixed and accurate bandwidth expansion proportion on Ethernet links. However, the bandwidth expansion proportion can be calculated based on the average packet length.

The VPWS bandwidth expansion proportion parameters can be configured. The default value is calculated based on the average parameter values:

1. User packet length

The average packet length of 300 bytes is used as the default packet length.

2. Public network link type

The public network interface uses Ethernet encapsulation and sends packets carrying double VLAN tags.

The default bandwidth expansion proportion is: $30/300 = 10\%$

NOTE

The bandwidth expansion proportion varies depending on the POS and Ethernet encapsulation lengths and the number of VLAN tags on the Ethernet network.

Parent Topic: [Understanding IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.9.2.6 Hard Pipe Reliability

IP hard pipe provides reliability for both services and public network tunnels.

Deploy TP OAM and PW APS for primary and secondary static VPWS PWs to provide a service switchover of 50 ms.

Deploy TP OAM and PW APS for primary and secondary static VPLS PWs to provide a service switchover of 50 ms.

Deploy TP OAM and TE APS for static TE tunnel protection groups to provide granular tunnel protection of 50 ms.

The reliability mechanism for IP hard pipe is the same as that for common services.

Parent Topic: [Understanding IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.9.2.7 Hard Pipe Service Quality Monitoring

IP FPM can be deployed to implement service flow-based real-time performance monitoring. TP OAM or Y.1731 can be deployed to perform packet loss and delay measurement for VLL/PWE3 and TE tunnels of IP hard pipe.

The reliability mechanism for IP hard pipe is the same as that for common services.

Parent Topic: [Understanding IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.9.3 Application Scenarios for IP Hard Pipe

Hard pipe applies to P2P leased line services of high-end enterprise users.

[Hard-Pipe-based Enterprise Leased Line Application](#)

[Hard-Pipe-based Enterprise Leased Line Protection](#)

[Hard-Pipe-based Leased Line Services Implemented by Huawei and Non-Huawei Devices](#)

Parent Topic: [IP Hard Pipe Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

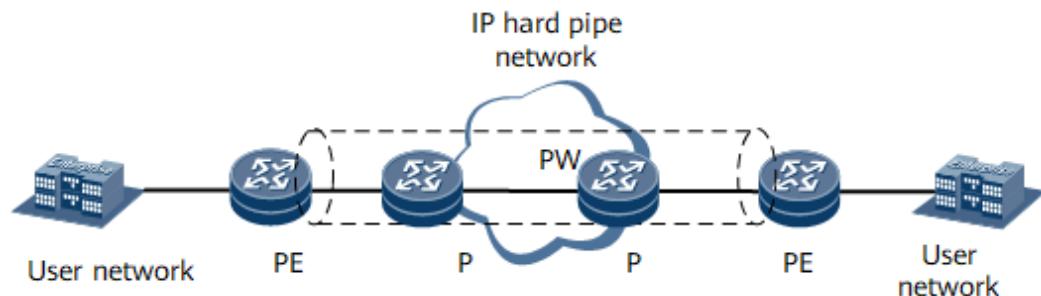
1.9.3.1 Hard-Pipe-based Enterprise Leased Line Application

The carrier plans the logical hard network on the existing IP bearer network and reserves the hard pipe bandwidth on physical interfaces.

On the network shown in [Figure 1](#), an enterprise user wants to establish a leased line between two sites over a carrier's network. The carrier first plans a path and then establishes a static bidirectional LSP

dedicated to the hard pipe over the path. The PEs then establish a hard-pipe-based static PW. Bandwidth is reserved hop by hop along the path. If Ethernet links that share bandwidth are used for user access, configure QoS on AC interfaces to limit the access bandwidth. Subsequently, a hard-pipe-based enterprise leased line is established. If the hard pipe bandwidth is inadequate, the NMS does not allow the static PW to be established.

Figure 1 Hard-pipe-based enterprise leased line application



Parent Topic: [Application Scenarios for IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.

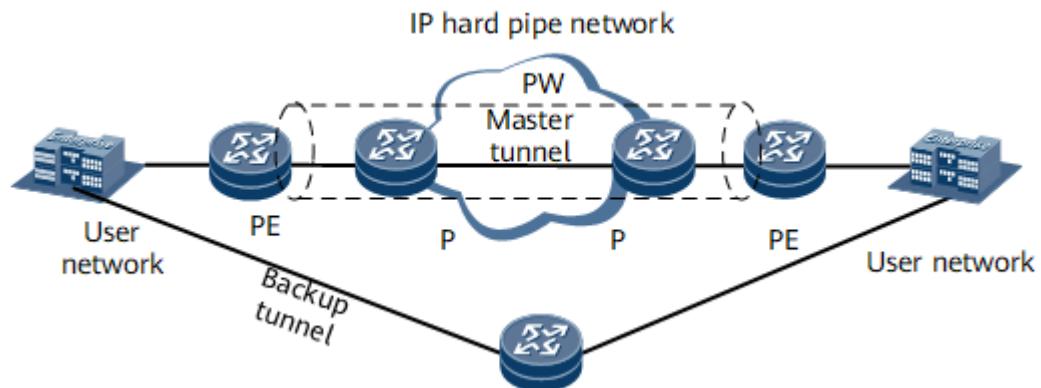
Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.9.3.2 Hard-Pipe-based Enterprise Leased Line Protection

Tunnel protection groups can be configured for hard-pipe-based enterprise leased lines to protect the network against faults. On the network shown in [Figure 1](#), master and backup tunnels are established using hard pipes between the PEs, forming a tunnel protection group. Hard-pipe-based PWs are established over the tunnel protection group. TP-OAM is deployed for the master tunnel. If the master tunnel link or a P fails, traffic can be quickly switched to the backup tunnel, implementing protection switching.

Figure 1 Hard pipe enterprise leased line tunnel protection



Parent Topic: [Application Scenarios for IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.

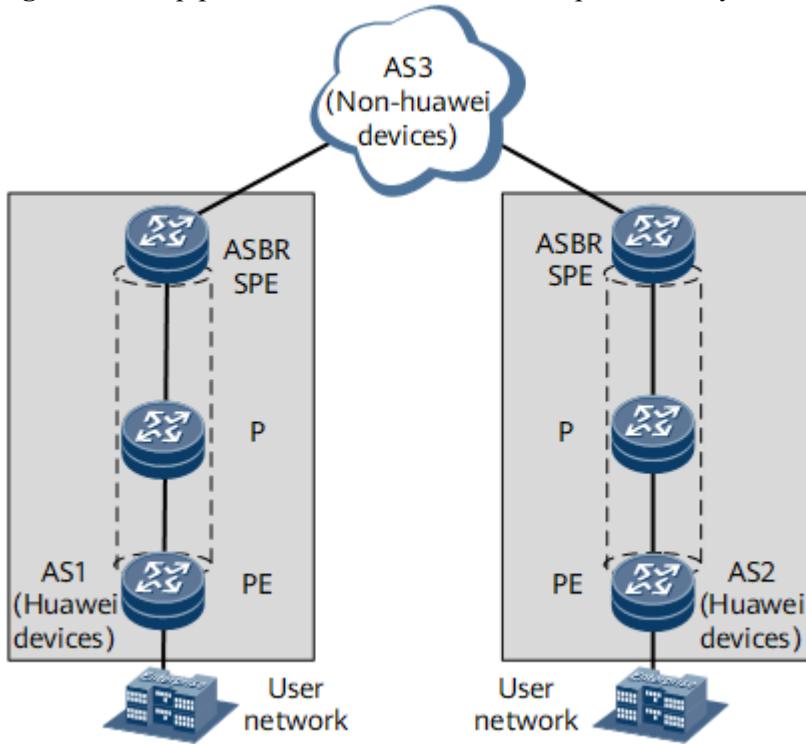
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.9.3.3 Hard-Pipe-based Leased Line Services Implemented by Huawei and Non-Huawei Devices

Hard pipe deployment requires network-wide devices. When Huawei devices are connected over a network constructed by non-Huawei devices to implement E2E hard pipe services, if non-Huawei devices also support PWs in implementing hard pipe, multi-segment PWs (MS-PWs) can be used.

Figure 1 Hard-pipe-based leased line services implemented by Huawei and non-Huawei devices



Parent Topic: [Application Scenarios for IP Hard Pipe](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.9.4 Terminology for IP Hard Pipe

Terms

Term	Description
IP hard pipe	A technology that provides IP leased line services with strict bandwidth guarantee and low delay.
SLA	Service level agreement. A service agreement between a customer and a service provider, defining the service type and quality and payment for a customer.
SDH	Synchronous digital hierarchy. A comprehensive transmission network that integrates multiplexing, line transmission, and switching function operated by the NMS.

Parent Topic: [IP Hard Pipe Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.10 VPLS Description

[Overview of VPLS](#)

[Understanding VPLS](#)

[Application Scenarios for VPLS](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

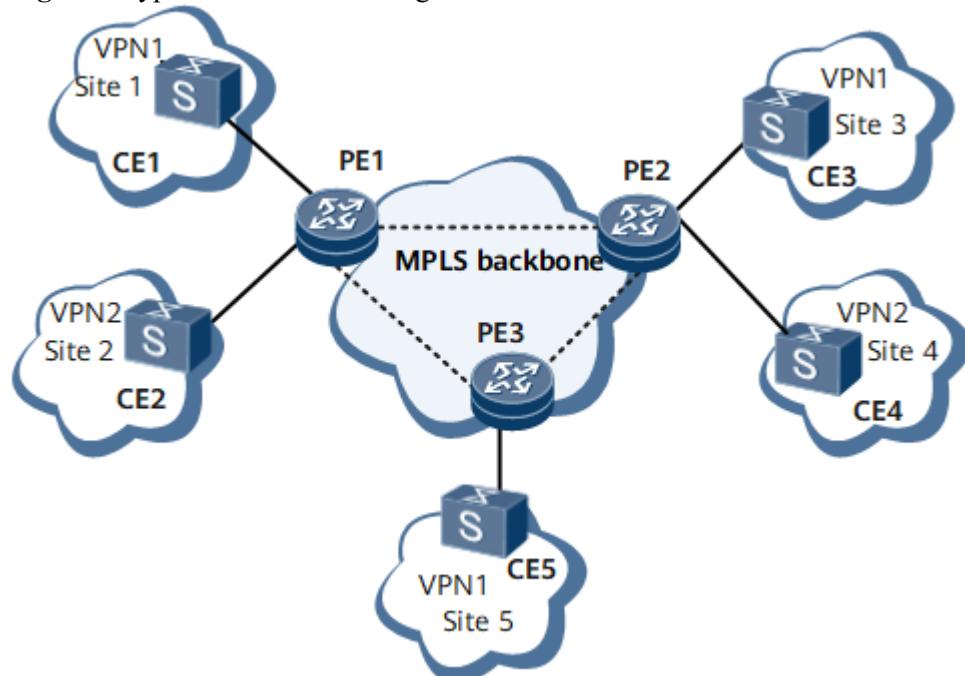
1.10.1 Overview of VPLS

Definition

The virtual private LAN service (VPLS) is an MPLS-based Ethernet point-to-multipoint (P2MP) L2VPN service provided over a public network. VPLS ensures that geographically isolated user sites can communicate over MANs and WANs as if they were on the same LAN. VPLS is also called transparent LAN service (TLS).

[Figure 1](#) shows a typical VPLS networking mode. In this networking, users located in different geographical regions communicate with each other over different PEs. From the perspective of users, a VPLS network is a Layer 2 switched network that allows them to communicate with each other in a way similar to communication over a LAN.

Figure 1 Typical VPLS networking



Purpose

As enterprises set up more and more branches in different regions and office flexibility increases, applications such as voice over IP (VoIP), instant messaging, and teleconferencing are increasingly widely used. This imposes high requirements for end-to-end (E2E) datacom technologies. A network capable of providing P2MP services is the key to datacom function implementation.

Traditional ATM and FR technologies provide only Layer 2 P2P connections. In addition, those network types have disadvantages, such as high construction costs, low speed, and complex deployment. The development of IP has led to the MPLS VPN technology, which can provide VPN services over an IP network and offer advantages such as easy configuration and flexible bandwidth control. MPLS VPNs can be classified into MPLS L2VPNs and MPLS L3VPNs.

- Traditional MPLS L2VPNs, such as VPWS networks, can provide P2P services but not P2MP services over a public network.
- MPLS L3VPNs can provide P2MP services on the precondition that PEs keep routes destined for end users. This implementation requires high routing performance of PEs.

To solve the preceding problems, VPLS, an MPLS-based Ethernet technology, is introduced.

- Like Ethernet, VPLS supports P2MP communication.
- MPLS is a Layer 2 label switching technology. From the perspective of users, the entire MPLS IP backbone network is a Layer 2 switching device. PEs do not need to keep routes destined for end users.

VPLS provides a more complete multipoint communication solution, integrating the advantages provided by Ethernet and MPLS. By emulating traditional LAN functions, VPLS enables users on different LANs to communicate with each other over MPLS networks as if they were on the same LAN.

Benefits

VPLS offers the following benefits:

- VPLS networks can be constructed based on carrier's IP networks, reducing construction costs.
- VPLS networks inherit the high-speed advantage of the Ethernet.
- VPLS networks allow users to communicate over Ethernet links, regardless of whether these links are on WANs or LANs. This feature allows services to be rapidly and flexibly deployed.
- VPLS networks free carriers from configuring and maintaining routing policies, reducing operational expenditure.

Parent Topic: [VPLS Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.10.2 Understanding VPLS

[VPLS Description](#)

[VPLS Functions](#)

[LDP VPLS](#)

[BGP VPLS](#)

[HVPLS](#)

[BGP AD VPLS](#)

[Inter-AS VPLS](#)

[Flow-Label-based Load Balancing](#)

[VPLS PW Redundancy](#)

[Multicast VPLS](#)

[VPLS Multi-homing](#)

[VPLS Service Isolation](#)

Parent Topic: [VPLS Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.1 VPLS Description

Basic VPLS Transmission Structure

[Figure 1](#) shows an example of a VPLS network. The entire VPLS network is similar to a switch. PWs are established over MPLS tunnels between VPN sites to transparently transmit Layer 2 packets between sites. When forwarding packets, PEs learn the source MAC addresses of these packets and create MAC entries, mapping MAC addresses to ACs and PWs.

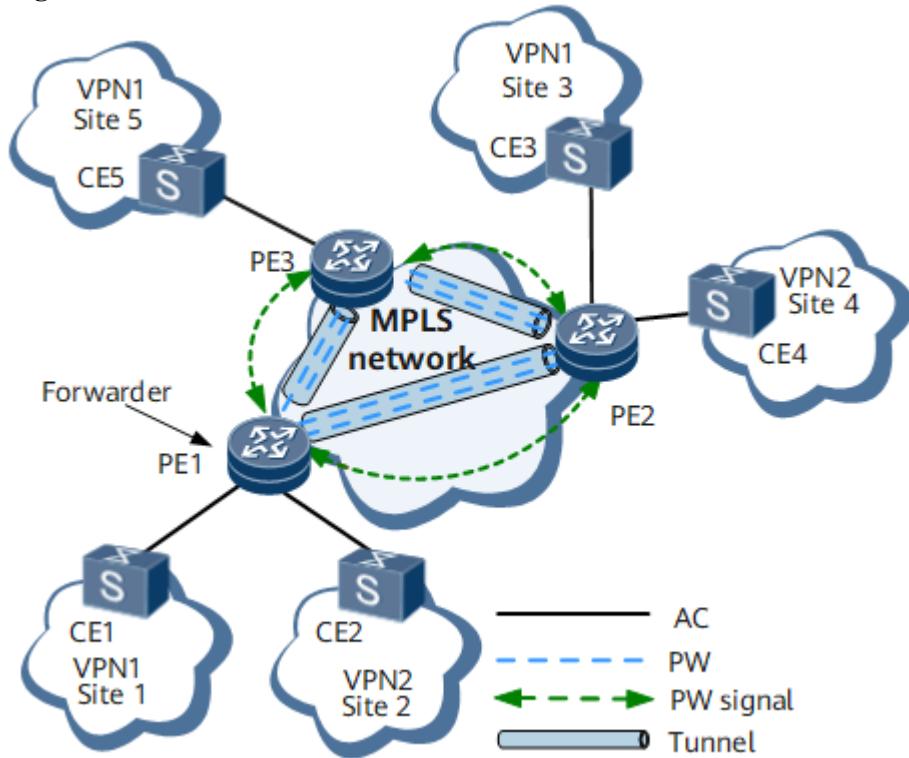
[Table 1](#) describes the concepts related to VPLS networks.

Table 1 Description of VPLS concepts

Name	Full Name	Concept
AC	Attachment circuit	A link between a CE and a PE. The interface must be an Ethernet interface.
PW	Pseudowire	A bidirectional virtual connection between two virtual switch instances (VSIs) residing on two PEs. A PW consists of a pair of unidirectional MPLS VCs in opposite directions.
VSI	Virtual switch instance	A type of instance used to map ACs to PWs. A VSI independently provides VPLS services and forwards Layer 2 packets based on MAC addresses and VLAN tags. A VSI has the Ethernet bridge function and can terminate PWs.
PW signaling	PW signaling protocol	A type of signaling used to create and maintain PWs. PW signaling is the foundation for VPLS implementation. Typically, LDP and BGP are used as the PW signaling protocols.

Name	Full Name	Concept
Tunnel	Tunnel	A tunnel can carry multiple PWs. A tunnel is a direct channel that transparently transmits data between the local and remote PE devices. It can be an MPLS or a GRE tunnel.
Forwarder	Forwarder	After a PE receives packets from an AC, the forwarder of the PE selects a PW to forward these packets. It is similar to a VPLS forwarding table.

Figure 1 Basic VPLS transmission structure



The forwarding of a packet from CE1 to CE3 on VPN1 is used as an example:

1. CE1 sends a Layer 2 packet to PE1 over an AC.
2. After PE1 receives the packet, the forwarder of PE1 selects a PW for forwarding the packet.
3. PE1 then adds two MPLS labels to the packet based on the PW forwarding entry and tunnel information and sends the packet to PE2. The inner private label identifies the PW, and the outer public network label identifies the tunnel between PE1 and PE2.
4. After PE2 receives the packet sent along the public tunnel, PE2 removes the private network label from the packet.
5. The forwarder of PE2 selects an AC and forwards the packet to CE3 over the AC.

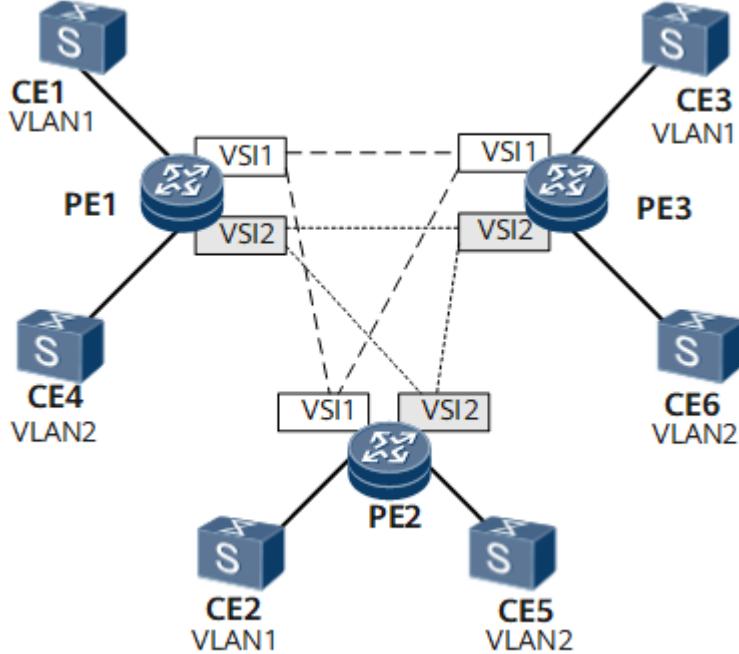
VPLS Implementation Process

Transmission of packets between CEs relies on VSIs configured on PEs, and PWs established between the VSIs. [Figure 1](#) shows transmission of Ethernet frames over a full mesh of PWs between PEs. [Figure 2](#) shows transmission of Ethernet frames through a full mesh of PWs between PEs.

The Ethernet often uses the Spanning Tree Protocol (STP) to prevent loops. VPLS networks, however, use a full mesh of PWs and split horizon to avoid loops:

- PEs on a VPLS network must be fully meshed. That is, a PE must create a tree path to every other PE on the VPLS network.
- Each PE must support split horizon to prevent loops. Split horizon requires that packets received from a PW in a VSI should not be forwarded to other PWs in the VSI. Any two PEs on a VPLS network must communicate over a direct PW, which explains why a VPLS network requires a full mesh of PWs between PEs.

Figure 2 VPLS packet forwarding model



A PE on a VPLS network consists of a control plane and a forwarding plane.

- The control plane of a VPLS PE is responsible for PW establishment, including:
 - Member discovery: a process in which a PE in a VSI discovers other PEs in the same VSI. You can manually configure or use a protocol to automatically complete the configuration. The latter is called automatic discovery.
 - Signaling mechanism: PWs between PEs with the same VSI ID are established, maintained, or torn down using signaling protocols, such as LDP and BGP.
- The forwarding plane of a VPLS PE is responsible for data forwarding over the PW, including:
 - Encapsulation: After receiving Ethernet frames from a CE, a PE encapsulates the frames into packets and sends the packets to a PSN.
 - Forwarding: A PE determines how to forward a packet based on the inbound interface and destination MAC address of the packet.
 - Decapsulation: After receiving packets from a PSN, a PE decapsulates these packets into Ethernet frames and sends the frames to a CE.

VPLS Encapsulation Types

- Packet encapsulation types on AC interfaces

Packet encapsulation on AC interfaces depends on the user access mode, which can be VLAN or Ethernet access. The default user access mode is VLAN access.

Table 2 Packet encapsulation types on AC interfaces

Packet Encapsulation Type on AC Interfaces	Description
VLAN access	Each Ethernet frame transmitted between CEs and PEs carries a VLAN tag called a Provider-tag (P-tag). This is a service delimiter identifying users on an ISP network.
Ethernet access	Ethernet frames transmitted between CEs and PEs do not necessarily carry VLAN tags. If an Ethernet frame carries a VLAN tag, the tag is an internal VLAN tag called a user-tag (U-tag) in user packets. The U-tag is carried in a packet before the packet is sent to a CE and is not added by the CE. The U-tag is used by the CE to identify to which VLAN the packet belongs and is meaningless to PEs.

- **Packet encapsulation on PWs**

The PW ID and PW encapsulation type together uniquely identify a PW. The PW IDs and PW encapsulation types configured on both end PEs of a PW must be the same. The packet encapsulation types of packets on PWs can be raw or tagged. By default, packets on PWs are encapsulated in tagged mode.

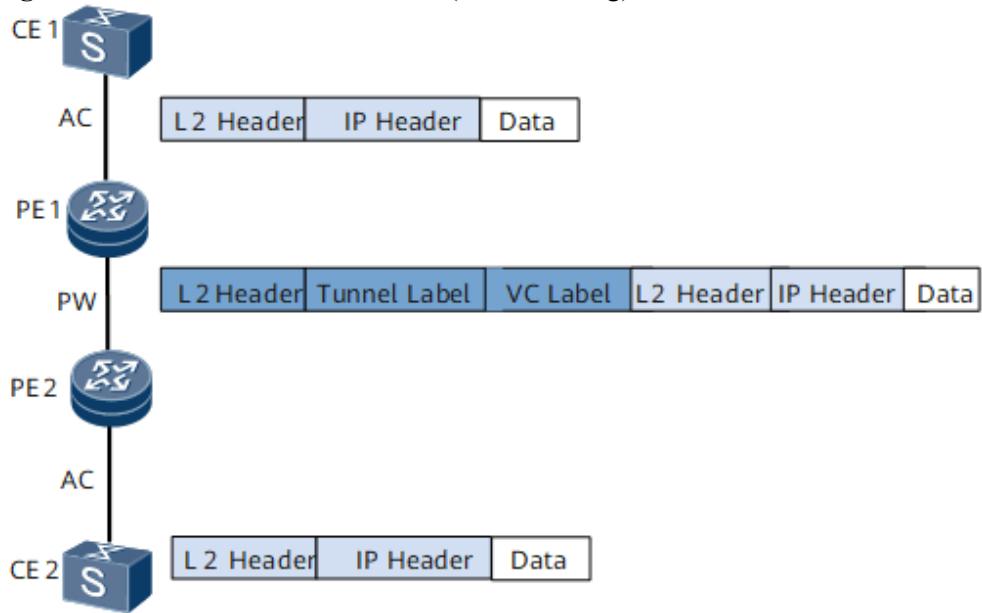
Table 3 Packet encapsulation types

Packet Encapsulation on PWs	Description
Raw	Packets transmitted over a PW cannot carry P-tags. If a PE receives a packet with the P-tag from a CE, the PE strips the P-tag and adds double labels (outer tunnel label and inner VC label) to the packet before forwarding it. If a PE receives a packet with no P-tag from a CE, the PE directly adds double labels (outer tunnel label and inner VC label) to the packet before forwarding it. The PE determines whether to add the P-tag to a packet, depending on the configuration, before sending it to a CE. The PE is not allowed to rewrite or remove an existing U-tag.
Tagged	Packets transmitted over a PW must carry P-tags. If a PE receives a packet with the P-tag from a CE, the PE directly adds double labels (outer tunnel label and inner VC label) to the packet before forwarding it. If a PE receives a packet with no P-tag from a CE, the PE adds a null P-tag and double labels (outer tunnel label and inner VC label) to the packet before forwarding it. The PE determines whether to rewrite, remove, or preserve the P-tag of a packet, depending on the configuration, before forwarding it to a CE. In a scenario where a sub-interface is connected to an L2VPN, to ensure normal communication between a Huawei device and a non-Huawei device, you can configure the PE to change the P-tag of packets entering a PW in tagged mode. In a scenario where a main interface accesses an L2VPN, the main interface adds an empty tag to each packet by default, adds two MPLS labels to each packet, and then forwards the packets.

Encapsulation modes of packets transmitted over ACs and PWs can be used together. The following uses Ethernet access in raw mode (without the U-tag) and VLAN access in tagged mode (with the U-tag) as examples to describe the packet exchange process.

- Ethernet access in raw mode (without U-tag)

Figure 3 Ethernet access in raw mode (without U-tag)



As shown in [Figure 3](#), ACs use Ethernet encapsulation and PWs use raw encapsulation; packets transmitted from CEs to PEs do not carry U-tags.

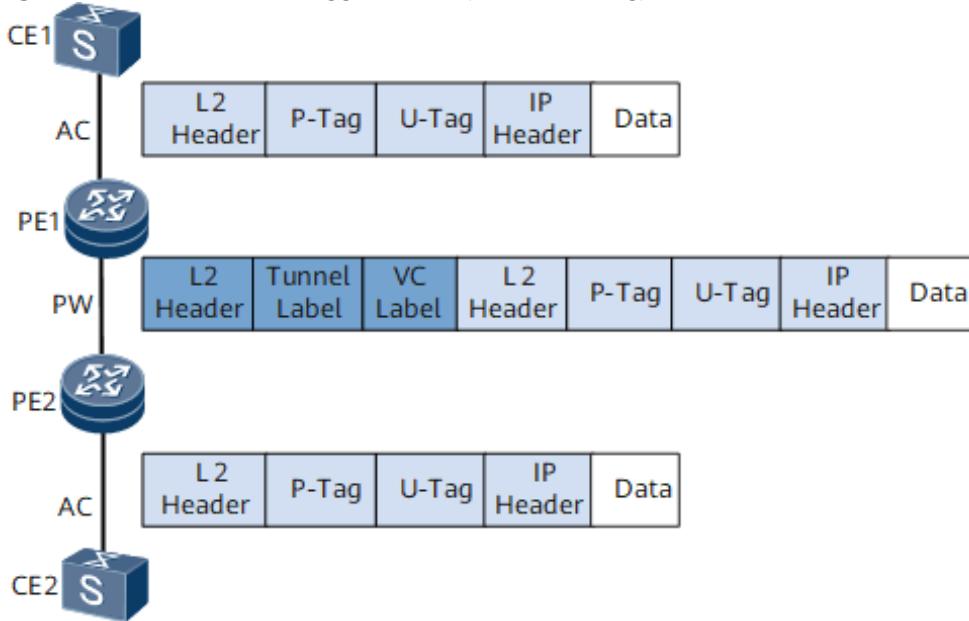
The packet exchange process in Ethernet access in raw mode is as follows:

1. CE1 sends to PE1 a packet that is encapsulated at Layer 2 and does not carry any U-tag or P-tag.
2. After receiving the packet, PE1 searches the corresponding VSI for the entry, and selects a tunnel and a PW for the packet. PE1 adds double MPLS labels (outer tunnel label and inner VC label) to the packet based on the selected tunnel and PW, performs Layer 2 encapsulation, and forwards the packet to PE2.
3. PE2 receives the packet from PE1 and decapsulates the packet to remove Layer 2 encapsulation information and two MPLS labels.
4. PE2 sends the original Layer 2 packet to CE2.

The processing of sending a packet from CE2 to CE1 is similar to this process.

- VLAN access in tagged mode (with the U-tag)

Figure 4 VLAN access in tagged mode (with the U-tag)



As shown in [Figure 4](#), ACs use VLAN encapsulation and PWs use tagged encapsulation; packets transmitted from CEs to PEs carry U-tags and P-tags.

The packet exchange process in VLAN access in tagged mode (with the U-tag) is as follows:

1. CE1 sends to PE1 a packet that is encapsulated at Layer 2 and carries both a U-tag and a P-tag.
2. Upon receipt of the packet, PE1 does not process the two tags. PE1 retains the U-tag because it treats the U-tag as service data.
3. PE1 retains the P-tag because a packet sent to a PW with the tagged packet encapsulation mode must carry a P-tag.
4. PE1 queries entries in the VSI, and selects a tunnel and a PW for the packet.
5. PE1 adds double MPLS labels (outer tunnel label and inner VC label) to the packet based on the selected tunnel and PW, performs Layer 2 encapsulation, and forwards the packet to PE2.
6. PE2 receives the packet from PE1 and decapsulates the packet to remove Layer 2 encapsulation information and two MPLS labels.
7. PE2 sends the original Layer 2 packet that is decapsulated from CE1 to CE2. The packet contains the U-tag and the replaced P-tag.

The processing of sending a packet from CE2 to CE1 is similar to this process.

VPLS Access Modes

- VLANIF interface in switching or routing mode

There are two types of VLANIF interfaces:

- A VLANIF interface in routing mode is multiplexed from a physical interface. For example, a GE interface can be divided into multiple sub-interfaces, with each sub-interface acting as a VLANIF interface in routing mode.
- A VLANIF interface in switching mode is a logical interface, but not the sub-interface of a physical interface. A VLANIF interface in switching mode can

contain multiple physical interfaces and receive VLAN packets from these physical interfaces.

A physical interface in a VLANIF interface in switching mode can send VLAN packets in the following modes:

- Access mode: allows only VLAN packets with the default VLAN ID to pass through.
- Trunk mode: allows only VLAN packets with the VLAN ID of the local VLANIF interface to pass through.
- CE-to-PE access mode

A CE can access a PE in the following modes:

- Through an access port: An access port allows only default VLAN packets of this port to pass. The VLAN packets on this physical port are untagged.
- You can assign multiple access ports of the PE to a VLAN for user access.
- You can assign multiple access ports of the PE to a VLAN for user access. Through a trunk port: A trunk port allows the packets of multiple VLANs to pass. Packets of the default VLAN (one of these VLANs) are untagged, whereas packets of other VLANs are tagged. You can connect the trunk port of the PE to the Ethernet switch to allow the access of multiple VLAN users.

Derivative VPLS Functions

Traffic statistics collection

Traffic statistics can be collected based on VSIs or VSI peers, and the status of various types of traffic can be viewed in real time.

VPLS service isolation

Users of different services can be isolated using different VSIs. Users in the same VSI also need to be isolated. VPLS service isolation allows you to prevent communication between users who have the same service and are bound to the same VSI. For example, high-speed Internet (HSI) users bound to the same VSI cannot communicate with each other. For details, see section [VPLS Service Isolation](#).

Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.10.2.2 VPLS Functions

Background

A characteristic of the Ethernet is that a port sends unicast packets with unknown destination MAC addresses, broadcast packets, and multicast packets to all other ports on the Ethernet. As an Ethernet-based technology, VPLS emulates an Ethernet bridge for user networks. To forward packets on a VPLS network, PEs must establish MAC address tables and forward packets based on MAC addresses or MAC addresses and VLAN tags.

Related Concepts

- MAC address learning

[Table 1](#) describes MAC address learning modes.

Table 1 MAC address learning modes

MAC Address Learning Mode	Description	Characteristic
Qualified	A PE learns the MAC addresses and VLAN tags of received Ethernet frames. In this mode, each user VLAN is an independent broadcast domain and has an independent MAC address space.	The broadcast domain is confined to each user VLAN. Qualified learning can result in large forwarding information base (FIB) table sizes, because the logical MAC address is now a VLAN tag + MAC address.
Unqualified	A PE learns only the MAC addresses of Ethernet frames. In this mode, all user VLANs share the same broadcast domain and MAC address space. The MAC address of each user VLAN must be unique.	If an AC interface is associated with multiple user VLANs, this AC interface must be a physical interface bound to a unique VSI.

- MAC address aging

An aging mechanism removes MAC entries that a PE no longer needs. If a MAC entry is not updated within a specified period of time, this entry will be aged.

Implementation

PEs establish MAC address tables based on dynamic MAC address learning and associates destination MAC addresses with PWs. [Table 2](#) describes the MAC address learning process.

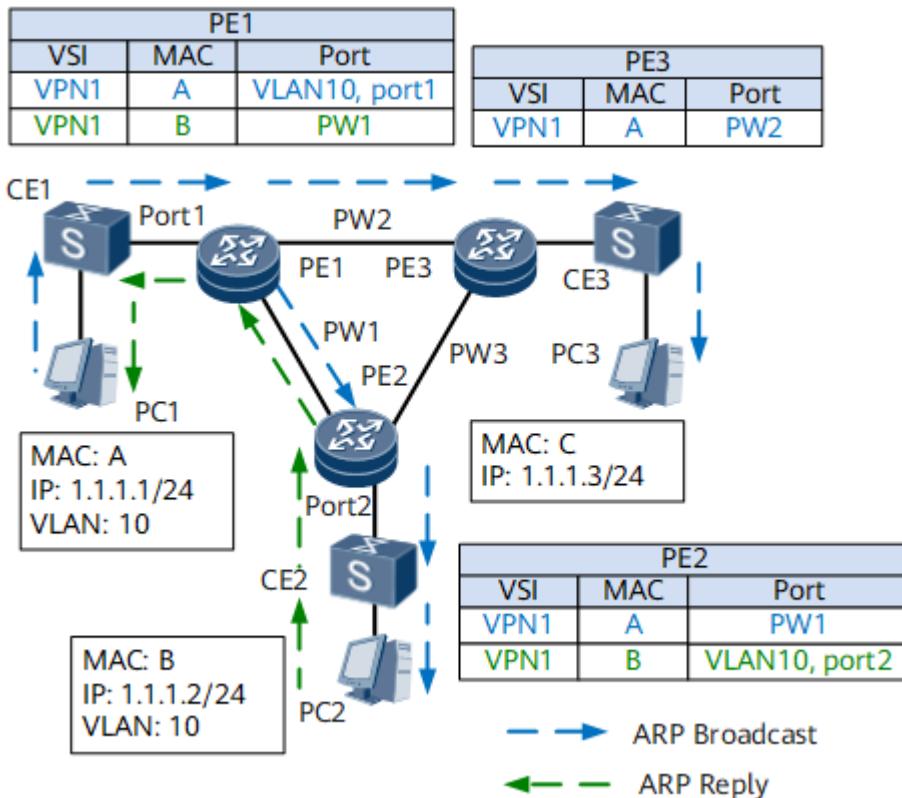
Table 2 MAC address learning process

MAC Address Learning Process	Description
Learning MAC addresses from user-side packets	After receiving packets from a CE, a PE maps their source MAC addresses to AC interfaces. Figure 1 shows a mapping example with Port1.
Learning MAC addresses from PW-side packets	A PW consists of a pair of MPLS VCs transmitting in opposite directions. A PW will go Up only after the two MPLS VCs are established. After a PE receives a packet with an unknown source MAC address from a PW, the PE maps the source MAC address to the PW receiving the packet.

Unqualified MAC address learning is similar to qualified MAC address learning. The major difference is that unqualified MAC address learning is based on the key set of VSI IDs and MAC addresses whereas qualified MAC address learning is based on the key set of VSI IDs, MAC addresses, and VLAN IDs.

[Figure 1](#) shows the process of MAC address learning and flooding on a PE. PC1 and PC2 both belong to VLAN10. When PC1 pings IP address 1.1.1.2, PC1 does not know the MAC address corresponding to this IP address and advertises an ARP Request packet. The following uses the unqualified mode as an example to describe the specific MAC address learning process.

Figure 1 MAC address learning process



1. After receiving the ARP Request packet sent by PC1 from Port1 that connects to CE1, PE1 adds the MAC address of PC1 to its own MAC address table, as shown in the blue section of the MAC entry.
2. PE1 advertises the ARP Request packet to its other ports (PW1 and PW2 can be viewed as ports).
3. After receiving the ARP Request packet from PW1, PE2 adds the MAC address of PC1 to its own MAC address table, as shown in the blue section of the MAC entry.
4. Based on split horizon, PE2 sends the ARP Request packet to only the port connecting to CE2 (as indicated by the blue dashed line), but not to PW1. This ensures that only PC2 receives the ARP Request packet. VPLS split horizon ensures that packets received from public network PWs are forwarded to only private networks, not to other public network PWs.
5. After PC2 receives the ARP Request packet and finds that it is the destination of this packet, PC2 sends an ARP Reply packet to PC1 (as indicated by the green dashed line).
6. After receiving the ARP Reply packet from PC2, PE2 adds the MAC address of PC2 to its own MAC address table, as shown in the green section of the MAC entry. The destination MAC address of the ARP Reply packet is the MAC address of PC1 (MAC A). After searching its MAC address table, PE2 sends the ARP Reply packet to PE1 over PW1.
7. After receiving the ARP Reply packet from PE2, PE1 adds the MAC address of PC2 to its own MAC address table, as shown in the green section of the MAC entry. After searching its MAC address table, PE1 sends the ARP Reply packet to PC1 through Port1.
8. After receiving the ARP Reply packet from PC2, PC1 completes MAC address learning.
9. While advertising the ARP Request packet to PW1, PE1 also advertises the ARP Request packet to PE3 over PW2. After receiving the ARP Request packet from PW2, PE3 adds the MAC address of PC1 to its own MAC address table, as shown in the blue section of the

MAC entry. Based on split horizon, PE3 sends the ARP Request packet to only PC3. Because PC3 is not the destination of the ARP Request packet, PC3 does not send any ARP Reply packet.

Derivative Functions

Traffic Restriction

On a VPLS network, you can limit the rates of broadcast, multicast, and unknown unicast packets to:

- Enhance traffic management and appropriately allocate user bandwidth.
- Prevent traffic attacks and enhance network security.

Processing of Unknown Packets

After receiving a packet, if a VSI cannot find a MAC entry that matches the destination address of the packet in its MAC address table, the packet is considered an unknown packet.

Unknown packets can be unknown unicast or multicast packets. Unknown packets are dropped, locally processed, or broadcast based on network security requirements. Similar to Ethernet, a VPLS network broadcasts unknown packets by default.

In broadcast mode, a VPLS network processes unknown packets in the following ways:

- After receiving an unknown packet from an AC interface in a VSI, a PE floods the packet to all its AC interfaces that connect to local CEs and remote PEs in the VSI.
- After receiving an unknown packet from a PW in a VSI, a PE floods the packet to all its AC interfaces that connect to local CEs, but not to remote PEs.

PEs can be configured to learn the MAC addresses of unknown unicast packets when dropping these packets. This function prevents the access of unauthorized users and enables PEs to identify the sources of unknown unicast packets.

Limit on the Number of Learned MAC Addresses

After the number of MAC entries or MAC address learning time reaches the set threshold, a device forwards or drops newly received packets and decides whether to report an alarm to the network management system (NMS).

This function applies to networks with relatively fixed users but insufficient security, such as residential access networks and enterprise intranets without security management.

Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.3 LDP VPLS

Background

LDP VPLS uses a static discovery mechanism to discover VPLS members using LDP signaling. VPLS information is carried in extended TLV fields (type 128 and type 129 FEC TLVs) of LDP signaling packets. Here, FEC stands for forwarding equivalence class. During the establishment of a PW, the label distribution mode is downstream unsolicited (DU) and the label retention mode is liberal.

Related Concepts

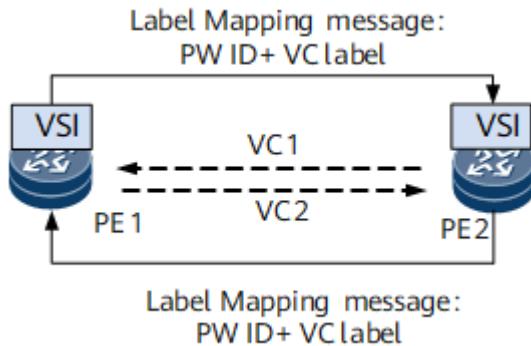
LDP VPLS involves the following concepts:

- FEC: A set of packets with similar or identical characteristics and forwarded in the same way by LSRs. Characteristics determining the FEC of a packet include the destination address, service type, and QoS attribute.
- TLV: A highly efficient and expandable coding mode for protocol packets. To support new features, you only need to add new types of TLVs to carry information required by the features.
- DU: A label distribution mode in which a label switching router (LSR) distributes labels to FECs without having to receive Label Request messages from its upstream LSR.
- Liberal: A label retention mode in which an LSR retains the label mapping received from a neighboring LSR, regardless of whether the neighboring LSR is its next hop. In liberal label retention mode, an LSR can use the labels sent from neighboring LSRs that are not at the next hop to re-establish an LSP. This mode requires more memory and label space than the conservative mode.

Implementation Process

[Figure 1](#) shows the process of establishing a PW using LDP signaling.

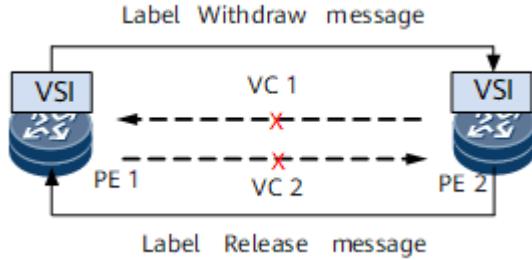
Figure 1 Establishing a PW using LDP signaling



- After PE1 is associated with a VSI, and PE2 is configured as a peer of PE1, PE1 sends a Label Mapping message to PE2 in DU mode if an LDP session exists between PE1 and PE2. The Label Mapping message carries information required to establish a PW, such as the PW ID, VC label, and interface parameters.
- Upon receipt, PE2 checks whether itself has been associated with the VSI. If PE2 has been associated with the VSI and PW parameters on PE1 and PE2 are consistent, PE1 and PE2 belong to the same VSI. In this case, PE2 establishes a unidirectional VC named VC1 immediately after PE2 receives the Label Mapping message. Meanwhile, PE2 sends a Label Mapping message to PE1. After receiving the message, PE1 takes a sequence of actions similar to those taken by PE2 and establishes VC2.

[Figure 2](#) shows the process of tearing down a PW using LDP signaling.

Figure 2 Tearing down a PW using LDP signaling



- After the peer configuration about PE2 is deleted from PE1, PE1 sends a Label Withdraw message to PE2. After receiving the Label Withdraw message, PE2 withdraws its local VC label, tears down VC1, and sends a Label Release message to PE1.
- After receiving the Label Release message, PE1 withdraws its local VC label and tears down VC2.

Derivative Functions

MAC Withdraw Loop Detection

On a dual-homing VPLS or hierarchical VPLS (HVPLS) network, data packets and MAC Withdraw messages can be forwarded between hub and spoke PWs and between spoke PWs. If hub or spoke PWs are not configured correctly, a data packet or MAC Withdraw message loop may occur.

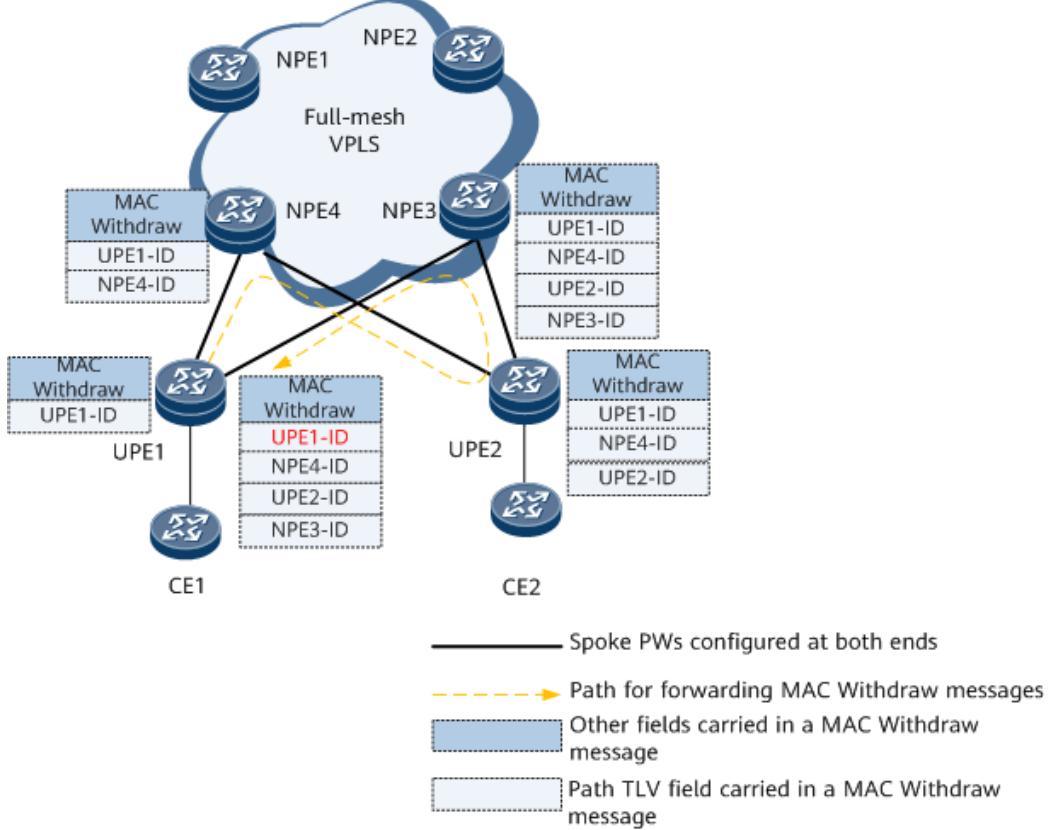
Techniques, such as Virtual Router Redundancy Protocol (VRRP), Spanning Tree Protocol (STP), and MAC flapping, can be used to prevent a data packet loop. MAC Withdraw loop detection prevents a MAC Withdraw message loop. MAC Withdraw loop detection enables a PE to add the Path TLV field to a MAC Withdraw message before the message is forwarded. The Path TLV field records the forwarding path of the message. [Figure 3](#) shows MAC Withdraw loop detection. The rules for a PE to forward a MAC Withdraw message are as follows:

- When a PE forwards a MAC Withdraw message, the PE adds its own LSR ID to the Path TLV field carried in the message.
- After the PE receives the message, it checks whether the message includes its own LSR ID and whether the number of LSR IDs carried in the Path TLV field exceeds 255. If the message includes its own LSR ID or the number of LSR IDs carried in the Path TLV field exceeds 255, the PE immediately discards the message.

NOTE

After you configure MAC Withdraw loop detection on a PE, the PE adds the Path TLV field to a MAC Withdraw message before forwarding the message. If a PE is not configured with MAC Withdraw loop detection, the PE directly forwards the MAC Withdraw message that it receives.

Figure 3 MAC Withdraw loop detection



Receiving of Group Messages by PWs

The IETF defines the usage scenario of this function. If multiple PWs, belonging to the same group and having the same status, are configured on a physical interface, group messages can be used to notify PWs of the interface status change when the physical interface goes Up or Down, reducing the number of Notification messages required.

PW Reliability

LDP VPLS ensures PW reliability using the following mechanisms:

- Association between VPLS and management Virtual Router Redundancy Protocol (mVRRP): In the IP radio access network (RAN) solution, after a cell site gateway (CSG) obtains the active/standby device status information from an mVRRP group, it determines the primary/secondary status of PWs itself.
- Manual configuration: LDP VPLS allows the primary/secondary status of PWs to be configured manually.

Usage Scenario

The LDP mode applies to VPLS networks that do not have many sites, do not span multiple ASs, or with PEs that do not run BGP.

Benefits

LDP VPLS offers the following benefits:

- Easy configuration
- Label resource saving

1.10.2.4 BGP VPLS

Background

BGP VPLS uses a dynamic discovery mechanism to discover VPLS members using BGP signaling. BGP VPLS uses MP-BGP Update packets to transmit VPLS member information. In an MP-BGP Update packet, the MP-REACH and MP-UNREACH attributes carry VPLS label information, and the extended community attributes carry interface parameters, RDs, and VPN targets. The RDs and VPN targets are used to identify VPN member relationships.

Related Concepts

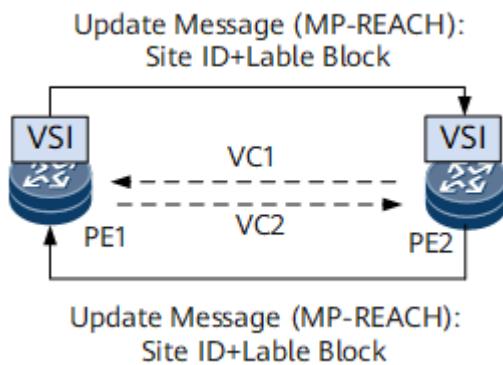
BGP VPLS involves the following concepts:

- MP-BGP: A multi-protocol extension of BGP-4. MP-BGP supports multiple network-layer protocols and identifies protocols based on address families. MP-BGP transmits VPN member information and VPN-IPv4 routes between PEs. Compared with conventional BGP, MP-BGP introduces the following path attributes:
 - MP_REACH: advertises reachable routes and their next hops.
 - MP_UNREACH: instructs a peer to delete unreachable routes.

Implementation

[Figure 1](#) shows the process of establishing a PW using BGP signaling.

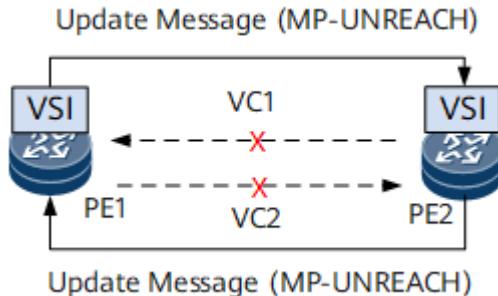
Figure 1 Establishing a PW using BGP signaling



1. PE1 is configured with a local label block. If a BGP session has been established between PE1 and PE2, PE1 sends an MP-BGP Update packet carrying the MP-REACH attribute, site ID, and label block information to PE2.
2. Upon receipt of the MP-BGP Update packet, PE2 calculates a unique label as its local VC label based on its own site ID and the label block carried in the packet. Then PE2 establishes a unidirectional VC named VC1. PE2 also calculates the local VC label of PE1 based on the site ID carried in the packet and its local label block and sends an MP-BGP Update packet to PE1. Upon receipt of the MP-BGP Update packet, PE1 takes similar actions as PE2 does and establishes a unidirectional VC named VC2.

[Figure 2](#) shows the process of tearing down a PW using BGP signaling.

Figure 2 Tearing down a PW using BGP signaling



1. After the local label block is deleted from PE1, PE1 withdraws its local VC label and tears down VC2. Additionally, PE1 sends an MP-BGP Update packet carrying the MP-UNREACH attribute to PE2.
2. Upon receipt of the MP-BGP Update packet, PE2 withdraws its local VC label and tears down VC1.

Derivative Functions

Ignoring MTU Match Check Results

By default, a VSI MTU check is performed on both ends of a VC in a VSI. If VSI MTUs on the two ends do not match, the VC cannot go Up.

When Huawei devices interwork with non-Huawei devices that do not support the VSI MTU check, configure the Huawei devices to ignore VSI MTU check results and encapsulate VPLS packets in the standard type 19 format.

RRs Not Filtering Received VPN Routes or Label Blocks Based on VPN Targets

Full-mesh connections need to be established between IBGP peers to ensure interconnectivity between them. If there are n BGP speakers in an AS, $n(n-1)/2$ IBGP connections need to be established. Establishing full-mesh connections for a large number of IBGP peers consumes lots of network resources.

In an AS, one router serves as the RR, although; whereas the other routers serve as clients. The clients establish IBGP connections with the RR. The RR and its clients form a cluster. The RR reflects routes among the clients, eliminating the need for the clients to establish BGP connections with each other. In an AS with n routers, if one router serves as the RR and others serve as clients, only $(n-1)$ IBGP connections need to be established. This implementation greatly reduces the consumption of network and CPU resources.

After an RR is configured, VPNs or VPN targets no longer need to be configured on RR, eliminating the need for the RR to save VPN routes or label blocks. In this situation, VPN-target-based filtering must be disabled for received VPN routes or label blocks.

Support for Intercommunication Between the L2VPN-AD Address Family and the BGP VPLS Address Family on the Peer Device

Both the BGP VPLS address family and L2VPN-AD address family use the standard 25/65 address family. When the L2VPN-AD address family receives the encapsulation type of 4 or 5 from the peer end, it determines the BGP VLL if the TLV sent by the peer end carries the CSV flag (identifying the PW status) and BGP VPLS if the TLV does not carry the CSV flag.

Usage Scenario

BGP VPLS applies to the core layers of large-scale networks with BGP-running PEs.

Benefits

BGP VPLS offers the following benefits:

- Provides a dynamic discovery mechanism to discover VPN members, thereby simplifying user operations
- Improves network expansibility by using RRs to reduce the number of BGP connections

Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.5 HVPLS

Definition

In a VPLS solution, all PEs that provide VPLS services must be fully meshed using label switched paths (LSPs). $N \times (N - 1)/2$ PWs must be established using signaling protocols between PEs for each VPLS service. The preceding solution cannot be applied on a large scale, because the PEs that provide virtual circuits (VCs) must copy packets, and each provider edge (PE) must broadcast the first unicast, broadcast and multicast packets to all the peers. This is a waste of bandwidth. In this situation, you can use hierarchical connections to reduce the burden of signaling protocols and packet replication and apply VPLS on a large scale.

The core of hierarchical virtual private LAN service (HVPLS) is to hierarchize the network. The network of each level is fully meshed. Devices of different levels are connected using PWs and forward data to each other without complying with the split horizon principle.

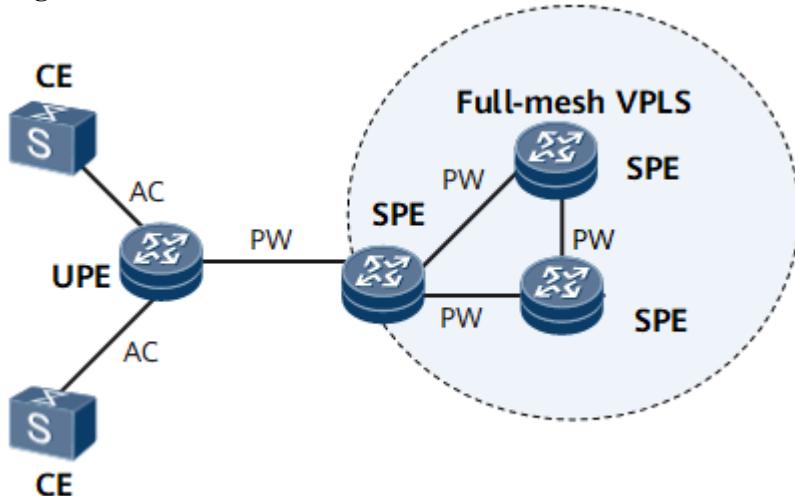
Purpose

HVPLS is introduced to cope with the problem of VPLS full mesh and enhance the expansibility of VPLS networks.

Principles

[Figure 1](#) shows a basic HVPLS model.

Figure 1 HVPLS model



In a basic HVPLS model, PEs can be classified into the following types:

- User-end PE (UPE): A customer convergence device that directly connects to CEs. A UPE needs to be connected to only one PE on a full-mesh VPLS network. A UPE supports routing and MPLS encapsulation. If a UPE connects to multiple CEs and possesses the basic bridge function, frame forwarding is performed only on the UPE. This implementation reduces the burden on SPEs.
- SPE: A device that connects to UPEs and is located in the core of a full-mesh VPLS network. An SPE connects to all the devices on a full-mesh VPLS network.

From the perspective of an SPE, a UPE functions like a CE. In data forwarding, an SPE uses the PW established between itself and a UPE as an AC. The UPE adds double MPLS labels to packets sent by CEs. The outer label is an LSP label that is switched when a packet passes through devices on the access network. The inner label is a VC label that identifies a VC. The inner label remains unchanged when a packet is transmitted along an LSP. After receiving double-tagged packets, an SPE directly removes the outer label, a statically configured public network label and determines the VSI which the AC accesses based on the inner label.

Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.6 BGP AD VPLS

Definition

BGP AD VPLS, short for Border Gateway Protocol Auto-Discovery virtual private LAN service, is a new technology for automatically deploying VPLS services.

BGP AD VPLS-enabled devices exchange extended BGP Update packets to automatically discover BGP peers in a VPLS domain. After BGP peer relationships are established, these devices use LDP FEC 129 to negotiate and establish VPLS PWs. In addition, BGP AD HVPLS is deployed with split horizon disabled. This allows all BGP peers in an AS to function as UPEs on an HVPLS network.

Purpose

The wide use of VPLS technologies leads to the growing scale of VPLS networks and configurations. BGP AD VPLS is introduced to simplify configurations, enable automatic service deployment, and reduce operating expense (OPEX).

BGP AD VPLS has the advantages of both BGP and LDP VPLS. BGP AD VPLS-enabled devices exchange extended BGP Update packets to automatically discover BGP peers in a VPLS domain. After BGP peer relationships are established, these devices use LDP FEC 129 to negotiate and establish VPLS PWs. VPLS services are automatically deployed after PWs are established.

Automatic VPLS member discovery and PW establishment simplify the configurations required by VPLS networks, enable automatic service deployment, and reduce OPEX for carriers.

Concepts

Acronym and Abbreviation	Full Name	Description

Acronym and Abbreviation	Full Name	Description
VPLS ID	virtual private LAN service ID	Identifier of a VPLS domain
VSI ID	virtual switching instance ID	Identifier of a VSI in a VPLS domain
RD	route distinguisher	Route distinguisher in a BGP packet which carries VSI information
RT	route target	Route attribute carried in a BGP packet used to advertise VSI information
AGI	attachment group identifier	Domain identifier used during PW negotiation between PEs in a VPLS domain
AII	attachment individual identifier	VSI identifier used during PW negotiation between PEs in a VPLS domain
SAII	source attachment individual identifier	Local IP address used by BGP AD VPLS to negotiate PW establishment
TAII	target attachment individual identifier	Remote IP address used by BGP AD VPLS to negotiate PW establishment
FEC 129	forwarding equivalence class 129	New type of FEC used by LDP signaling

Principles

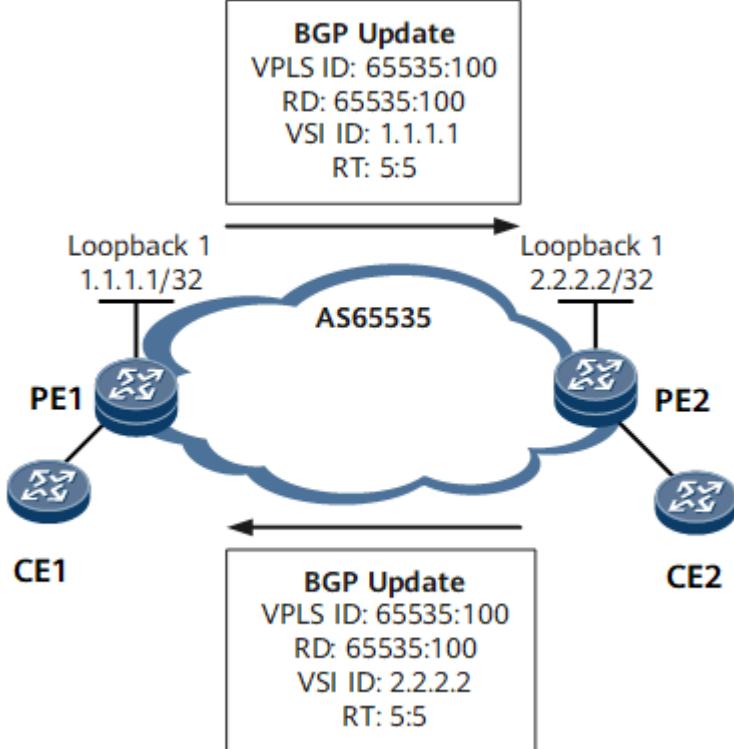
BGP AD VPLS has the advantages of both BGP and LDP VPLS. BGP AD VPLS automatically discovers VPLS BGP peers, simplifying the configurations and saving labels.

BGP AD VPLS-enabled devices exchange extended BGP Update packets carrying VSI member information to automatically discover BGP peers in a VPLS domain. After BGP peer relationships are established, these devices use LDP Mapping (FEC 129) messages to negotiate and establish VPLS PWs. VPLS services are automatically deployed after PWs are established.

Automatically Discovering PEs in a VPLS Domain

Automatically discovering PEs in a VPLS domain is the first phase of VPLS service deployment. BGP is used to automatically discover PEs in a VPLS domain. [Figure 1](#) shows the process and information used for automatically discovering PEs in a VPLS domain.

Figure 1 Networking diagram for automatically discovering PEs in a VPLS domain



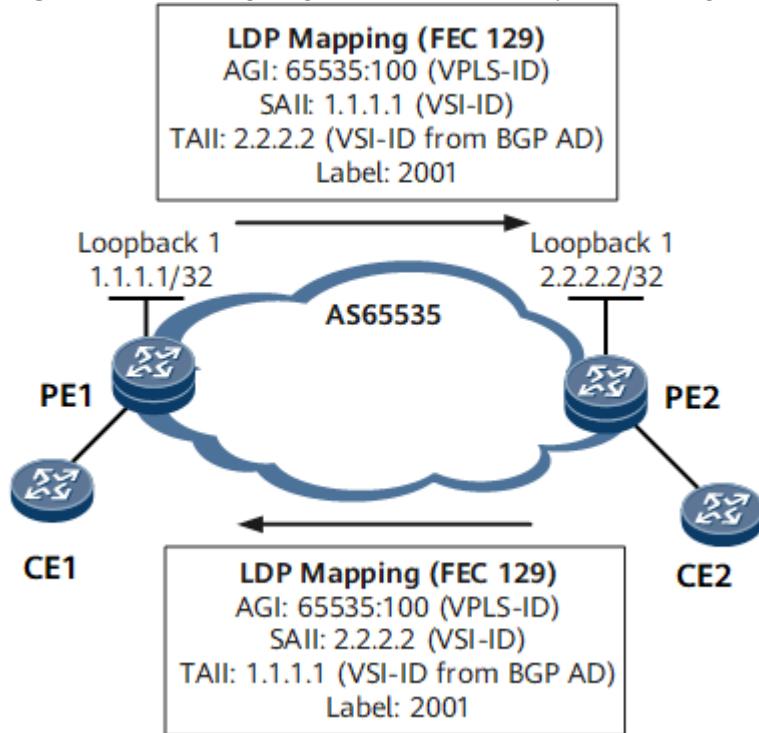
The process of automatically discovering PEs in a VPLS domain is as follows:

1. After the information, such as the VPLS ID, RD, RT, and VSI ID, is set on PE1 and PE2, the two PEs encapsulate the information into BGP Update messages and send these messages as BGP AD packets to all peer PEs in the BGP domain.
2. After receiving a BGP AD packet, a PE checks whether the BGP AD packet matches its RT policy. If they match, the PE obtains the information carried in the packet and compares the obtained information with local configurations:
 - If the VPLS IDs of VSIs on the two PEs are the same, the two VSIs are in the same VPLS domain and a PW can be established between them.
 - If the VPLS IDs of VSIs on the two PEs are different, the two VSIs are in different VPLS domains and no PW can be established between them.

Automatically Establishing a PW

After a PE discovers a remote PE in the same VPLS domain, the two PEs use LDP Mapping (FEC 129) messages to negotiate PW establishment. [Figure 2](#) shows the process of automatically establishing a PW.

Figure 2 Networking diagram for automatically establishing a PW



The process of automatically establishing a PW is as follows:

1. Two PEs connected by an LDP session in a VPLS domain exchange LDP Mapping (FEC 129) messages that carry VSI information, including the AGI, SAI, TAII, and label information. If no LDP session has been established between two PEs in a VPLS domain, the two PEs initiate negotiation upon the creation of an LDP session.

NOTE

After BGP AD VPLS members are discovered, BGP AD VPLS proactively triggers LDP to establish LDP sessions, facilitating the establishment of PWs for VPLS services. When a VPLS service is deleted, BGP AD VPLS proactively triggers LDP to delete the corresponding LDP sessions. This implementation simplifies LDP session maintenance, improves system resource usage, and optimizes network performance.

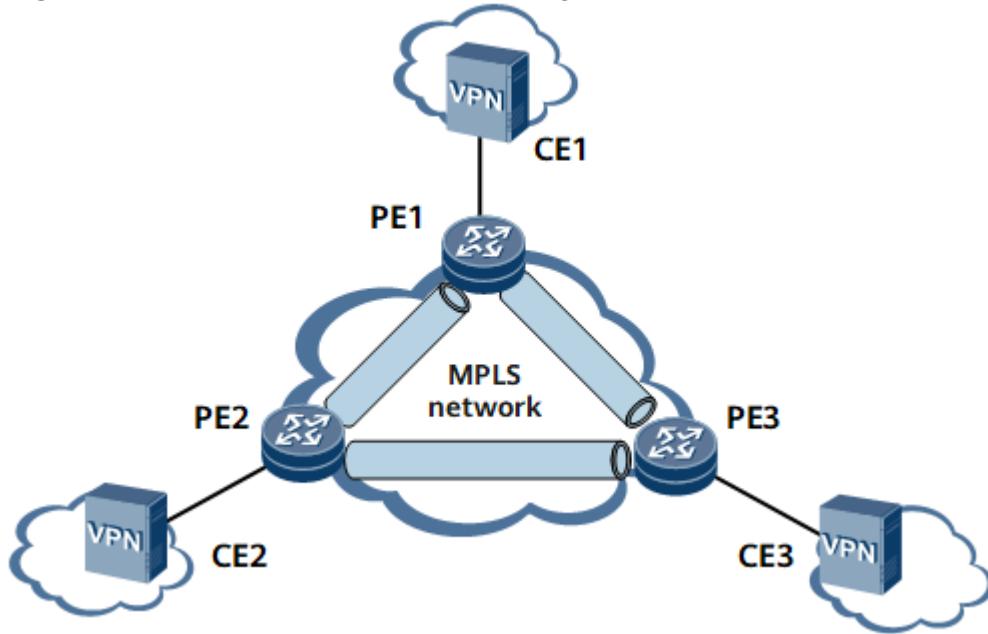
2. After a PE receives an LDP Mapping message, the PE obtains VSI information, such as the VPLS ID, PW type, MTU, and TAII, carried in the message and compares the obtained VSI information with local VSI information. If the information matches, the PE establishes a PW to the remote PE.

Application of BGP AD VPLS on a Full-Mesh Network

On the network shown in [Figure 3](#):

- BGP peer relationships are established between PE1, PE2, and PE3.
- BGP AD VPLS is configured on PE1 and PE2 in a VPLS domain.
- PE3 is assigned the same VPLS ID as that on PE1 and PE2, which allows PE3 to join the VPLS domain. (PE3 is to be added to the VPLS domain as the network expands.)
- BGP AD VPLS is enabled on PE3, allowing PWs to be automatically established between PE3 and PE1 and between PE3 and PE2.

Figure 3 Full-mesh BGP AD VPLS networking



Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.7 Inter-AS VPLS

Definition

Inter-AS VPLS refers to VPLS applications across more than one AS. Inter-AS VPLS Option A, Option B, and Option C are supported.

NOTE

Except the learning and forwarding functions of VSIs, the fundamentals and implementation of PW establishment in inter-AS VPLS are similar to those in inter-AS L2VPN.

Purpose

Inter-AS VPLS enables VPLS users in different ASs to communicate.

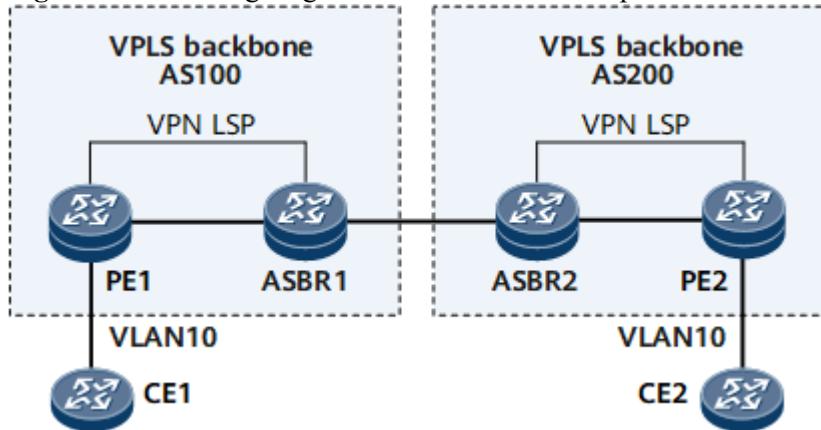
Implementation of Inter-AS VPLS Option A

On the network shown in [Figure 1](#), the implementation of inter-AS VPLS Option A is as follows:

- An IGP is configured on the backbone network to allow communications between devices in the same AS.
- MPLS is enabled on the backbone network and a dynamic LSP is established between the PE and ASBR in the same AS.
- An IBGP peer relationship and a VPN LSP are established between the PE and ASBR in the same AS.

- VSIs are configured on the PE and ASBR in the same AS, and the AC interfaces of the devices are bound to these VSIs. On an ASBR in an AS, configure its peer ASBR in the other AS as if it was a CE.

Figure 1 Networking diagram for inter-AS VPLS Option A



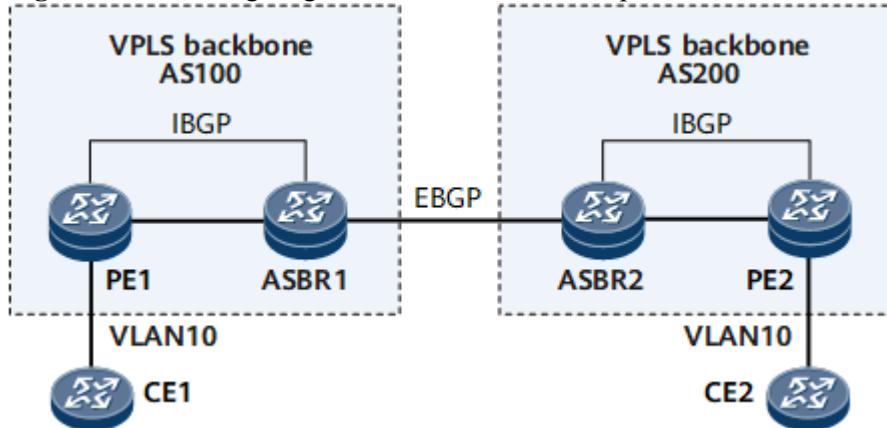
Implementation of Inter-AS VPLS Option B

According to the BGP VPLS standards, in Option B, when L2VPN A-D routes are transmitted between ASs, a new label block is allocated, and the mapping between the new label block and the original label block is created.

On the network shown in [Figure 2](#), the implementation of inter-AS VPLS Option B is as follows:

- Establish an EBGP peer relationship between ASBRs, configure an L2VPN-AD address family, and trigger the establishment of a local IFNET tunnel.
- Establish an IBGP peer relationship between a PE and an ASBR in the same AS and configure VPN LSPs.
- Configure BGP VPLS on the PEs on both ends, configure a VSI on a PE and an ASBR in the same AS, and bind the AC interface to the VSI of each device.

Figure 2 Networking diagram for inter-AS VPLS Option B



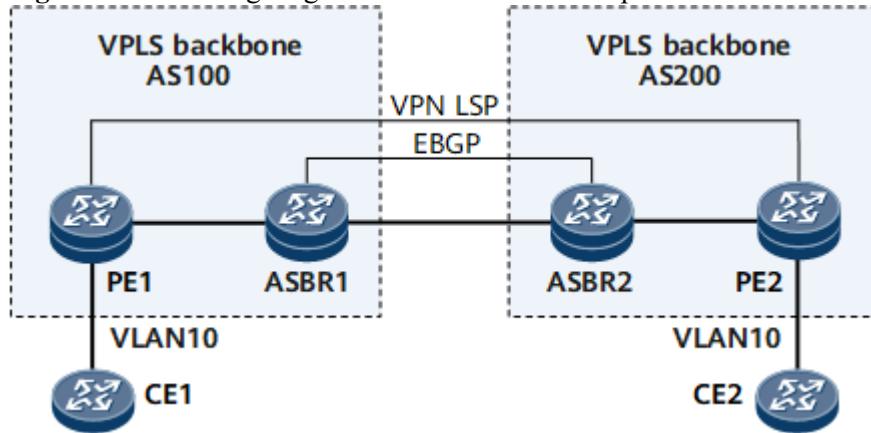
Implementation of Inter-AS VPLS Option C

On the network shown in [Figure 3](#), the implementation of inter-AS VPLS Option C is as follows:

- Configure an IGP on the backbone network to allow communications between devices within the same AS.

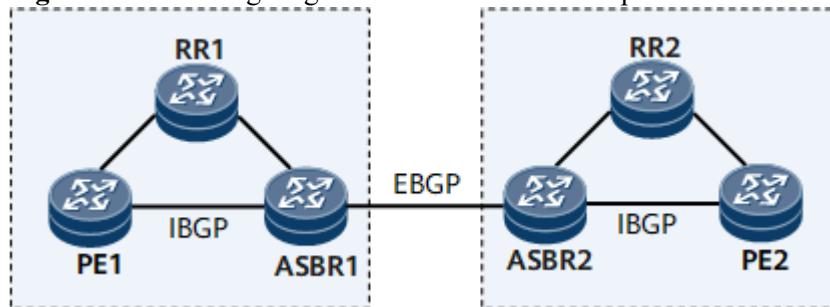
- Enable MPLS on the backbone network and establish a dynamic LSP between the PE and ASBR within the same AS.
- Establish an IBGP peer relationship between the PE and ASBR within the same AS and an EBGP peer relationship between ASBRs.
- Configure routing policies on ASBRs and enable the ASBRs to exchange labeled routes.
- Establish an MP-EBGP peer relationship between PE1 and PE2.
- Configure a VSI on each of PE1 and PE2 and bind the AC interface of each PE to a corresponding VSI.

Figure 3 Networking diagram for inter-AS VPLS Option C



To improve scalability, you can specify a route reflector (RR) in each AS. The RR stores all L2VPN routes and exchanges L2VPN routing information with the PE within an AS.

Figure 4 Networking diagram for inter-AS VPLS Option C with RRs



Application Scenarios of Inter-AS VPLS Options

- Inter-AS Option A: Its advantage is that configurations are easy. There is no need to run MPLS between ASBRs or perform particular configurations for inter-AS communications. The disadvantage of inter-AS Option A is that this mode has poor expansibility and poses high requirements for PEs. This mode is applicable to the early service deployment phase when the number of inter-AS VPNs is small.
- Inter-AS Option B: Its advantage is that ASBRs exchange information over routes, rather than along dedicated links. The disadvantage of inter-AS Option B is that label mappings need to be configured on ASBRs, and consequently, a great number of labels are consumed, leading to a label waste. In addition, an ASBR needs to establish an LSP with each PE in the same AS, which results in the shortage of label resources on the ASBR and easily leads to a performance bottleneck.
- Inter-AS Option C: Its advantage is that ASBRs only forward packets but do not maintain VPN routes, and intermediate devices need to support only MPLS forwarding. This mode

prevents the ASBRs from becoming performance bottlenecks. The disadvantage of this mode is that the management cost is high for maintaining an end-to-end BGP LSP. This mode is applicable when the number of inter-AS VPNs is large, many ASs are crossed, and services grow on a large scale.

Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.8 Flow-Label-based Load Balancing

Background

Packets of multiple data flows on the same L2VPN carry the same VC labels, which are encapsulated on a PE. When these packets reach a P device, they can be forwarded only over one path.

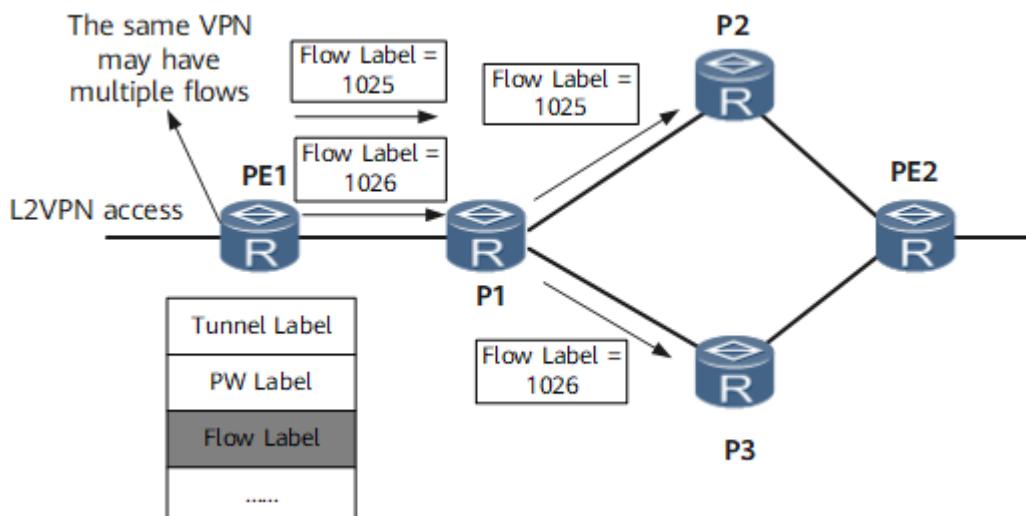
To load-balance different data flows, configure flow-label-based load balancing on the PE. After you have completed the configuration, the PE encapsulates a data packet and adds a flow label following the PW label. The P device load-balances different data flows based on flow labels.

Implementation

On the L2VPN shown in [Figure 1](#) where two data flows exist:

1. PE1 calculates flow labels based on the source and destination IP addresses of the two data flows. In this scenario, the flow labels are calculated as 1025 and 1026.
2. PE1 adds the flow labels following the PW labels of the packets in the two data flows.
3. When the two data flows reach P1, P1 performs a hash calculation based on the flow labels, and the two data flows are mapped onto different paths. In this example, the next hop of the data flow with the flow label 1025 is P2, and the next hop of the data flow with the flow label 1026 is P3.
4. When the two data flows reach PE2, the PW and flow labels are sequentially removed. PE2 then forwards the two data flows to their destination CEs based on their PW labels.

Figure 1 Flow-label-based load balancing



Usage Scenario

Flow-label-based load balancing applies to an L2VPN on which multiple links exist between P devices.

Benefits

Flow-label-based load balancing allows data flows on the same VPN to be load-balanced along different paths based on flow labels, improving resource usage.

Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.9 VPLS PW Redundancy

Background

A redundant provider edge (PE) is often deployed to enhance service reliability. If a virtual private wire service (VPWS) or virtual private LAN service (VPLS) network uses a redundant PE, two pseudo wires (PWs) have to be deployed for service protection. This mechanism is called PW redundancy.

PW redundancy is a technique so widely used that it has developed into a reliability standard. PW redundancy improves service switchover efficiency and minimizes impact of device faults on services.

PW redundancy is best suited for point-to-point services, such as VPWS. VPLS, a point-to-multipoint service, can be viewed as a collection of point-to-point services. Therefore, VPLS can also use PW redundancy.

In VPLS, PW redundancy can speed up VPLS network convergence to reduce service downtime.

Related Concepts

Some key concepts for VPLS PW redundancy are described by using service traffic protection between CE1 and CE2 on the VPLS network shown in [Figure 1](#) as an example.

Currently, VPLS PW redundancy can operate in either of the following modes (the operating mode is specified on PE1):

- Master/slave mode: PE1 determines whether a local PW is in the active or standby state based on the PW's preset forwarding priority.
- Independent mode: PE1 determines whether a local PW is in the active or standby state based on the master/backup status of PE2 and PE5.

The endpoint PEs of a PW protection group must negotiate PW status to ensure that they select the same PW to transmit packets. The following concepts are introduced for PW status negotiation inside a PW protection group:

- Primary/secondary: used to describe the forwarding priority of a PW and can be configured. A smaller value indicates a higher priority. A PW with the higher forwarding priority is the primary PW.

NOTE

The forwarding priorities take effect only if PE1 uses the master/slave mode as the PW redundancy mode. In master/slave mode, PE1 instructs PE2 and PE5 to change the forwarding status of PWs to be the same as that on PE1. In independent mode, the master/backup status of PE2 and PE5 determines the forwarding status of PWs on PE1.

- Active/standby: used to describe the forwarding status of a PW and cannot be configured. Only the active PW can be used to forward traffic. The standby PW may be used to receive traffic.
-

NOTE

In some documents, Huawei uses active/inactive or primary/backup to describe PW status. These terms have the same meaning as term active/standby defined in draft. They all indicate the PW forwarding status.

Implementation

To keep the original forwarding behavior, the endpoint PEs of a PW protection group must use the same PW to transmit traffic and ensure that only the PW used to transmit traffic is in the active state. To achieve this goal, the PW protection group must use a signaling mechanism.

Relevant standards specify the PW Status TLV to transmit the PW forwarding status. The PW Status TLV, a 32-bit status code field, is carried in a Label Mapping or LDP Notification message. PW redundancy uses a new PW status code, 0x00000020, to indicate that a PW is in the standby state.

NOTE

Only PWE3 VPLS supports PW redundancy.

The forwarding priorities must be specified for PWs in a PW protection group. The PW with the higher forwarding priority is preferentially selected as the active PW to forward traffic. The remaining PW stays in the standby state to protect the primary PW.

The forwarding status of a PW determines whether the PW is used to forward traffic. The PW forwarding status depends on:

- Local and remote PW signaling status: A PE monitors its local signaling status and uses PW redundancy signaling to obtain the remote signaling status from a remote PE.
- PW redundancy mode: The PW redundancy mode, which can be master/slave or independent, is specified on PE1.
- PW forwarding priority: The PW forwarding priority, which determines whether a PW is the primary or secondary PW, is specified on PE1.

On the network shown in [Figure 1](#), VPLS PW redundancy is configured on PE1. In normal cases, the local and remote PW signaling status on PE1 are both Up. The following describes how the endpoint PEs of a PW protection group choose the same PW to transmit traffic based on PW redundancy modes:

- In master/slave mode, PE1 determines the local PW forwarding status based on preset forwarding priorities and informs PE2 and PE5 of the PW forwarding status; PE2 and PE5 determine their PW forwarding status based on the received PW forwarding status.

- In independent mode, PE1 determines the local PW forwarding status based on the PW forwarding status learned from PE2 and PE5; PE2 and PE5 determine their PW forwarding status based on signaling, which can be enhanced trunk (E-Trunk) or Virtual Router Redundancy Protocol (VRRP) signaling, and notify PE1 of the forwarding status.

In master/slave or independent mode, if the primary PW becomes faulty, it becomes inactive and the corresponding secondary PW becomes active. PW-side faults do not affect the AC status. In independent mode, if an AC-side fault occurs, for example, a PE or AC is faulty, the PW forwarding status will change, because the status is determined by the master/backup status of the dual-homing devices. In master/slave mode, if an AC-side fault occurs, the PW forwarding status will not change, because the status is determined by preset forwarding priorities.

NOTE

VPLS PW redundancy is similar to VPWS PW redundancy. The difference lies in that a virtual switching instance (VSI) has multiple PWs destined for different PEs. These PWs may form multiple PW protection groups. The forwarding status changes of PWs in one group do not affect the forwarding status of PWs in other groups.

Currently, the VPLS PW redundancy mode can only be master/slave.

Derivative Function

In addition to providing real-time service protection against network faults, VPLS PW redundancy also allows you to manually switch traffic between PWs in a PW protection group during network operation and maintenance. For example, if you want to maintain a device that serves as the endpoint of a primary PW, you can switch traffic to the secondary PW before maintenance and switch traffic back to the primary PW after maintenance.

Usage Scenario

VPLS PW redundancy can be used on hierarchical virtual private LAN service (HVPLS) networks and virtual leased line (VLL) accessing VPLS networks. These two types of networks can bear many types of services. It is recommended that you determine which types of services to deploy based on the networking characteristics of the two types of networks if they are newly planned or deployed:

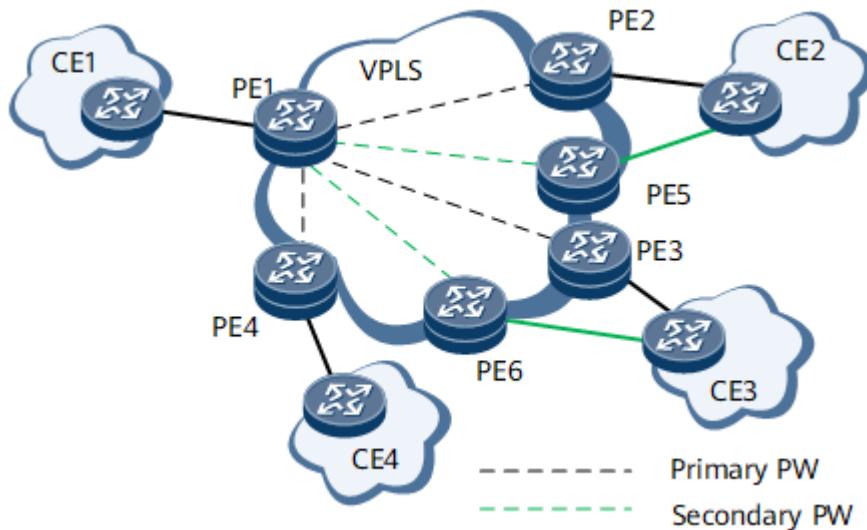
- HVPLS networks are best suited for multicast services, such as Internet Protocol television (IPTV) services, because HVPLS networks can save VPLS core network bandwidth. For more information, see [VPLS PW Redundancy for Protecting Multicast Services](#).
- VLL accessing VPLS networks are best suited for unicast services, such as high-speed internet (HSI) and Voice over Internet Protocol (VoIP) services, because VLL PEs do not need to learn user MAC addresses. For more information, see [VPLS PW Redundancy for Protecting Unicast Services](#).

VPLS PW redundancy can also be used to improve the reliability of existing networks. On the VPLS network shown in [Figure 1](#), CE1 communicates with CE2, CE3, and CE4 over PWs between PE1 and PE2, PE3, and PE4. As services develop, services between CE1 and CE2 and between CE1 and CE3 raise higher reliability requirements. To meet the reliability requirements, PE5 and PE6 are deployed on the VPLS network to provide VPLS PW redundancy protection for PE2 and PE3, respectively. In addition, multiple PW protection groups are configured on PE1. As a result, VPLS PW redundancy protects services against network-side, AC-side, and PE failures without affecting existing services, improving network reliability.

NOTE

VPLS PW redundancy can be configured for desired services without affecting services on other PWs, reducing costs and maximizing profits.

Figure 1 VPLS PW redundancy networking



Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.10 Multicast VPLS

Background

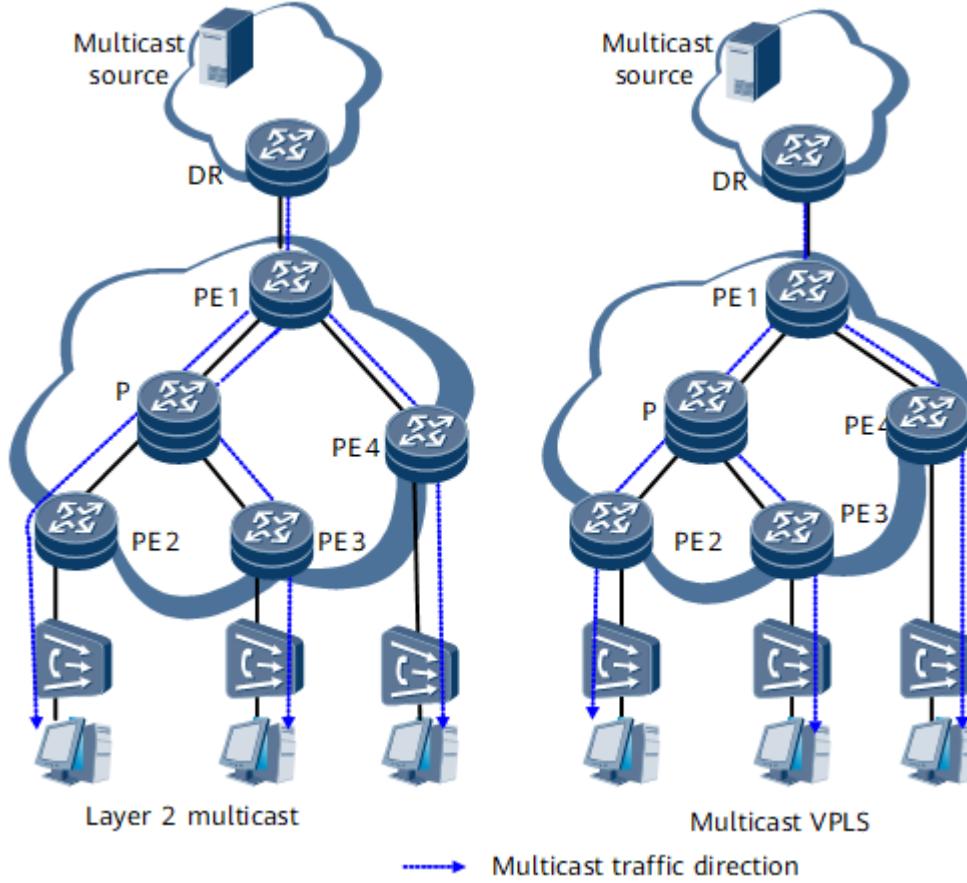
IP/MPLS backbone networks carry an increasing number of multicast services, such as IPTV, video conferences, and massively multiplayer online role-playing games (MMORPGs). These services require bandwidth assurance, QoS guarantee, and high network reliability. The following commonly used multicast solutions cannot meet the increasing requirements of multicast services and network carriers:

- IP multicast: An IP multicast network is complex to deploy and maintain. The network does not have QoS or TE capabilities and provides low reliability.
- Layer 2 multicast: A large Layer 2 multicast network must use the HVPLS technology and solve routing loop problems. The network is complex to deploy and its reliability scheme is hard to design.

To provide better multicast services, IETF proposed the multicast VPLS solution. On a multicast VPLS network, the ingress directly transmits multicast traffic to multiple egresses over a P2MP MPLS tunnel. This solution eliminates the need to deploy PIM and HVPLS on the transit nodes of tunnels, simplifying network deployment. In addition, multicast VPLS can utilize the advantages of MPLS in TE, QoS guarantee, and reliability assurance.

Multicast VPLS reduces redundant multicast traffic on the network by replicating multicast traffic on demand. [Figure 1](#) shows the differences between Layer 2 multicast and multicast VPLS in multicast traffic replication. On the Layer 2 multicast network, the multicast traffic is replicated into three copies right at the ingress PE1. On the multicast VPLS network, the multicast traffic is replicated on demand at each node. Compared with Layer 2 multicast, multicast VPLS reduces the burden of links.

Figure 1 Comparison of Layer 2 multicast and multicast VPLS in multicast traffic replication



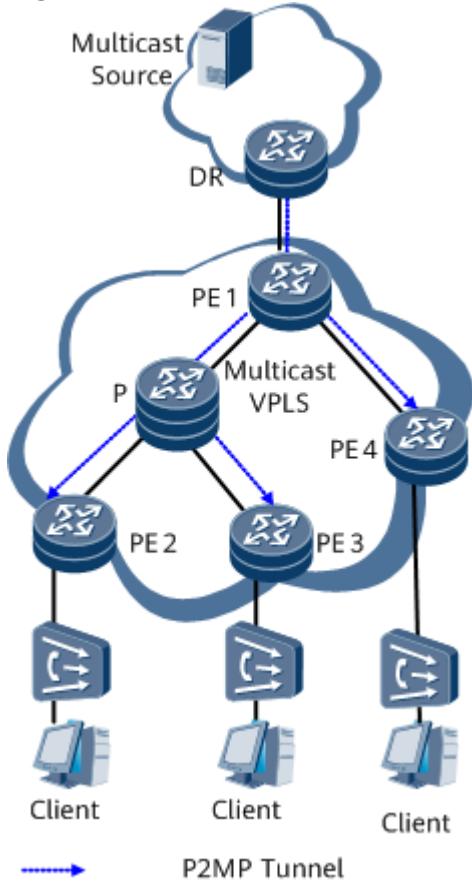
Related Concepts

[Table 1](#) describes some important concepts used in multicast VPLS.

Table 1 Important concepts used in multicast VPLS

Name	Description	Corresponding Device
Root node	Ingress of a P2MP tunnel.	PE1 in Figure 2
Branch node	A type of transit node. A branch node replicates each incoming packet and swaps the label in the incoming packet with another label before forwarding the packet to each leaf node.	P in Figure 2
Leaf node	Egress of a P2MP tunnel.	PE2, PE3, and PE4 in Figure 2

Figure 2 Multicast VPLS



Implementation

Tunnel establishment

On a multicast VPLS network, multicast traffic can be carried over either P2MP TE tunnels or P2MP mLDP tunnels. For the establishment of P2MP TE tunnels, see [P2MP TE](#). For the establishment of P2MP mLDP tunnels, see [mLDP](#). [Table 2](#) lists the differences between P2MP TE and P2MP mLDP tunnels.

Table 2 Differences between P2MP TE and P2MP mLDP tunnels

Compared Aspect	P2MP TE Tunnel	P2MP mLDP Tunnel
Usage scenario	Networks that require control over destination nodes	Networks that do not require control over destination nodes
Creation mode	The root node initiates LSP setup.	The leaf nodes initiate LSP setup.
Signaling	The P2MP tunnel is maintained by periodically sent signaling packets. If a large number of leaf nodes exist, network congestion is likely to occur.	Signaling packets do not need to be periodically sent, reducing network pressure.

The establishment process of the P2MP tunnel shown in [Figure 2](#) is described as follows:

- Tunnel type being P2MP TE

- After the tunnel type is configured as P2MP TE for the root node VSI, the VSI applies for tunnel FEC information from root node TE.
 - The root node VSI notifies root node TE of the IP addresses of all VSI peers (leaf nodes).
 - Root node TE sends TE signaling packets to all leaf nodes to trigger P2MP tunnel establishment.
 - Leaf node TE establishes the P2MP tunnel after receiving TE signaling packets from root node TE.
- Tunnel type being P2MP mLDP
 - After the tunnel type is configured as P2MP mLDP for the root node VSI, the VSI applies for tunnel FEC information from root node mLDP.
 - The root node VSI sends tunnel FEC information to all VSI peers (leaf nodes) using BGP AD or BGP Multi-homing signaling packets.
 - Leaf node VSIs parse BGP AD or BGP Multi-homing signaling packets and notify leaf node mLDP of tunnel FEC information.
 - Leaf node mLDP sends mLDP signaling packets to the root node to trigger P2MP tunnel establishment based on tunnel FEC information.
 - Root node mLDP establishes the P2MP tunnel after receiving mLDP signaling packets from leaf node mLDP.

After the P2MP tunnel is established, PE1 sends the multicast traffic received from the DR to each leaf node over the P2MP tunnel. The leaf nodes replicate the multicast traffic on demand before sending the traffic to multicast receivers.

Data forwarding

P2MP mLDP and P2MP TE tunnels have the same data forwarding process. A branch node replicates MPLS packets, swaps existing labels with outgoing labels in the MPLS packets, and sends the same MPLS packets over every sub-LSP. This process increases the efficiency of network bandwidth resource usage. [Figure 3](#) shows the data forwarding process.

Figure 3 Multicast VPLS

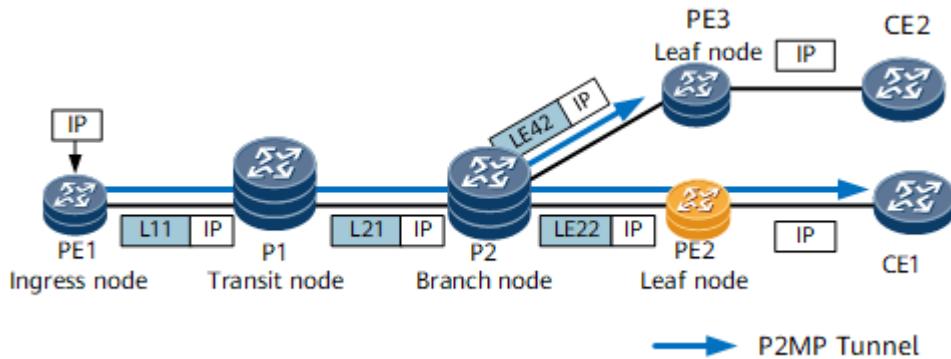


Table 3 P2MP TE data forwarding

Node	Forwarding Entry		Forwarding Behavior
	Incoming Label	Outgoing Label	

Node	Forwarding Entry		Forwarding Behavior
	Incoming Label	Outgoing Label	
PE1	N/A	L11	Pushes an outgoing label with the value of 11 into an IP multicast packet and forwards the packet to P1.
P1	L11	L21	Swaps the incoming label with an outgoing label with the value of 21 in an MPLS packet and forwards the packet to P2.
P2 (branch node)	L21	LE22	Replicates the IP multicast packet, swaps the incoming label with an outgoing label in each packet, and forwards each packet to a next hop through a specific outbound interface.
		LE42	
PE2	LE22	None	Removes the label from the packet so that this MPLS packet becomes an IP multicast packet.
PE3	LE42	None	Removes the label from the packet so that this MPLS packet becomes an IP multicast packet.

Benefits

Deploying multicast VPLS on an IP/MPLS backbone network offers the following benefits:

- Optimizes bandwidth usage.
- Provides bandwidth assurance for multicast services.
- Simplifies network deployment by eliminating the need to deploy multicast protocols, such as PIM, on the core nodes of the backbone network.

Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.11 VPLS Multi-homing

Background

To deliver high-reliability services over a VPLS network, carriers usually dual-home a CE to two PEs through redundant links. While providing link-level protection, the dual-homing mechanism also brings in the risk of routing loops. Currently, E-Trunk multi-homing or STP over VPLS is used to prevent routing loops. However, E-Trunk multi-homing is not widely supported by non-Huawei devices, and STP over VPLS introduces high routing costs during real-time loop detection. VPLS multi-homing adjusts link priorities to prevent routing loops. It ensures that one access link of a multi-homed CE is in the active state and the other access links are in the blocked state.

Related Concepts

VPLS multi-homing uses BGP to transmit multi-homing site information. It enhances standard BGP VPLS and improves reliability and applies to scenarios where a CE is multi-homed to PEs.

Table 1 VPLS multi-homing concepts

Concept	Description
MH-ID	A multi-homing ID that uniquely identifies a multi-homing site
Optimal site	Preferred multi-homing site used for PW establishment

Implementation

In VPLS multi-homing implementation, after a multi-homing site is configured on a PE accessed by a multi-homed CE, the PE advertises a BGP Update message carrying its multi-homing site information (including the MH-ID) to the other PEs in the VPLS domain. Upon receipt of the BGP Update message, PEs with the same MH-ID as that carried in the BGP Update message start an election process to determine which PE is the preferred one. After the preferred PE is elected, the access links between the multi-homed CE and the non-preferred PEs are blocked to prevent routing loops.

A multi-homed CE determines which PE to access based on link priorities. The link-priority-affecting factors are listed in descending order of influence as follows:

1. AC status (ACS): For PEs with the same MH-ID, the ACS value may be either of the following:
 - 1: The AC between a multi-homed CE and a PE is Down.
 - 0: The AC between a multi-homed CE and a PE is Up.

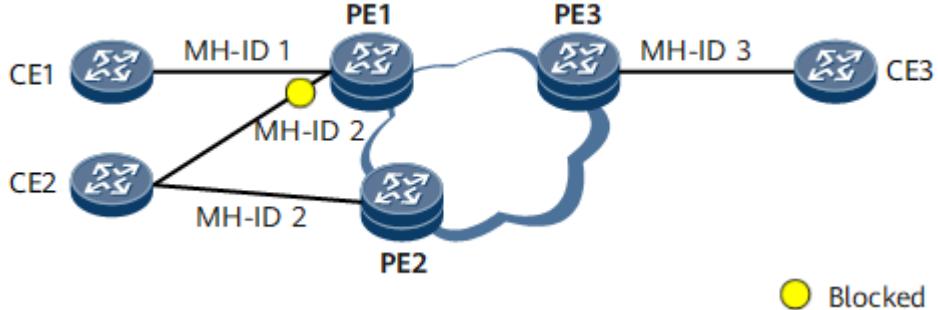
The link whose ACS is 0 has a higher priority than the link whose ACS is 1.

2. Preference (PREF): A larger PREF value indicates a higher link priority. The value must be configured.
3. PE's router ID (PE-ID): If no BGP router ID is configured, the system router ID is used by default. If a BGP router ID is configured, it is used as the PE-ID.

A lower PE-ID indicates a higher link priority.

If a link between a multi-homed CE and a PE fails, the PE advertises a BGP Update message that carries ACS, PREF, and PE-ID information to the other PEs in the VPLS domain. Then, PEs with the same MH-ID as that carried in the BGP Update message start an election process to select a preferred link.

Figure 1 VPLS multi-homing networking



A PE assigns a multi-homing site and MH-ID to each accessed CE in a VSI, and shares a default multi-homing site with a single-homed CE. If a CE is dual-homed to PE1 and PE2, they must assign the same MH-ID to the CE. On the network shown in [Figure 1](#), PE1 assigns a multi-homing site and MH-ID 1 to CE1, and PE3 assigns a multi-homing site and MH-ID 3 to CE3. CE2 is dual-homed to

PE1 and PE2, which both assign MH-ID 2 to CE2. The following example uses PE1 to illustrate how to select an optimal site for PW establishment, with priority 100 for PE1's MH-ID 2 and priority 200 for PE2's MH-ID 2.

1. Blocks the AC interface to prevent routing loops. After PE1 receives a BGP Update message from PE2, PE1 finds PE2's multi-homing site with the same MH-ID (2) as a multi-homing site on itself. Therefore, PE1 compares the ACS, PREF, and PE-ID values of the two sites. Because PE1's MH-ID 2 has a lower priority than PE2's MH-ID 2, PE1 blocks the AC interface of MH-ID 2. After PE2 receives a BGP Update message from PE1, PE2 also finds PE1's multi-homing site with MH-ID 2 and compares the ACS, PREF, and PE-ID values of the two sites. Because PE1's MH-ID 2 has a lower priority than PE2's MH-ID 2, PE2 does not take any action.
2. PE1 selects MH-ID 1 as the optimal site, PE2 MH-ID 2, and PE3 MH-ID 3.
3. Establishes PWs between the optimal sites.

Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.2.12 VPLS Service Isolation

Users of different services can be isolated using different VSIs. Users in the same VSI also need to be isolated.

Service Isolation Modes

VPLS networks, however, use a full mesh of PWs and split horizon to prevent loops. Split horizon means that if a packet is received along a PW of VSI, the packet is not forwarded along other PWs associated with the same VSI. VPLS supports either the hub or spoke service isolation mode. In hub mode, traffic forwarding must comply with split horizon rules. In spoke mode, traffic forwarding does not comply with split horizon rules. As described in [Table 1](#), traffic cannot be exchanged between hub AC interfaces or between hub PWs in a VSI. ("T" indicates that traffic can be exchanged between AC interfaces or between PWs, and "F" indicates that traffic cannot be transmitted between AC interfaces or between PWs.)

Table 1 Interworking in default mode

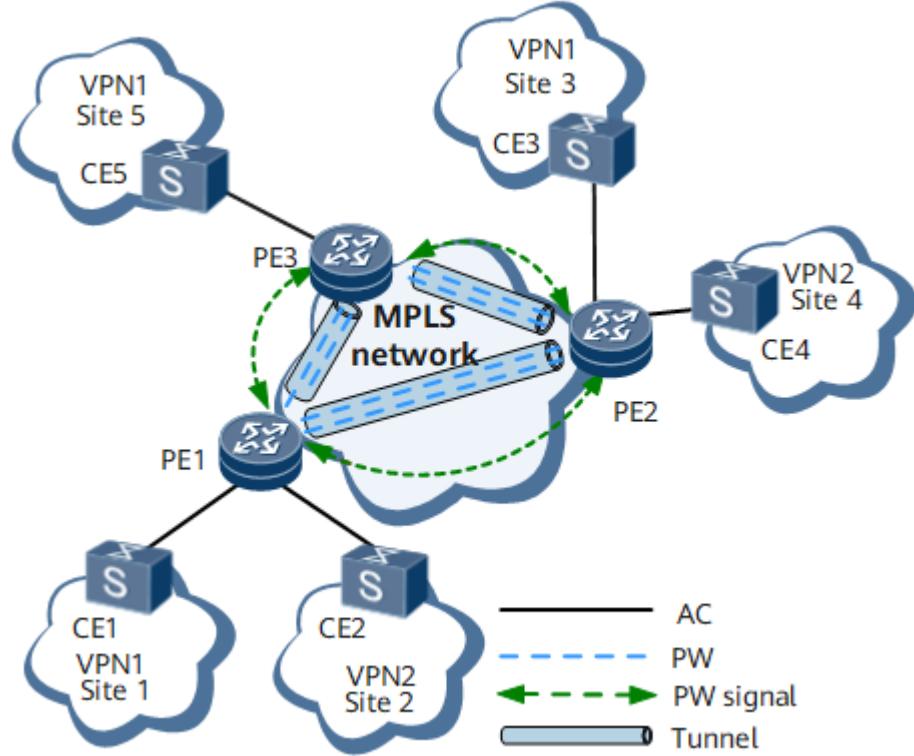
Name	Hub AC	Spoke AC	Hub PW	Spoke PW
Hub AC	F	T	T	T
Spoke AC	T	T	T	T
Hub PW	T	T	F	T
Spoke PW	T	T	T	T

Isolation of Traffic of Users Through Different VSIs

If PE resources are sufficient and the network structure is clear, you can use different VSIs to isolate traffic of different users. In this way, users are grouped and allocated to different VPLS VSIs. Users in a VSI cannot communicate with users in another VSI.

As shown in [Figure 1](#), CE1, CE2, CE3, CE4, and CE5 use the same type of service. CE1, CE3, and CE5 need to communicate with one another; CE2 and CE4 need to communicate with each other; CE1, CE3, and CE5 do not need to communicate with CE2 and CE4. To meet the requirements, different VSIs can be configured to isolate user traffic.

Figure 1 Networking diagram for using different VSIs to isolate user traffic



This method has the following advantages:

- The logical network structure is clear, facilitating management and control.
- MAC address learning and resource usage of different VSIs are reduced.
- If a fault occurs, this feature facilitates fault locating and maintenance and reduces the fault locating access.

The disadvantage is that the modification poses a great impact if mutual access requirements are adjusted.

Isolation of Different Users of the Same Service in the Same VSI

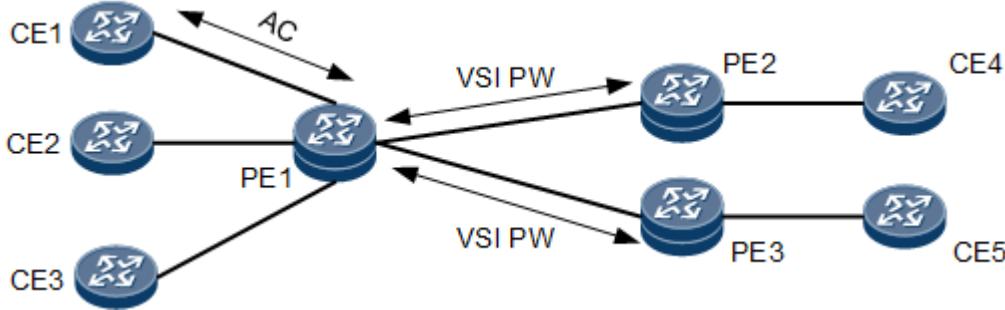
Service isolation requirements of a VSI are classified into the following types:

- Local access users in the same VSI are isolated as needed.
- Local access users and remote access users in the same VSI are isolated.

In a common VPLS scenario, the default attribute of an AC interface is spoke, and the default attribute of a PW is hub.

On the network shown in [Figure 2](#), CE1, CE2, CE3, CE4, and CE5 belong to the same VPN. All local CEs (CE1, CE2, and CE3) connected to PE1 can communicate with one another and with the remote CE4 connected to PE3 and CE5 connected to PE3. However, CE4 connected to PE2 and CE5 connected to PE3 cannot communicate because their VSI PW attribute is hub.

Figure 2 Isolation of common VPLS services



In this case, a VSI is configured on PE1 and the VSI is bound to PE1's AC interface. Then, you can disable the traffic forwarding in spoke mode to prevent all local users on PE1 from communicating with each other. As shown in [Table 2](#), services on spoke ACs are isolated from one another. The AC attribute of the VSI is changed from spoke to hub and the traffic exchange between the hub AC and hub PW is disabled. In this way, the communication between some local users on PE1 and between local access users on PE1 and remote users is isolated, implementing isolation of different users of the same service in the same VSI.

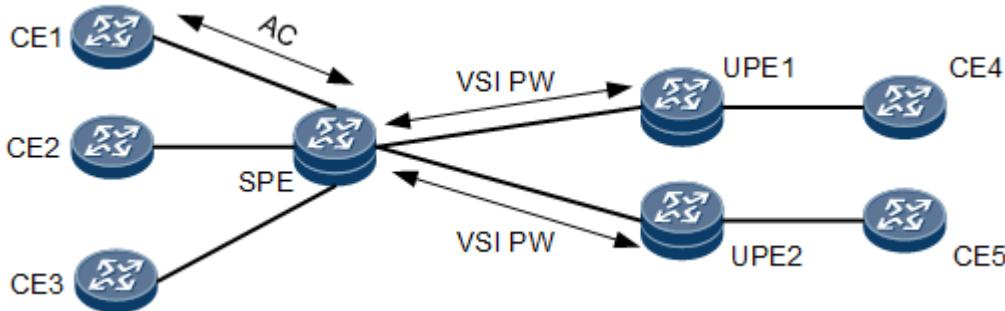
Table 2 Interworking after traffic forwarding in spoke mode is disabled in VPLS

Name	Hub AC	Spoke AC	Hub PW	Spoke PW
Hub AC	F	T	T	T
Spoke AC	T	F	T	F
Hub PW	T	T	F	T
Spoke PW	T	F	T	F

In an HVPLS scenario, the default attributes of AC interfaces and PWs between SPEs and UPEs is spoke, and the default attribute of PWs between SPEs is hub.

On the network shown in [Figure 3](#), when the SPE designates the UPEs as peers, the attribute of the PWs between the SPE and the UPEs changes to spoke. In this case, all local CEs (CE1, CE2, and CE3) connected to the SPE can communicate with one another, and with the remote CE4 connected to UPE1 and remote CE5 connected to UPE2. In addition, CE4 connected to UPE1 and CE5 connected to UPE2 can communicate with each other. In this case, disabling traffic interworking in spoke mode means disabling traffic interworking between spoke ACs, between spoke ACs and UPE PWs, and between UPE PWs.

Figure 3 HVPLS service isolation



Parent Topic: [Understanding VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.10.3 Application Scenarios for VPLS

[Application of VPLS in Residential Services](#)

[Application of VPLS in Enterprise Services](#)

[VPLS PW Redundancy for Protecting Multicast Services](#)

[VPLS PW Redundancy for Protecting Unicast Services](#)

[Application of Multicast VPLS](#)

[VPWS Accessing VPLS](#)

[VPLS Multi-Homing Application](#)

Parent Topic: [VPLS Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.10.3.1 Application of VPLS in Residential Services

Service Overview

Residential services, such as HSI, VoIP, and broadband TV (BTV) are usually carried over carriers' MANs.

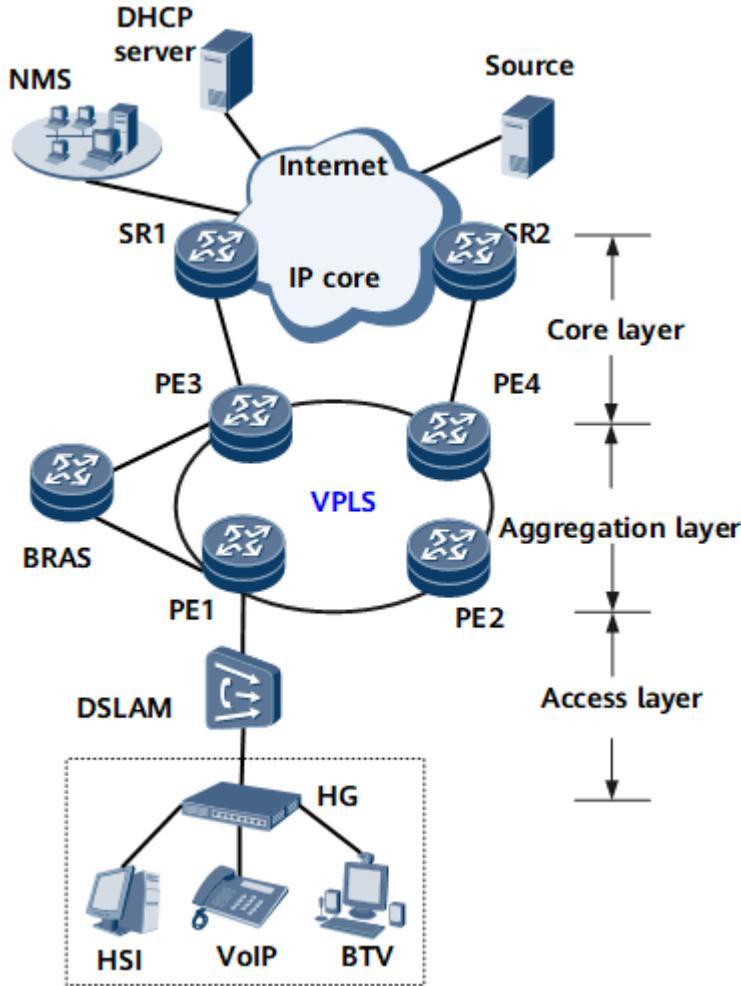
Traditional ATM and FR technologies provide only P2P connections. In addition, those network types have disadvantages, such as high construction costs, low speed, and complex deployment. The development of IP has led to the Ethernet-based VPLS technology that can provide P2MP connections to transparently transmit residential services. Meanwhile, VPLS networks have advantages, such as low construction costs, high speed, and simple configuration. Therefore, current MANs generally use VPLS to transmit user traffic.

Networking Description

Residential services are transmitted to the Internet over the access layer, aggregation layer, and core layer of a MAN. [Figure 1](#) shows the typical networking for residential services. On this network:

- HSI services access the Internet over the MAN.
- VoIP services request IP addresses from the Dynamic Host Configuration Protocol (DHCP) server over the MAN.
- BTV multicast members apply for BTV services from multicast sources over the MAN.

Figure 1 Typical networking for residential services



Feature Deployment

VPLS is configured on PEs to transparently transmit traffic between them. From the perspective of residential users, the public network is like a Layer 2 switch. [Figure 1](#) uses LDP VPLS as an example to show VPLS configuration:

- Access-layer devices

VLANs are configured to differentiate different types of users.

Multicast VLAN and Internet Group Management Protocol (IGMP) snooping are configured to transmit multicast services.

- Aggregation-layer devices

IGPs are configured on PEs so that these PEs can communicate with each other.

Basic MPLS functions are configured on PEs so that these PEs can establish remote LDP sessions. MPLS TE tunnels are established between PEs, and TE fast reroute (FRR) is configured on these PEs.

MPLS L2VPN and VSIs are configured on PEs.

Authentication and accounting features are configured on BRASs so that BRASs can terminate HSI services.

A VPLS daisy chain is deployed on PEs to transmit multicast services.

- Core-layer devices

IGPs are configured on SRs so that these SRs can communicate with each other.

Basic MPLS functions are configured on SRs.

DHCP relay is configured on SRs, allowing VoIP users to obtain IP addresses from DHCP servers.

Layer 3 multicast features are configured on SRs so that these SRs can communicate with multicast sources.

Parent Topic: [Application Scenarios for VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.10.3.2 Application of VPLS in Enterprise Services

Service Overview

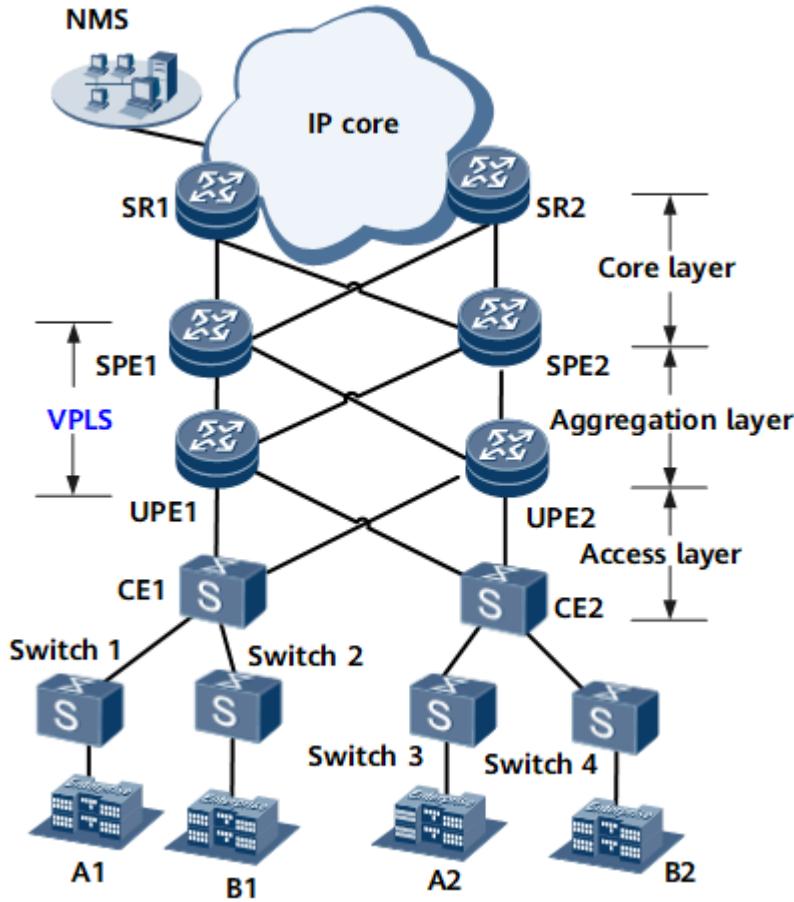
As enterprises set up more and more branches in different regions and office flexibility increases, applications such as instant messaging and video conference are increasingly widely used. This trend imposes high requirements for E2E datacom technologies. A network capable of providing P2MP services is the key to datacom function implementation. To ensure enterprise data security, secure, reliable, and transparent data channels must be provided for multipoint transmission.

To meet the preceding requirements, VPLS is used on carriers' MANs to implement communication between different branches of an enterprise.

Networking Description

Enterprise services are transmitted to the Internet over the access layer, aggregation layer, and core layer of a MAN. [Figure 1](#) shows the typical networking for enterprise services. An enterprise has one branch in city 1 and one branch in city 2. A1 and A2 are R&D departments whereas B1 and B2 are financial departments. VPLS is configured to ensure that the R&D departments can communicate with each other, the financial departments can communicate with each other, but the R&D departments cannot communicate with the financial departments.

Figure 1 Typical networking for enterprise services



Feature Deployment

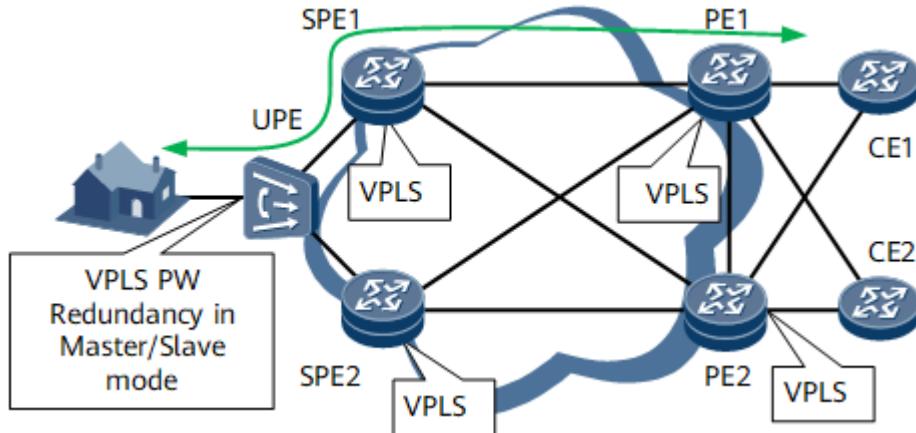
VPLS is configured on PEs to transparently transmit traffic between them. From the perspective of enterprise users, the public network is like a Layer 2 switch. [Figure 1](#) uses LDP VPLS as an example to show VPLS configuration:

- Access-layer devices
 - QinQ is configured to differentiate different types of enterprise users.
- Aggregation-layer devices
 - An Interior Gateway Protocol (IGP) is configured on PEs for these PEs to communicate with each other.
- Basic MPLS functions are configured on PEs so that these PEs can establish remote LDP sessions. MPLS TE tunnels are established between PEs, and TE FRR is configured on these PEs.
- MPLS L2VPN and VSIs are configured on PEs. Dual-homing is used on a VPLS network to protect traffic.
- Limit on the number of learnt MAC addresses and traffic suppression are configured on PEs to protect data.
- Core-layer devices
 - IGPs are configured on SRs so that these SRs can communicate with each other.
 - Basic MPLS functions are configured on SRs.

1.10.3.3 VPLS PW Redundancy for Protecting Multicast Services

[Figure 1](#) illustrates an application of VPLS PW redundancy for protecting multicast services, such as Internet Protocol television (IPTV) services, on a hierarchical virtual private LAN service (HVPLS) network.

Figure 1 VPLS PW redundancy for protecting multicast services



Multicast sources CE1 and CE2 are each dual-homed to PE1 and PE2 using the E-Trunk mechanism; PEs connect to SPEs over common PWs. A gateway UPE connects the user end to SPEs. The link between the UPE and SPE1 and the link between the UPE and SPE2 back up each other.

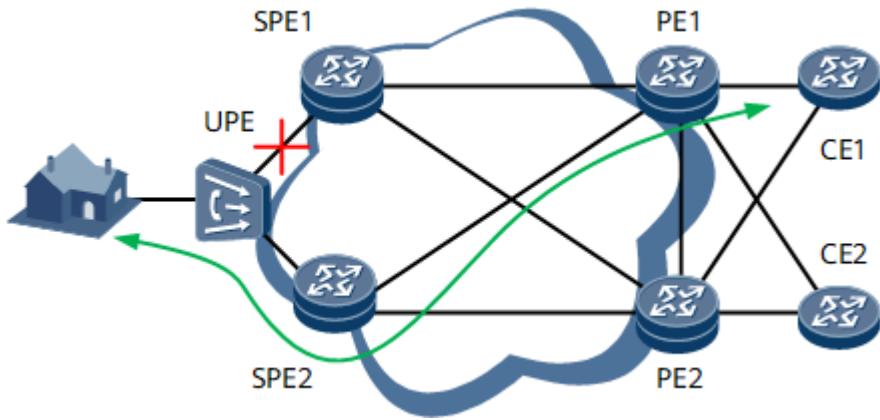
In this networking, the UPE must use the master/slave PW redundancy mode because SPE1 and SPE2 do not exchange signaling to determine which one is the master SPE. Upon detecting that the primary PW fails, the UPE rapidly switches traffic to the secondary PW, instructs SPE2 to work as the primary SPE, and sends MAC Withdraw messages to SPE2, instructing SPE2 to delete the MAC addresses learned from SPE1. SPE2 transmits the MAC Withdraw messages to PE1 and PE2, instructing PE1 and PE2 to clear the MAC addresses learned from SPE1. After deleting the MAC addresses learned from SPE1, PE1 will relearn MAC addresses by broadcasting upon receiving multicast traffic from CE1 and CE2 and switch received traffic to the secondary PW.

[Figure 1](#) shows how traffic transmits when no fault occurs. The following describes how VPLS PW redundancy protects traffic after a fault occurs.

Failure of the Primary PW Between the UPE and SPE1

[Figure 2](#) shows how traffic is switched if the primary PW between the UPE and SPE1 fails.

Figure 2 VPLS PW redundancy protecting services against a failure of the primary PW between the UPE and SPE1



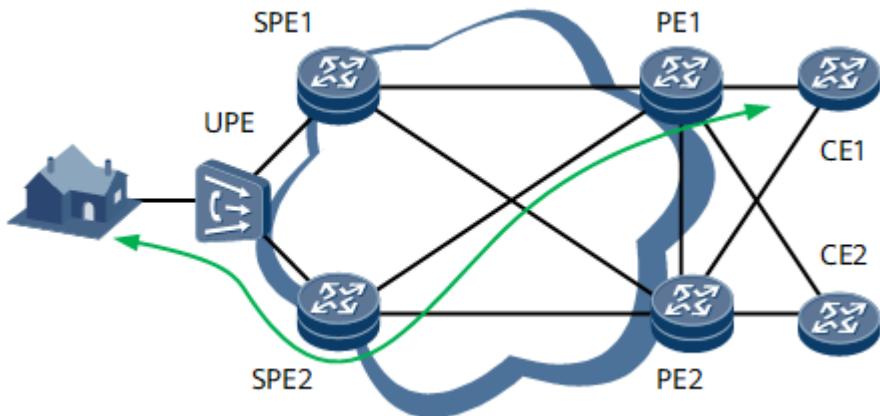
Label switched path (LSP) Down events or BFD for PW may cause a PW to go Down. Upon detecting that the primary PW fails, the UPE switches traffic to the secondary PW and sends MAC Withdraw messages in which the PE ID field carries the SPE1 LSR ID to SPE2. SPE2 transparently transmits the MAC Withdraw messages to PE1 and PE2. Then, SPE2, PE1, and PE2 clear the MAC addresses learned from SPE1.

Switchback: After the primary PW recovers, the UPE instructs SPE2 to change its PW forwarding status to standby and SPE1 to change its PW forwarding status to active. The UPE sends MAC Withdraw messages in which the PE ID field carries the SPE2 LSR ID to SPE1. SPE1 transparently transmits the MAC Withdraw messages to PE1 and PE2. SPE1, PE1, and PE2 clear the MAC addresses learned from SPE2. PE1 and PE2 then relearn MAC addresses by broadcasting upon receiving multicast packets from the primary PW.

SPE1 Failure

[Figure 3](#) shows how traffic is switched if SPE1 fails.

Figure 3 VPLS PW redundancy protecting services against an SPE1 failure



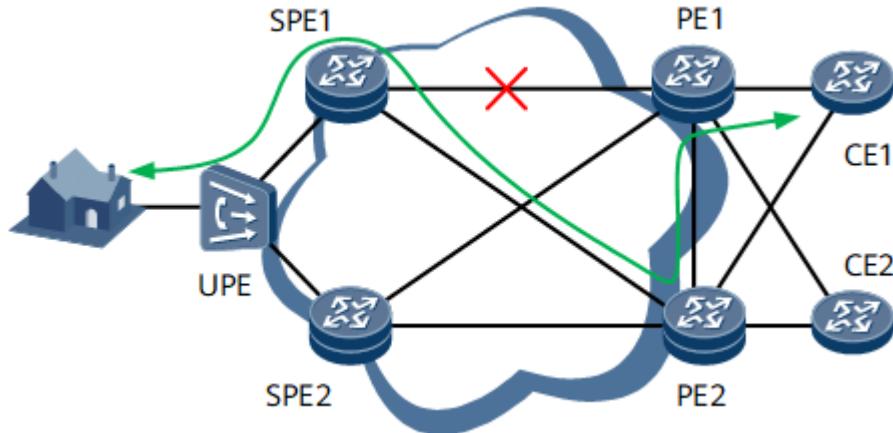
After detecting that SPE1 fails, the UPE switches traffic to the secondary PW and sends MAC Withdraw messages in which the PE ID field carries the SPE1 LSR ID to SPE2. SPE2 transparently transmits the MAC Withdraw messages to PE1 and PE2. Then, SPE2, PE1, and PE2 clear the MAC addresses learned from SPE1. Sometimes, PE1 and PE2 detect that the PW passing through SPE1 is faulty before receiving the MAC Withdraw messages and directly clear the MAC addresses learned from SPE1.

Switchback: After the primary PW recovers, the UPE instructs the PW passing through SPE2 to work as the standby PW and the PW passing through SPE1 to work as the master PW. The UPE sends MAC Withdraw messages in which the PE ID field carries the SPE2 LSR ID to SPE1. SPE1 transparently transmits the MAC Withdraw messages to PE1 and PE2. SPE1, PE1, and PE2 clear the MAC addresses learned from SPE2. PE1 and PE2 then relearn MAC addresses by broadcasting upon receiving multicast packets from the primary PW.

Link Failure Between SPE1 and PE1

[Figure 4](#) shows how traffic is switched if the link between SPE1 and PE1 fails.

Figure 4 VPLS PW redundancy protecting services against a link failure between SPE1 and PE1



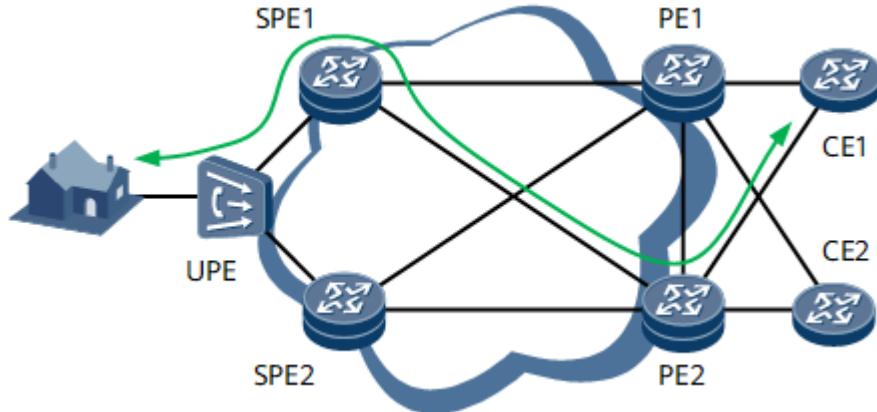
If Label Distribution Protocol (LDP) fast reroute (FRR) is deployed on SPE1 and PE1, LDP FRR ensures the availability of traffic between SPE1 and PE1. If LDP FRR is not deployed, the LDP LSP ensures the availability of traffic between SPE1 and PE1 by means of route convergence.

Switchback: Traffic will not be switched between the primary and secondary PWs. After route convergence, the primary PW is carried by a new LSP.

PE1 Failure

[Figure 5](#) shows how traffic is switched if PE1 fails.

Figure 5 VPLS PW redundancy protecting services against a PE1 failure



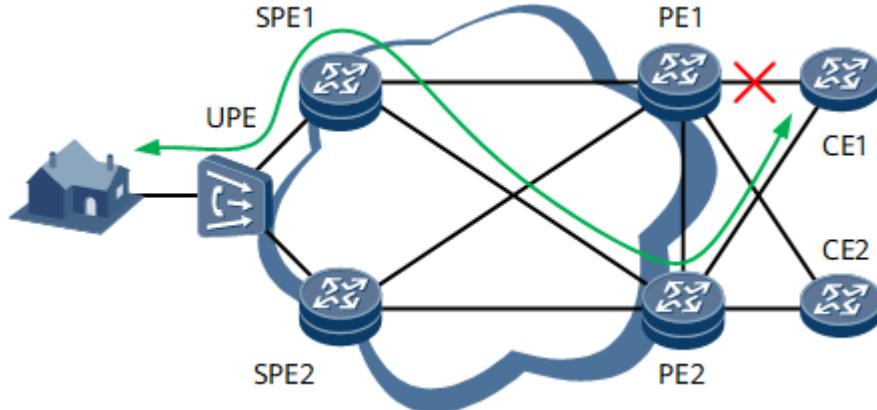
CE1 and CE2 are each dual-homed to PE1 and PE2 using the E-Trunk mechanism. If PE1 fails, a master/backup E-Trunk switchover occurs. PE2 detects the AC interface status change and sends MAC Withdraw messages to SPE1 and SPE2, instructing SPE1 and SPE2 to clear MAC addresses in the corresponding VSIs. Sometimes, SPE1 and SPE2 detect that the PW passing through PE1 is faulty before receiving the MAC Withdraw messages and directly clear MAC addresses associated with the PW.

Switchback: If PE1 recovers, traffic switches back to PE1 after a default E-Trunk switchback delay. Upon detecting the AC status changes, PE1 and PE2 send MAC Withdraw messages to SPE1 and SPE2, instructing SPE1 and SPE2 to clear MAC addresses learned from PE2.

Failure of the Primary AC Link

[Figure 6](#) shows how traffic is switched if the link between CE1 and PE1 fails.

Figure 6 VPLS PW redundancy protecting services against a link failure between CE1 and PE1



CE1 and CE2 are each dual-homed to PE1 and PE2 using the E-Trunk mechanism. After the link between CE1 and PE1 fails, a master/backup E-Trunk switchover occurs. Upon detecting AC interface status changes, PE1 and PE2 send MAC Withdraw messages to SPE1, instructing SPE1 to clear all MAC addresses in the corresponding VSI.

Switchback: After the link between CE1 and PE1 recovers, a master/backup E-Trunk switchback occurs. Upon detecting AC interface status changes, PE1 and PE2 send MAC Withdraw messages to SPE1, instructing SPE1 to clear MAC addresses.

Parent Topic: [Application Scenarios for VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

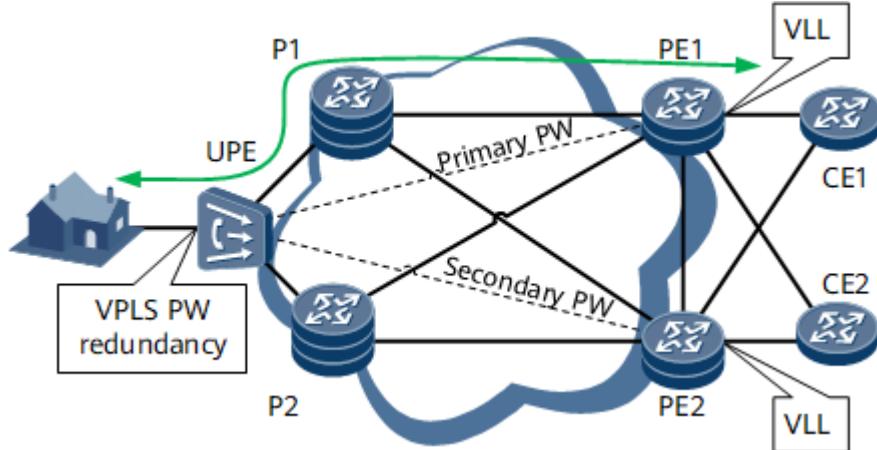
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.3.4 VPLS PW Redundancy for Protecting Unicast Services

[Figure 1](#) illustrates an application of VPLS PW redundancy for protecting unicast services, such as high-speed internet (HSI) or Voice over Internet Protocol (VoIP) services, on a virtual leased line (VLL) accessing virtual private LAN service (VPLS) network.

Figure 1 VPLS PW redundancy for protecting unicast services



Authentication servers CE1 and CE2 are each dual-homed to PE1 and PE2 using the E-Trunk mechanism. A UPE connects the user end to PEs. The link between the UPE and PE1 and the link between the UPE and PE2 back up each other.

In this networking, PE1 and PE2 can determine their master/backup status through E-Trunk negotiation. Therefore, the UPE can use the independent PW redundancy mode to determine the active/standby PW status based on the master/backup status of PE1 and PE2. Upon detecting that the primary PW fails, the UPE rapidly switches traffic to the secondary PW and instructs PE2 to work as the master PE. After the E-Trunk mechanism detects that the primary PW fails, it switches traffic to the secondary AC link.

[Figure 1](#) shows how service traffic transmits when no fault occurs. The following describes how VPLS PW redundancy protects traffic after a fault occurs.

NOTE

The master/slave and independent VPLS PW redundancy modes protect services in different ways. The following describes the differences between the two modes in terms of service protection.

Network-Side Failure

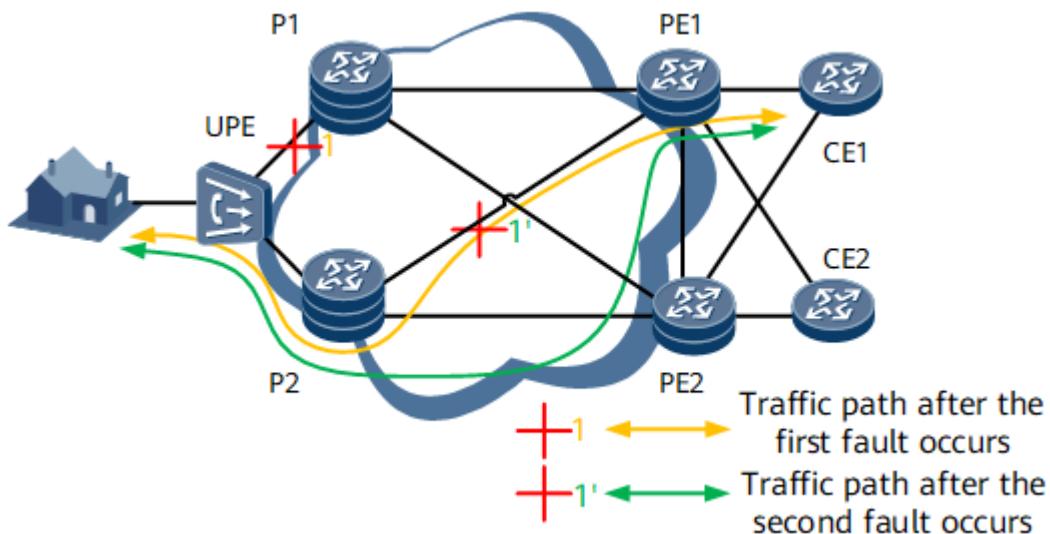
Two levels of protection can be provided to protect services against network-side faults:

NOTE

If a network-side fault occurs, LSPs or TE tunnels first detect the fault and switch traffic to other tunnels. If tunnel protection is unavailable or fails, PW redundancy is required to protect traffic. A bypass PW needs to be configured between PE1 and PE2 for PW redundancy.

- LSP or TE tunnel protection: After a network-side fault occurs, routes converge, and the LSP carrying the primary PW switches to a new route. [Figure 2](#) shows how traffic is switched. After the fault is rectified, routes re-converge, and the LSP carrying the primary PW switches to a new route.

Figure 2 Protecting services against an LSP failure (bypass PW not configured)

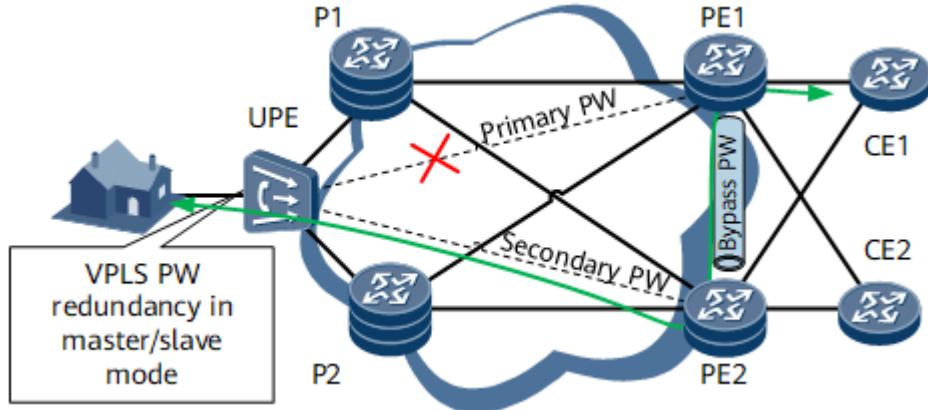


- PW redundancy: If LSP or TE tunnel switching fails, traffic is switched to the secondary PW. [Figure 3](#) shows how traffic is switched. After the fault is rectified, traffic will be switched back based on preset switchback policies.

NOTE

Bypass PWs are required for PW redundancy to transmit traffic between PE1 and PE2.

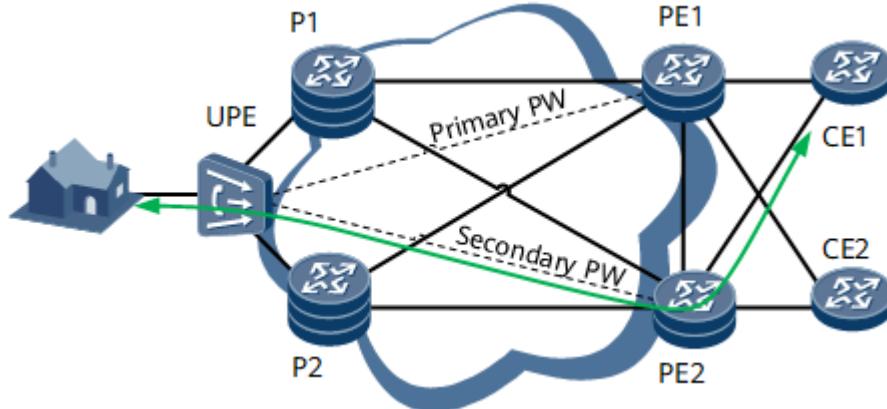
Figure 3 Protecting services against failures of the primary PW (bypass PW configured)



PE1 Failure

[Figure 4](#) shows how traffic is switched if PE1 fails.

Figure 4 Protecting services against a PE1 failure in VPLS PW redundancy mode



PE2 becomes the master and PE1 becomes the backup after E-Trunk negotiation. The UPE is informed of the switchover. Upon detecting that the primary PW fails, the UPE clears MAC addresses learned from the primary PW and switches traffic to the secondary PW.

Switchback: After PE1 recovers, PE1 becomes the master through E-Trunk negotiation. Upon detecting PE1 and PE2 status changes, the UPE clears MAC addresses learned from PE2 and relearns MAC addresses by broadcasting.

Failure of the Primary AC Link

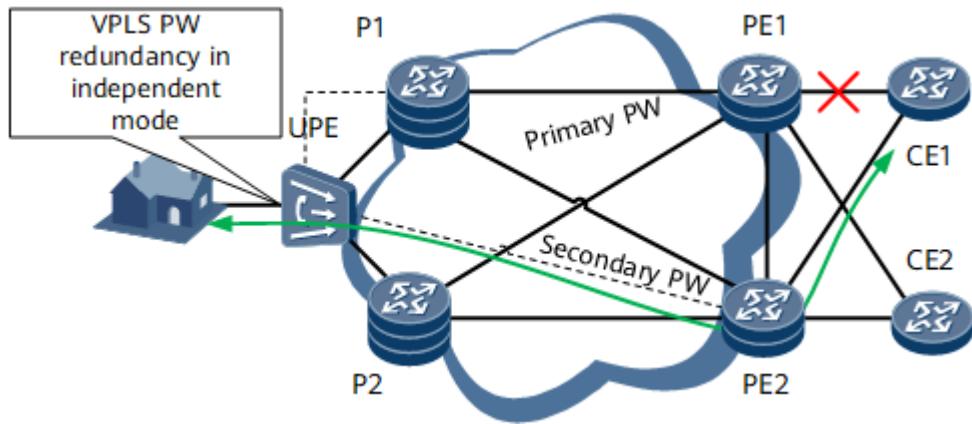
The following describes how traffic is switched if the link between PE1 and CE1 fails:

- In independent mode

After the primary AC link between CE1 and PE1 fails, PE2 works as the master after E-Trunk negotiation. The UPE is informed of the switchover. The UPE detects that the primary AC link fails and switches traffic to the secondary PW.

After the link between CE1 and PE1 recovers, PE1 becomes the master after E-Trunk negotiation. Upon detecting PE1 and PE2 status changes, the UPE clears MAC addresses learned from PE2 and relearns MAC addresses by broadcasting.

Figure 5 Protecting services against failures of the primary AC link in independent VPLS PW redundancy mode

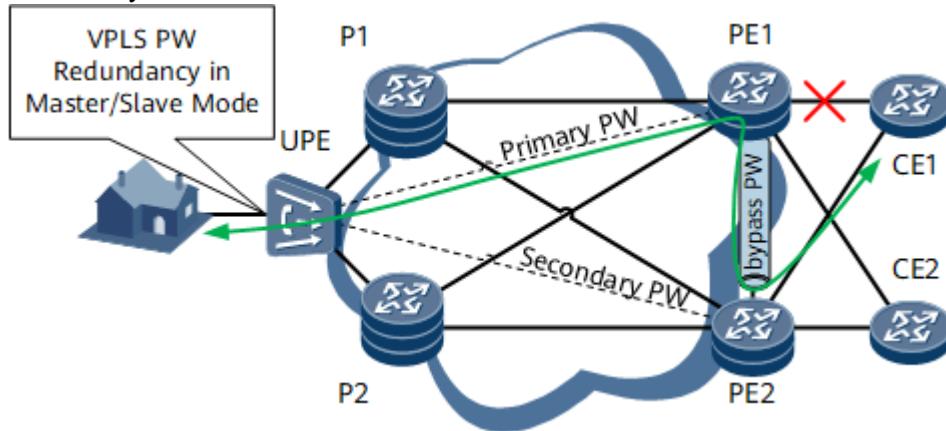


- In master/slave mode

After the primary AC link between CE1 and PE1 fails, PE2 works as the master after E-Trunk negotiation. The PW forwarding status on the UPE remains unchanged.

After the fault is rectified, PE1 becomes the master after the master/backup status is negotiated in the E-Trunk. The UPE, however, can detect that PE1 becomes the master and PE2 becomes the backup. However, the PW status is not switched on the UPE.

Figure 6 Protecting services against failures of the primary AC link in master/slave VPLS PW redundancy mode



Parent Topic: [Application Scenarios for VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

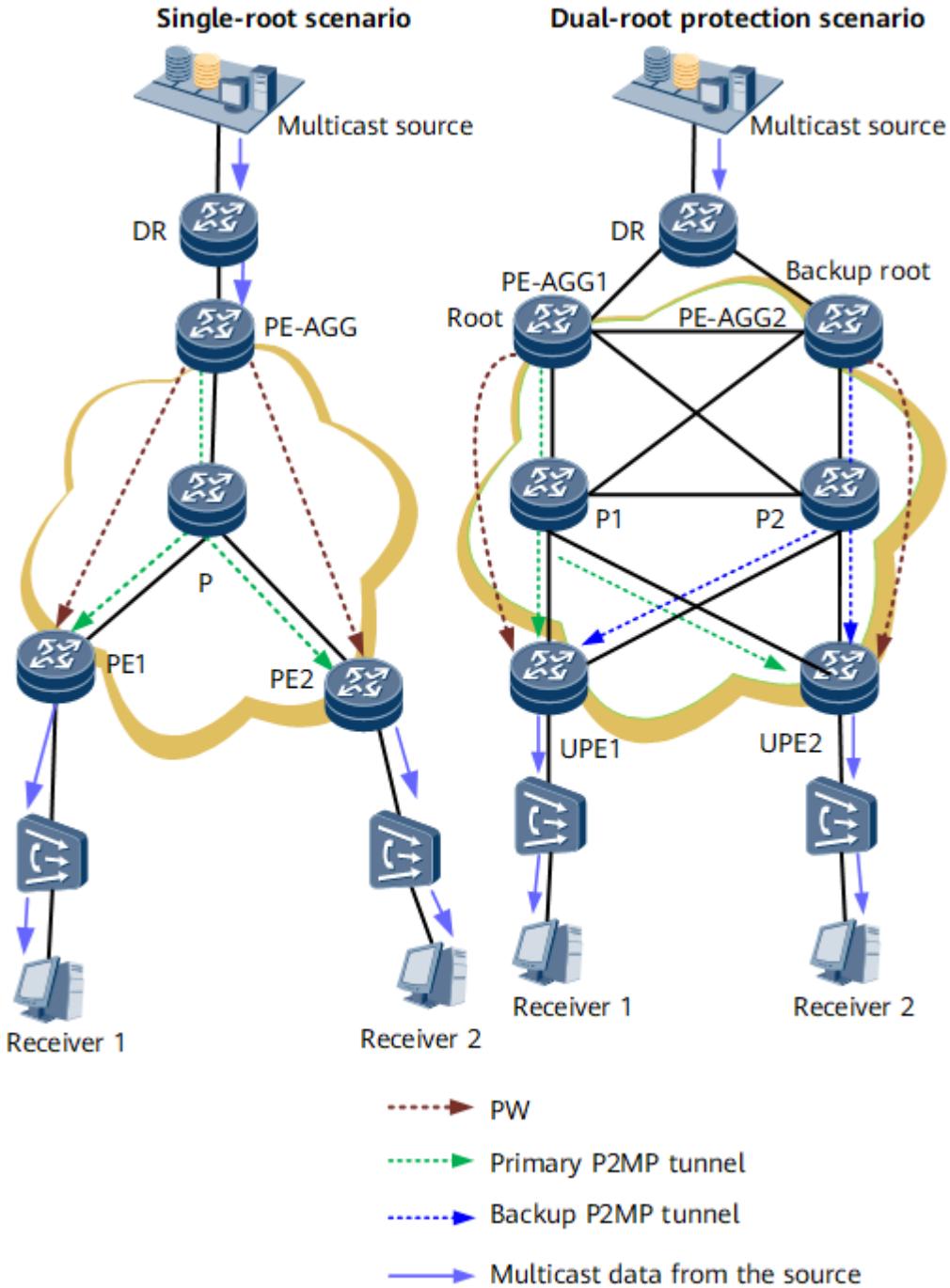
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.3.5 Application of Multicast VPLS

Multicast VPLS is deployed on IP/MPLS backbone networks to solve multicast service issues related to traffic congestion, transmission reliability, and data security. [Figure 1](#) shows the application of multicast VPLS on an IP/MPLS backbone network.

Figure 1 Application of multicast VPLS on an IP/MPLS backbone network



To carry multicast traffic over P2MP tunnels by means of multicast VPLS:

- Configure BGP and multicast VPLS on the IP/MPLS backbone network. BGP is used to exchange BGP routes between PEs on the same multicast VPLS network.
- Configure MPLS for the establishment of P2MP tunnels on the IP/MPLS backbone network.

Parent Topic: [Application Scenarios for VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.3.6 VPWS Accessing VPLS

L2VPN PWs have different encapsulation types, such as Ethernet and VLAN. These encapsulation types are essentially irrelevant to services. A service can be transmitted over VPWS and VPLS PWs in succession so long as these PWs use the same encapsulation type.

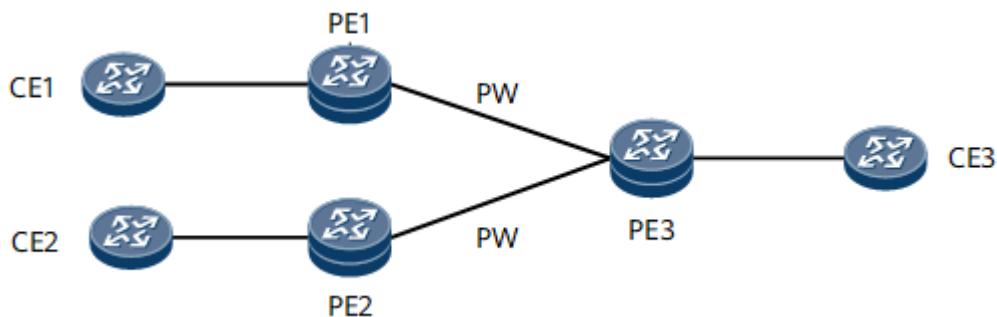
VPWS Accessing VPLS in Basic Mode

On the network shown in [Figure 1](#), CE1, CE2, and CE3 belong to a broadcast domain. CE1 and CE2 are required to communicate only with CE3. To avoid imposing heavy pressure on the capacities of MAC address tables on PE1 and PE2, you can deploy a VPWS PW from PE1 to PE3 and from PE2 to PE3 to carry traffic from CE1 and CE2 respectively, and deploy a VPLS PW from PE3 to PE1 and from PE3 to PE2 to carry traffic from CE3.

NOTE

Because VPLS uses split horizon by default, after PE3 receives traffic from CE1 and CE2, it does not forward the traffic over VPLS PWs back to PE1 or PE2. This implementation ensures that CE1 and CE2 can communicate with only CE3. If you want the three CEs to communicate with each other, change the VPLS PW attributes on PE3 to spoke.

Figure 1 VPWS accessing VPLS in basic mode



VPWS Accessing VPLS in Dual-homed mode

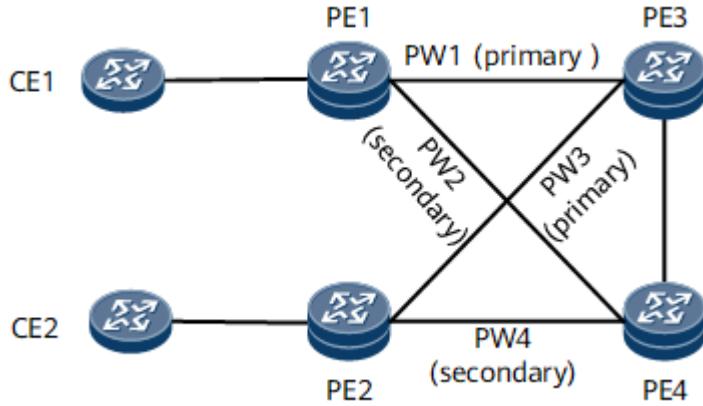
On the network shown in [Figure 2](#), PE1 and PE2 each is dual-homed to PE3 and PE4. CE1 and CE2 are required to communicate. To avoid imposing heavy pressure on the capacities of MAC address tables on PE1 and PE2, you can use VPWS PWs to dual-home PE1 and PE2 each to PE3 and PE4.

Specifically, configure VPWS PWs from PE1 to PE3 and PE4 and from PE2 to PE3 and PE4. Then, configure spoke VPLS PWs from PE3 to PE1 and PE2 and from PE4 to PE1 and PE2 and a hub PW between PE3 and PE4. After that, configure the PWs between PE1 and PE3 and between PE2 and PE3 as primary PWs, and the PWs between PE1 and PE4 and between PE2 and PE4 as secondary PWs. Configure PW protection on PE1 and PE2 based on the signaling mode used by PE1 and PE2:

- If PE1 and PE2 use LDP VPWS, configure master/slave PW redundancy.
- If PE1 and PE2 use static VPWS, configure PW APS.

After receiving traffic from CE1, PE1 forwards the traffic to PE3 over PW1. Upon receipt, PE3 broadcasts the traffic to PE2 and PE4 (because a VPLS network is a broadcast domain by default). Because the PW between PE2 and PE3 is a primary PW, PE2 can successfully receive the broadcast traffic and forward the traffic to CE2.

Figure 2 VPWS accessing VPLS in dual-homed mode



Parent Topic: [Application Scenarios for VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

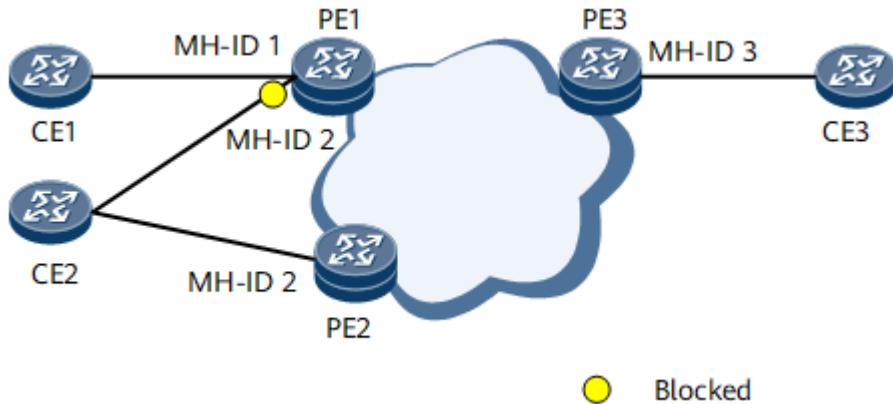
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.10.3.7 VPLS Multi-Homing Application

On VPLS networks, carriers usually dual-home CE2 to PE1 and PE2 through redundant links to deliver high-reliability services. To prevent routing loops caused by link redundancy, VPLS multi-homing can be configured on PE1 and PE2.

Figure 1 VPLS multi-homing networking



On the network shown in [Figure 1](#), an MPLS backbone network is deployed to provide VPLS services for enterprise users. CE1 is single-homed to PE1, CE2 is dual-homed to PE1 and PE2, and CE3 is single-homed to PE3. VPLS multi-homing can be configured so that PE1 and PE2 will assign the same multi-homing site ID to CE2. A preferred PE is then selected for PW establishment. PE1, PE2, and PE3 can establish PWs between each other to allow communication between CE1, CE2, and CE3.

Feature Deployment

- Deploy IP and an IGP on the carrier network to ensure reachability between the PEs and Ps at the network layer.
- Deploy MPLS on the carrier network and establish LDP LSPs or TE LSPs between the PEs.
- Deploy BGP between the PEs.

- Deploy VPLS on the PEs, create a VSI for each user, and bind a user-side interface to the L2VPN.
- Allow each PE to assign a site and priority to each CE.

Parent Topic: [Application Scenarios for VPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.11 L2VPN Accessing L3VPN Description

[Overview of L2VPN Accessing L3VPN](#)

[Understanding L2VPN Accessing L3VPN](#)

[Application Scenarios for L2VPN Accessing L3VPN](#)

[Terminology for L2VPN Accessing L3VPN](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

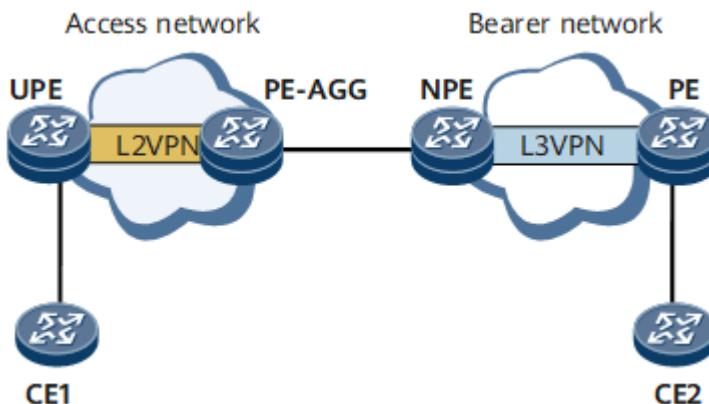
[< Previous topic](#) [Next topic >](#)

1.11.1 Overview of L2VPN Accessing L3VPN

Definition

Multiprotocol Label Switching (MPLS) is widely applied in Metropolitan Area Networks (MANs) because it features high reliability, high security, and sound IP-based operation and maintenance capabilities, and supports quality of service (QoS). A Layer 2 virtual private network (L2VPN) provides MPLS-based L2VPN services to transparently transmits Layer 2 user data over tunnels on an MPLS network. This reduces the number of label switched paths (LSPs) maintained by transit nodes.

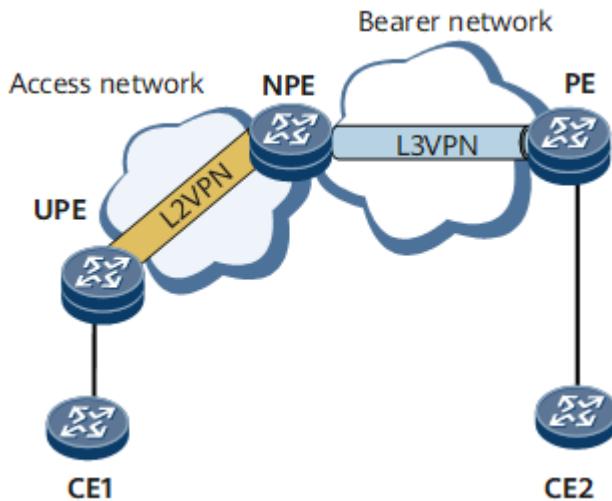
Figure 1 Traditional L2VPN accessing L3VPN



On a traditional network, a provider edge aggregation (PE-AGG) and a network provider edge (NPE) are required to connect the access network to the bearer network. Then, the L2VPN can access the public network or Layer 3 virtual private network (L3VPN).

On the network shown in [Figure 1](#), the user provider edges (UPEs) are responsible for providing access services for user sites by creating an L2VPN tunnel between the access network and PE-AGG. The PE-AGG terminates the L2VPN and connects to the NPE. An L3VPN is set up between the NPE and other common PEs on the bearer network of the carrier. As a CE of the L2VPN, the NPE connects to the PE-AGG. For the L3VPN on the bearer network, CE1 accesses the L3VPN over the leased line emulated by the L2VPN.

Figure 2 L2VPN accessing L3VPN supported by the NE40E



If an NPE can provide the functions of a PE-AGG and an NPE, it helps lower the networking costs and simplify the networking. As shown in [Figure 2](#), the NPE terminates an L2VPN and accesses an L3VPN over a virtual Ethernet (VE) group. The NPE provides the functions of the PE-AGG and NPE in the traditional networking.

In a VE-Group, the VE interface used to terminate an L2VPN is called a Layer 2 Virtual Ethernet (L2VE) interface, and the VE interface used to access an L3VPN is called a Layer 3 Virtual Ethernet (L3VE) interface.

An NPE supports the division of multiple VSs. To implement L2VPN and L3VPN service interworking between different VSs, you can allocate the L2VE and L3VE interfaces in the same VE-Group to different VSs.

Parent Topic: [L2VPN Accessing L3VPN Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.11.2 Understanding L2VPN Accessing L3VPN

[L2VPN Accessing L3VPN Fundamentals](#)

[Classification of L2VPN Accessing L3VPN](#)

Parent Topic: [L2VPN Accessing L3VPN Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.11.2.1 L2VPN Accessing L3VPN Fundamentals

Basic Concepts

- Virtual Ethernet (VE) interface

A VE interface has the common features of an Ethernet interface, and supports services such as maximum transmission unit (MTU) and QoS.

- Layer 2 Virtual Ethernet (L2VE) interface

An L2VE interface supports VPWS and VPLS services.

- Layer 3 Virtual Ethernet (L3VE) interface

An L3VE interface supports the termination of IP services or the access to L3VPNs. An L3VE interface also supports sub-interfaces.

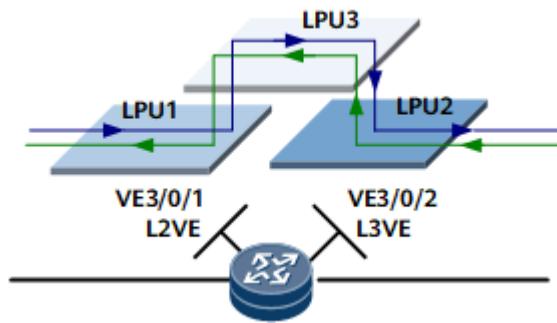
- VE-Group

As a connector of L2VE and L3VE interfaces, a VE-Group provides a type of virtual connection to associate L2VE interfaces with L3VE interfaces by using the same group ID.

One VE-Group corresponds to one L2VPN accessing L3VPN service. By creating multiple VE-Groups with VE interfaces bound to different L2VPNs and L3VPNs, you can implement multiple L2VPN accessing L3VPN services.

Implementation

Figure 1 L2VPN accessing L3VPN



On the network shown in [Figure 1](#), L2VE and L3VE interfaces are bound by means of a VE-Group. L2VPN accessing L3VPN is implemented through the loopback between the L2VE and L3VE interfaces of the same VE-Group. Logically, the principle of the loopback between the L2VE and L3VE interfaces is similar to that of connecting two physical interfaces through fibers. One of the interfaces is bound to an L2VPN, and the other is bound to an L3VPN.

A VE-Group is associated with a tag. By setting up different VE-Groups and bind them to different L2VPNs and L3VPNs, you can configure multiple L2VPN accessing L3VPN services.

Parent Topic: [Understanding L2VPN Accessing L3VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

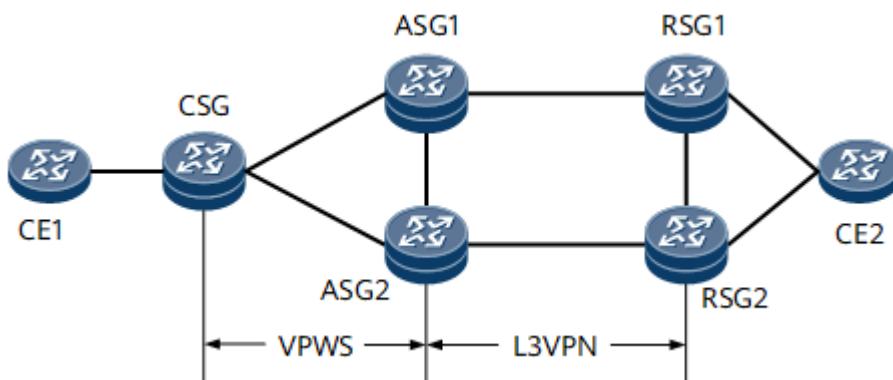
1.11.2.2 Classification of L2VPN Accessing L3VPN

VPWS Accessing an L3VPN or Public Network

Virtual Private Wire Service (VPWS) provides MPLS network-based L2VPN services and transparently transmits Layer 2 data of users over an MPLS network. From the perspective of users, the MPLS access network is a Layer 2 switching network that provides Layer 2 connections between users and the carrier network. Users can access a public network or L3VPN services on the carrier network over VPWS connections.

On the network shown in [Figure 1](#), L2VPN accessing L3VPN is deployed for CE1 and CE2 to communicate. The CSG is dual-homed to two ASGs, and the ASGs connect to the RSG over the L3VPN. PW APS-protected VPWS is deployed on the CSG, ASG1, and ASG2, and VE groups are created on ASGs to terminate VPWS and provide L3VPN access. VPN FRR is deployed on the L3VPN side to enhance service connection reliability.

Figure 1 PW APS-protected VPWS accessing L3VPN



VPLS Accessing a Public Network or L3VPN

VPLS connects multiple Ethernet LAN segments over a packet switched network (PSN). These Ethernet LAN segments function like a LAN.

Unlike point-to-point services of L2VPNs, VPLS can connect multiple Ethernet sites of a user to a public network or an L3VPN of the carrier's bearer network over an MPLS access network.

L2VPN Accessing Multiple L3VPNs Through L3VE QinQ VLAN Tag Termination Sub-interfaces

802.1Q-in-802.1Q (QinQ) is a Layer 2 tunnel protocol based on IEEE 802.1Q. Packets transmitted by means of QinQ carry double 802.1Q tags. QinQ distinguishes different types of services for different users.

Ethernet packets carried over VPWS or VPLS can carry either one or two tags. Currently, the NE40E can transmit service packets to different L2VPNs by adding different outer VLAN tags to these packets. When the packets with double VLAN tags reach an L3VE interface through an L2VPN, the L3VE sub-interfaces terminate QinQ packets with specified inner tags and transmit services of different types to their destination L3VPNs on the bearer network.

This allows the carrier to provide differentiated QoS guarantee for different services on the bearer network, helping the carrier to fully utilize network resources and provide Differentiated Services (DiffServ) for users.

1.11.3 Application Scenarios for L2VPN Accessing L3VPN

[VPWS Accessing L3VPN](#)

[VPLS Accessing L3VPN](#)

Parent Topic: [L2VPN Accessing L3VPN Description](#)

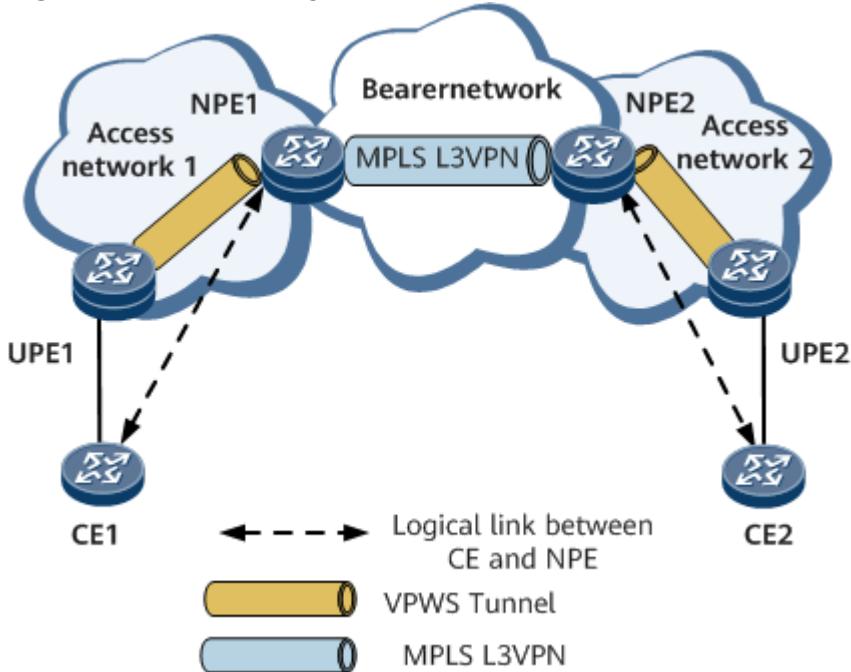
Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.11.3.1 VPWS Accessing L3VPN

On the network shown in [Figure 1](#), the NPE functions as an exterior gateway. When hosts attached to the CE need to access the Layer 3 network, the MAC address of the gateway is required. If the MAC address does not exist, the hosts send Address Resolution Protocol (ARP) requests to the NPE over VPWS. The NPE terminates VPWS, analyzes the ARP requests, and generates related ARP entries. Later, when receiving a packet, the NPE checks the MAC address carried in the packet. If the MAC address carried in the packet is the MAC address of the NPE, the NPE forwards the packet to the L3VPN.

Figure 1 VPWS accessing L3VPN



Parent Topic: [Application Scenarios for L2VPN Accessing L3VPN](#)

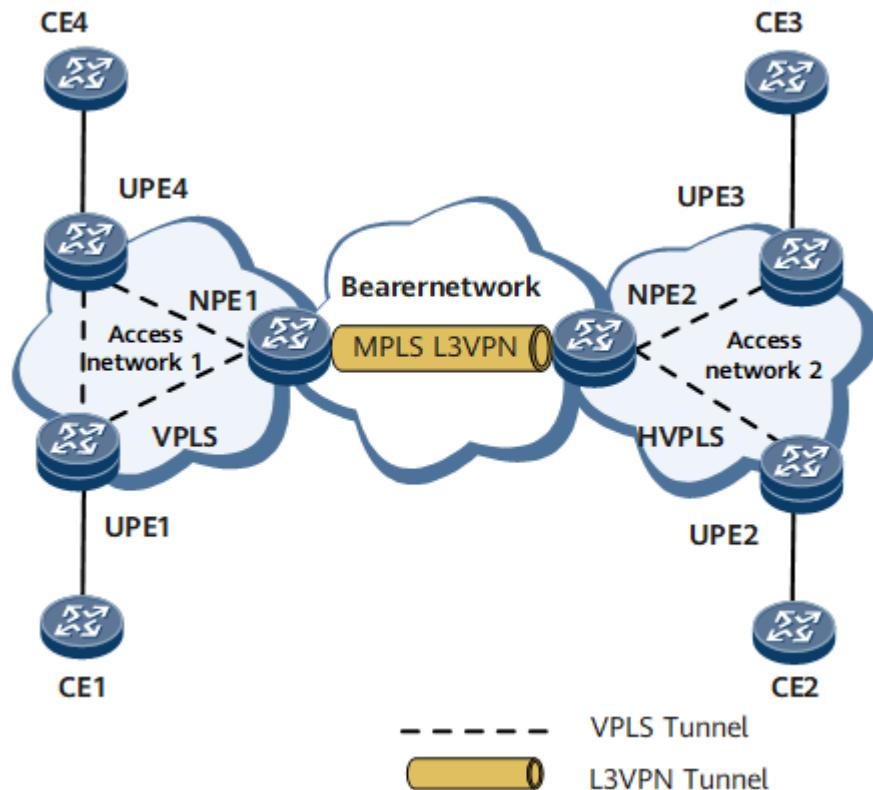
Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.
[Next topic >](#)

1.11.3.2 VPLS Accessing L3VPN

On the network shown in [Figure 1](#), the NPE functions as the PE of both the access network and the bearer network. In addition to VPLS, the NPE needs to support the gateway function, including the configuration of IP addresses, access to L3VPNs, ARP, and packet forwarding.

When a CE needs to access the L3VPN, the CE sends an ARP request to the gateway interface of the NPE. The NPE then forwards traffic between the L2VPN and L3VPN. Meanwhile, the traffic forwarding on the original L2VPN is not affected. The NPE needs to broadcast ARP packets locally and in the VSI.

Figure 1 VPLS accessing L3VPN



Parent Topic: [Application Scenarios for L2VPN Accessing L3VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.11.4 Terminology for L2VPN Accessing L3VPN

Acronyms and Abbreviations

Acronym and Abbreviation	Full Name
L2VE	Layer 2 Virtual Ethernet Interfaces
L3VE	Layer 3 Virtual Ethernet Interfaces

Acronym and Abbreviation	Full Name
VE	Virtual Ethernet Interfaces

Parent Topic: [L2VPN Accessing L3VPN Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.12 EVPN Feature Description

[Overview of EVPN](#)

[EVPN Fundamentals](#)

[EVPN-MPLS](#)

[EVPN-VXLAN](#)

[EVPN VPWS](#)

[PBB-EVPN](#)

[EVPN E-Tree](#)

[MAC Duplication Suppression for EVPN](#)

[EVPN ORF](#)

[IGMP Snooping over EVPN MPLS](#)

[Application Scenarios for EVPN](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.1 Overview of EVPN

Definition

Ethernet Virtual Private Network (EVPN) is a next-generation full-service bearer VPN solution. It unifies the control planes for various VPN services and uses BGP extensions to transmit Layer 2 or Layer 3 reachability information, separating the forwarding plane from the control plane.

Purpose

EVPN was initially proposed to overcome the drawbacks of traditional L2VPN. The following describes these drawbacks using VPLS as an example:

- 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 every two PEs. The amount of network resources consumed for PW establishment increases as the number of PEs increases. Furthermore, a significant 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 no need to learn the MAC addresses carried in them.

EVPN integrates the following characteristics to overcome the preceding drawbacks:

- EVPN uses BGP extensions to implement MAC address learning and advertisement on the control plane instead of 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. This is because PEs on an EVPN communicate using BGP, which provides the route reflection function. As such, a route reflector (RR) can be deployed on the carrier backbone network to reflect EVPN routes to PEs with which the RR has established peer relationships. This significantly reduces network complexity and the number of network signaling messages.
- EVPN enables PEs to learn local MAC addresses using ARP and learn remote MAC and IP addresses using MAC/IP advertisement routes. The PEs can then cache these addresses locally. After receiving an ARP request, a PE searches its locally cached MAC and IP address information based on the destination IP address in the ARP request, and then returns an ARP reply when it finds the corresponding information. This reduces consumption of network resources because the PE does not broadcast ARP requests to other PEs.

Benefits

EVPN offers the following benefits:

- Improved link utilization and transmission efficiency: EVPN supports load balancing, fully utilizing network resources and alleviating 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 respond to ARP requests from connected sites using locally cached MAC addresses, minimizing the amount of broadcast ARP requests.

Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.12.2 EVPN Fundamentals

Typical EVPN Networking

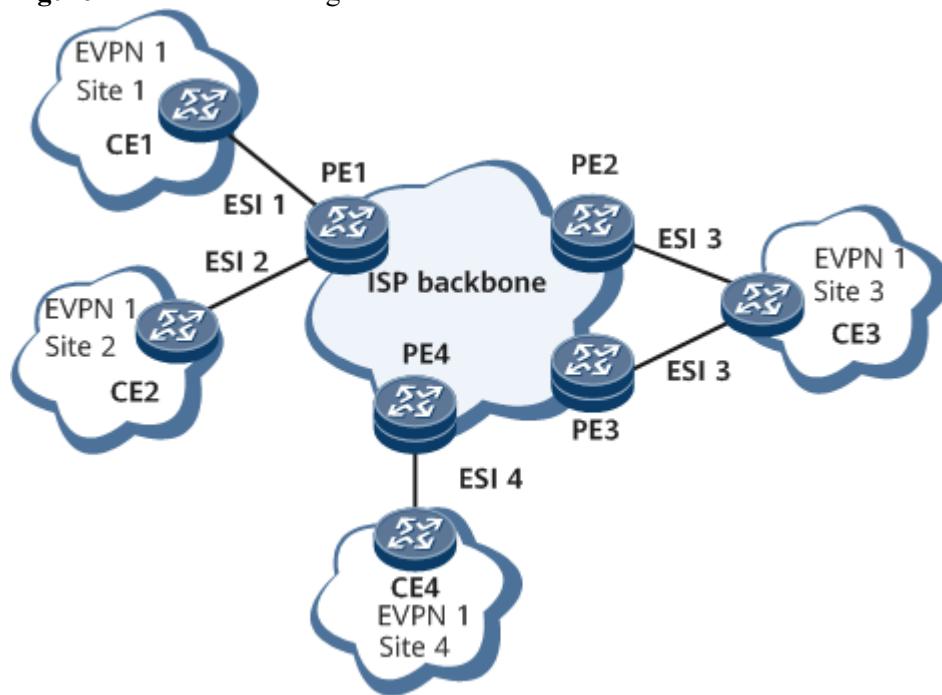
As shown in [Figure 1](#), an EVPN has a similar network structure to a BGP/MPLS IP VPN. In EVPN networking, to implement interconnection between sites, PEs have an EVPN instance created on a carrier backbone network, connect to CEs at different sites, and establish EVPN BGP peer relationships and MPLS/SR tunnels with each other. Different from a BGP/MPLS IP VPN, an EVPN uses Layer 2 networks within sites. As such, a PE learns MAC addresses rather than IP routes from

the CEs at a site, and then advertises these MAC addresses to the other sites within the same EVPN instance using EVPN-specific routes.

In EVPN networking, a CE can be single-homed to one PE or multi-homed to several PEs. As shown in [Figure 1](#), CE1, CE2, and CE4 are single-homed to PE1, while CE3 is multi-homed to PE2 and PE3. Load balancing is supported in CE multi-homing networking.

EVPN technology defines a unique Ethernet Segment Identifier (ESI) on PEs to identify connections to the same CE. The PE interfaces must use the same ESI to connect to a CE and different ESIs to different CEs. During route advertisement between PEs, a PE can be aware of the other PEs connecting to the same CE after receiving routes that carry the same ESI.

Figure 1 EVPN networking



NOTE

A PE can use both IPv4 and IPv6 addresses to establish EVPN peer relationships with the other PEs. MPLS/VXLAN/SR tunnels can be deployed between IPv4 EVPN peers to carry services, and SRv6 tunnels need to be deployed between IPv6 EVPN peers to carry services. A PE sends the EVPN routes that carry SIDs only to IPv6 EVPN peers and the EVPN routes that do not carry SIDs only to IPv4 EVPN peers.

EVPN Routes

To enable sites to learn MAC addresses from each other, EVPN defines a new type of BGP network layer reachability information (NLRI), also known as the EVPN NLRI. An EVPN NLRI can be one of the following EVPN routes:

- Ethernet auto-discovery route: also known as the Ethernet A-D route. PEs advertise Ethernet auto-discovery routes after establishing an EVPN BGP peer relationship. A local PE advertises such routes to other PEs to notify the reachability of MAC addresses of sites connected to the local PE. Ethernet A-D routes are classified into per-ES routes and per-EVI routes. Ethernet A-D per-ES routes are used for fast convergence, redundancy mode, and split horizon. Ethernet A-D per-EVI (EVPN Instance) routes are used for aliasing. [Figure 2](#) shows the NLRI of an Ethernet A-D route.

Figure 2 NLRI of an Ethernet A-D route

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:

- Route Distinguisher: In an Ethernet A-D per-ES route, this field contains the source IP address set on a PE, for example, X.X.X.X:0. In an Ethernet A-D per-EVI route, this field is the RD of an EVPN instance.
- Ethernet Segment Identifier: uniquely identifies connections between PEs and a CE.
- Ethernet Tag ID: The value of the field is all Fs in Ethernet A-D per-ES routes. In an Ethernet A-D per-EVI route, this field identifies a sub-broadcast domain in an ES. If this field is set to all 0s, the EVI contains only one broadcast domain.
- MPLS Label: The value is all 0s for Ethernet A-D per-ES routes, in compliance with the standard. The value of this field for Ethernet A-D per-EVI routes is the MPLS label used to forward unicast traffic in load balancing mode.

NOTE

Although the MPLS Label field of a per-ES route is all 0s according to the standard, by default, a device sets the MPLS Label field to an ESI label value. After the **peer esad-route-compatible enable** command is run on a device, the device advertises the Ethernet A-D per-ES routes with the MPLS Label field changed to all 0s.

- MAC/IP advertisement route: also known as the MAC/IP route. A MAC/IP advertisement route can carry the RD and ESI of an EVPN instance configured on the local PE and the VPN label assigned to the EVPN instance. [Figure 3](#) shows the NLRI of a MAC/IP advertisement route. The MAC/IP advertisement route contains information, such as the RTs and next hop of the EVPN instance, in addition to the NLRI. This type of route can be used for the local PE to advertise unicast MAC/IP address reachability to the other PEs. For details, see [Unicast MAC Address Transmission](#).

Figure 3 NLRI of a 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: RD of an EVPN instance.
- Ethernet Segment Identifier: uniquely identifies connections between PEs and a CE.
- Ethernet Tag ID: The value is all 0s in regular scenarios, the same as the local service ID in an EVPN VPWS scenario, or the same as the BD tag value in VLAN-aware BD EVPN access scenarios.
- MAC Address Length: length of a MAC address advertised in the route.
- MAC Address: MAC address advertised in the route.
- IP Address Length: mask length of a host IP address advertised in the route.
- IP Address: host IP address advertised in the route.
- MPLS Label1: label used for Layer 2 service traffic forwarding.
- MPLS Label2: label used for Layer 3 service traffic forwarding.

The functions of MAC/IP advertisement routes on the control plane are as follows:

- Host MAC address advertisement

To implement Layer 2 service exchanges between hosts connected to two PEs, the two PEs need to learn host MAC addresses 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 MAC Address Length and MAC Address fields identify a host MAC address.

- Host ARP advertisement

A MAC/IP advertisement route carries both the MAC address and IP address of a host. Therefore, this route can be used to transmit host ARP entries between PEs. The MAC Address and MAC Address Length fields identify a host MAC address, and the IP Address and IP Address Length fields identify a host IP address. In this case, MAC/IP advertisement routes are also called ARP routes.

- Host IP route advertisement

To implement Layer 3 service exchanges between IPv4 hosts connected to two 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 a MAC/IP advertisement route identify a host destination address, and the MPLS Label2 field must carry a label used for Layer 3 service forwarding. In such a scenario, MAC/IP advertisement routes are also called Integrate Routing and Bridge (IRB) routes.

 **NOTE**

An ARP route carries the following valid information: host MAC address, host IP address, and Layer 2 traffic forwarding label. IRB routes carry the following valid information: host MAC address, host IP address, Layer 2 traffic forwarding label, and Layer 3 traffic forwarding label. As a result, IRB routes include ARP routes and can be used to advertise both the host IP routes and host ARP entries.

- Host ND information advertisement

A MAC/IP advertisement route can carry both a host MAC address and a host IPv6 address. Such routes can be used to transmit and advertise host ND entries between PEs. The MAC Address and MAC Address Length fields identify a host MAC address, and the IP Address and IP Address Length fields identify a host IPv6 address. In such a scenario, MAC/IP advertisement routes are also called ND routes.

- Host IPv6 route advertisement

To implement Layer 3 service exchanges between IPv6 hosts connected to two 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 IPv6 addresses to each other. The IP Address Length and IP Address fields carried in the MAC/IP advertisement route identify a host IPv6 destination address, and the MPLS Label2 field must carry a label used for Layer 3 service traffic forwarding. In such a scenario, 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 traffic forwarding label. An IRBv6 route carries the following valid information: host MAC address, host IPv6 address, Layer 2 traffic forwarding label, and Layer 3 traffic forwarding label. As such, IRBv6 routes include ND routes and can be used to advertise both a host IPv6 route and ND entry.

- Inclusive multicast Ethernet tag route: also known as the IMET route. After a BGP peer relationship is established between PEs, the PEs exchange inclusive multicast routes. An inclusive multicast route carries the RD and route target (RT) of the EVPN instance on the local PE, source IP address (loopback address of the local PE) and provider multicast service interface (PMSI) information. The PMSI tunnel is used to carry the tunnel type (ingress replication or mLDP) and tunnel label used to transmit multicast packets. The PMSI and RT values are carried in routes as attributes, and the RD and source IP address are contained in NLRI information. [Figure 4](#) shows the NLRI of an inclusive multicast route. BUM traffic includes broadcast, multicast, and unknown unicast traffic. Upon receipt of BUM traffic, a PE forwards it to the other PEs in P2MP mode. The PEs use the inclusive multicast routes to establish tunnels. For details, see [BUM Packet Transmission](#).

Figure 4 NLRI of an 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: RD of an EVPN instance.
- Ethernet Tag ID: The value is all 0s in regular scenarios, the same as the local service ID in an EVPN VPWS scenario , or the same as the BD tag value in VLAN-aware BD EVPN access scenarios.
- IP Address Length: length of a 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.

NOTE

Currently, the EVPN source address on a PE can only be an IPv4 address. As such, this field is 4 bytes long.

- Ethernet segment route: carries the ESI, source IP address, and RD (source IP address:0) of the local PE. PEs connecting to the same CE use Ethernet segment routes to discover each other. This type of route is used in designated forwarder (DF) election. [Figure 5](#) shows the NLRI of an Ethernet segment route.

Figure 5 NLRI of an 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: in the format of X.X.X.X:0. X.X.X.X indicates the EVPN source IP address configured on the local PE.
- Ethernet Segment Identifier: uniquely identifies connections between PEs and a CE.
- IP Address Length: length of a 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.

NOTE

Currently, the EVPN source address on a PE can only be an IPv4 address. As such, this field is 4 bytes long.

- IP prefix route: used to advertise a host IP address received from an access network or the network segment where the host IP address resides. [Figure 6](#) shows the NLRI of an IP prefix route.

Figure 6 NLRI of an IP prefix route

Route Distinguisher (8 bytes)
Ethernet Segment Identifier (10 bytes)
Ethernet Tag ID (4 bytes)
IP Prefix Length (1 byte)
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:

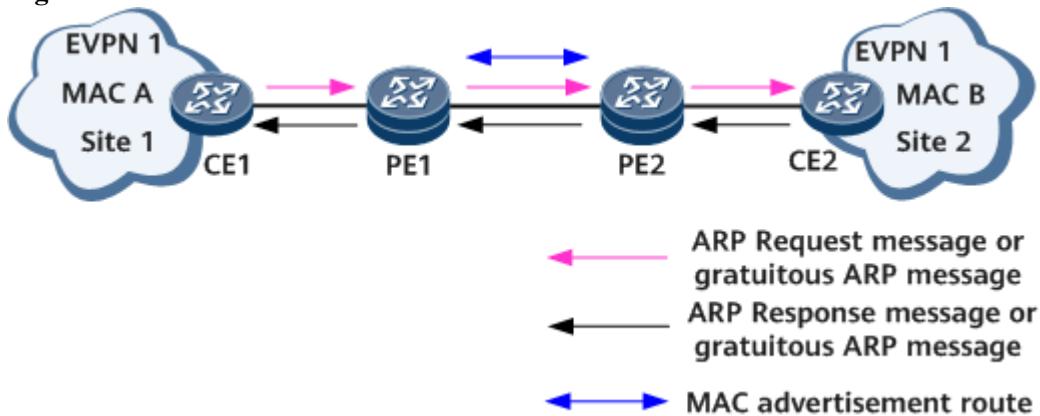
- Route Distinguisher: RD of an EVPN instance.
- Ethernet Segment Identifier: uniquely identifies connections between PEs and a CE.
- Ethernet Tag ID: Currently, this field can only be set to 0.
- IP Prefix Length: mask length of an IP prefix carried in the route.
- IP Prefix: IP prefix address.
- GW IP Address: default gateway IP address.
- MPLS Label: label used for Layer 3 service traffic forwarding.

Unicast MAC Address Advertisement

As shown in [Figure 7](#), the process of advertising a unicast MAC address is as follows:

1. Site 1 sends an ARP Request message or a gratuitous ARP message to advertise its MAC address (MAC A) and IP address to site 2. After the message arrives at PE1, PE1 generates a [MAC/IP advertisement route](#) for MAC A.
2. Site 2 responds to site 1 with an ARP Response message or a gratuitous ARP message carrying site 2's MAC address (MAC B) and IP address. After the message arrives at PE2, PE2 generates a [MAC/IP advertisement route](#) for MAC B.
3. PE1 and PE2 exchange [MAC/IP advertisement routes](#) that carry MAC addresses, next hops, and EVPN instance-based extended community attributes (such as RTs).
4. After PE1 and PE2 receive [MAC/IP advertisement routes](#) from each other, PE1 and PE2 find the corresponding EVPN instances based on the RT values carried in the routes. Then, PE1 and PE2 generate traffic forwarding entries in the EVPN instances based on the NLRI values carried in the routes for traffic transmission.

Figure 7 Unicast MAC address advertisement



Unicast packet transmission

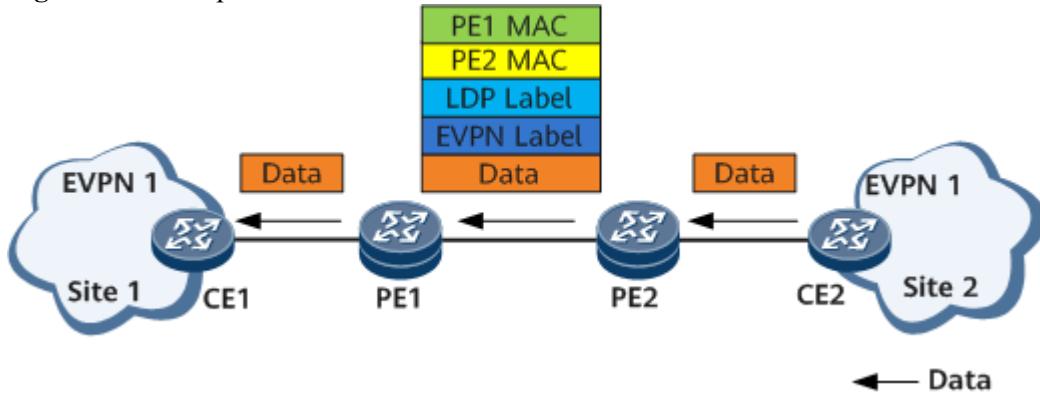
As shown in [Figure 8](#), after a local PE learns a MAC address from another site and successfully establishes a tunnel to the site over the public network, the local PE can transmit unicast packets to the site. The detailed transmission process is as follows:

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 the order into the unicast

packets. PE2 then forwards the encapsulated unicast packets to PE1.

3. After receiving the unicast packet after encapsulation, PE1 decapsulates the packet, locates the EVPN instance based on the EVPN label, searches the MAC table of the EVPN instance for an outbound interface mapped to the destination MAC address in the original packet, and forwards the unicast packet to the corresponding CE through the outbound interface.

Figure 8 Unicast packet transmission

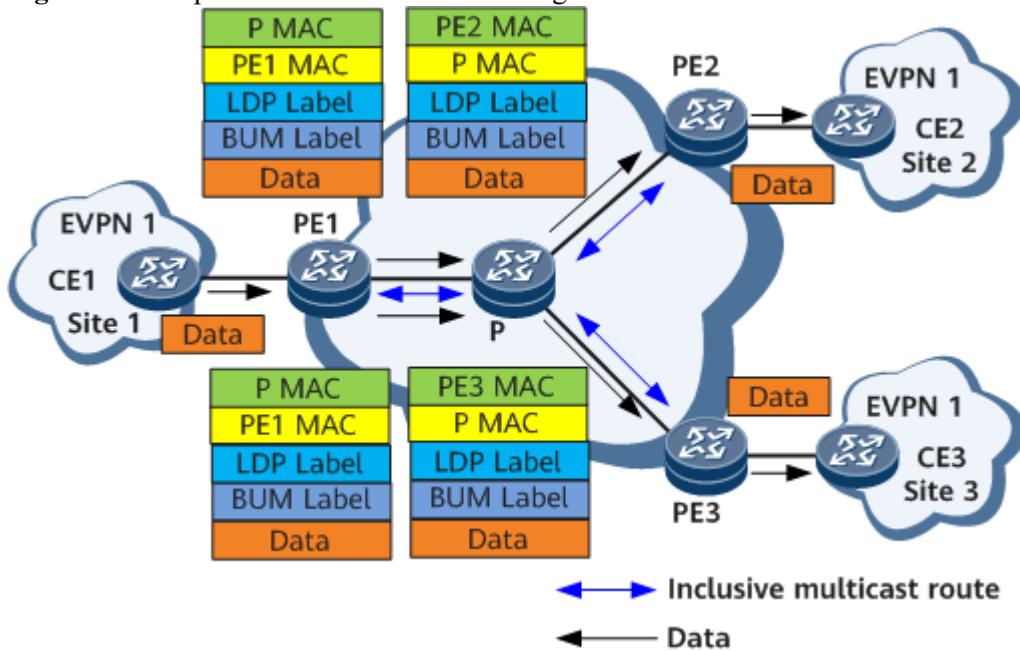


BUM Packet Transmission

After two PEs establish an EVPN BGP peer relationship, they exchange inclusive multicast routes. A PE can discover PEs that belong to the same EVPN instance as itself after matching RTs in the received inclusive multicast routes against the local EVPN instance, which enables the PE to obtain information about reachability to these PEs. This PE then automatically establishes MPLS tunnels with these PEs to carry BUM packets. On the network shown in [Figure 9](#), BUM packets are transmitted as follows:

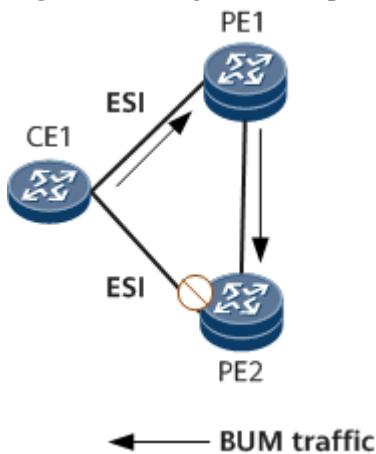
1. CE1 sends BUM packets to PE1.
2. Upon receipt of the BUM packets, PE1 forwards them to PE2 and PE3 that belong to the same EVPN instance. Specifically, PE1 replicates each BUM packet and encapsulates an EVPN BUM label, public network tunnel label, PE1's MAC address, and P's MAC address into each packet, and sends the packets to the remote PE.
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.

Figure 9 BUM packet transmission networking



In the case where a CE is dual-homed to two PEs, based on the split horizon mechanism, an EVPN ESI label will be encapsulated into the BUM packets exchanged between the two PEs to prevent loops. As shown in [Figure 10](#), CE1 is dual-homed to PE1 and PE2. After receiving a BUM packet from CE1, PE1 encapsulates the packet with an ESI and forwards the packet to PE2. When PE2 receives the BUM packet and finds that the ESI carried in the BUM packet is the same as the local ESI, PE2 discards the BUM packet to prevent a loop.

Figure 10 Using an ESI to prevent loops



Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.3 EVPN-MPLS

[EVPN Multi-Homing](#)

[Fundamentals of EVPN Seamless MPLS](#)

[EVPN Service Modes](#)

1.12.3.1 EVPN Multi-Homing

Related Concepts

Interface-based and VLAN-based DF election

As shown in [Figure 1](#), CE1 is dual-homed to PE1 and PE2. CE2 sends BUM traffic to PE1 and PE2. To prevent CE1 from receiving duplicate traffic from both PE1 and PE2, the EVPN DF election mechanism is introduced to specify either PE1 or PE2 to forward BUM traffic to CE1. 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.

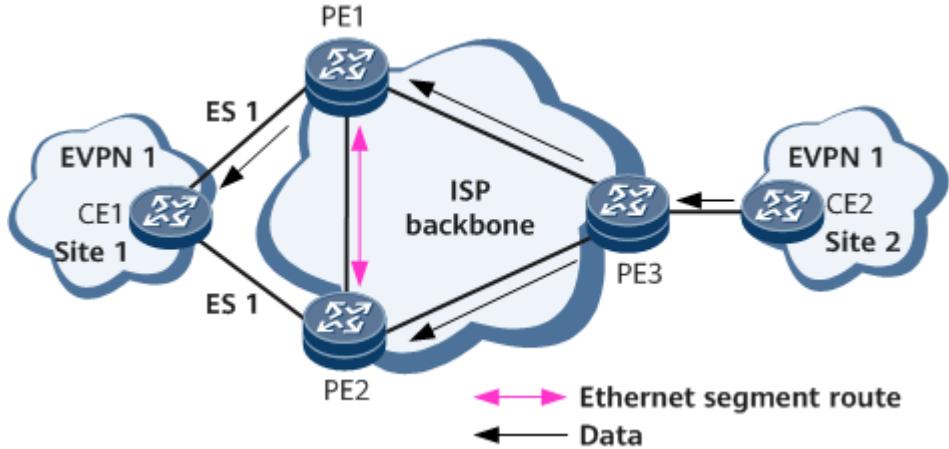
If a PE interface connecting to a CE goes down, the PE becomes a backup DF. If a PE interface connecting to a CE goes up, the PE and other PEs with up interfaces elect a primary DF. The DF election process is as follows:

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 the 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 in ascending order 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

If an Ethernet segment has multiple VLANs bound to, the smallest VLAN ID is used as the value for V .

Figure 1 DF election diagram

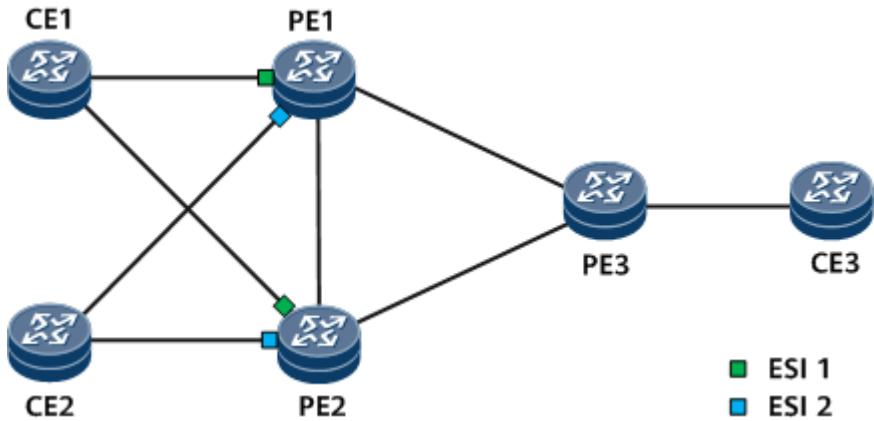


ESI-based DF priority election

On the network shown in [Figure 2](#), CE1 and CE2 are each dual-homed to PE1 and PE2. The PE interfaces that connect to CE1 have ESI 1 (shown in green), and the PE interfaces that connect to CE2 have ESI 2 (shown in blue). If you want CE3-to-CE1 BUM traffic and CE3-to-CE2 BUM traffic to be transmitted in load balancing and non-load balancing mode, respectively, you can set a redundancy mode per ESI instance. Specifically, you can set the redundancy mode of the ESI 1 instance to all-active and that of the ESI 2 instance to single-active.

If you want CE3-to-CE2 BUM traffic to be forwarded by PE1, configure ESI-based DF priority election on PE1 and PE2 and specify PE1 as the primary DF.

Figure 2 ESI-based DF priority election diagram



AC interface influenced DF election

In CE dual-homing networking, an AC interface's sub-interfaces on the access side are bound to an EVPN instance. If an ESI is set on the interface and one of the sub-interfaces goes down due to a fault or some other reason, the ESI remains valid because the other sub-interfaces bound to the EVPN instance remain up. As a result, the PE does not regenerate ES routes to trigger DF election, which may prevent the BUM traffic from being forwarded.

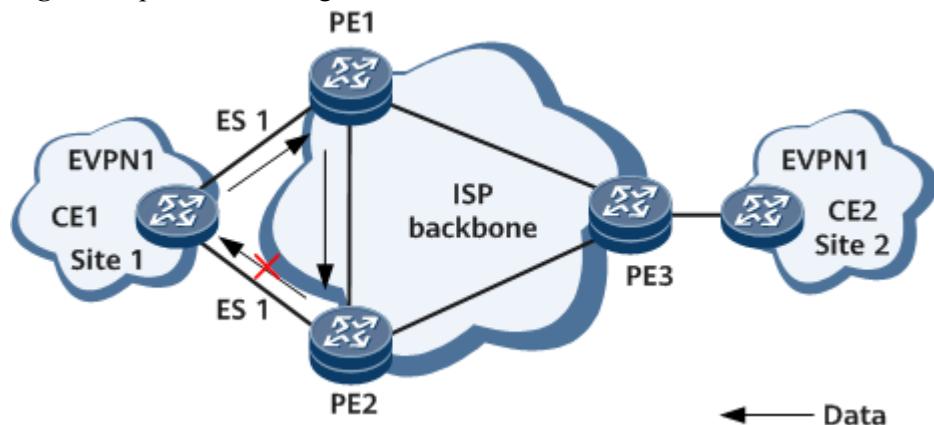
To resolve this issue, enable the function that the AC interface status influences DF election. This configuration helps check whether the PE has received the auto discovery (AD) routes from all PEs during DF election to determine whether these PEs are qualified for DF election. If a PE has not received the AD routes from a peer PE, the peer PE cannot participate in DF election.

Split horizon

On the network shown in [Figure 3](#), CE1 is dual-homed to PE1 and PE2 and has load balancing enabled. If PE1 and PE2 have established an EVPN BGP peer relationship, after PE1 receives BUM traffic from CE1, it forwards the BUM traffic to PE2. 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 3 Split horizon diagram

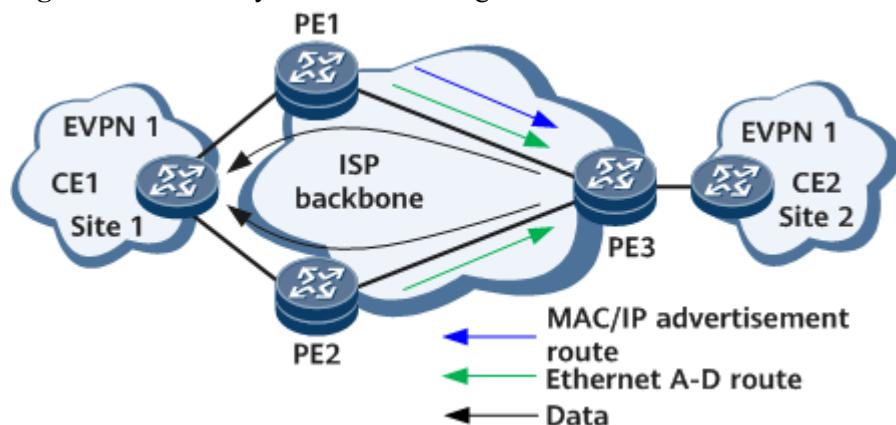


Redundancy mode and aliasing

If a CE is multi-homed to several PEs, a redundancy mode can be set to all-active or single-active for PEs connecting to the same CE. The redundancy mode determines whether load balancing is implemented for unicast traffic in CE multi-homing scenarios. If PE1 and PE2 are configured to work in all-active mode, after PE1 and PE2 send [Ethernet A-D routes](#) carrying the redundancy mode information to PE3, PE3 sends unicast traffic destined for CE1 to both PE1 and PE2 in load balancing mode.

EVPN also supports aliasing. When a CE is multi-homed to several PEs that work in all-active mode, some PEs may fail to learn the MAC addresses on the CE side. Aliasing enables remote PEs to learn the reachability to CE-side MAC addresses based on the ESIs carried in [Ethernet A-D routes](#) received from the multi-homing PEs. On the network shown in [Figure 4](#), only PE1 sends [MAC/IP advertisement routes](#) carrying CE1-side MAC addresses to PE3. However, PE3 can use [Ethernet A-D routes](#) to detect that PE2 can also reach CE1, implementing load balancing.

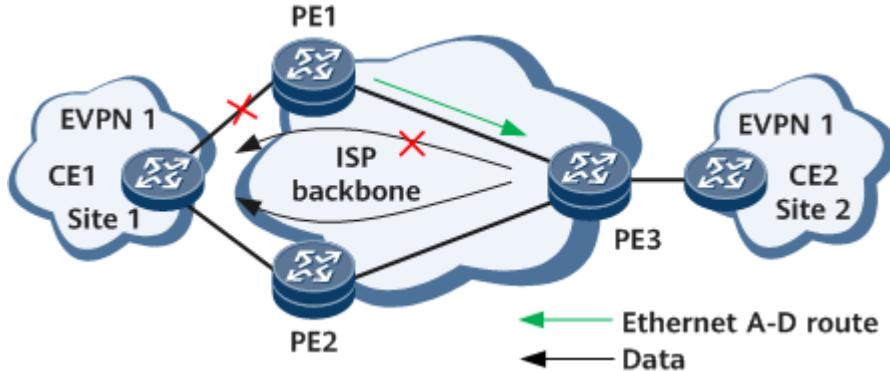
Figure 4 Redundancy mode and aliasing



Rapid convergence

On the network shown in [Figure 5](#), if the link between CE1 and PE1 fails, PE1 advertises an [Ethernet A-D route](#) to PE3 to inform that PE1 has become unreachable to site 1. Upon receipt of the route, PE3 withdraws the corresponding route and sends traffic to site 1 only through PE2, implementing rapid route convergence.

Figure 5 Rapid route convergence



Parent Topic: [EVPN-MPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.12.3.2 Fundamentals of EVPN Seamless MPLS

EVPN seamless MPLS establishes a BGP LSP across the access, aggregation, and core layers and transmits services along this BGP LSP in E2E mode. Service traffic can be transmitted between any two points over the LSP. The EVPN 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 converge services in case of any network-side faults, without affecting service transmission.

Background

The popularity of EVPN MPLS networks poses increasing requirements for the service scalability of the network architecture. The different metro networks of a service provider or the collaborative backbone networks of different service providers often span multiple ASs. In this case, EVPN seamless MPLS can be used to establish an inter-AS E2E BGP LSP to carry EVPN services.

Implementation

On a seamless MPLS network, the EVPN services to be transmitted need to be encapsulated using signaling only at service access points. In addition, if a network-side fault triggers EVPN service convergence, the same transport layer convergence technology is used to converge the services, without the service layer being aware of the fault.

Application Scenarios

EVPN seamless MPLS supports the following networking solutions:

- EVPN intra-AS seamless MPLS: The access, aggregation, and core layers are deployed within a single AS. This solution mainly applies to mobile transport networks.
- EVPN inter-AS seamless MPLS: The access and aggregation layers are deployed in a single AS, whereas the core layer is deployed in a different AS. This solution mainly applies to

enterprise networks.

EVPN Intra-AS Seamless MPLS

Table 1 EVPN intra-AS seamless MPLS networking

Network Deployment	Description
Control plane	<p>Deploying routing protocols</p> <p>In Figure 1, routing protocol deployment on devices is as follows:</p> <ul style="list-style-type: none"> An IGP (IS-IS or OSPF) is enabled on devices at each of the access, aggregation, and core layers to implement intra-AS network connectivity. An IBGP peer relationship is established between each of the following pairs of devices: <ul style="list-style-type: none"> CSG and AGG AGG and core ABR Core ABR and MASG <p>The AGGs and core ABRs are configured as RRs to reflect routes to CSGs and MASGs, respectively, so that the CSGs and MASGs can obtain the route to each other's loopback address.</p> <ul style="list-style-type: none"> The AGGs and core ABRs set the next-hop IP addresses in BGP routes to their own addresses to prevent the public routes of other IGP areas from being advertised. <p>Figure 1 Deploying routing protocols for the EVPN intra-AS seamless MPLS networking</p>
BGP EVPN peer relationship establishment and route advertisement	<p>As shown in Figure 1, AGGs and core ABRs function as RRs. BGP EVPN peer relationships need to be established between CSGs and AGGs, between MASGs and core ABRs, and between AGGs and core ABRs. Then, EVPN MAC/IP routes (Type 2) and IP prefix routes (Type 5) need to be transmitted between the peers to transmit MAC and IP routing information.</p>

Network Deployment	Description
Deploying tunnels	<p>On the network shown in Figure 2, tunnels are deployed as follows:</p> <ul style="list-style-type: none"> • A public network tunnel is established using LDP, TE, or LDP over TE in each IGP area. • An IBGP peer relationship is established between each of the following pairs of devices: <ul style="list-style-type: none"> ▪ CSG and AGG ▪ AGG and core ABR ▪ Core ABR and MASG <p>These devices are enabled to advertise labeled routes and assign labels to BGP routes that match a specified route-policy. After the devices exchange labeled BGP routes, an E2E BGP LSP is established between each pair of a CSG and MASG.</p> <p>Figure 2 Deploying tunnels for the EVPN intra-AS seamless MPLS networking</p>

Network Deployment	Description
Forwarding plane	<p>Figure 3 illustrates the forwarding plane of EVPN intra-AS seamless MPLS networking. Seamless MPLS is mainly used to transmit EVPN packets. The following example demonstrates how EVPN packets, including labels and packet content, are transmitted from a CSG to an MASG along the path CSG1->AGG1->core ABR1->MASG1.</p> <ol style="list-style-type: none"> 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. 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) 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. 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 core ABR pushes a core-layer MPLS tunnel label to each packet and forwards the packets to the MASG. 4. 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. <p>Figure 3 Forwarding plane for the EVPN intra-AS seamless MPLS networking</p>

EVPN Inter-AS Seamless MPLS

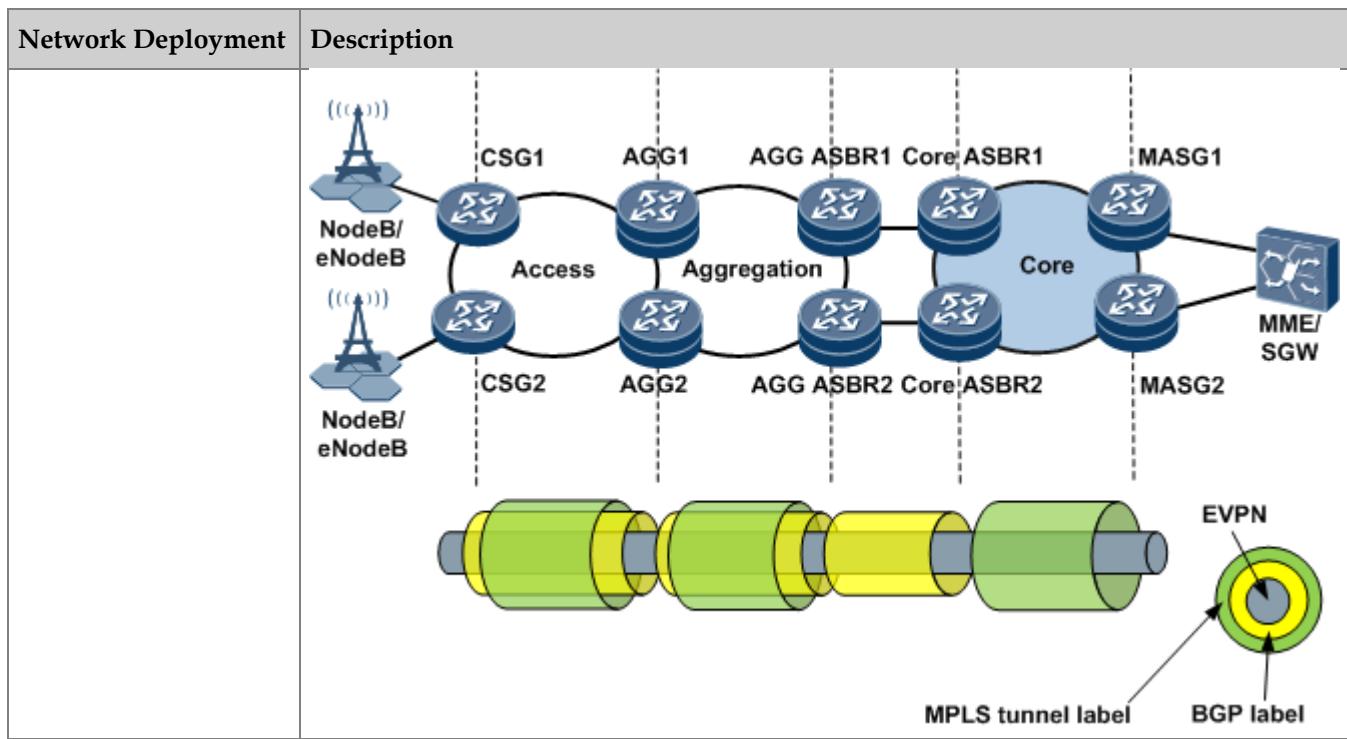
Table 2 EVPN inter-AS seamless MPLS networking

Network Deployment	Description

Network Deployment	Description
Control plane Deploying routing protocols	<p>In Figure 4, routing protocol deployment on devices is as follows:</p> <ul style="list-style-type: none"> An IGP (IS-IS or OSPF) is enabled on devices at each of access, aggregation, and core layers to implement intra-AS network connectivity. A BGP peer relationship is established between each of the following pairs of devices: <ul style="list-style-type: none"> CSG and AGG AGG and AGG ASBR AGG ASBR and core ASBR Core ASBR and MASG <p>An EBGP peer relationship is established between the AGG ASBR and core ASBR, and IBGP peer relationships are established between the other pairs of devices.</p> <ul style="list-style-type: none"> The AGGs are configured as RRs to reflect routes so that IBGP peers can exchange BGP routes, and the CSGs and MASGs can obtain BGP routes destined for each other's loopback addresses. If the AGG ASBR and core ASBR are connected indirectly, an IGP neighbor relationship between them must be established to implement inter-area connectivity. <p>Figure 4 Deploying routing protocols for the EVPN inter-AS seamless MPLS networking</p>
BGP EVPN peer relationship establishment and route advertisement	<p>On the network shown in Figure 4, BGP EVPN peer relationships need to be established between the following pairs of devices:</p> <ul style="list-style-type: none"> CSG and AGG AGG and AGG ASBR AGG ASBR and core ASBR Core ASBR and MASG <p>The peers exchange MAC/IP routes (Type 2) and IP prefix routes (Type 5) to advertise MAC and IP routing information.</p>

Network Deployment	Description
Deploying tunnels	<p>On the network shown in Figure 5, tunnels are deployed as follows:</p> <ul style="list-style-type: none"> • A public network tunnel is established using LDP, TE, or LDP over TE in each IGP area. An LDP LSP or a TE LSP must be established if more than one hop exists between each pair of an AGG ASBR and core ASBR. • The CSGs, AGGs, AGG ASBRs, and core ASBRs are enabled to advertise labeled routes and assign labels to BGP routes that match a specified route-policy. After the devices exchange labeled BGP routes, a BGP LSP is established between each pair of a CSG and core ASBR. • Either of the following tunnel deployment methods can be used in the core area: <ul style="list-style-type: none"> ▪ A BGP LSP between a core ASBR and MASG is combined with the BGP LSP between the CSG and core ASBR to form an E2E BGP LSP. The route to the MASG's loopback address is installed 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. ▪ 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 installs the route to the routing table. The core ASBR assigns a BGP label to the route and associates the route with an intra-AS MPLS tunnel. The BGP LSP between the CSG and core ASBR and the MPLS tunnel in the core area are combined into an E2E tunnel.
Forwarding plane	<p>Figure 5 Deploying tunnels for the EVPN inter-AS seamless MPLS networking</p> <p>Figure 6 illustrates the forwarding plane of the EVPN inter-AS seamless MPLS networking with a core-layer BGP LSP established. EVPN seamless MPLS is mainly used to transmit EVPN packets. The following example demonstrates how EVPN packets, including VPN labels and packet data, are transmitted from a CSG to an MASG along the path CSG1->AGG1->AGG ASBR1->core ASBR1->MASG1.</p> <ol style="list-style-type: none"> 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. 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

Network Deployment	Description
	<p>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"> 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 therefore, the AGG ASBR receives packets without MPLS tunnel labels. 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. 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 MASG, the core ASBR has removed the MPLS tunnel labels from the packets, and therefore, the MASG receives packets without MPLS tunnel labels. The EVPN packet transmission along the inter-AS seamless MPLS LSP is complete. <p>Figure 6 Forwarding plane for the EVPN inter-AS seamless MPLS networking with a core-layer BGP LSP established</p> <p>Figure 7 illustrates the forwarding plane for the EVPN 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 in the core area. The difference is that without a BGP LSP in the core area, the core ASBR removes (rather than swaps) BGP labels from packets and pushes MPLS tunnel labels into these packets.</p> <p>Figure 7 Forwarding plane for the EVPN inter-AS seamless MPLS networking without a BGP LSP established in the core area</p>



Reliability

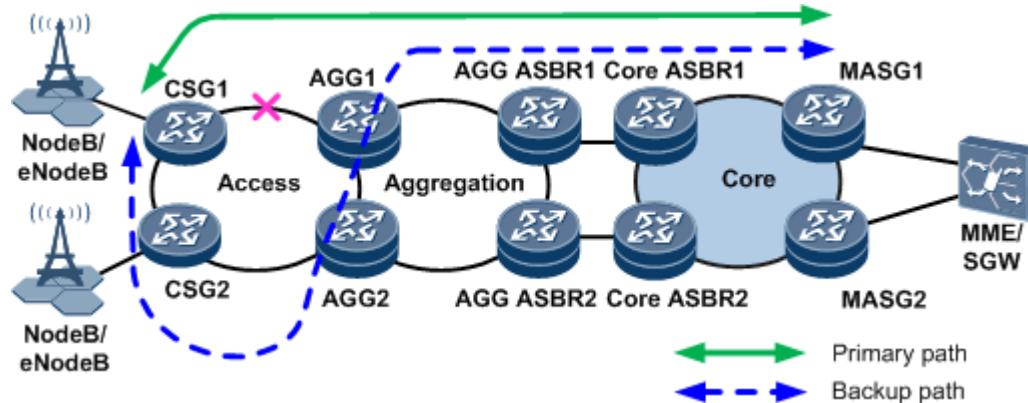
EVPN seamless MPLS network reliability can be improved using a variety of functions. If a network fault occurs, devices with reliability functions enabled immediately detect the fault and switch traffic from the active link to the standby link.

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

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

On the EVPN inter-AS seamless MPLS network shown in [Figure 8](#), 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.

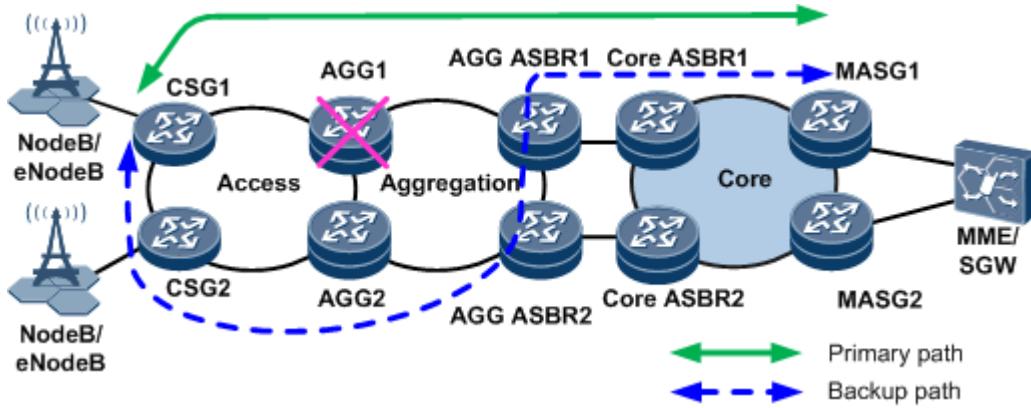
Figure 8 Traffic protection triggered by a fault of the CSG-AGG link on the EVPN inter-AS seamless MPLS network



- A fault occurs on an AGG.

On the EVPN inter-AS seamless MPLS network shown in [Figure 9](#), 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 an AGG1 fault, the BFD module switches traffic from the primary path to the backup path.

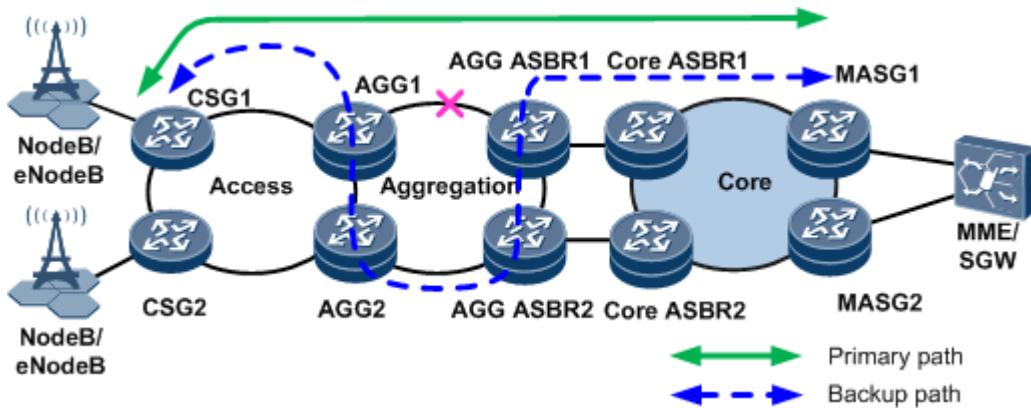
Figure 9 Traffic protection triggered by an AGG fault on the EVPN inter-AS seamless MPLS network



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

On the EVPN inter-AS seamless MPLS network shown in [Figure 10](#), 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 instructs LDP FRR, TE hot standby, or BGP FRR to switch traffic from the primary path to the backup path.

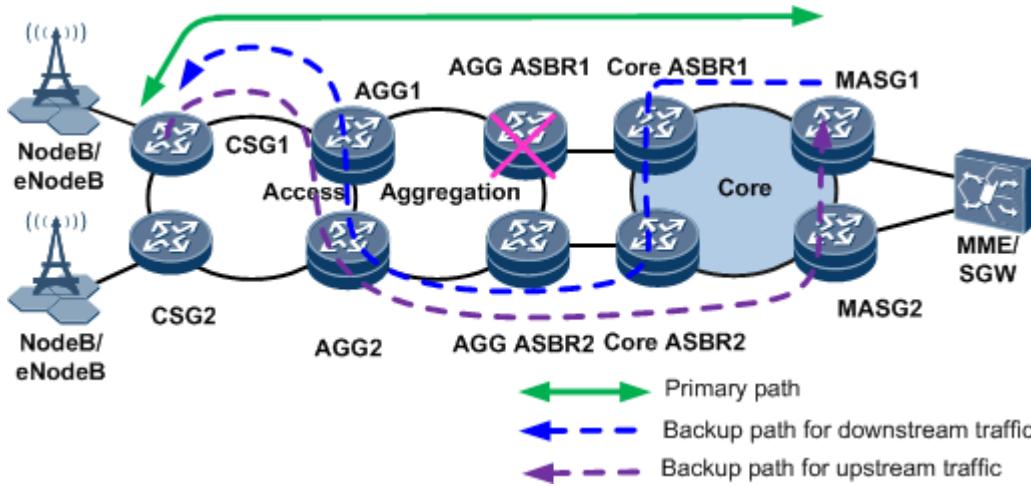
Figure 10 Traffic protection triggered by a fault of the link between an AGG and an AGG ASBR on the inter-AS seamless MPLS network



- A fault occurs on an AGG ASBR.

As shown in [Figure 11](#), 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 the BGP Auto FRR function. BGP Auto FRR switches both upstream and downstream traffic from the primary path to backup paths.

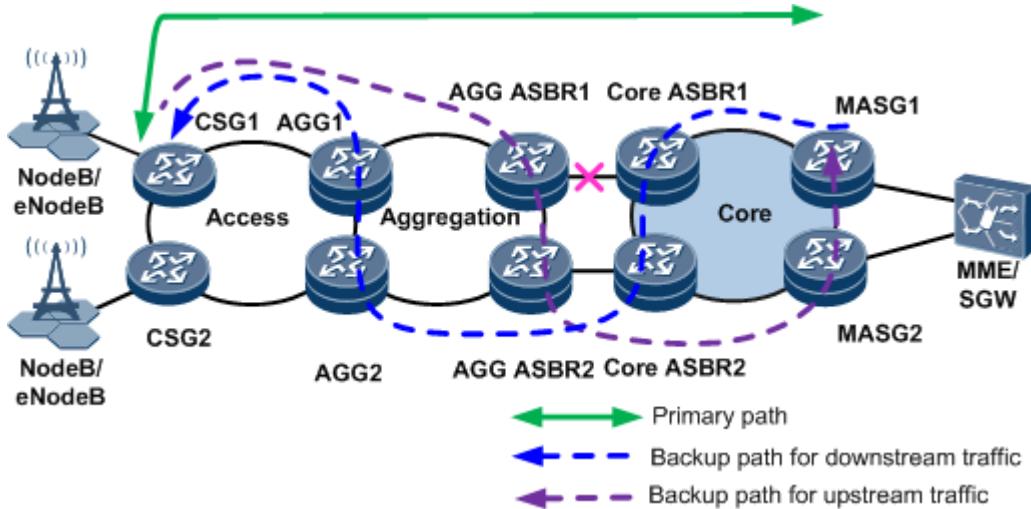
Figure 11 Traffic protection triggered by a fault of an AGG ASBR on the EVPN inter-AS seamless MPLS network



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

As shown in [Figure 12](#), BFD for interface is configured on AGG ASBR1 and core ASBR1. If the BFD module detects a fault of the link between AGG ASBR1 and core ASBR1, the BFD module triggers the BGP Auto FRR function. BGP Auto FRR switches both upstream and downstream traffic from the primary path to backup paths.

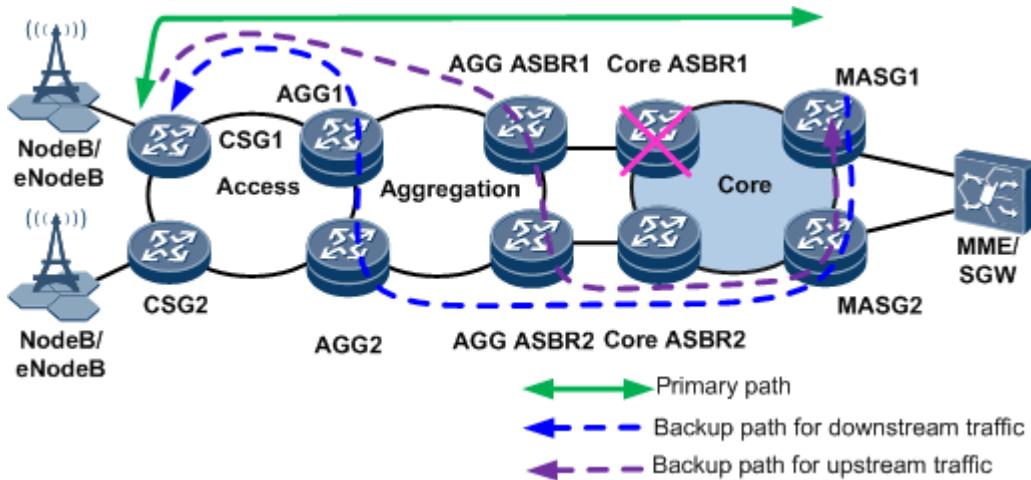
Figure 12 Traffic protection triggered by a fault of the link between an AGG ASBR and a core ASBR on the EVPN inter-AS seamless MPLS network



- A fault occurs on a core ASBR.

On the EVPN inter-AS seamless MPLS network shown in [Figure 13](#), 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 switches both upstream and downstream traffic from the primary path to backup paths.

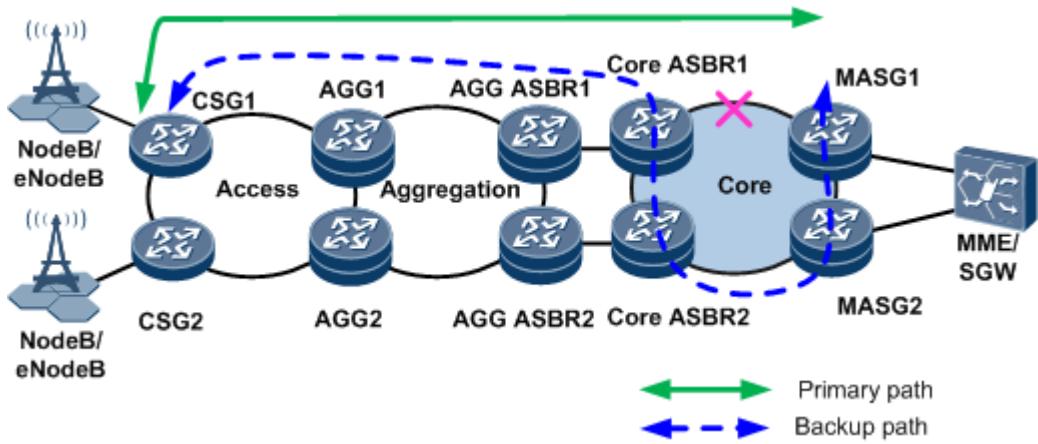
Figure 13 Traffic protection triggered by a fault of a core ASBR on the EVPN inter-AS seamless MPLS network



- A link fault occurs in the core area.

On the EVPN inter-AS seamless MPLS network shown in [Figure 14](#), 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 instructs the LDP FRR, TE hot standby, or BGP FRR function to switch both upstream and downstream traffic from the primary path to the backup paths.

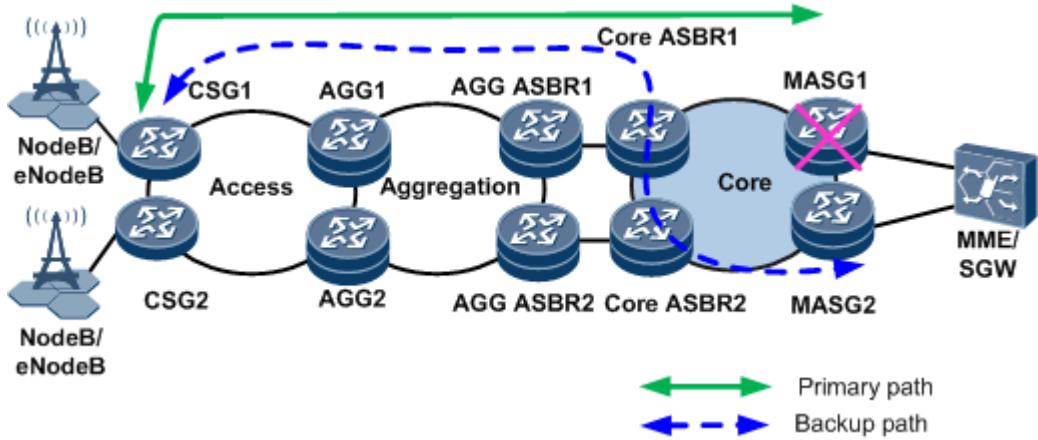
Figure 14 Traffic protection from a link fault in the core area on the EVPN inter-AS seamless MPLS network



- A fault occurs on an MASG.

As shown in [Figure 15](#), BFD for BGP tunnel is configured on CSG1. BFD for BGP tunnel is implemented in compliance with a standard titled "Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)." BFD for BGP tunnel monitors E2E BGP LSPs, including a BGP LSP stitched with an LDP LSP. If MASG1 that functions as a remote PE fails, BFD for BGP LSP can rapidly detect the fault and trigger VPN FRR switching. The BFD module then switches both upstream and downstream traffic from the primary path to the backup path.

Figure 15 Traffic protection triggered by a fault of an MASG on the EVPN inter-AS seamless MPLS network

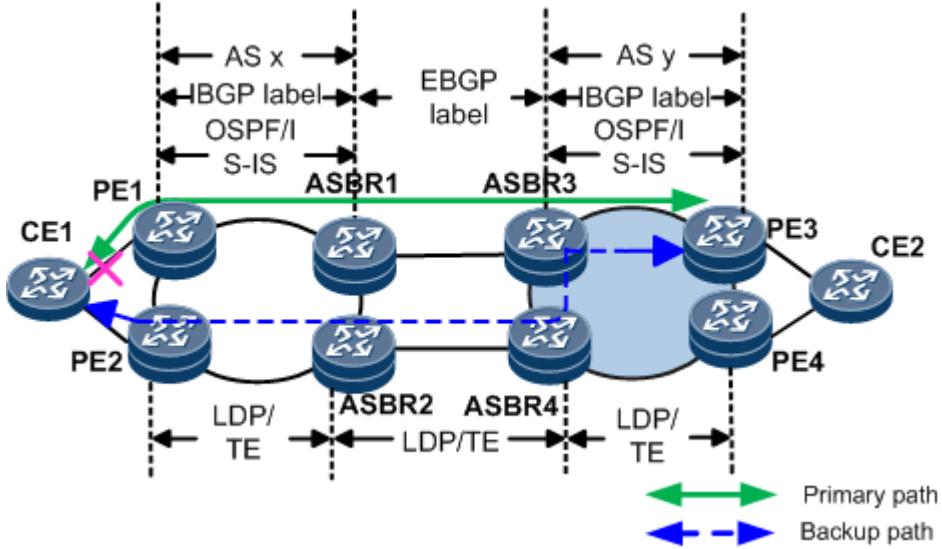


- A fault occurs on an access-side link.

On the inter-AS seamless MPLS network shown in [Figure 16](#), if an E-Trunk in single-active mode detects a link failure, the E-Trunk switches traffic from the primary path to the backup path and PE2's interface connected to CE1 is unblocked. Then upstream traffic on CE1 is switched to PE2. For BUM traffic on the network side, PE1 sends a per-ES A-D route withdraw message to PE2, and PE2 is elected as the DF to forward BUM traffic. After receiving the MAC route advertised by PE2, PE3 switches unicast traffic to PE2.

If an E-Trunk in active-active mode detects a link failure, PE1 sends a per-ES A-D route withdraw message to PE3, and PE3 switches unicast traffic to PE2.

Figure 16 Traffic protection triggered by an access-side link fault on the EVPN inter-AS seamless MPLS network



- A PE on the access side fails.

If PE1 fails, the original EVPN detection mechanism is triggered, which is similar to that triggered when an access-side link fails. Other PEs switch traffic after detecting PE1's down state rather than receiving a route withdraw request.

Parent Topic: [EVPN-MPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.3.3 EVPN Service Modes

Multiple Ethernet VPN instances (EVIs) can be configured on PEs at the edge of an EVPN, with each EVI connecting to one or more user networks. EVPN allows user network access in various service modes, as described in the following table.

Table 1 EVPN service modes

Service Mode	Application Scenario
Port Based	The physical interface connected to a user network is directly bound to a common EVI. This service mode is used to carry only Layer 2 services.
VLAN Based	The physical interface connected to a user network is divided into different sub-interfaces. Each sub-interface is added to a specific BD, and each BD is bound to a specific EVI. One EVI is required per user. This service mode is used to carry Layer 2 or Layer 3 services.
VLAN Bundle	Users are divided based on VLANs. Each VLAN is added to a specific BD, and each BD is bound to a specific EVI. This service mode is used to carry Layer 2 or Layer 3 services.
VLAN-Aware Bundle	Users are divided based on VLANs. Each VLAN is added to a specific BD, and all these BDs are bound to the same EVI. This service mode is used to carry Layer 2 or Layer 3 services.

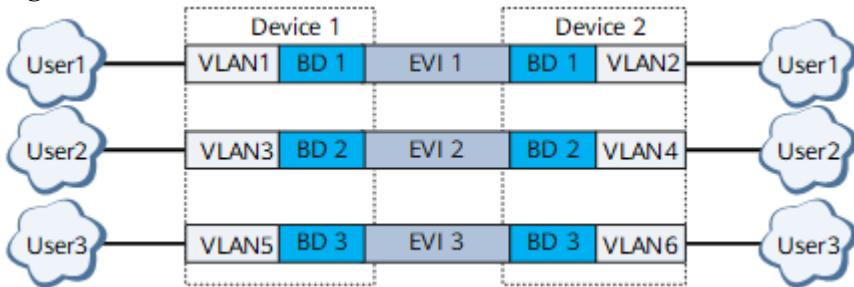
Port Based

In port-based mode, an entire interface is used for service access. Specifically, the physical interface connected to a user network is directly bound to a common EVI (not an EVI in BD or VPWS mode) and has no sub-interfaces created. This service mode is used to carry Layer 2 services.

VLAN Based

On the network shown in [Figure 1](#), in VLAN-based mode, the physical interfaces connected to user networks each have different sub-interfaces created. Each sub-interface is associated with a unique VLAN and added to a specific BD, and each BD is bound to a specific EVI. In this service mode, the sub-interface, VLAN, BD, and EVI are exclusively used by a user to access the network, and a separate MAC forwarding table is used on the forwarding plane for each user. Although this mode effectively ensures service isolation, it consumes a large amount of EVI resources because each user requires one EVI. This service mode is used to carry Layer 2 or Layer 3 services.

Figure 1 VLAN-based mode



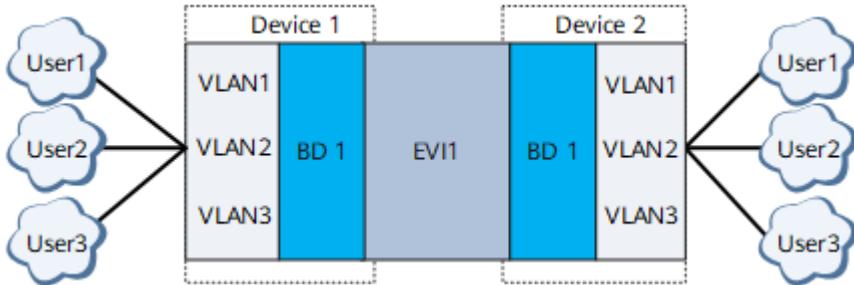
VLAN Bundle

On the network shown in [Figure 2](#), in VLAN bundle mode, an EVI connects to multiple users that 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 EVC VLAN tag termination sub-interfaces support both Layer 2 and Layer 3 interfaces. Other EVC sub-interfaces support only Layer 2 services.

Figure 2 VLAN bundle mode

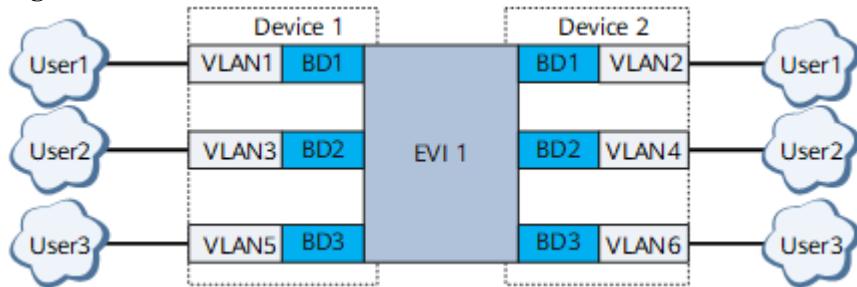


VLAN-Aware Bundle

On the network shown in [Figure 3](#), in VLAN-aware bundle mode, an EVI connects to multiple users divided by VLAN. Additionally, the EVI can be bound to multiple BDs. In this case, the EVI must have a different BD tag configured when being bound to a BD. When EVPN peers send routes to each other, a BD tag is encapsulated into the Ethernet Tag ID field of Ethernet auto-discovery route, MAC/IP advertisement route, and inclusive multicast route packets. In this service mode, users connected to the same EVI use separate forwarding entries. During traffic forwarding, the system uses

the BD tag carried in packets to locate the corresponding BD MAC forwarding table and searches the table for a forwarding entry based on a MAC address.

Figure 3 VLAN-aware bundle mode



Unlike other service modes, the VLAN-aware bundle mode is implemented based on BDs in terms of load balancing, designated forwarder (DF) election, host migration, and route re-origination.

- 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, load balancing cannot be implemented because the routes belong to different BDs.
- DF election:
 - For interface-based DF election, the system chooses the first interface to go up in a BD for DF election.
 - If AC interfaces are enabled to influence DF election, a PE cannot participate in DF election if the system does not receive any Ethernet auto-discovery route advertised by the PE. In this scenario, if the VLAN-aware bundle mode is enabled, an Ethernet auto-discovery route is generated for each BD tag. As such, a 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 an identical route from the remote end. If it has, the system adds the MAC address transfer attribute to the locally generated route or increments the value of the Sequence field in the MAC address transfer attribute 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 when checking MAC/IP advertisement routes. This prevents host migration failures caused by MAC address conflicts between different BDs.
- Route re-origination: In the 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. If the VLAN-aware bundle mode is enabled on a DCI-PE, the DCI-PE can re-originate a MAC/IP advertisement route only if the Ethernet tag ID is consistent with the BD tag in the route.

Parent Topic: [EVPN-MPLS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.12.4 EVPN-VXLAN

1.12.4.1 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 a control plane, and VTEP discovery and host information (IP and MAC addresses, VNIs, and gateway VTEP IP address) 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.

In summary, EVPN introduces several new types of BGP EVPN routes through BGP extension to advertise VTEP addresses and host information. In this way, EVPN applied to VXLAN networks enables VTEP discovery and host information learning on the control plane instead of on the data 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 1](#) shows the format of a MAC/IP route.

Figure 1 Format of a 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)

[Table 1](#) describes the meaning of each field.

Table 1 Fields of a MAC/IP route

Field	Description
Route Distinguisher	RD value set in an EVI

Field	Description
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	Length of the host IP address carried in the route
IP Address	Host IP address carried in the route
MPLS Label1	L2VNI carried in the route
MPLS Label2	L3VNI 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 field identifies 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 field identifies the MAC address of the host, whereas the IP Address field identifies 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 field identifies the destination address of the IP route. In addition, the MPLS Label2 field must carry the L3VNI. 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 an L2VNI. An IRB route carries host MAC and IP addresses, an L2VNI, and an L3VNI. 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 functioning as BGP EVPN peers exchange MAC/IP routes to advertise host IPv6 routes to each other. The IP Address field carried in the MAC/IP routes indicates the destination addresses of host IPv6 routes, and the MPLS Label2 field must carry an L3VNI. MAC/IP routes in this case are also called IRBv6 routes.

NOTE

An ND route carries host MAC and IPv6 addresses and an L2VNI. An IRBv6 route carries host MAC and IPv6 addresses, an L2VNI, and an L3VNI. Therefore, IRBv6 routes carry ND routes and can be used to advertise both host IPv6 routes and ND entries.

Type 3 Route: Inclusive Multicast Route

An inclusive multicast route comprises a prefix and a PMSI attribute. [Figure 2](#) shows the format of an inclusive multicast route.

Figure 2 Format of an inclusive multicast route

Prefix

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

PMSI attribute

Flags (1 byte)
Tunnel Type (1 byte)
MPLS Label (3 bytes)
Tunnel Identifier (variable)

[Table 2](#) describes the meaning of each field.

Table 2 Fields of an inclusive multicast route

Field	Description
Route Distinguisher	RD value set in an EVI.
Ethernet Tag ID	VLAN ID, which is all 0s in this type of route.
IP Address Length	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 the tunnel. This field is inapplicable in VXLAN scenarios.

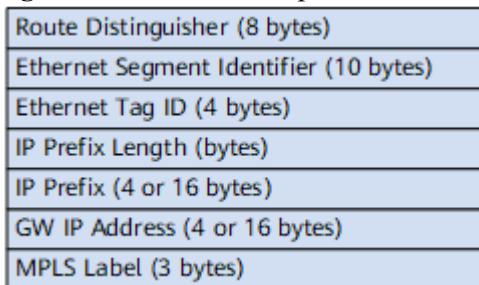
Field	Description
Tunnel Type	Tunnel type carried in the route. The value can only be 6, representing Ingress Replication in VXLAN scenarios. It is used for BUM packet forwarding.
MPLS Label	L2VNI carried in the route.
Tunnel Identifier	Tunnel identifier carried in the route. This field is the local VTEP's IP address in VXLAN scenarios.

Inclusive multicast routes are used on the VXLAN control plane for automatic VTEP discovery and dynamic VXLAN tunnel establishment. VTEPs that function as BGP EVPN peers transmit L2VNIs 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 an L2VNI. If the remote VTEP's IP address is reachable at Layer 3, a VXLAN tunnel to the remote VTEP is established. In addition, the local end creates a VNI-based ingress replication list and adds the peer VTEP IP address to the list for subsequent BUM packet forwarding.

Type 5 Route: IP Prefix Route

[Figure 3](#) shows the format of an IP prefix route.

Figure 3 Format of an IP prefix route



[Table 3](#) describes the meaning of each field.

Table 3 Fields of an IP prefix route

Field	Description
Route Distinguisher	RD value set in a VPN instance
Ethernet Segment Identifier	Unique ID for defining the connection between local and remote devices
Ethernet Tag ID	Currently, this field can only be set to 0
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
MPLS Label	L3VNI carried in the route

An IP prefix route can carry either a host IP address or a network segment address.

- When carrying 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.
- When carrying a network segment address, the route can be advertised to allow hosts on a VXLAN network to access the specified network segment or external network.

Parent Topic: [EVPN-VXLAN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

1.12.5 EVPN VPWS

[EVPN VPWS Fundamentals](#)

Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.5.1 EVPN VPWS Fundamentals

Introduction

The EVPN virtual private wire service (VPWS) solution provides a P2P L2VPN service based on the EVPN service architecture. This solution reuses and simplifies the original EVPN technology, uses the MPLS tunnel technology to traverse the backbone network, and provides a Layer 2 packet forwarding mode for connections between access circuits (ACs) without searching for MAC forwarding entries.

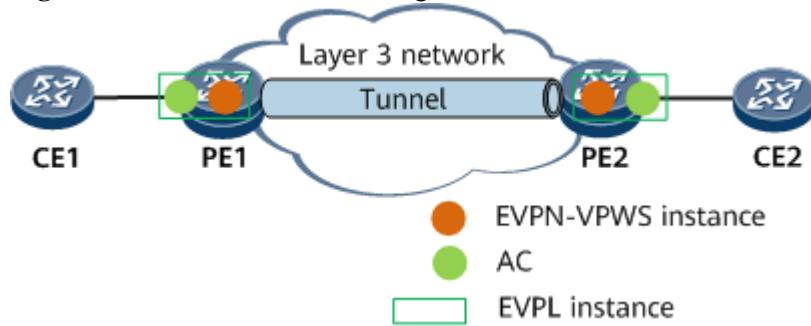
As shown in [Figure 1](#), the basic EVPN VPWS architecture consists of the following parts:

- AC: an independent link or circuit that connects a CE to a PE. An AC interface can be a physical or virtual interface. The AC attributes include the encapsulation type, maximum transmission unit (MTU), and interface parameters of the specified link type.
- EVPL instance: An EVPL instance corresponds to an AC. Each EVPL instance has a service ID. An EVPL instance on the local PE corresponds to an EVPL instance on the remote 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 P2P 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: network-side MPLS tunnel or SR tunnel.

Compared with the conventional L2VPN VPWS solution, the EVPN VPWS solution simplifies the control and data models and uses BGP as the control plane where BGP route selection and the BGP

next hop recursion are used to select traffic paths over backbone networks. This eliminates the need of specifying PWs.

Figure 1 EVPN VPWS networking



Routes Used in EVPN VPWS

On the basis of BGP, EVPN defines a new type of NLRI, which is called the EVPN NLRI. EVPN VPWS supports the following types of EVPN NLRI:

- Ethernet auto-discovery route: also known as the Ethernet A-D route. Ethernet A-D routes are classified into per-ES routes and per-EVI 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 2](#) shows the NLRI format of Ethernet A-D per-EVI routes.

Figure 2 NLRI of an Ethernet A-D route

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:

- Route Distinguisher: 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 connections between PEs and a CE.
- Ethernet Tag ID: local service ID of the EVPL instance on the local PE.
- MPLS Label: For a non-SRv6 tunnel, this field is the EVPL label assigned based on each Ethernet A-D per-EVI route. For an SRv6 tunnel, this field is the SRv6 SID of this tunnel.

In addition to the NLRI, Ethernet A-D per-EVI routes also carry carries the Layer 2 extended community attribute that includes the following control fields:

- C: a control word identifier. If this field is set to 1, packets sent by the local PE must carry control information.

- P: indicates whether the local PE is the master PE. In all-active scenarios, this control field must be set to 1.
- B: indicates whether a PE is a backup PE in dual-homing single-active scenarios.
- Ethernet Segment (ES) route: carries the RD, ESI, and source IP address of the local PE to implement automatic discovery and DF election between PEs connecting to the same CE. [Figure 3](#) shows the NLRI of an Ethernet segment route.

Figure 3 NLRI of an 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: in the format of X.X.X.X:0. X.X.X.X indicates the EVPN source IP address configured on the local PE.
- Ethernet Segment Identifier: uniquely identifies connections between PEs and a CE.
- IP Address Length: length of a source IP address configured on the local PE.
- Originating Router's IP Address: source IP address configured on the local PE.

NOTE

Currently, the EVPN source address on the PE supports only IPv4. Therefore, the field contains only 4 bytes.

Protocol Packet Exchange Process in the Single-Homing Scenario

[Figure 1](#) shows the protocol packet exchange process in the EVPN VPWS single-homing scenario.

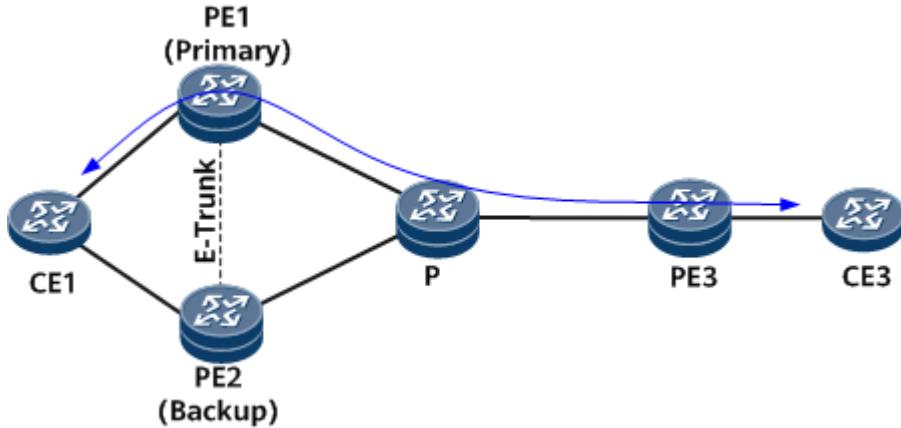
1. PE1 and PE2 are each configured with an EVPL instance and an EVPN VPWS instance. The EVPL instance must be bound to an AC interface and an EVPN VPWS instance, and each EVPL instance must be assigned a local service ID and a remote service ID. After the configuration, the local PE generates a forwarding entry indicating the association between the AC interface and EVPL instance.
2. PE1 and PE2 each send Ethernet A-D per-EVI routes to each other. An Ethernet A-D per-EVI route carries the RD, RTs, next-hop information, local service ID, and EVPL label or SRv6 SID.
3. PE1 and PE2 receive the Ethernet A-D per-EVI routes from each other, match the RTs of the routes, and import the routes to the corresponding EVPN VPWS instances. The routes then recure to an MPLS or SRv4 tunnel based on next hop information or to SRv6 tunnels based on SRv6 SIDs. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, a forwarding entry indicating the association between the MPLS or SRv4/v6 tunnel and local EVPL instance is generated.

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

On the network shown in [Figure 4](#), a CE is dual-homed to PE1 and PE2. PE1 and PE2 work in single-active mode and are configured with an E-Trunk. In this case, the master/backup relationship between PE1 and PE2 is determined by the E-Trunk configured between PE1 and PE2. The protocol 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 assigned a local service ID and a remote service ID. After the configuration, the local PE generates a forwarding entry 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 exchange ES routes that carry RDs, RTs, ESIs, and source IP addresses. After receiving the ES routes, PE1 and PE2 trigger DF election. The active/standby status of PE1 and PE2 is determined by the E-Trunk configured between PE1 and PE2. In this example, PE1 is the active device, and PE2 is the standby device.
3. PE1 and PE2 send PE3 the Ethernet A-D per-ES routes that carry the RD, RTs, next-hop information, and single-active mode information.
4. The PEs send each other the Ethernet A-D per-EVI routes that carry the RD, RTs, next-hop information, local service ID, EVPL label or SRv6 SID, and active/standby role.
5. Upon receipt of Ethernet A-D per-EVI 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 or select an SRv6 tunnel to perform traffic recursion based on SRv6 SIDs. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, a forwarding entry indicating the association between the MPLS or SRv4/v6 tunnel and local EVPL instance is 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 or select an SRv6 tunnel to perform traffic recursion based on SRv6 SIDs. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, an FRR entry indicating the association between the MPLS or SRv4/v6 tunnel and local EVPL instance is generated. The entry pointing to PE1 is the master entry, and the entry pointing to PE2 is the backup entry.
7. PE1 and PE2 each receive EVI Ethernet AD routes from each other and match the RTs of the corresponding EVPN VPWS instance. PE1 and PE2 then select an MPLS or SRv4 tunnel to perform traffic recursion based on the next-hop information or select an SRv6 tunnel to perform traffic recursion based on SRv6 SIDs. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, a bypass entry indicating the association between the MPLS or SRv4/v6 tunnel and local EVPL instance is generated.

Figure 4 EVPN VPWS dual-homing single-active networking (with an E-Trunk deployed)



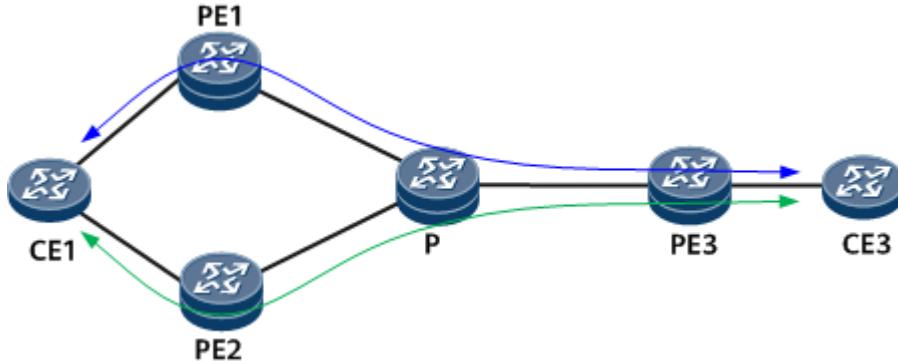
Protocol Packet Exchange Process in Dual-Homing Active-Active Scenarios

On the network shown in [Figure 5](#), a CE is dual-homed to PE1 and PE2. PE1 and PE2 work in active-active mode, and E-Trunk is deployed between PE1 and PE2. The protocol 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 assigned a local service ID and a remote service ID. After the configuration, the local PE generates a forwarding entry 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 PE2 are assigned the same ESI.
2. PE1 and PE2 exchange ES routes that carry RDs, RTs, ESIs, and source IP addresses. PE1 and PE2 send ES routes that carry the RD, RT, ESI, and source IP address. Upon receipt of ES routes, PE1 and PE2 trigger DF election. The E-Trunk between PE1 and PE2 determines the master/backup relationship between PE1 and PE2. In an active-active scenario, the E-Trunk between PE1 and PE2 is in the master state.
3. PE1 and PE2 send PE3 the Ethernet A-D per-ES routes that carry the RD, RTs, next-hop information, and all-active mode information.
4. PEs send Ethernet A-D per-EVI routes to each other. An Ethernet A-D per-EVI route carries the RD, RTs, next hop, local service ID, EVPL label or SRv6 SID, and active/standby role.
5. Upon receipt of Ethernet A-D per-EVI 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 or select an SRv6 tunnel to perform traffic recursion based on SRv6 SIDs. If the service ID in the received routes is the same as the remote service ID configured for the local EVPL instance, a forwarding entry indicating the association between the MPLS or SRv4/v6 tunnel and local EVPL instance is generated.
6. PE3 receives the Ethernet A-D per-EVI routes from PE1 and PE2, matches the RTs of the routes with the corresponding EVPN VPWS instances, and recurses the routes to MPLS or SRv4 tunnels based on next hop information or to SRv6 tunnels based on SRv6 SIDs. If the service ID of the received route is the same as the remote service ID configured for the local EVPL instance, load balancing entries are generated for the MPLS or SRv4/v6 tunnel and the local EVPL instance.
7. PE1 and PE2 receive Ethernet A-D per-EVI routes from each other, match the RTs of the routes, and import the routes to the corresponding EVPN VPWS instances. Then the routes

recurse to MPLS or SRv4 tunnels based on next hop information or to SRv6 tunnels based on SRv6 SIDs. If the service ID of the received route is the same as the remote service ID and ESI configured for the local EVPL instance, a bypass entry is generated, indicating the association between the MPLS or SRv4/v6 tunnel and the local EVPL instance.

Figure 5 EVPN VPWS dual-homing active-active networking



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 forwards the packets to the corresponding AC interface based on the association entries.

Parent Topic: [EVPN VPWS](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

1.12.6 PBB-EVPN

[PBB-EVPN Fundamentals](#)

[Migration from an HVPLS Network to a PBB-EVPN](#)

Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.6.1 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.

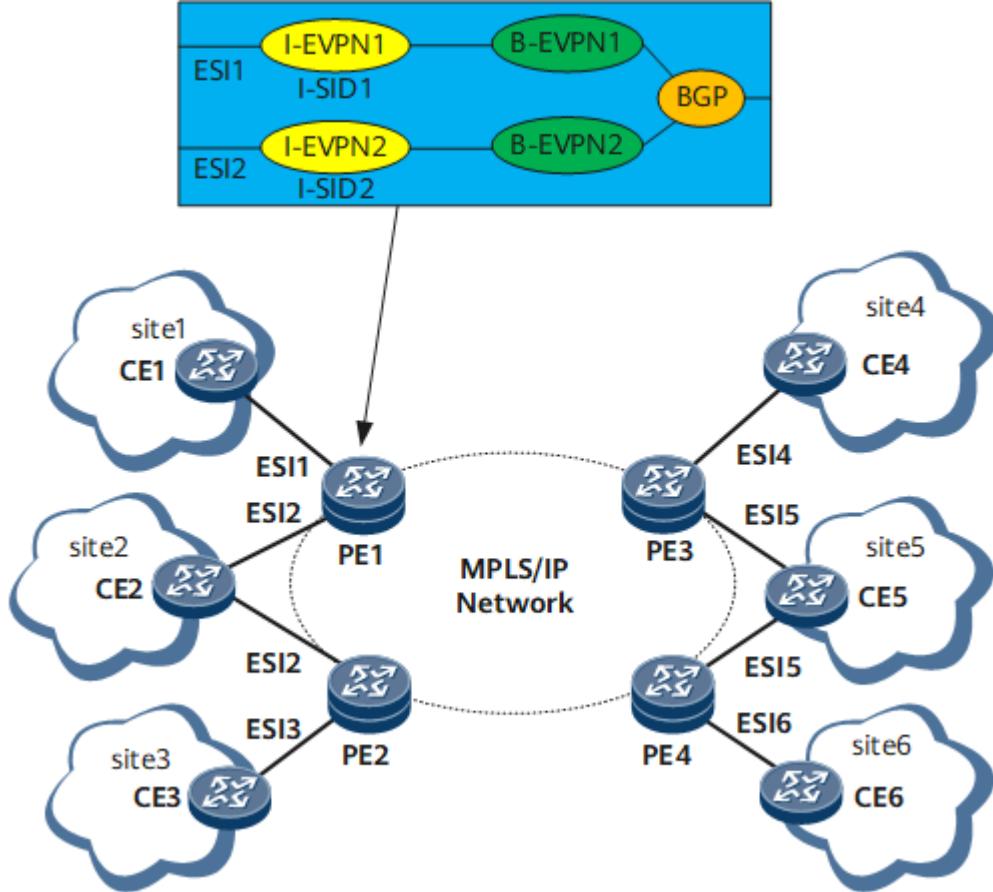
As shown in [Figure 1](#), a PBB-EVPN has similar architecture as an EVPN. Compared with EVPN, PBB-EVPN introduces some new concepts. Understanding these basic concepts is the prerequisite for learning the working principles of PBB-EVPN.

- 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 1 PBB-EVPN networking



[Table 1](#) describes the key points in PBB-EVPN implementation.

Table 1 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"> • MAC advertisement route • Inclusive multicast route Unicast MAC address advertisement BUM packet transmission
	PBB-EVPN supports fast convergence.	Fast convergence

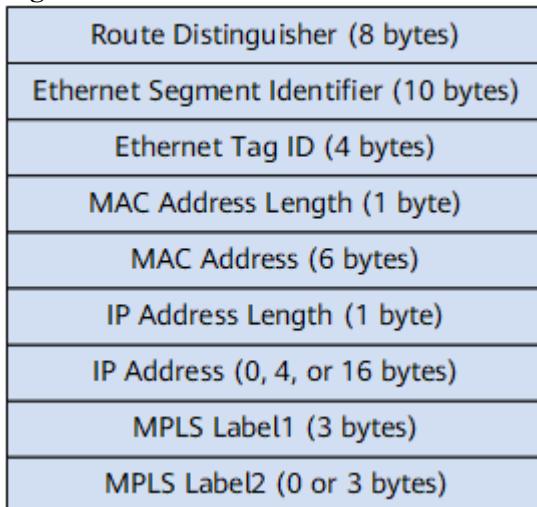
Plane	Key Points in Implementation	Related Concepts
	On a multi-homing network, PBB-EVPN uses DF election to prevent bandwidth waste.	Ethernet segment routeDF election
	PBB-EVPN uses split horizon to prevent routing loops.	Split horizon
	On a multi-homing network, PBB-EVPN supports load balancing. Currently, PBB-EVPN supports only per-flow load balancing, and does not support VLAN-based (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 2](#) 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 2 Prefix format of a MAC advertisement route packet



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. In a dual-homing single-active scenario, the value is all 0s. In a dual-homing active-active scenario, the value is 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 3](#) 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 3 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.

NOTE

Currently, the EVPN source address on the PE supports only IPv4. Therefore, the field contains only 4 bytes.

- 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 4](#) shows the prefix format of an Ethernet segment route packet.

Figure 4 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.

NOTE

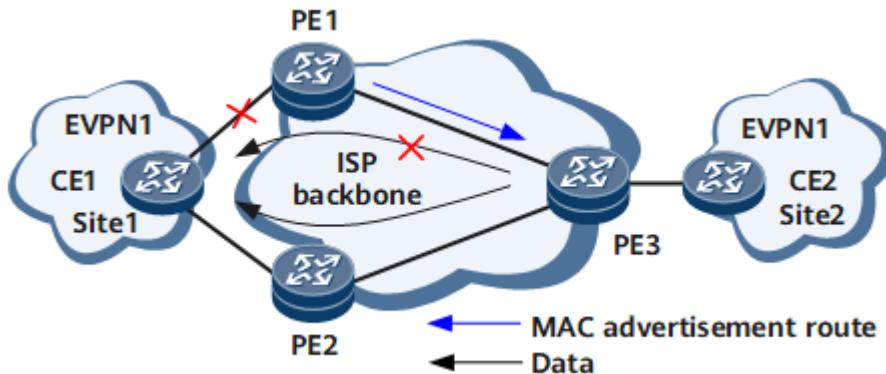
Currently, the EVPN source address on the PE supports only IPv4. Therefore, the field contains only 4 bytes.

Other Concepts

- Fast convergence

On the network shown in [Figure 5](#), 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 5 Fast convergence networking



- DF election

On the network shown in [Figure 6](#), 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.

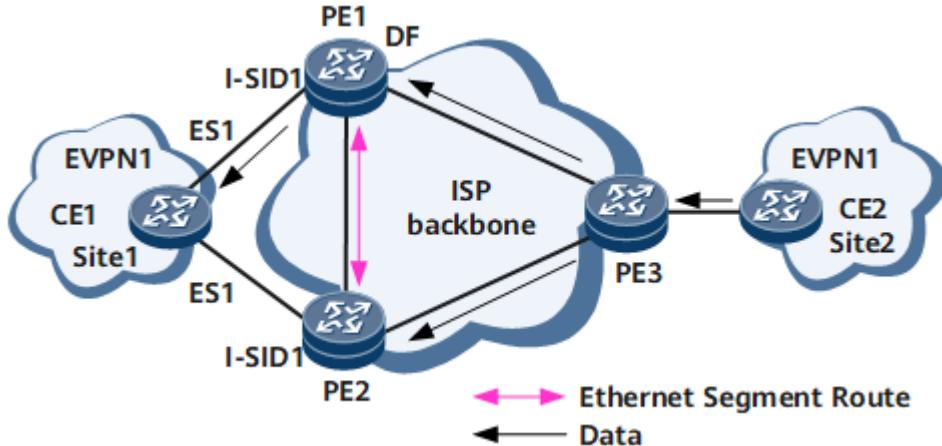
If a PE interface connecting to a CE goes Down, the PE functions as a backup DF. If a PE interface connecting to a CE goes Up, the PE and other PEs with Up interfaces elect a primary DF using the following procedure:

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 these 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.

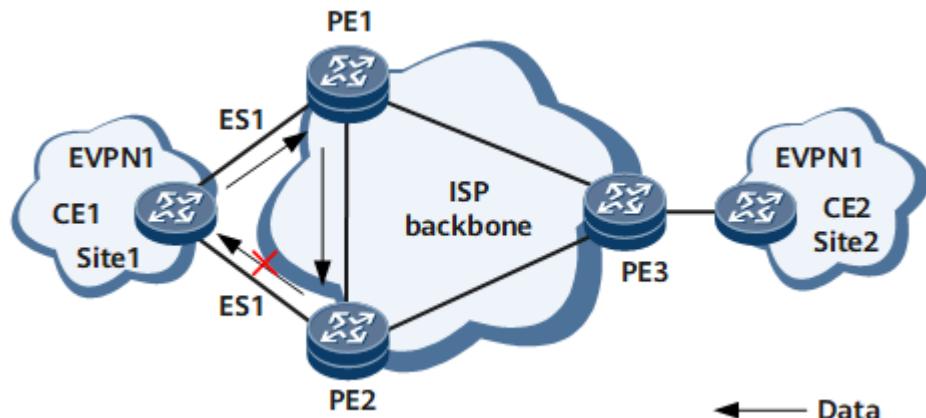
Figure 6 DF election networking



- Split horizon

On the network shown in [Figure 7](#), 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 7 Split horizon networking

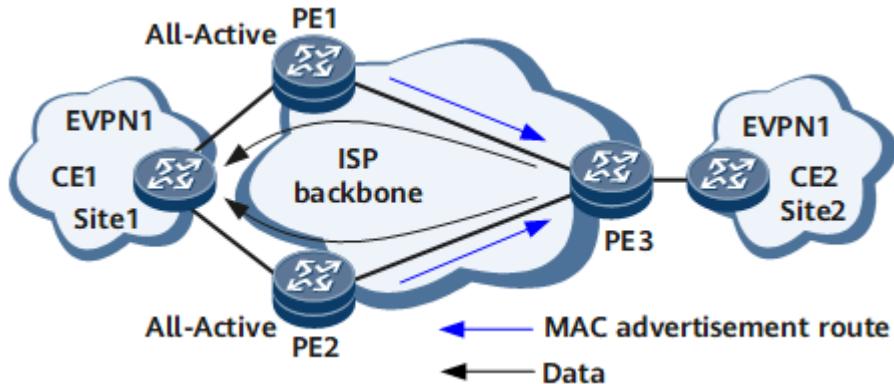


- Redundancy mode

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 8](#), if PE1 and PE2 are both configured to work in All-Active mode, after PE1 and PE2 send MAC advertisement routes carrying the same B-MAC

address to PE3, PE3 sends unicast traffic destined for CE1 to both PE1 and PE2 in load balancing mode.

Figure 8 Redundancy mode networking

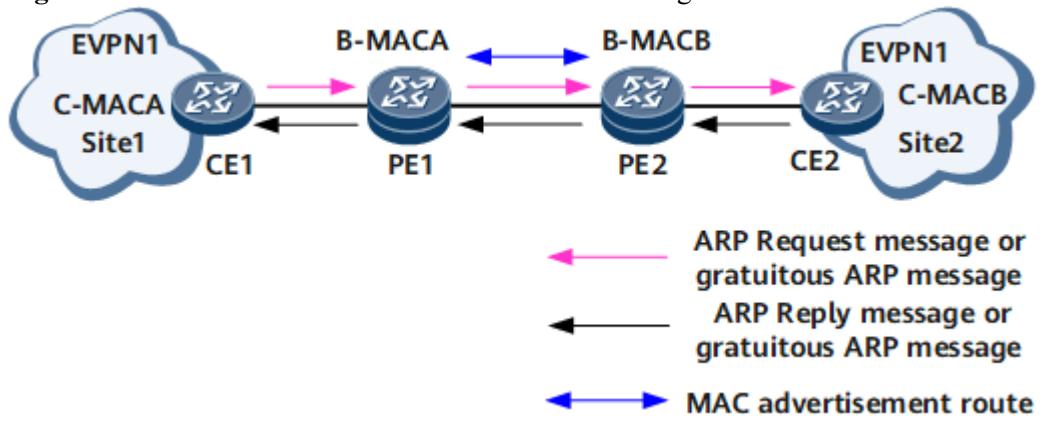


Unicast MAC Address Advertisement

On the network shown in [Figure 9](#), 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 9 Unicast MAC address advertisement networking



BUM Packet Transmission

BUM packets are a collection of broadcast packets, unknown unicast packets, and multicast packets. 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 10](#), BUM packets are transmitted as follows:

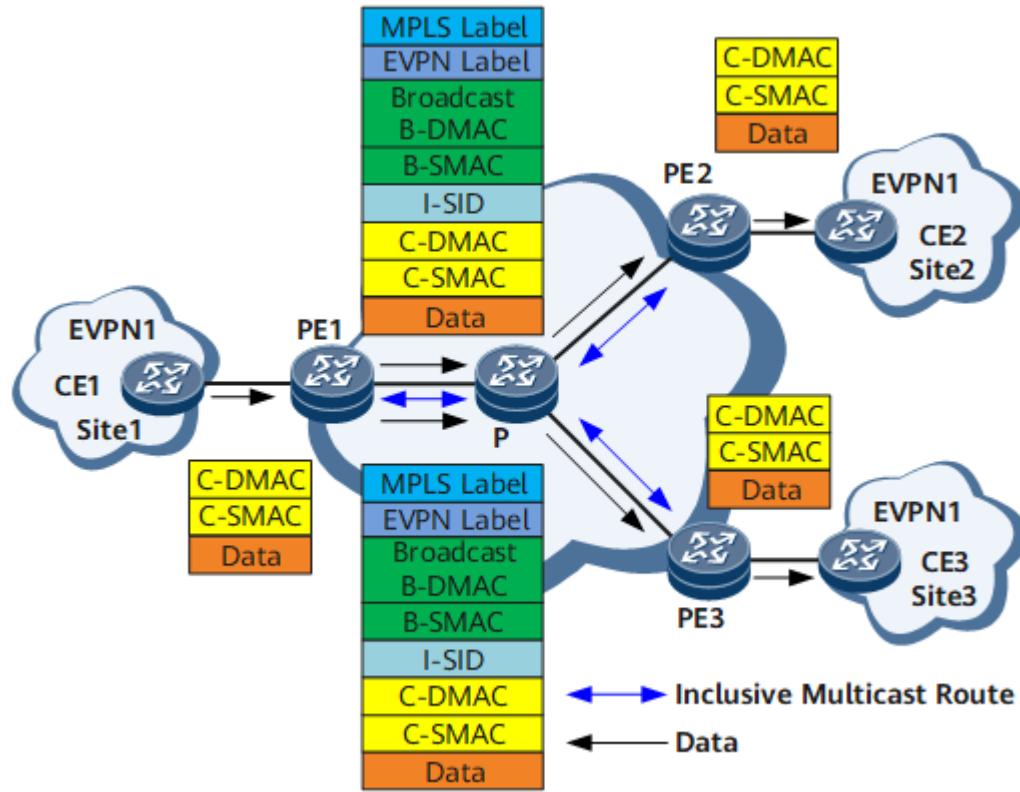
1. CE1 sends BUM packets to PE1.

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 remote PEs, respectively. The B-DMAC address carried in the PBB header is a broadcast MAC address.
3. 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 8](#) 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 10 BUM packet transmission networking



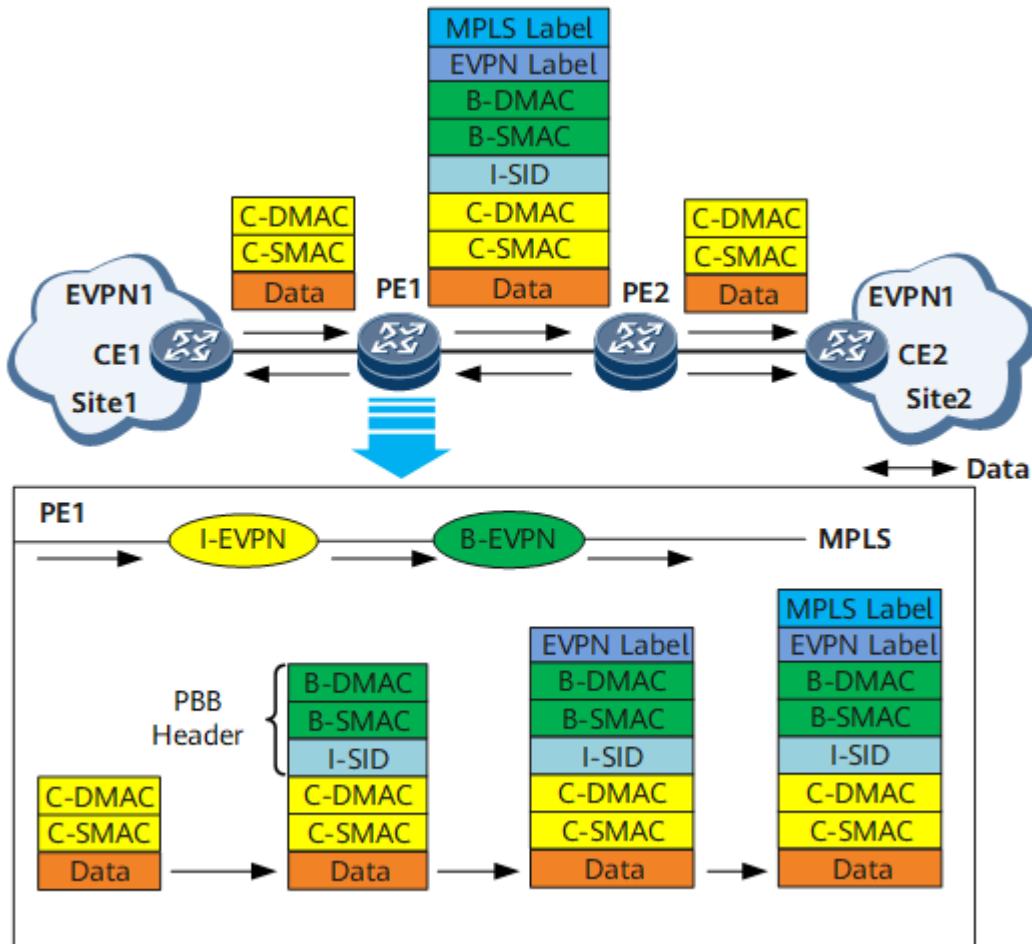
Unicast packet transmission

On the network shown in [Figure 11](#), unicast packets are transmitted as follows:

1. CE1 forwards unicast packets that carry the source C-MAC (C-SMAC) and destination C-MAC (C-DMAC) addresses to PE1 at Layer 2.
2. Upon receipt of the packets, the I-EVPN instance on PE1 searches its C-MAC address table for a matching forwarding entry based on the destination C-MAC address in these packets. After finding such an entry, PE1 encapsulates a PBB header, a public network MPLS 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.

3. 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 11 Unicast packet transmission networking



Parent Topic: [PBB-EVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.12.6.2 Migration from an HVPLS Network to a PBB-EVPN

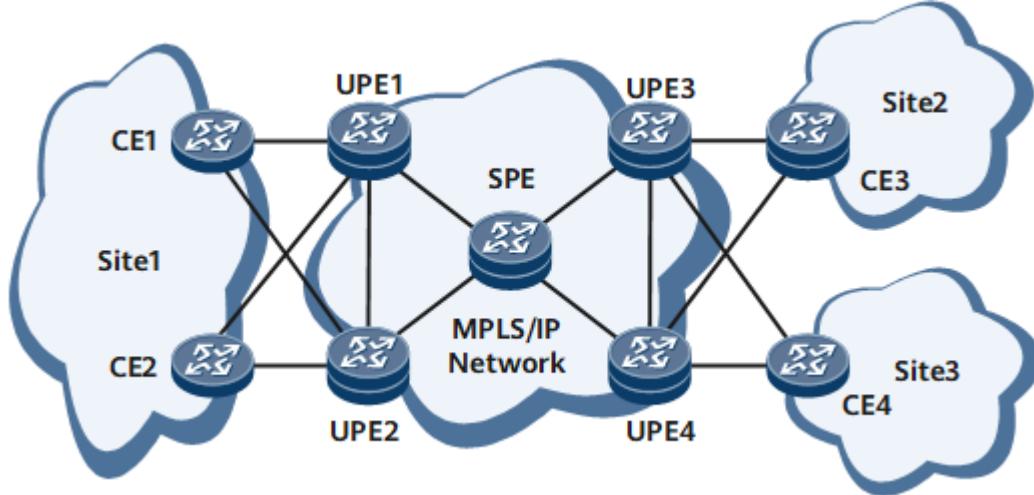
On the network shown in [Figure 1](#), VPLS is deployed to allow services of the same private network to access VSIs over different PEs. To avoid establishment of full-mesh PWs, SPEs are deployed on the network to form an HVPLS.

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 1 Typical networking



Parent Topic: [PBB-EVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.12.7 EVPN E-Tree

Background

As the number of services carried on an EVPN increases, the number of user MAC addresses managed by the EVPN is also increasing. 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.

Fundamentals

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 the 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 1](#) shows the format of a packet with the E-Tree extended community attribute. The packet includes the Leaf Label field and the Flags field. The Flags field contains 8 bits, in which the first 7 bits are all zeros and the last identifies whether an EVPN MAC route is from a leaf AC interface. Value 1 indicates that the MAC route comes from this interface. The extended community attribute can be advertised through Ethernet 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 1 Format of a packet with the EVPN E-Tree extended community attribute

0	7	15	23	31
Type=0x06	Sub-Type=0x05	Flags	Reserved=0	
Reserved=0		Leaf Label		

Flags: 0000000L

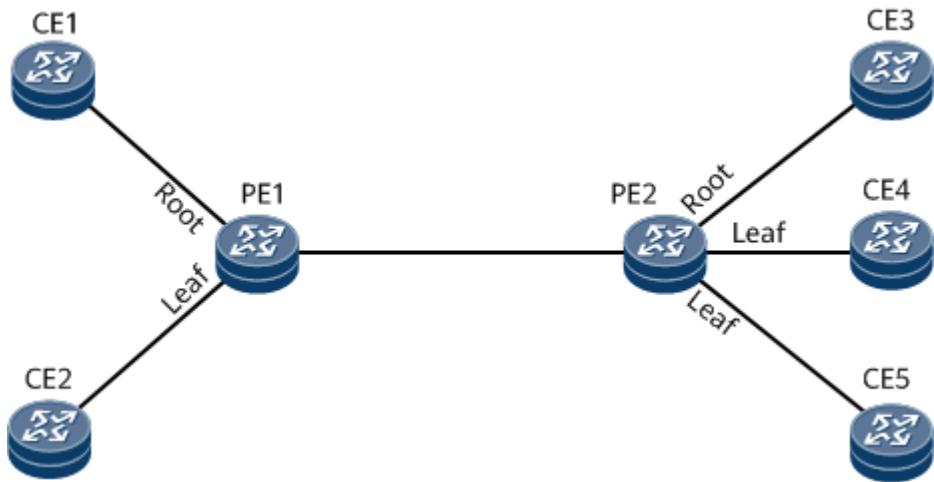
Take the network shown in [Figure 2](#) deployed as an example. Known unicast traffic is isolated through the following process:

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 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. 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 or CE5. Instead, PE2 sends the traffic only to CE3, implementing BUM traffic isolation between leaf AC interfaces.

Figure 2 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.

Application Scenarios

Currently, EVPN E-Tree can be used in two scenarios: EVPN E-Tree over MPLS and EVPN E-Tree over SRv6. The differences between the two modes are as follows:

Table 1 Implementation differences between EVPN E-Tree over MPLS and EVPN E-Tree over SRv6

Application Scenarios	Public Network Tunnel	BUM Traffic Forwarding
EVPN E-Tree over MPLS	MPLS LDP/MPLS TE	When a PE sends BUM traffic from a leaf interface to the peer PE, the PE encapsulates the leaf label value in the traffic. After receiving BUM traffic, the peer PE finds the leaf label allocated by itself and does not forward the traffic to the interface with the leaf attribute.
EVPN E-Tree over SRv6	SRv6 BE/SRv6 TE Policy	When a PE sends BUM traffic from a leaf interface to the peer PE, the PE encapsulates the Arguments value (a field in a prefix SID) sent by the peer PE into the traffic. After receiving BUM traffic, the remote PE parses the valid Arguments value and does not forward the traffic to the interface with the Leaf attribute.

Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.

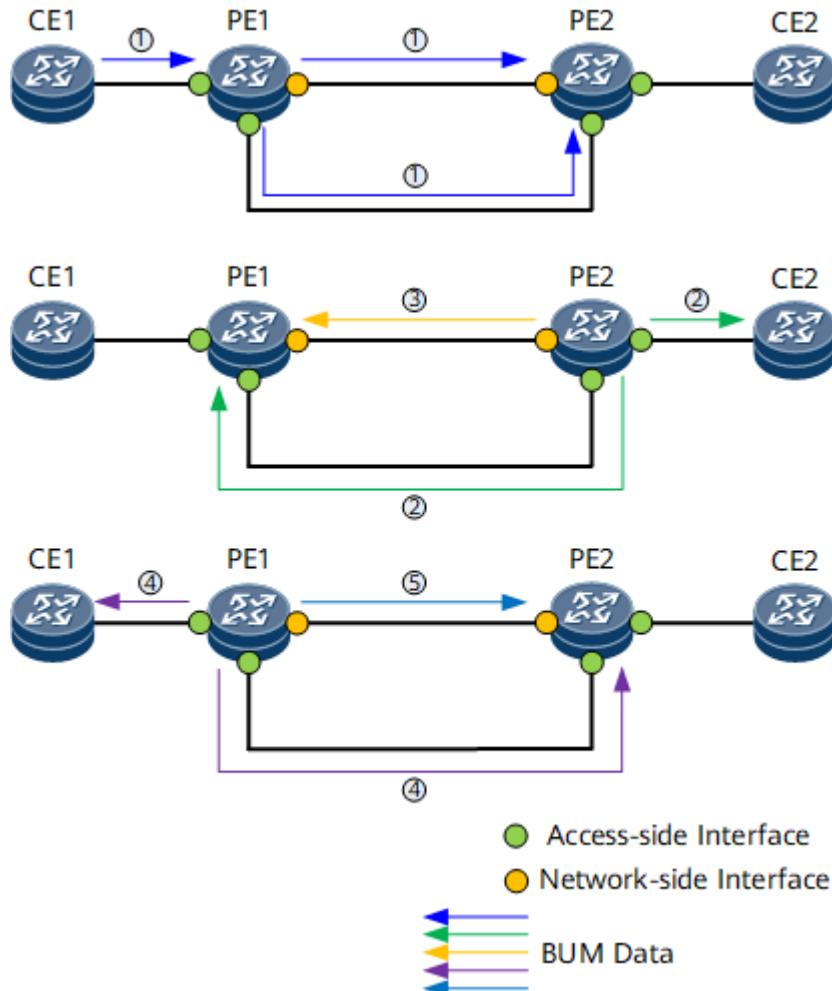
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.8 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.

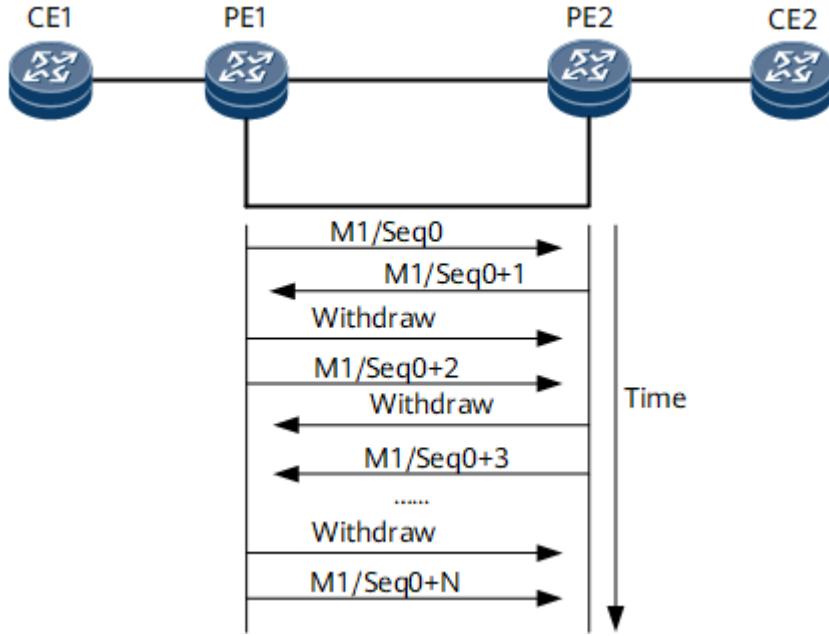
Figure 1 BUM traffic loop over an EVPN



On the network shown in [Figure 1](#), 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 1](#)).
2. PE2 forwards the BUM traffic received from PE1 through the network-side link to the access-side link (traffic 2 in [Figure 1](#)). Equally, PE2 forwards the BUM traffic received from the access side to the network side (traffic 3 in [Figure 1](#)).
3. PE1 forwards the BUM traffic received from PE2 through the network-side link to the access-side link (traffic 4 in [Figure 1](#)). Equally, PE1 forwards the BUM traffic received from the access side to the network side (traffic 5 in [Figure 1](#)).
4. As steps 2 and 3 are repeated, BUM traffic is continuously transmitted between PE1 and PE2.

Figure 2 Route flapping over the EVPN



On the network shown in [Figure 2](#), 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 recurses 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 recurses 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 recurses 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.

Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.9 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.

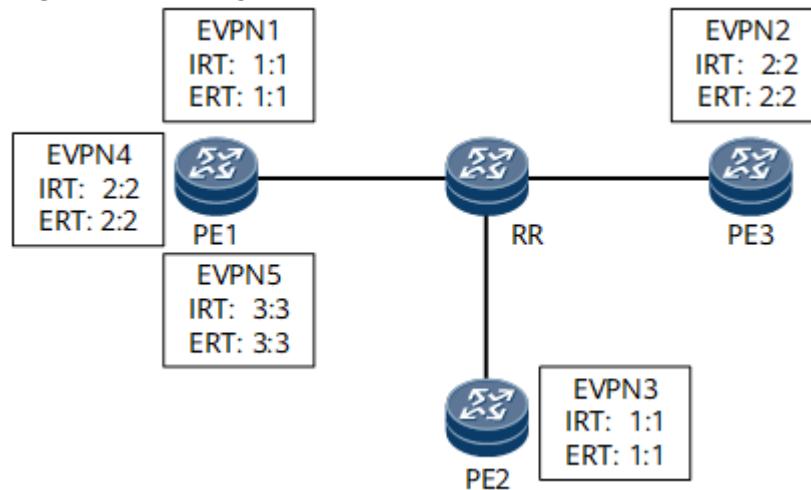
[Figure 1](#) shows the basic EVPN ORF network on which each device supports EVPN ORF. PE1, PE2, and PE3 establish BGP-EVPN peer relationships with the RR, and are also clients of the RR. An EVPN instance with a specific RT is configured on each PE.

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.

Figure 1 Basic usage scenario of EVPN ORF



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. However, the BGP-VT address family can resolve this problem.

On the network shown in [Figure 1](#), 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.

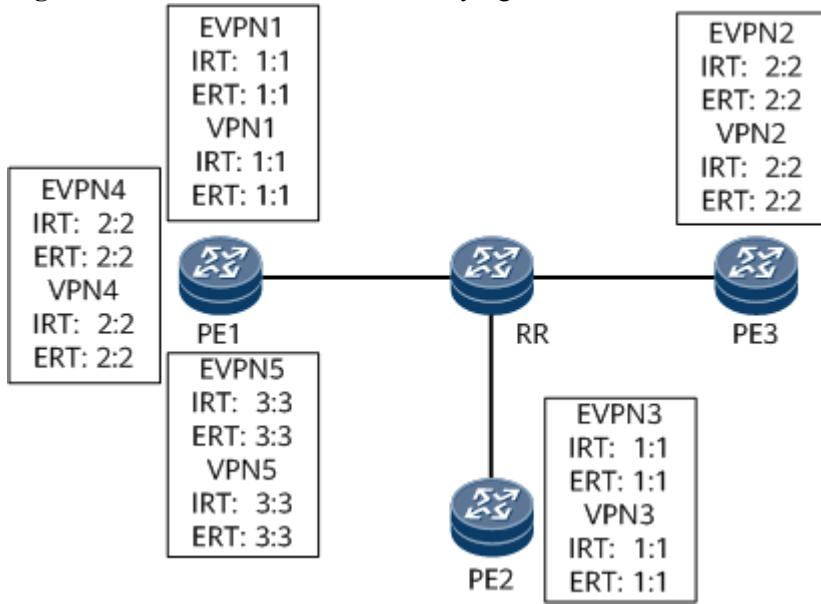
NOTE

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 preceding network, 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 2](#), 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 2 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.

Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.10 IGMP Snooping over EVPN MPLS

If the Ethernet virtual private network (EVPN) function is deployed on a network to carry multicast services but Internet Group Management Protocol (IGMP) snooping is not 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

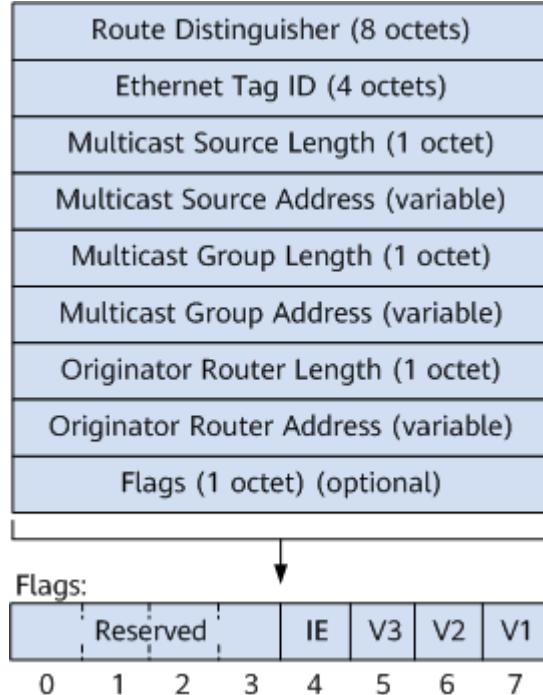
EVPN routes used by IGMP snooping over EVPN MPLS include Selective Multicast Ethernet Tag (SMET) and IGMP Join Sync routes.

- 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 1](#), the fields in the routing information are described as follows:

- Route Distinguisher: RD configured 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 8 bits. The first 4 most significant bits are reserved, and the last 3 least significant bits are used to identify IGMP versions. If bit 5 is set to 1, the IGMP version of the multicast entry carried in the route is IGMPv3. Only one of these 3 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 group types, respectively.

Figure 1 SMET route format



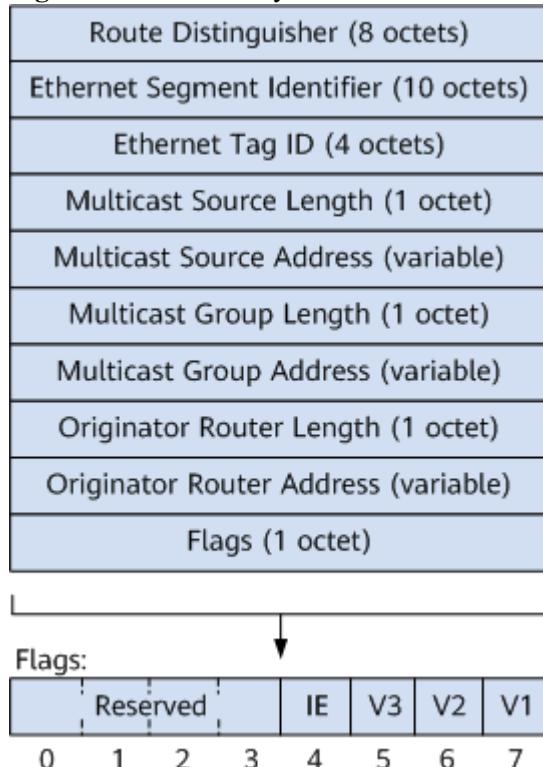
- IGMP Join Synch route

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 2](#), the fields in the routing information are described as follows:

- Route Distinguisher: route distinguisher (RD) configured 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 8 bits. The first 4 most significant bits are reserved, and the last 3 least significant bits are used to identify IGMP versions. If bit 5 is set to 1, the IGMP version of the multicast entry carried in the route is IGMPv3. Only one of these 3 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 group types, respectively.

Figure 2 IGMP Join Synch route format



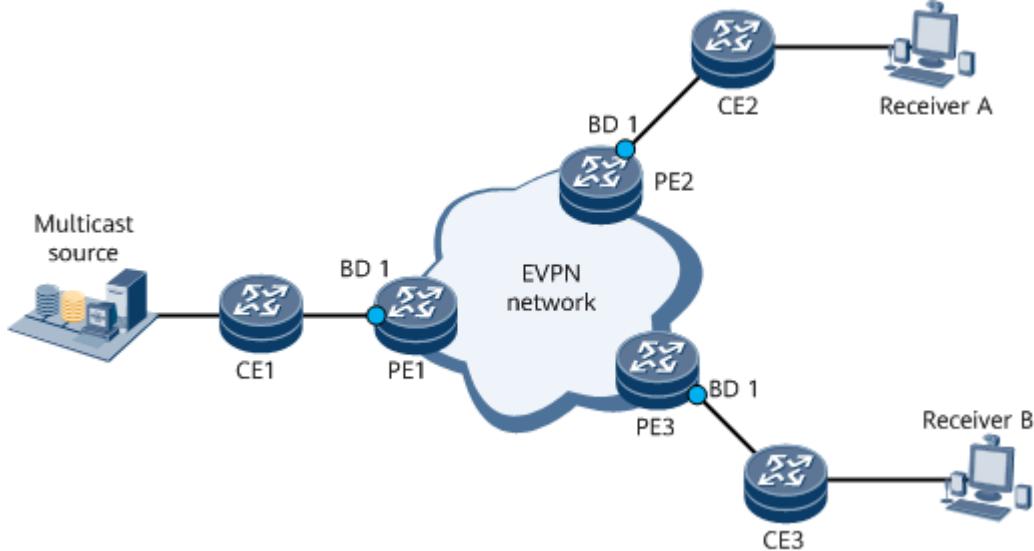
Route Advertisement and Traffic Forwarding

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

Single-homing access for IGMP snooping over EVPN MPLS

[Figure 3](#) 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 on CE1's interface connected to a multicast source and IGMP on CE1's interface connected to PE1.

Figure 3 Single-homing access for IGMP snooping over EVPN MPLS



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

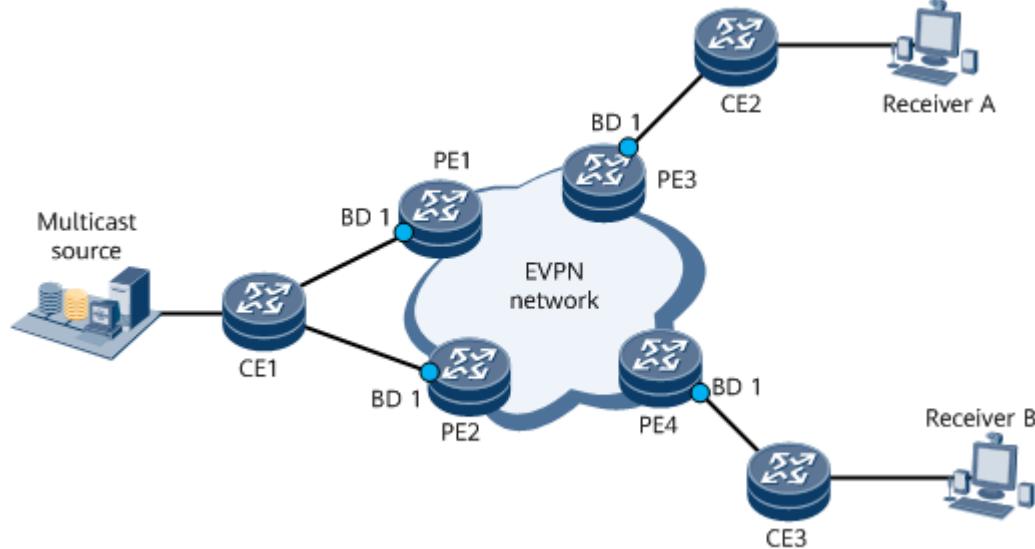
1. PE1, PE2, and PE3 periodically send IGMP Query messages to the access side in BD1.
2. Receiver A and Receiver B send IGMP Report messages to CE2 and CE3, respectively. For example, Receiver A sends an IGMPv3 (S, G) Report message, and Receiver B sends an IGMPv2 (*, G) Report message.
3. After receiving 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 the 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 the 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.

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

[Figure 4](#) 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.

Figure 4 Dual-homing access for IGMP snooping over EVPN MPLS on the multicast source side



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 the 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 the 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 add 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.

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

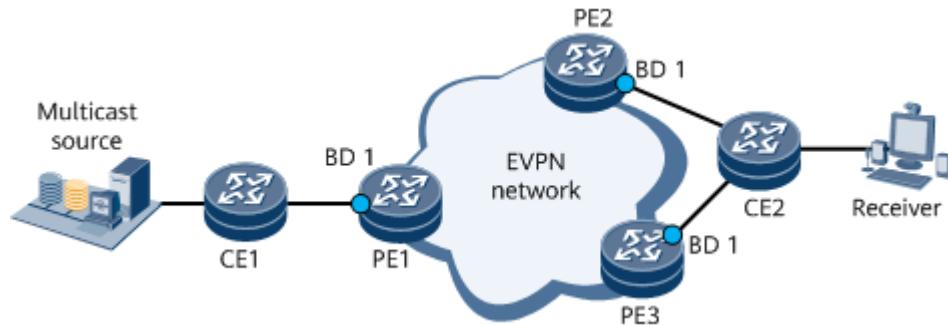
[Figure 5](#) shows dual-homing access for IGMP snooping over EVPN MPLS on the access side.

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.

Figure 5 Dual-homing access for IGMP snooping over EVPN MPLS on the access side



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 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

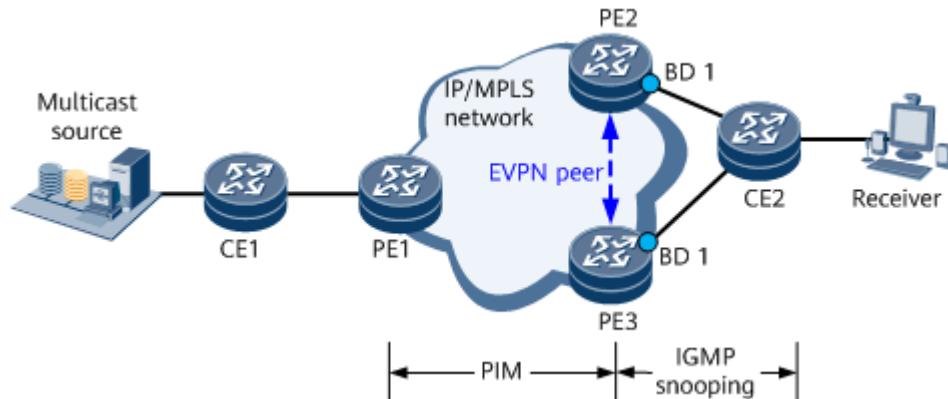
is a non-DF. Additionally, PE3 does not add the Eth-Trunk interface to CE2 as the outbound interface, because the Eth-Trunk interface is down.

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 mLDPE 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) entry based on the IGMP Report message received from CE2 and then sends an IGMP Join Synch route withdraw message to PE3. This ensures that PE3 deletes the (*, G) entry corresponding to the receiver, so that the (*, G) entry on PE3 is the same as that 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.

IGMP snooping over EVPN MPLS access-side dual-homing access (Layer 2 to Layer 3 in active-active mode)

As shown in [Figure 6](#), CE2 is dual-homed to PE2 and PE3 on the IGMP Snooping over EVPN MPLS access side. Configure PE1 as a sender PE and PE2 and PE3 as receiver PEs. Configure PIM-SM on PE1, PE2, and PE3. Configure an EVPN instance on PE2 and PE3, bind BD1 to the EVPN instance, create VBDIF1 for BD1, and enable PIM-SM and IGMP on VBDIF1. IGMP snooping and IGMP snooping proxy are automatically enabled in BD1. Establish a BGP EVPN peer relationship between PE2 and PE3. CE1 is connected to PE1 through a common Layer 3 interface, and CE2 is connected 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. The E-Trunks on PE2 and PE3 are configured to work in active-active mode, and the Eth-Trunks on PE2 and PE3 are both up. In this example, PE2 is selected as the DF, and PE3 is not selected as the DF.

Figure 6 Dual-homing access for IGMP snooping over EVPN MPLS on the access side (Layer 2 accessing Layer 3 in active-active mode)



In a scenario of dual-homing access for IGMP snooping over EVPN on the access side, devices are dual-homed to receivers, and Layer 2 multicast services access Layer 3 multicast services. Then the IGMP snooping over EVPN MPLS working process is as follows:

1. PE2 and PE3 periodically send 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. CE2's Eth-Trunk connected to PE2 and PE3 is up. CE2 sends an IGMP Report message to PE2 or PE3 based on the hash rule.
3. Upon receipt of the IGMP Report message, PE2 or PE3 creates an IGMP snooping (*, G) entry and sends an IGMP Join Synch route to each other. The IGMP Join Synch route carries the ESI on the access side of PE2 or PE3, in addition, the IGMP version and source filtering mode are included.
4. After PE2 or PE3 receives the IGMP Join Synch route, it creates an IGMP snooping (*, G) entry in BD1. PE2 functions as a DF to advertise Layer 3 IGMP (*, G) entries to VBDIF1 and uses the Eth-Trunk interface connected to CE2 as an outbound interface. As a non-DF device, PE3 does not need to advertise Layer 3 IGMP (*, G) entries to VBDIF1 or use the Eth-Trunk connected to CE2 as an outbound interface.
5. PE2 creates a Layer 3 PIM (*, G) entry, adds VBDIF1 as an outbound interface, and sends a PIM Join message to PE1.
6. After receiving the PIM Join message from PE2, PE1 creates an IGMP (*, G) entry 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 (*, G) entry to PE2 and PE3 through the mLDP tunnel interfaces.
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 or PE3 updates the (*, G) entry based on the IGMP Report message received from CE2 and then sends IGMP Join Synch route withdrawal messages to each other, which ensures that the (*, G) entries on PE2 and PE3 are the same.
12. If a fault occurs on the access side of PE2, the EVPN instance selects PE3 as the DF and adds the Eth-Trunk interface connecting PE3 to CE2 as the outbound interface so that multicast traffic can be forwarded from PE3 to CE2.

Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11 Application Scenarios for EVPN

[Using EVPN to Interconnect Other Networks](#)

[EVPN L3VPN HVPN](#)

[EVPN 6VPE](#)

[EVPN Interworking Scenarios](#)

[Inter-AS EVPN Option C](#)

[DCI Scenarios](#)

[NFVI Distributed Gateway \(SR Tunnels\)](#)

[NFVI Distributed Gateway Function \(BGP VPNv4/v6 over E2E SR Tunnels\)](#)

[NFVI Distributed Gateway Function \(BGP EVPN over E2E SR Tunnels\)](#)

[Application Scenarios for EVPN E-LAN Accessing L3VPN](#)

Parent Topic: [EVPN Feature Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

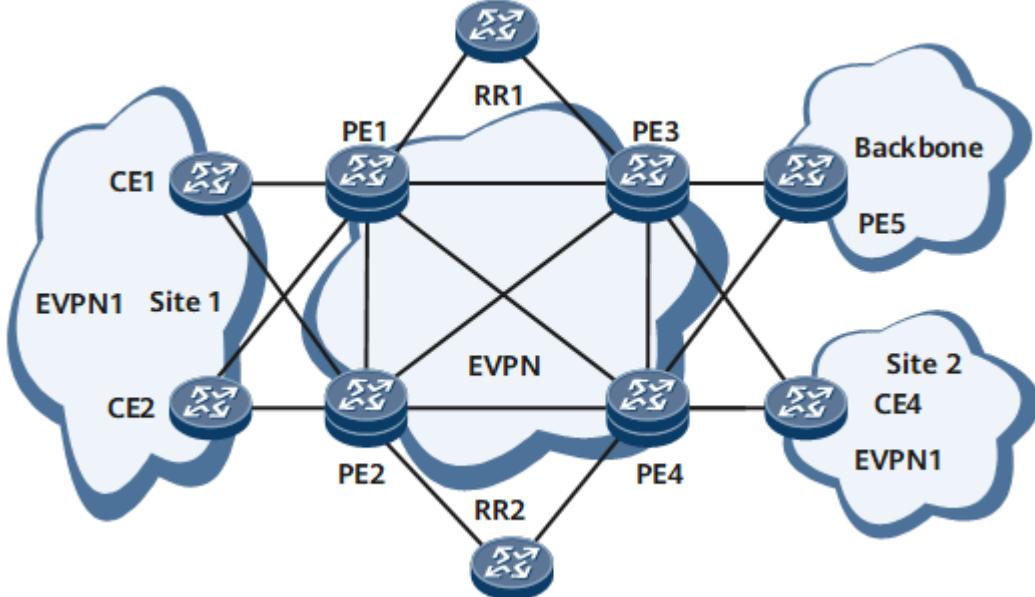
[< Previous topic](#)

1.12.11.1 Using EVPN to Interconnect Other Networks

On the network shown in [Figure 1](#), 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 1 EVPN application networking



Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

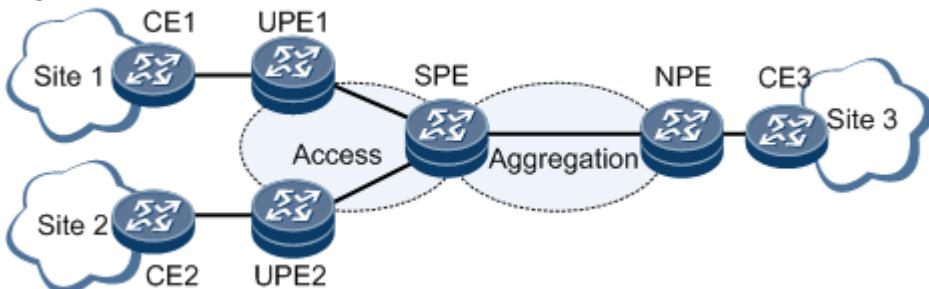
1.12.11.2 EVPN L3VPN HVPN

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 1](#) 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 1 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 interworking scenarios occur:

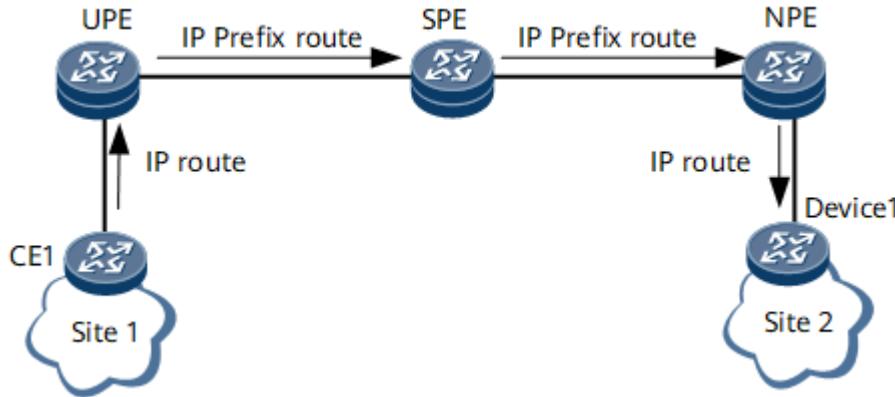
- Interworking between EVPN L3VPN HoVPN and common L3VPN: EVPN L3VPN HoVPN is deployed between the UPEs and SPE, and L3VPN is deployed between the SPE and NPE. The SPE advertises only default routes or summarized routes to the UPEs. After receiving specific routes (EVPN routes) from the UPEs, the SPE encapsulates these routes into VPNv4 routes and advertises them to the NPE.
- Interworking 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 2](#) 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 2 Route advertisement from CE1 to Device 1 on an EVPN L3VPN HoVPN or EVPN L3VPN H-VPN

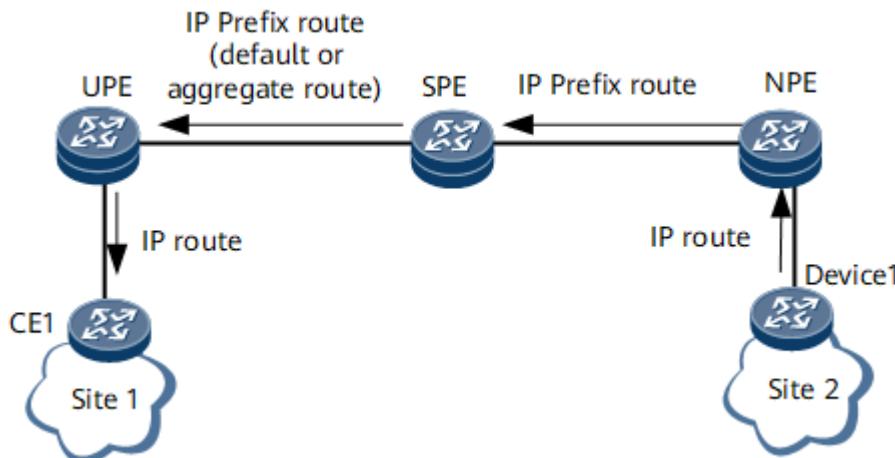


Route Advertisement from Device 1 to CE1 on an EVPN L3VPN HoVPN

[Figure 3](#) 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 3 Route advertisement from Device 1 to CE1 on an EVPN L3VPN HoVPN



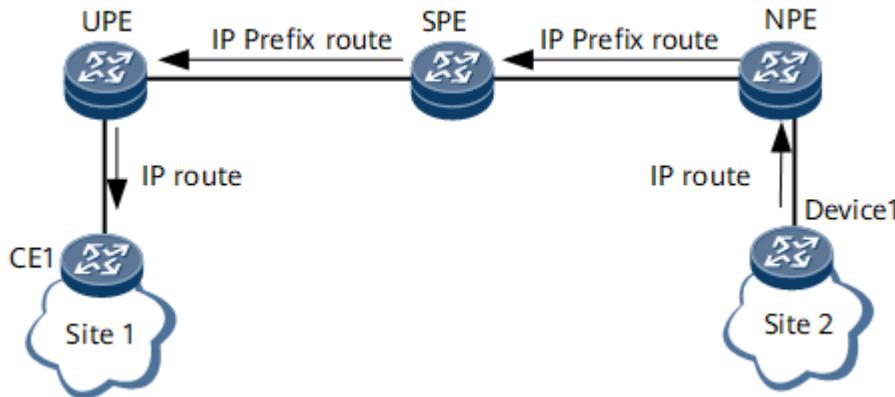
Route Advertisement from Device 1 to CE1 on an EVPN L3VPN H-VPN

[Figure 4](#) 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 4 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 interwork or L3VPN HoVPN and BD EVPN L3VPN interwork 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:

- Interworking 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.
- Interworking 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.

Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.

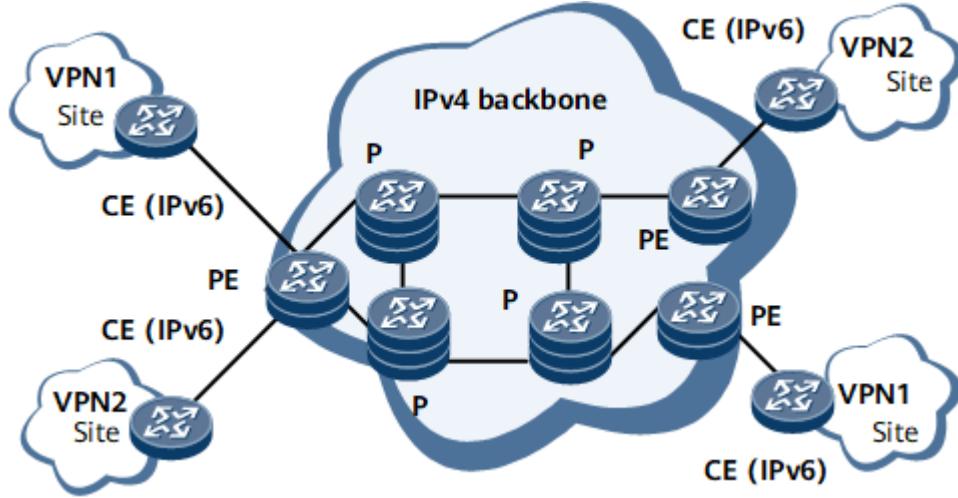
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11.3 EVPN 6VPE

6VPE is a VPN technology that allows connections to multiple IPv6 private sites over an IPv4 public network or IPv4 backbone network. This ensures service isolation between the IPv6 private sites of different users. On the network shown in [Figure 1](#), the components involved in the 6VPE function are the edge router (PE), customer edge router (CE), and core router (P) of the backbone network. The virtual routing and forwarding (VRF) tables stored on PEs are used to process the IPv6 routes of VPN sites. CEs that distribute user routes are connected to PEs through physical or logical interfaces. Ps are the backbone network devices used to forward VPN packets into which tunnel attributes are encapsulated.

Figure 1 6VPE model

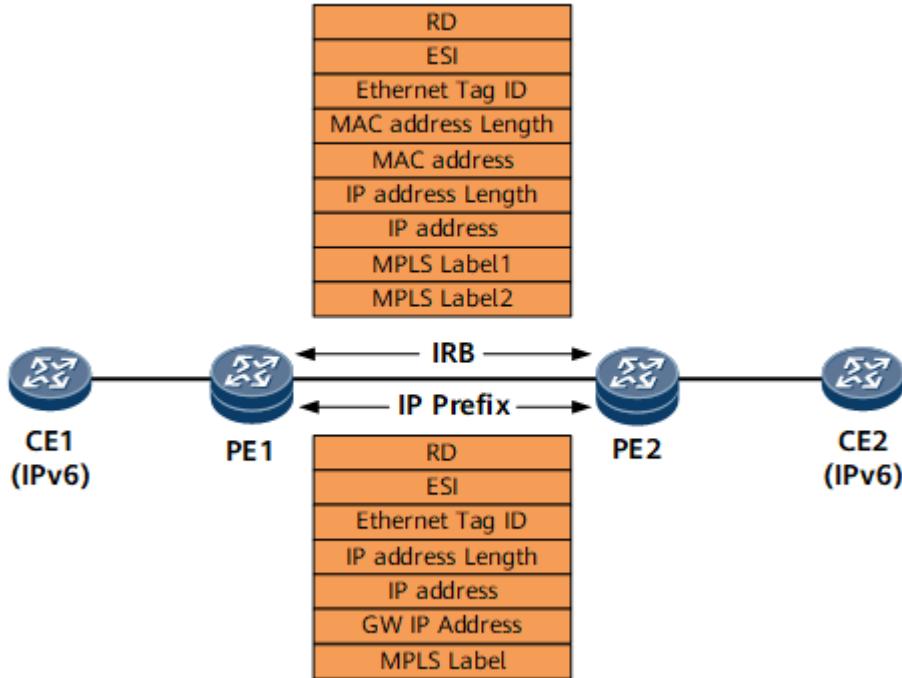


Currently, EVPN supports the 6VPE function. The processing mechanism of EVPN 6VPE is similar to that of EVPN L3VPN. On the network shown in [Figure 2](#), PEs can exchange host routes and IP prefix routes to transmit IPv6 route information of different sites in a VPN. The fields contained in the host routes and IP prefix routes are as follows:

- Fields in host routes:
 - Route Distinguisher (RD): Specifies the RD value set for an EVPN instance.
 - Ethernet Segment Identifier (ESI): Uniquely identifies a connection between a PE and a CE.
 - Ethernet Tag ID: In VLAN-Aware accessing BD EVPN scenarios, the value is **BD-Tag**. In other scenarios, the value contains all 0s.
 - MAC Address Length: Specifies the MAC address length in an ND entry.
 - MAC Address: Specifies a MAC address in an ND entry.
 - IPAddress Length: Specifies the IPv6 address length in an ND entry.
 - IP Address: Specifies an IPv6 address in an ND entry.
 - MPLS Label1: Specifies the VPN label of EVPN.
 - MPLS Label2: Specifies the label used for Layer 3 traffic forwarding.
- Fields in IP prefix routes:
 - RD: Specifies the RD value set for an EVPN instance.
 - ESI: Uniquely identifies a connection between a PE and a CE.

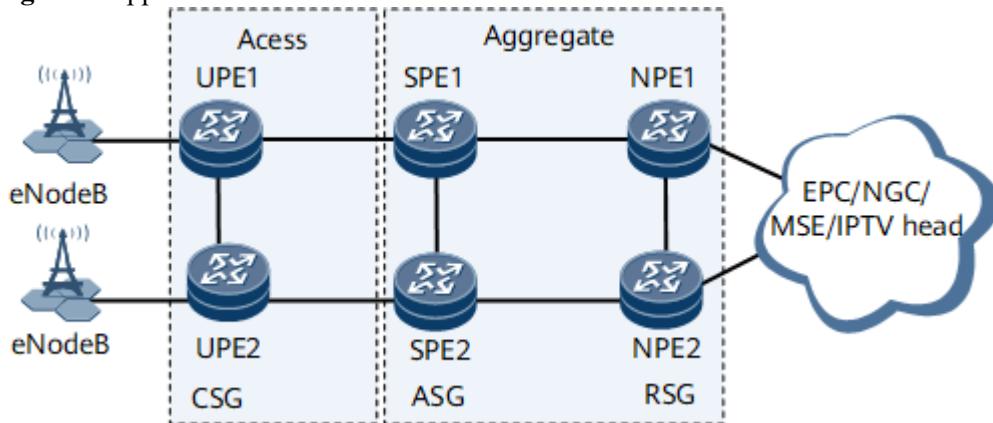
- Ethernet Tag ID: Currently, the value of this field contains all 0s.
- IPAddress Length: Specifies the IPv6 address length in an ND entry.
- IP Address: Specifies an IPv6 address in an ND entry.
- GW IP Address: Specifies the default gateway address.
- MPLS Label: Specifies the VPN label of EVPN.

Figure 2 EVPN 6VPE model



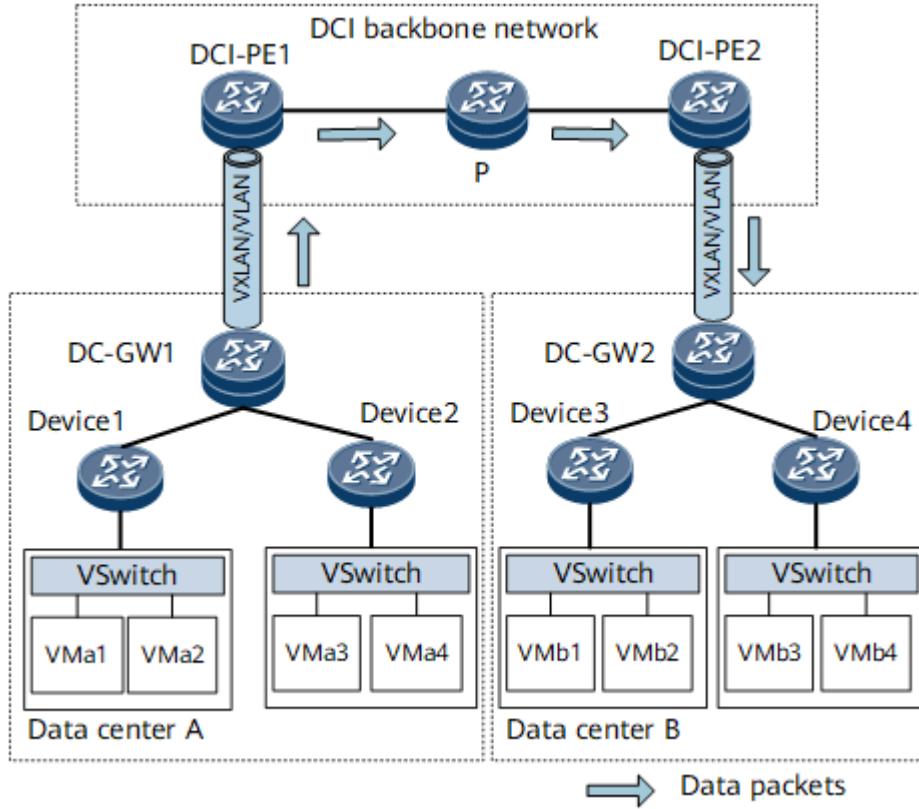
In [Figure 3](#), the network has been deployed with the IP RAN solution and a large number of IPv4-related services. If new services need to be deployed, IPv6 addresses may be used in consideration of IPv4 address exhaustion. In this case, deploy EVPN L3VPNv6 HVPN in the 6VPE model to carry IPv6 services. The processing mechanism of EVPN 6VPE is similar to that of L3VPN HVPN.

Figure 3 Application of EVPN 6VPE in IP RAN scenarios



Similarly, EVPN 6VPE can be deployed in DCI scenarios to carry IPv6-related services. On the network shown in [Figure 4](#), DCs are connected to the DCI backbone network over VXLAN tunnels or through VLAN sub-interfaces. The DCI backbone network uses the EVPN 6VPE function to transmit IPv6 routes between DCs. An MPLS or SR tunnel can be deployed between DCI-PEs to carry IPv6 services.

Figure 4 Application of EVPN 6VPE in DCI scenarios



Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11.4 EVPN Interworking Scenarios

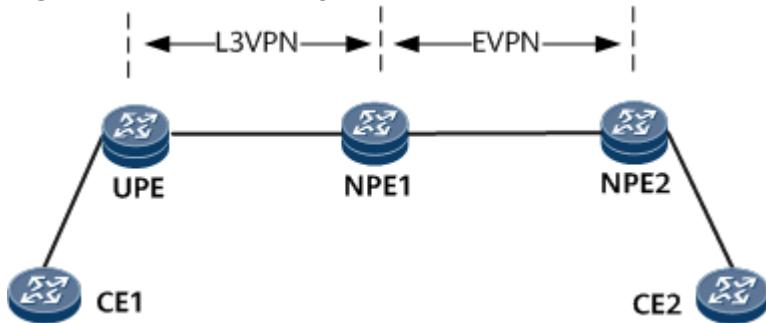
Background

Currently, metro networks are evolving towards EVPN. An existing network with a large number of aggregation devices cannot evolve to EVPN in an E2E mode at a time. During the transition, conventional L3VPN, VPWS, or VPLS is used at the aggregation layer, and the core network evolves to EVPN first. In this case, interworking between the EVPN and existing networks is required.

L3VPN Accessing EVPN

Devices between the UPE and NPE1 reside at the aggregation layer, and devices between NPE1 and NPE2 reside at the core layer. The L3VPN function is deployed at the aggregation layer, and the EVPN function is deployed at the core layer. After accepting access-side user routes, the UPE sends these routes to NPE1 through a BGP VPNv4 peer relationship. Upon receipt of the BGP VPNv4 routes, NPE1 with both an EVPN instance and an L3VPN instance configured imports these routes into the L3VPN instance, re-generates the routes as EVPN routes, and sends the EVPN routes to NPE2 through a BGP EVPN peer relationship. Then, network connectivity can be achieved in the L3VPN accessing EVPN scenario.

Figure 1 L3VPN accessing EVPN



VLL Accessing EVPN

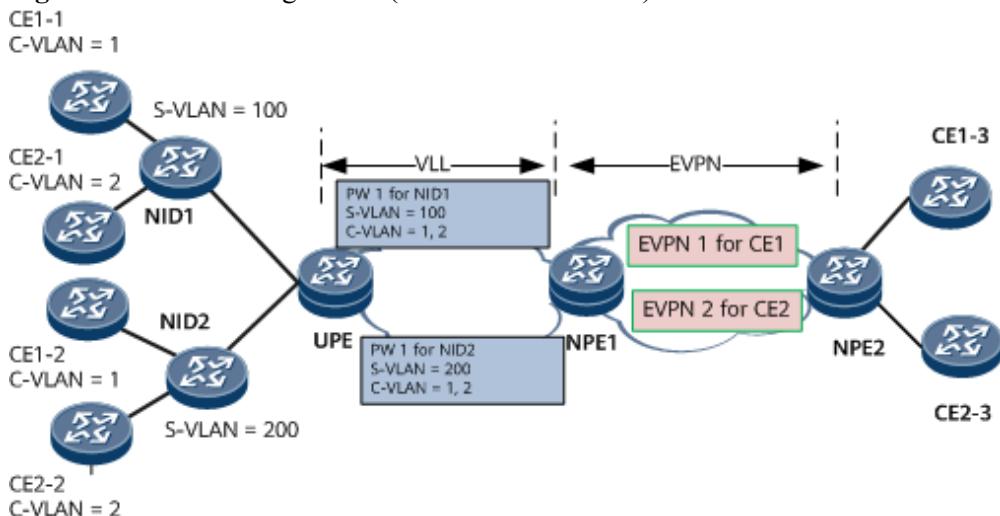
In the VLL accessing EVPN networking, a user named CE1 has three sites: CE1-1, CE1-2, and CE1-3; another user named CE2 also has three sites: CE2-1, CE2-2, and CE2-3. NIDs, which aggregate user site, add two VLAN tags to user packets before sending these packets over the aggregation network. In a double-tagged user packet, an S-VLAN ID identifies a NID, and a C-VLAN ID identifies a user site connected to the NID. Users access the VLL network, which is a native Layer 2 network, through the NIDs. An MPLS network is deployed at the aggregation layer between the UPE and NPE1, and uses VLL to carry services. Another MPLS network is deployed at the core layer between NPE1 and NPE2, and uses EVPN to carry services.

VLL accessing EVPN allows different sites of the same user to communicate in the following scenarios:

- Single-homing scenario

The access-side main interfaces on NIDs can be used to provide user access. A UPE establishes a PW to each NID. On the NPEs, an EVPN instance is created for each user and the VLL accesses the EVPN instance in PW VE mode. The VLL service is bound to the PW VE main interface, each EVPN instance is bound to a PW VE sub-interface. The sub-interfaces use QinQ encapsulation and forward packets based on S-VLAN and C-VLAN IDs, steering traffic to different EVPN instances.

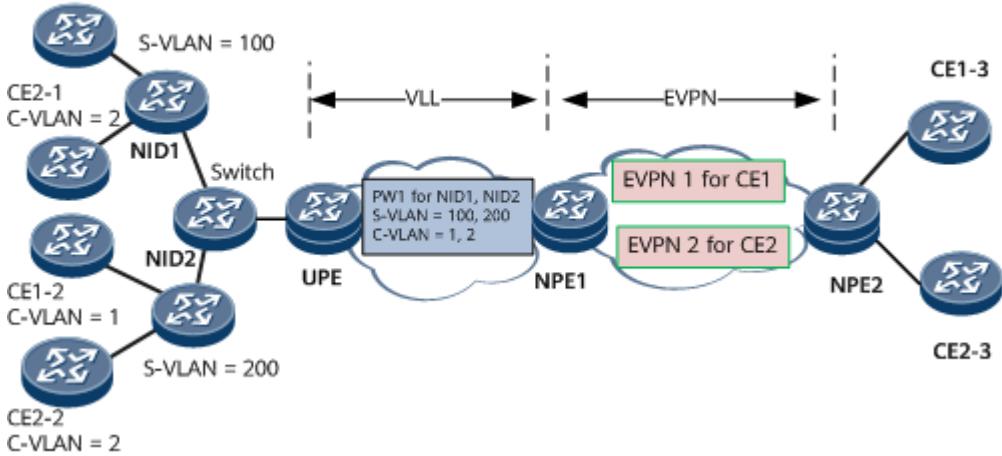
Figure 2 VLL accessing EVPN (one PW to each NID)



Similarly, the VLL accessing EVPN scenario also allows multiple NIDs to connect to a PW. Multiple NIDs are aggregated onto a switch and then connect to the same PW.

Figure 3 VLL accessing EVPN (multiple NIDs connected to one PW)

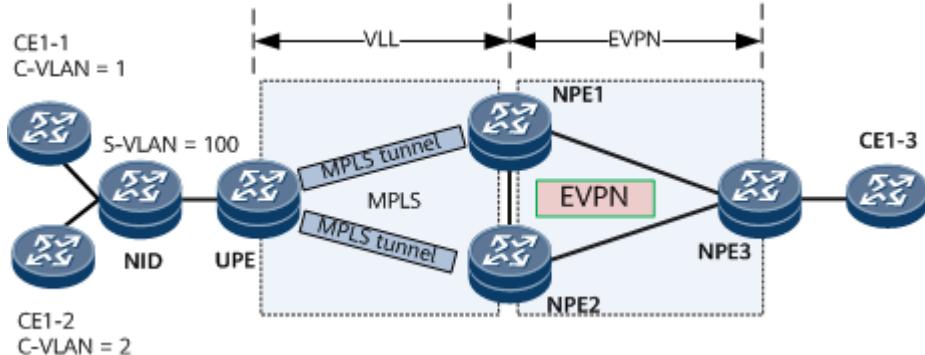
CE1-1
C-VLAN = 1



- Dual-homing networking

A UPE is dual-homed to the master and backup NPEs through the primary and secondary PWs, respectively, which improves access reliability. EVPN services on the NPEs can be configured to work in single-active or all-active mode. In all-active mode, load balancing can be implemented.

Figure 4 VLL accessing EVPN (dual-homing scenario)

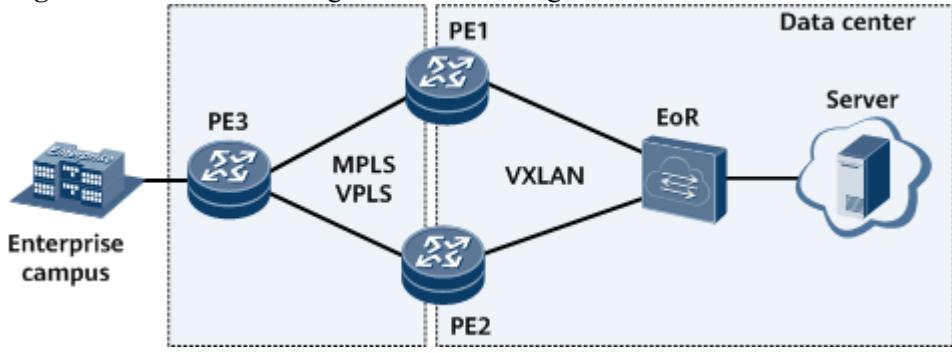


VXLAN Accessing VPLS

When an enterprise campus network is connected to a data center network, an EVPN VXLAN network needs to be deployed in the data center, and the data center network is connected to the enterprise campus network through an MPLS L2VPN. In this case, the functions of terminating VXLAN packets and accessing VPLS need to be deployed.

In [Figure 5](#), the EOR switch, functioning as a data center gateway, connects to the backbone network through egress PE1 and PE2 of the data center network. PE3, the egress of the enterprise campus network, is connected to PE1 and PE2 through the MPLS VPLS network. PE1 and PE2 are configured to terminate VXLAN packets and access the MPLS VPLS network to connect the data center to the campus network.

Figure 5 VXLAN accessing VPLS networking

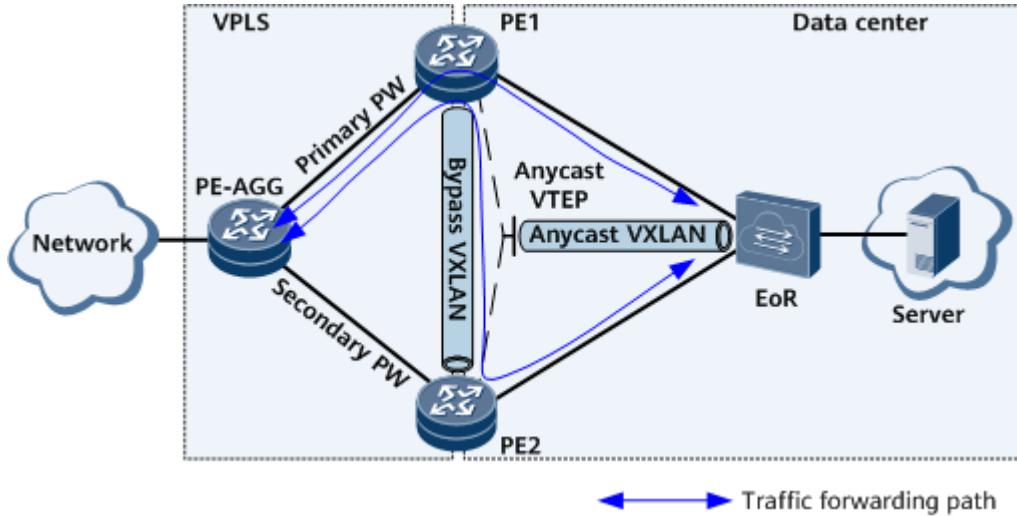


Interworking Primary and Secondary PWs with an Anycast VXLAN Tunnel in an EVPN Active-Active Scenario

In [Figure 6](#), PE1 and PE2 are egresses 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 EOR switch. In this case, PE1, PE2, and the EOR switch can communicate with each other. PE1 and PE2 communicate with the external network (an access network or the Internet) through the VPLS network. VPLS 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 EOR, the server in the data center can send traffic to PE1 and PE2. Traffic received by PE1 is sent directly to the PE-AGG through the primary PW. Traffic received by PE2 is forwarded to PE1 through the bypass VXLAN tunnel and also sent to the PE-AGG through the primary PW. Traffic from the PE-AGG to the server is transmitted along the reverse paths of the preceding forward paths.

Figure 6 Interworking the primary and secondary PWs with an anycast VXLAN tunnel in EVPN active-active scenario



Interworking VPLS with MPLS EVPN

VPLS technology has some inherent defects. For example, it does not support load balancing and consumes a large number of network resources (because MAC learning and ARP learning involve broadcasting on the entire network). With the application of EVPN on the live network, the VPLS network will gradually evolve to the EVPN network. However, a user network environment is complex and may not completely evolve to the EVPN network. VPLS is deployed on some devices, and EVPN is deployed on the other devices. In this case, you can configure interworking between VPLS and MPLS EVPN to implement connectivity of the entire network.

In [Figure 7](#), VPLS is deployed between the CSG and ASG, and the CSG is connected to ASGs through the primary and secondary PWs. EVPN is deployed between the ASGs and RSG. BDs are configured and bound to VSIs and EVPN instances on ASG1 and ASG2 so that all PWs in the VSIs access the EVPN instances through BDs. If a CSG is dual-homed to ASG1 and ASG2, the same ESI must be set for the primary and secondary PW interfaces on ASG1 and ASG2. The route advertisement and traffic forwarding process is as follows:

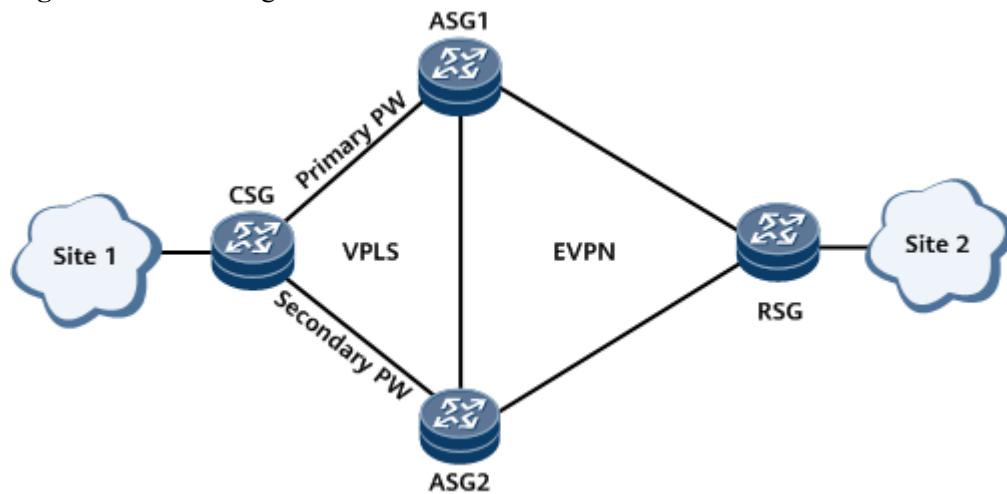
1. ASG1 sends an Ethernet Auto-Discovery route to the RSG as the PW interface is assigned an ESI and the PW is up on ASG1.
2. After the CSG forwards a Layer 2 packet sent by site 1 to ASG1, ASG1 generates a MAC route based on the MAC address carried in the Layer 2 packet and sends the route to the RSG through a BGP EVPN peer relationship. The RSG uses the MAC route and Ethernet Auto-Discovery route to generate a MAC forwarding entry. Similarly, the RSG sends a MAC route carrying site 2's MAC address to ASG1, and ASG1 generates a MAC forwarding entry.
3. After the forwarding entries are created, they can be used to guide unicast and BUM traffic forwarding. The process of forwarding unicast traffic from site 1 to site 2 is used as an example. After traffic is sent to ASG1 through the primary PW, ASG1 finds a next-hop MAC address based on a MAC route sent by the RSG and forwards the traffic to the RSG. The RSG then forwards the traffic to site 2.

NOTE

Although ASG1 and ASG2 exchange Ethernet segment routes, DF election between ASG1 and ASG2 is determined by the PW status, not by the Ethernet segment routes. ASG1 connected to the primary PW is the primary DF, and ASG2 connected to the secondary PW is the secondary DF.

In BUM traffic forwarding scenarios, the network provides the split horizon function, and the backup DF blocks traffic. Therefore, no traffic loop or excess packets occur on the network.

Figure 7 Interworking between VPLS and MPLS EVPN



Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11.5 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.

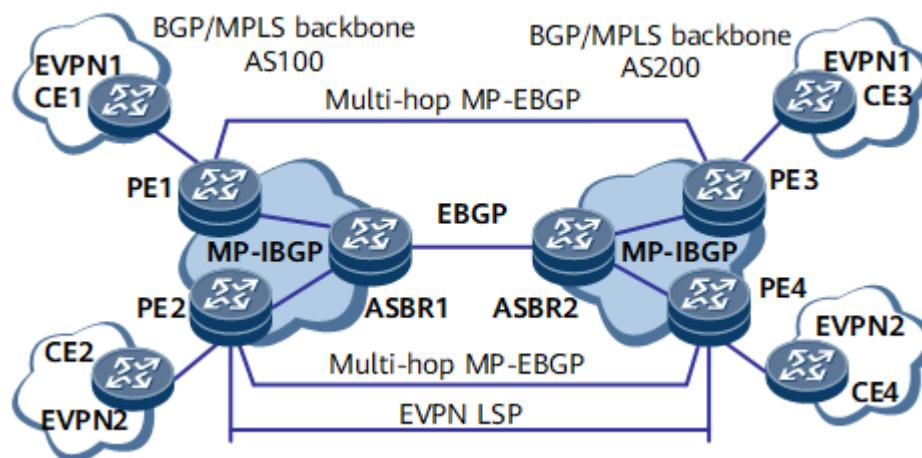
Implementation

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:

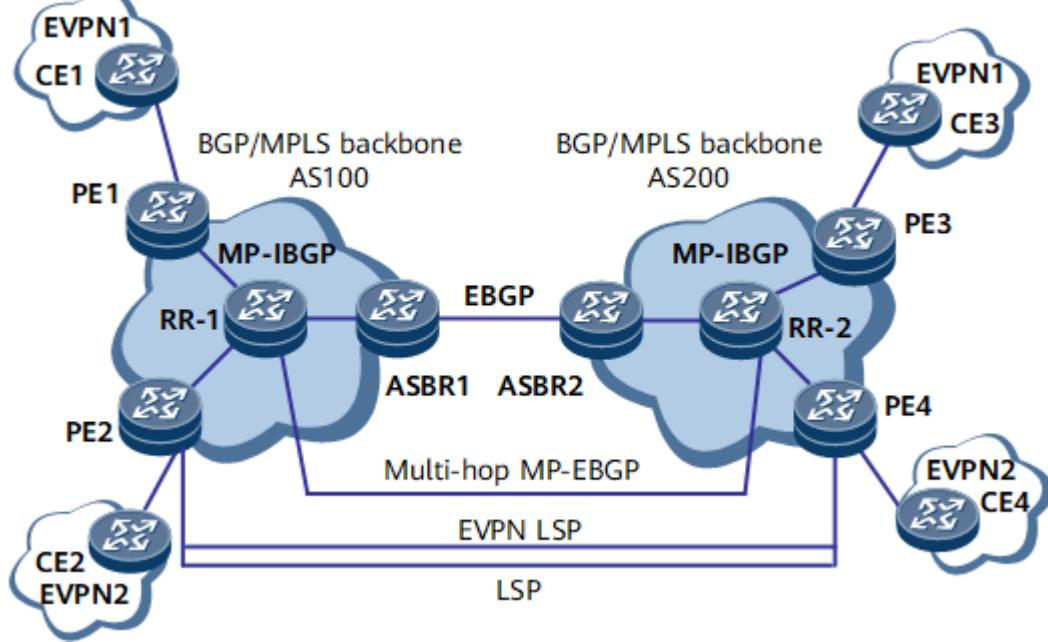
- Ethernet auto-discovery routes
- MAC and IP routes
- Inclusive multicast routes
- Ethernet segment routes
- ASBRs advertise labeled IPv4 routes to PEs in their respective ASs through MP-IBGP, and advertise labeled IPv4 routes received on PEs in the local AS to the ASBR peers in other ASs. ASBRs in 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 1 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 2 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.

Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11.6 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 1 Basic DCI concepts

Concept	Description
Overlay network	<ul style="list-style-type: none"> 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. The overlay network has an independent control plane and forwarding plane. 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.
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 1](#), 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 2](#).

Figure 1 Basic DCI scenario

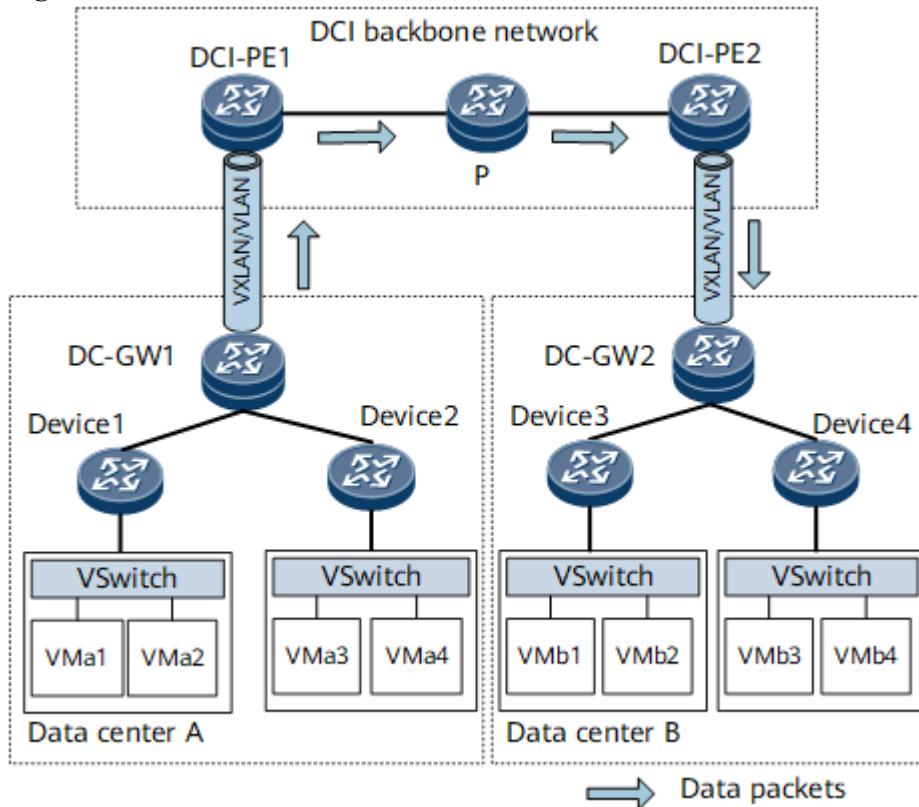


Table 2 Route information

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">• RD1: route distinguisher 1, indicating the route ID of an EVPN instance.• VM-MAC: MAC address of a VM.• VM-IP: IP address of a VM.• Label 1: L2VNI of a VXLAN tunnel or Layer 2 MPLS label.• Label 2: L3VNI of a VXLAN tunnel or Layer 3 MPLS label.• NHP: next hop of a route, usually a local IP address used to establish a BGP EVPN peer relationship.• ExtCommunity: extended community attributes of a route, including the VXLAN encapsulation mode, Router-MAC, and export route target (ERT) of a route.
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">• RD1: route distinguisher 1, indicating the route ID of an EVPN instance.• IP: VM's IP address or address of the network segment to which a VM's IP address belongs.• Label: L3VNI of a VXLAN tunnel or Layer 3 MPLS label.• NHP: next hop of a route, usually a local IP address used to establish a BGP EVPN peer relationship.• ExtCommunity: extended community attributes of a route, including the VXLAN encapsulation mode, Router-MAC, and ERT of a route.

Route	Function	Fields Carried in a Route
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"> RD2: route distinguisher 2, indicating the ID of a VPNv4 route. VM-IP: IP address of a VM. Label: VPN label carried in VPNv4 routes. NHP: next hop of a route, usually a local IP address used to establish a BGP VPNv4 peer relationship. ExtCommunity: extended community attribute of a route, only the ERT attribute.
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"> RD1: route distinguisher 1, indicating the route ID of an EVPN instance. VM-MAC: MAC address of a VM. VM-IP: IP address of a VM. This field is carried only in ARP routes. Label: L2VNI of a VXLAN tunnel or Layer 2 MPLS label. NHP: next hop of a route, usually a local IP address used to establish a BGP EVPN peer relationship. ExtCommunity: extended community attributes of a route, including the VXLAN encapsulation mode and ERT of a route.

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 3](#) describes Layer 3 route advertisement and Layer 2 route advertisement.

Table 3 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.	<p>DCI-PE1 re-encapsulates the EVPN route received from DC-GW1 into a BGP VPNv4 route, applying the following changes:</p> <ul style="list-style-type: none"> Changes the next hop to the local device's IP address used to establish a BGP VPNv4 peer relationship. Replaces the RD and RT values of the EVPN route with those of an L3VPN instance. Applies for and encapsulates a VPN label. <p>After re-encapsulation, DCI-PE1 sends the route to DCI-PE2.</p>	<p>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:</p> <ul style="list-style-type: none"> Changes the next hop to the VTEP address of DCI-PE2. Replaces the RD and RT values of the BGP VPNv4 route with those of the L3VPN instance and pads the route with an L3VNI. <p>After re-encapsulation, DCI-PE2 sends the IP prefix route to DC-GW2.</p>

Deployment Mode	Services	Advertisement Process		
		DC-GW1 to DCI-PE1	DCI-PE1 to DCI-PE2	DCI-PE2 to DC-GW2
EVPN-MPLS (VLAN access)	Layer 3 services	<p>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.</p>	<p>DCI-PE1 re-encapsulates the VPN route into an IP prefix route, applying the following changes:</p> <ul style="list-style-type: none"> Changes the next hop to the local device's IP address used to establish a BGP EVPN peer relationship. Adds the RD and RT attributes to the EVPN route. Applies for and encapsulates a VPN label. <p>After re-encapsulation, DCI-PE1 sends the route to DCI-PE2.</p>	<p>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.</p>

Deployment Mode	Services	Advertisement Process		
		DCI-GW1 to DCI-PE1	DCI-PE1 to DCI-PE2	DCI-PE2 to DCI-GW2
	Layer 2 services	<p>DCI-PE1 learns the source MAC address of service traffic received from DCI-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"> • Changes the next hop to the local device's IP address used to establish a BGP EVPN peer relationship. • Adds the RD and RT attributes to the EVPN route. • Applies for and encapsulates a VPN label. <p>After re-encapsulation, DCI-PE1 sends the route to DCI-PE2.</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
EVPN-MPLS (VXLAN access)	Layer 3 services	<p>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.</p>	<p>DCI-PE1 re-encapsulates the route into an IRB or IP prefix route. The encapsulation mode changes from VXLAN to MPLS:</p> <ul style="list-style-type: none"> Changes the next hop to the local device's IP address used to establish a BGP EVPN peer relationship. Adds the RD and RT attributes to the EVPN route. Applies for and encapsulates a VPN label. <p>After re-encapsulation, DCI-PE1 sends the route to DCI-PE2.</p>	<p>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.</p>
	Layer 2 services	<p>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.</p>	<p>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.</p>	<p>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.</p>

DCI Data Plane

[Table 4](#) describes Layer 2 traffic forwarding and Layer 3 traffic forwarding.

Table 4 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 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.	Upon receipt, DCI-PE1 removes the public MPLS tunnel label, and, based on the VPN label, finds the corresponding VPN instance. Based on the tenant's host IP address, DCI-PE1 searches the corresponding VPN instance forwarding table for the outbound interface to DC-GW1. Then, DCI-PE1 sends the data packet to DC-GW1 through the outbound interface.

Deployment Mode	Services	Forwarding Process		
		DC-GW2 to DCI-PE2	DCI-PE2 to DCI-PE1	DCI-PE1 to DC-GW1
	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 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 EVPN instance. Based on the MAC forwarding entry for the broadcast domain bound to the EVPN instance, DCI-PE1 finds the corresponding outbound interface and sends the data packet to DC-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.

Deployment Mode	Services	Forwarding Process		
		DC-GW2 to DCI-PE2	DCI-PE2 to DCI-PE1	DCI-PE1 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 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.	Upon receipt, DCI-PE1 removes the public MPLS tunnel label and, based on the 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.

Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11.7 NFVI Distributed Gateway (SR Tunnels)

The network function virtualization infrastructure (NFVI) telco cloud solution uses the data center interconnect (DCI)+data center network (DCN) networking. A large amount of mobile phone traffic is sent to virtual unified gateways (vUGWs) and virtual multiservice engines (vMSEs) on the DCN. After being processed by the vUGWs and vMSEs, the IPv4 or IPv6 mobile phone traffic is forwarded over the DCN to destination devices on the Internet. The destination devices send traffic to mobile phones in similar ways. To achieve these functions and ensure traffic load balancing on the DCN, you need to deploy the NFVI distributed gateway function.



A vUGW is a unified gateway developed for Huawei's CloudEdge solution, which can be used for 3GPP access in GPRS, UMTS, and LTE modes. A vUGW can be a GGSN, an S-GW, or a P-GW, which meets the networking requirements of carriers in different stages and operations scenarios.

A vMSE is a virtual type of an MSE. The current carrier network includes multiple functional boxes, including the firewall box, video acceleration box, header enhancement box, and URL filter box. All of these functions are enabled through patch installation, causing a more and more complex network and difficult service provisioning and maintenance. To address the problems, vMSEs incorporate the functions of these boxes, uniformly manage these functions, and implement value-added service processing for the data service initiated by users.

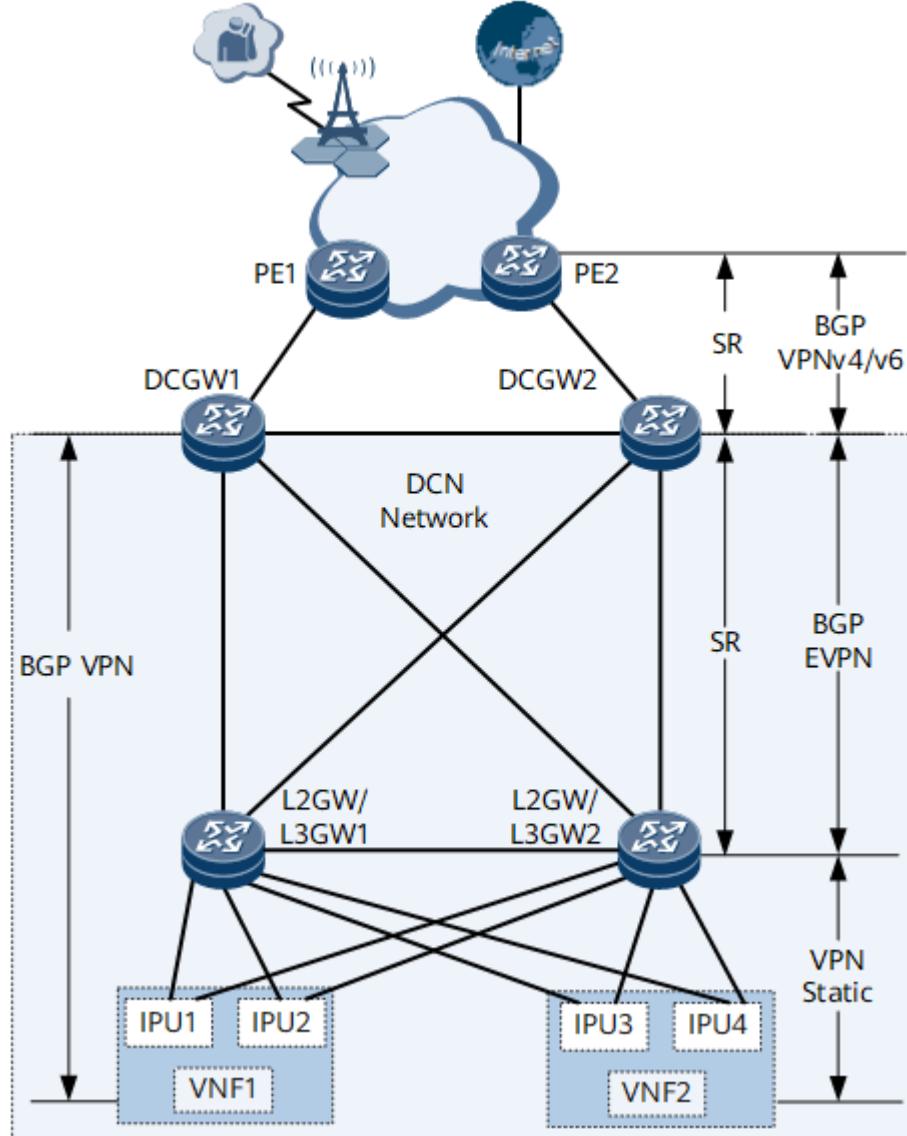
The NFVI distributed gateway function supports service traffic transmission over SR or VXLAN tunnels. SR tunnels are classified as SR tunnels or E2E SR tunnels. This section describes the implementation of the NFVI distributed gateway function for traffic transmission over SR tunnels.

Networking Introduction

[Figure 1](#) shows the networking of an NFVI distributed gateway (SR tunnels). DC-GWs, which are the border gateways of the DCN, exchange Internet routes with external devices over PEs.

L2GW/L3GW1 and L2GW/L3GW2 are connected to VNFs. VNF1 and VNF2 that function as virtualized NEs are deployed to implement the vUGW functions and vMSE functions, respectively. VNF1 and VNF2 are each connected to L2GW/L3GW1 and L2GW/L3GW2 through IPUs.

Figure 1 NFVI distributed gateway networking



Function Deployment

In NFVI distributed gateway networking (SR tunnels) shown in [Figure 1](#), users need to plan the number of bridge domains (BDs) based on the number of network segments corresponding to each IPU. An example assumes that the four IP addresses planned for four IPUs belong to four network segments. In this case, four BDs need to be planned. You need to configure the BDs and the corresponding VBDIF interfaces on all DC-GWs and L2GW/L3GWs and bind all the VBDIF interfaces to the same L3VPN instance. In addition, the following functions need to be deployed on the DCN:

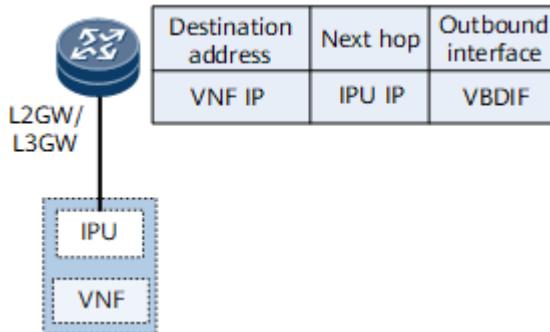
- Establish BGP VPN peer relationships between VNFs and DC-GWs so that the VNFs can advertise mobile phone routes (UE IP) to DC-GWs.
- On L2GW/L3GW1 and L2GW/L3GW2, configure static VPN routes with the IP addresses of VNFs as the destination addresses and the IP addresses of IPUs as next-hop addresses.
- Establish BGP EVPN peer relationships between any DC-GW and L2GW/L3GW. An L2GW/L3GW can flood static routes destined for VNFs to other devices through BGP EVPN peer relationships. A DC gateway can advertise local loopback routes and default routes to L2GWs/L3GWs through BGP EVPN peer relationships.
- Establish BGP VPNv4/v6 peer relationships between DC-GWs and PEs. DC-GWs can then advertise mobile phone routes to PEs and receive the Internet routes sent by external devices based on the BGP VPNv4/v6 peer relationships.
- Deploy SR tunnels between PEs and DC-GWs and between DC-GWs and L2GW/L3GWs to carry service traffic.
- The traffic transmitted between mobile phones and the Internet over VNFs is north-south traffic. The traffic transmitted between VNF1 and VNF2 is east-west traffic. To achieve load balancing of east-west traffic and north-south traffic, deploy the load balancing function on DC-GWs and L2GW/L3GWs.

Establishment of Forwarding Entries

In NFVI distributed gateway networking, to avoid route loops between DC-GWs and L2GW/L3GWs, both the mobile phone-to-Internet traffic and Internet-to-mobile-phone traffic are forwarded from DC-GWs to VNFs at Layer 2 when entering the DCN and forwarded from VNFs to DC-GWs at Layer 3 when leaving out of the DCN. The procedures for establishing forwarding entries on each device are as follows:

1. On L2GW/L3GWs, the number of BDs is planned based on the number of network segments corresponding to the IPUs, the BDs are bound to the links connecting to the corresponding IPUs, and VBDIF interfaces are configured as the gateways of IPUs. Static VPN routes are configured on L2GW/L3GWs so that the forwarding entries with the destination address as a VNF's address, the next-hop address as an IPU's IP address, and outbound interface as the VBDIF interface can be established on L2GW/L3GWs.

Figure 2 Forwarding entries of static routes on L2GW/L3GWs



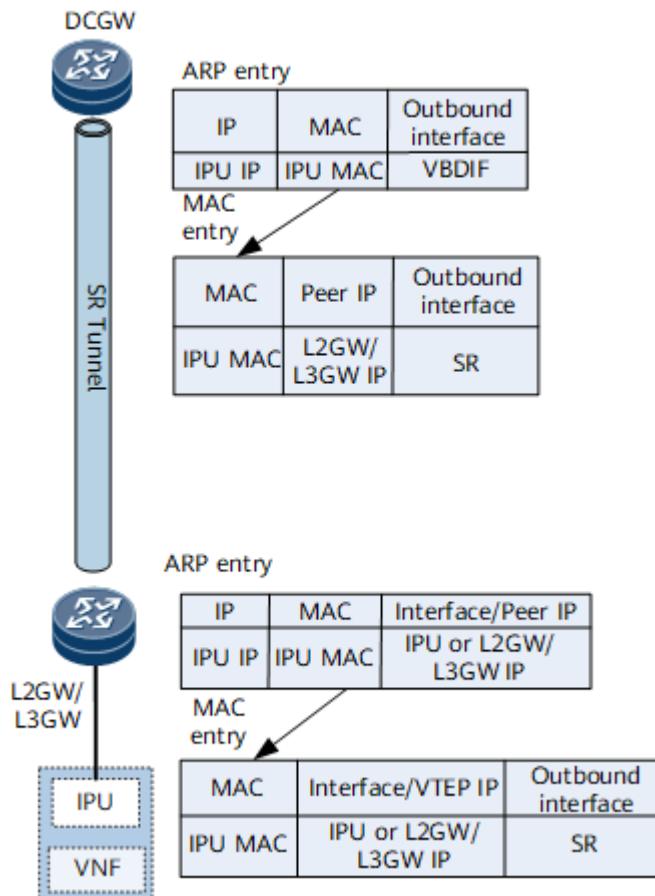
2. L2GW/L3GWs learn the MAC addresses and ARP information of IPUs through the data plane. Such information is advertised to DC-GWs through EVPN routes and can be used for establishment of ARP entries and MAC entries for Layer 2 forwarding.

- The destination MAC addresses in the MAC entries on L2GW/L3GWs are IPUs' MAC addresses. For the IPU directly connected to an L2GW/L3GW, this IPU is used as the outbound interface in MAC entries. For the IPUs (such as IPU3 and IPU4 corresponding to L2GW/L3GW2 in [Figure 1](#)) connecting to non-direct-link L2GW/L3GWs, MAC entries include the outbound interfaces of SR tunnels and the next-hop addresses, namely, the IP addresses of the BGP EVPN peers of these L2GW/L3GWs.
- In the MAC entries on DC-GWs, the destination MAC address is an IPU's MAC address, the IP address of the BGP EVPN peer of an L2GW/L3GW is the next-hop address, and the outbound interface is the one for traffic forwarding over SR tunnels.

NOTE

To allow inbound traffic to be forwarded only at Layer 2, configure these devices to advertise only ARP or ND routes. In this manner, DC-GWs and L2GW/L3GWs do not generate IP prefix routes based on IP addresses. If these devices are configured to advertise IRB or IRBv6 routes, you need to enable asymmetric IRB on the devices receiving routes.

Figure 3 MAC entries on DC-GWs and L2GW/L3GWs



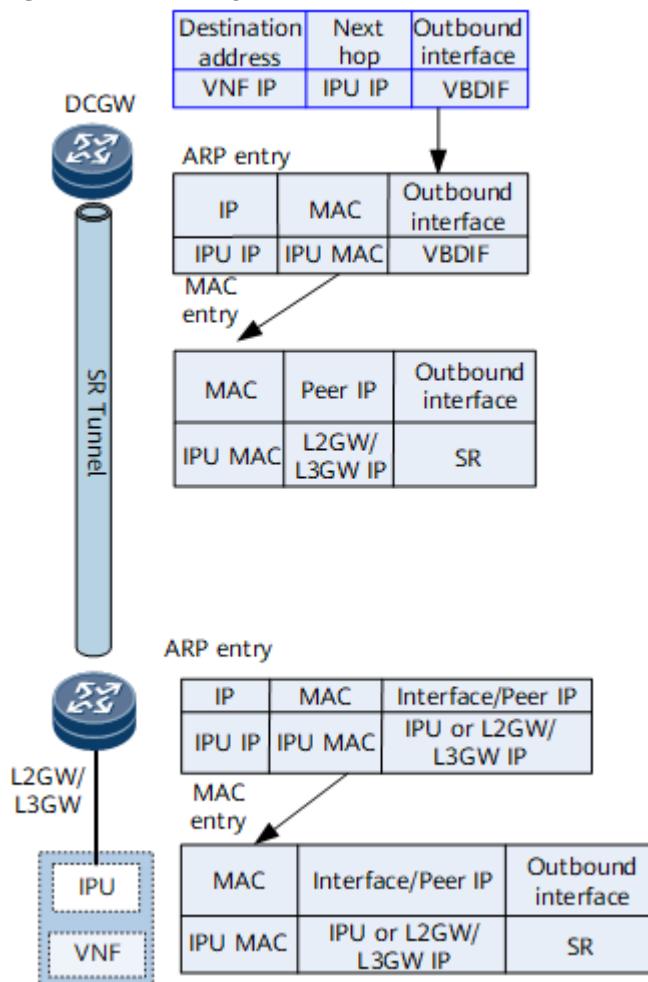
3. After static VPN routes are configured on L2GW/L3GWs, these static VPN routes are imported to the BGP EVPN routing table and sent to DC-GWs as IP prefix routes based on BGP EVPN peer relationships.

NOTE

Because multiple links and static routes exist between L2GW/L3GWs and VNFs, to achieve load balancing, the Add-Path function needs to be enabled during configuration of importing static routes to the BGP EVPN routing table.

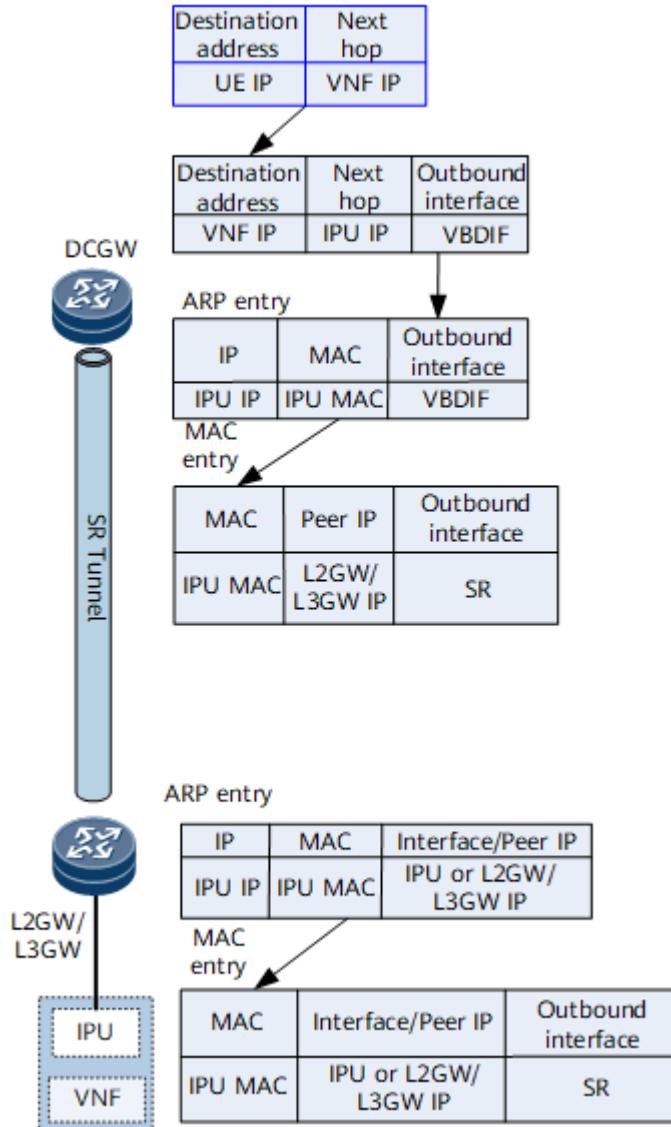
4. By default, the next-hop address of the IP prefix routes received by DC-GWs is the IP address of an L2GW/L3GW and routes are recursed over SR tunnels. This allows inbound traffic to be forwarded at Layer 3. To achieve Layer 2 forwarding of inbound traffic, a route-policy needs to be deployed on L2GW/L3GWs to add the Gateway IP attribute to the static routes destined for DC-GWs. The gateway IP attribute indicates the IP address of an IPU. Upon receipt of the IP prefix routes carrying the gateway IP attribute, DC-GWs recurse routes to next-hop IP addresses, instead of tunnels. In this manner, the destination IP address of the forwarding entries on DC-GWs is a VNF's IP address, the next-hop address is an IPU's IP address, and the outbound interface is the VBDIF interface corresponding to the network segment where the IPU resides. The VBDIF interface looks for a forwarding entry mapped to each VNF IP address and locates the outbound VBDIF interface. ARP entries and MAC entries can be located based on the VBDIF interface to implement Layer 2 forwarding.

Figure 4 Forwarding entries on DC-GWs and L2GW/L3GWs



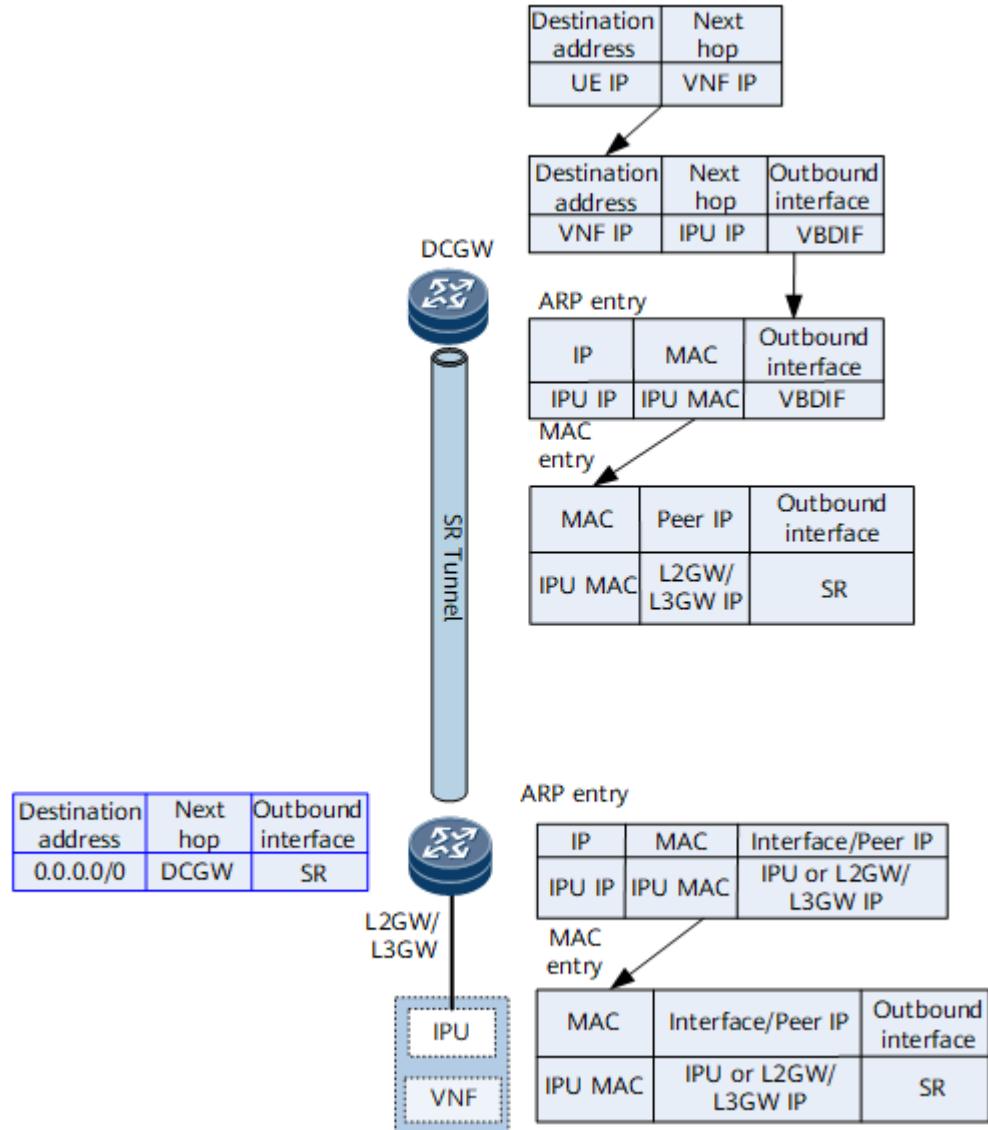
5. To establish BGP VPN peer relationships between DC-GWs and VNFs, DC-GWs need to advertise the routes destined for loopback addresses to L2GW/L3GWs. After BGP VPN peer relationships are established between DC-GWs and VNFs, VNFs can send DC-GWs the mobile phone routes whose next-hop address is a VNF's IP address.

Figure 5 Forwarding entries on DC-GWs and L2GW/L3GWs



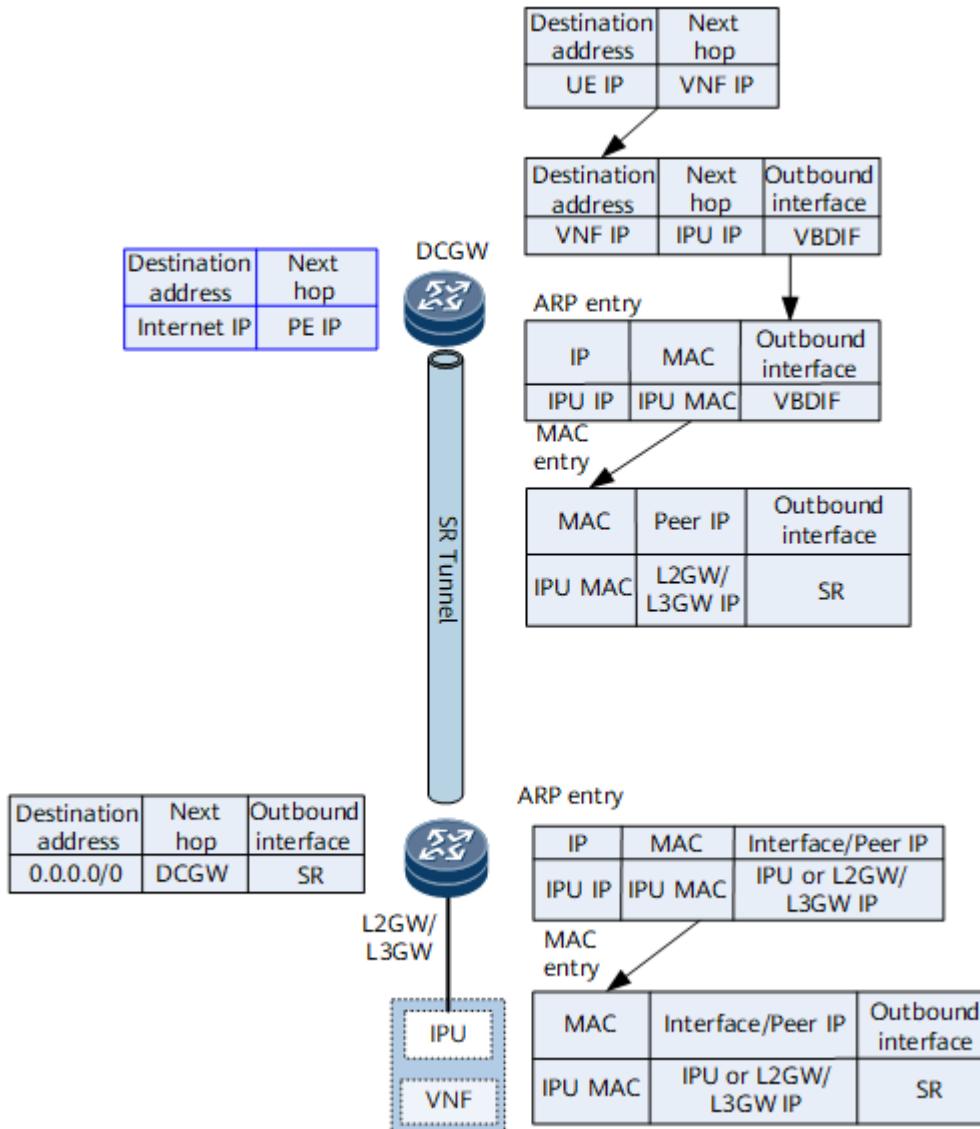
6. Devices on the DCN do not need to get aware of external routes. Therefore, a route-policy needs to be configured on DC-GWs to allow DC-GWs to send only default routes except for the loopback routes to L2GW/L3GWs.

Figure 6 Forwarding entries on DC-GWs and L2GW/L3GWs



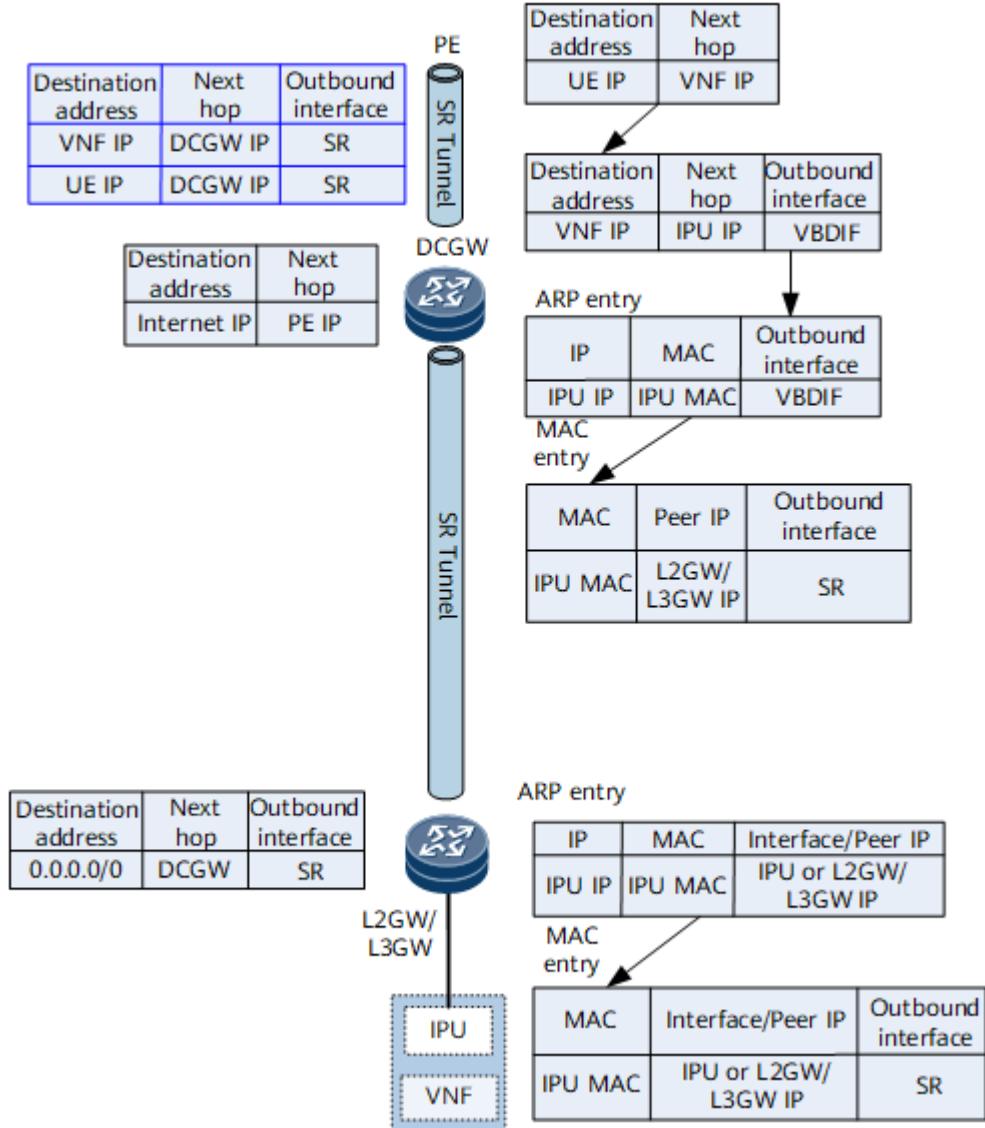
7. DC-GWs function as the border gateways of the DCN and can exchange Internet route information, such as the Internet server address, with PEs.

Figure 7 Forwarding entries on DC-GWs and L2GW/L3GWs



8. PEs can obtain mobile phone routes and the routes carrying VNFs' IP addresses from DC-GWs based on BGP VPNv4/v6 peer relationships. The next-hop addresses of these routes are DC-GWs' IP addresses, and the outbound interface is an SR tunnel connecting to a DC-GW.

Figure 8 Forwarding entries on PEs, DC-GWs and L2GW/L3GWs



9. To achieve load balancing of east-west traffic and north-south traffic, the load balancing function and Add-Path function need to be deployed on DC-GWs and L2GW/L3GWs.

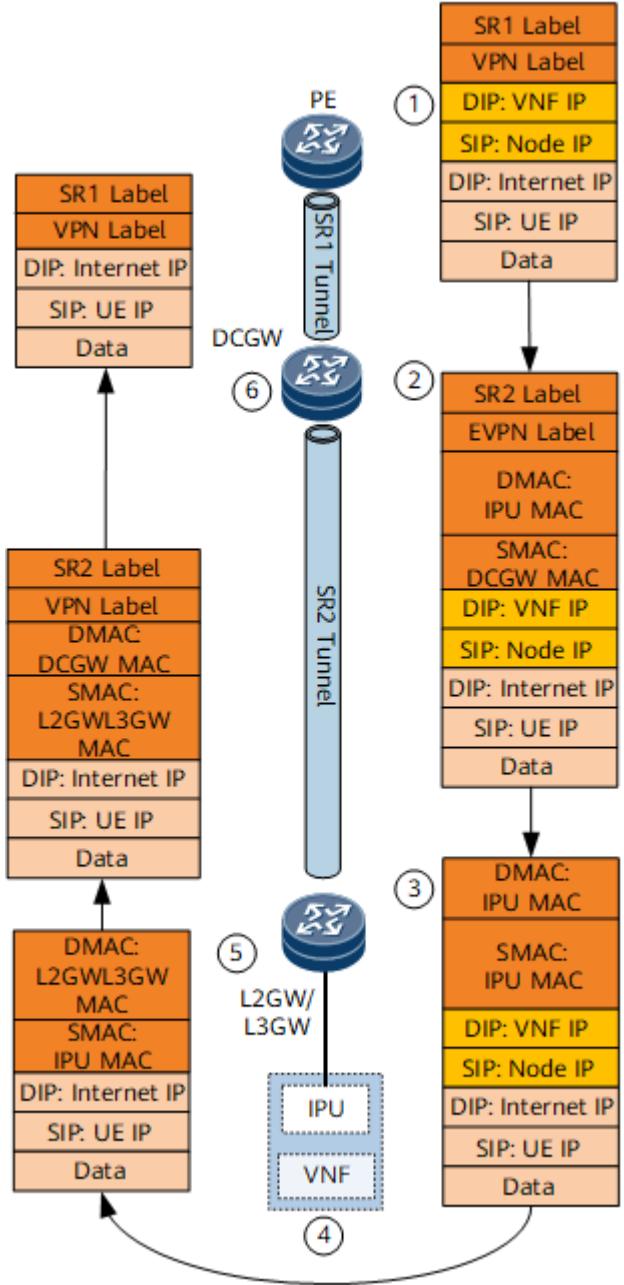
- Load balancing of north-south traffic: Taking DC-GW1 in [Figure 1](#) as an example, DC-GW1 can receive the EVPN routes destined for VNF2 from L2GW/L3GW1 and L2GW/L3GW2. By default, after the load balancing function is configured, DC-GW1 sends half of the traffic destined for VNF2 through L2GW/L3GW1 and the other half of the traffic through L2GW/L3GW2. However, L2GW/L3GW1 connects to VNF2 over one link and L2GW/L3GW2 connects to VNF2 over two links, causing the load balancing function failed to achieve the desired effect. Therefore, the Add-Path function needs to be deployed on L2GW/L3GWs. After the Add-Path function is deployed on L2GW/L3GWs, L2GW/L3GW2 sends two routes with the same destination address to DC-GW1, achieving load balancing.
- East-west traffic load balancing: Taking L2GW/L3GW1 in [Figure 1](#) as an example, because the Add-Path function is deployed on L2GW/L3GW2, L2GW/L3GW1 receives two EVPN routes from L2GW/L3GW2 and L2GW/L3GW1 has a static route with the next-hop address as IPU3's IP address. The destination addresses of all these routes are VNF2's IP address. Therefore, the load balancing function needs to be configured to balance traffic over static routes and EVPN routes.

Traffic Forwarding Process

As shown in [Figure 9](#), the procedure for forwarding north-south traffic from mobile phones to the Internet is as follows:

1. Mobile phone traffic is sent to base stations (Nodes) and encapsulated with a GPRS tunneling protocol (GTP) header. The destination IP address of the GTP tunnel is a VNF's IP address. PEs send the encapsulated packets to DC-GWs over SR tunnels based on the VPN routing table.
2. Upon receipt of the encapsulated packets, DC-GWs look for the VPN routing table and find that the next-hop addresses of the forwarding entries corresponding to VNFs' IP addresses are IPUs' IP addresses and the outbound interface is the VBDIF interface. Therefore, routes destined for the network segment corresponding to the VBDIF interface are hit. DC-GWs search the MAC address that belongs to the network segment in an ARP table, look for the MAC forwarding table based on the ARP information, and forward traffic to L2GW/L3GWs over SR tunnels based on the MAC forwarding table.
3. Upon receipt of packets, L2GW/L3GWs find the corresponding BDs based on EVPN labels, look for MAC forwarding entries in the BDs, and forward traffic to VNFs based on the MAC information.
4. Upon receipt of packets, VNFs decapsulate the GTP tunnel header, search the routing table based on the destination IP address in the decapsulated packets, and forward the packets to L2GW/L3GWs based on the default gateways of VNFs.
5. L2GW/L3GWs search for the VPN routing table on L2GW/L3GWs. The default routes advertised by DC-GWs to L2GW/L3GWs are recursed over SR tunnels and forwarded to DC-GWs after being encapsulated with a VPN label.
6. DC-GWs forward the packets to PEs using the VPN forwarding table found based on VPN labels. Specifically, the packets are re-encapsulated with a VPN label and an SR label before they are forwarded to PEs.

Figure 9 North-south traffic forwarding from mobile phones to the Internet



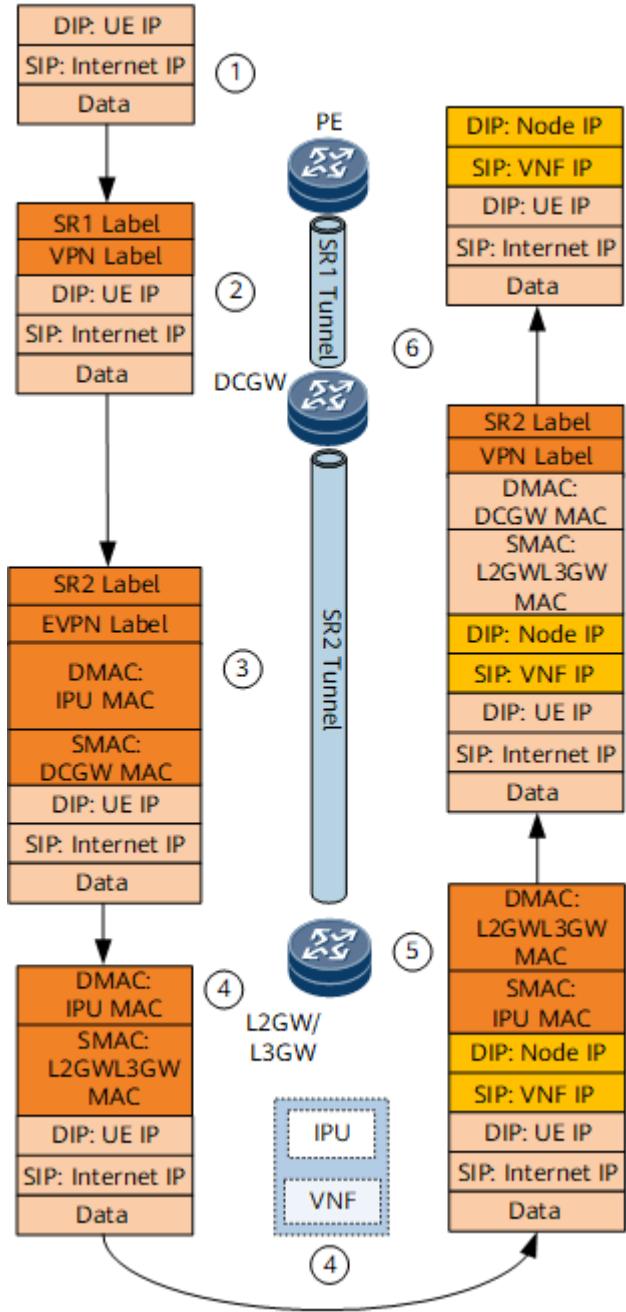
As shown in [Figure 10](#), the procedure for forwarding east-west traffic from the Internet to mobile phones over VNFs is as follows:

1. Devices on the Internet send mobile phones the reply packets whose destination IP addresses are the IP addresses of mobile phones. This is because mobile phone routes are advertised by VNFs to DC-GWs based on BGP VPN peer relationships and then advertised to the Internet by DC-GWs through PEs. Therefore, reply packets must be first transmitted to VNFs.
2. Upon receipt of reply packets, PEs search the VPN routing table for the forwarding entries corresponding to mobile phone routes whose next-hop addresses are DC-GWs' IP addresses and the outbound interface is the one for SR tunnels. The reply packets are sent to DC-GWs after being encapsulated with a VPN label and an SR label.
3. Upon receipt of the reply packets, DC-GWs search the VPN routing table for the forwarding entries corresponding to the mobile phone routes based on the BGP VPN peer relationships between DC-GWs and VNFs. These routes are recursed to one or more

VBDIF interfaces, and traffic is load balanced to these VBDIF interfaces. The VBDIF interfaces look for ARP information and MAC forwarding entries. Based on the MAC entries, the reply packets are forwarded to L2GW/L3GWs over SR tunnels after being encapsulated with an EVPN label.

4. Upon receipt of the reply packets, L2GW/L3GWs find the corresponding BDs based on the EVPN label and search for the MAC forwarding entries. The outbound interface information is then obtained based on the MAC forwarding entries. The reply packets are then forwarded to VNFs.
5. Upon receipt of the reply packets, VNFs search for the base stations corresponding to the destination IP addresses of mobile phones and add a tag of tunnel information with the destination IP address as the IP address of a base station. The reply packets are then forwarded to L2GW/L3GWs based on default gateways.
6. Upon receipt of the reply packets, L2GW/L3GWs search the VPN routing table, and the default routes advertised by DC-GWs to L2GW/L3GWs are hit. The reply packets are then forwarded to DC-GWs over SR tunnels after being encapsulated with a VPN label.
7. Upon receipt of the reply packets, DC-GWs find the corresponding VPN forwarding table based on the VPN label, and the default routes or specific routes are hit. The reply packets are then forwarded to PEs over SR tunnels and then from PEs to base stations. After being encapsulated by base stations, the reply packets are sent to the corresponding mobile phones.

Figure 10 East-west traffic forwarding from the Internet to mobile phones

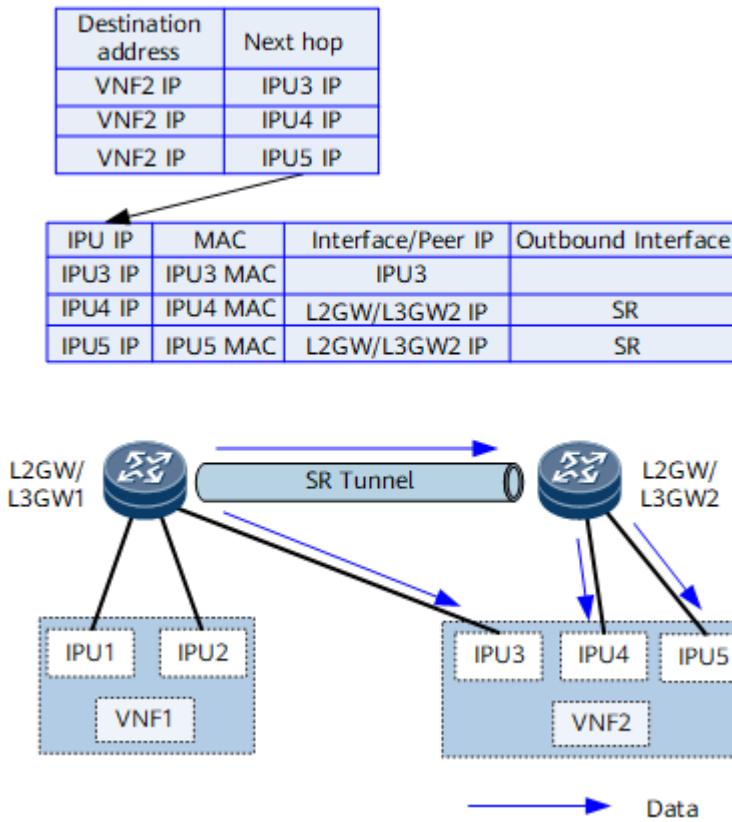


Upon receipt of user packets, a VNF finds that the packets need to be sent to another VNF for value-added service processing. In this case, east-west traffic occurs. On the network shown in [Figure 11](#), the difference between forwarding east-west traffic and forwarding north-south traffic lies in the processing of packets after they arrive VNF1.

1. Upon receipt of user packets, VNF1 finds that the packets need to be processed by VNF2 and then adds a tunnel label with the destination IP address as VNF2's IP address. The user packets are sent to L2GW/L3GWs based on default routes.
2. Upon receipt of user packets, an L2GW/L3GW searches the VPN forwarding table and finds that multiple load-balancing forwarding entries exist. In some entries, the outbound interface is an IPU or the next-hop address is the IP address of another L2GW/L3GW.
3. If the traffic hits the path of another L2GW/L3GW, an EVPN label is added to the user packets and the routes are recursed to L2GW/L3GW2 over an SR tunnel. L2GW/L3GW2 searches for the BD and destination MAC address based on the EVPN label before forwarding the packets to VNF2.

- Upon receipt of the user packets, VNF2 processes and forwards the packets to the server. The subsequent forwarding follows the north-south traffic forwarding procedure.

Figure 11 East-west traffic forwarding from VNF1 to VNF2



Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11.8 NFVI Distributed Gateway Function (BGP VPNv4/v6 over E2E SR Tunnels)

The NFVI telco cloud solution uses the DCI+DCN networking. A large amount of mobile phone traffic is sent to vUGWs and vMSEs on the DCN. After being processed by the vUGWs and vMSEs, the IPv4 or IPv6 mobile phone traffic is forwarded over the DCN to destination devices on the Internet. The destination devices send traffic to mobile phones in similar ways. To achieve these functions and ensure traffic load balancing on the DCN, you need to deploy the NFVI distributed gateway function.

NOTE

A vUGW is a unified gateway developed for Huawei's CloudEdge solution, which can be used for 3GPP access in GPRS, UMTS, and LTE modes. A vUGW can be a GGSN, an S-GW, or a P-GW, which meets the networking requirements of carriers in different stages and operations scenarios.

A vMSE is a virtual type of an MSE. The current carrier network includes multiple functional boxes, including the firewall box, video acceleration box, header enhancement box, and URL filter box. All of these functions are enabled through patch installation, causing a more and more complex network and difficult service provisioning and maintenance.

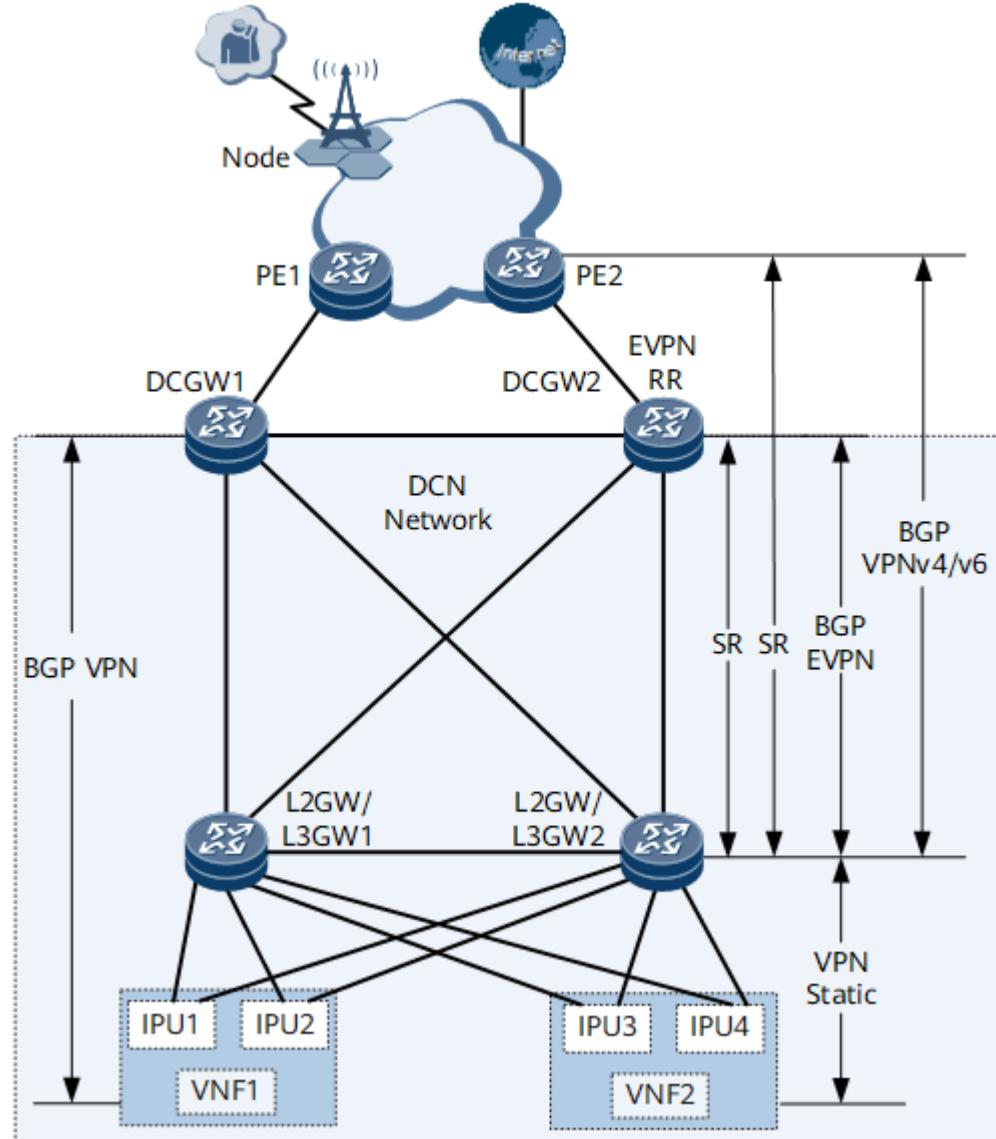
To address the problems, vMSEs incorporate the functions of these boxes, uniformly manage these functions, and implement value-added service processing for the data service initiated by users.

The NFVI distributed gateway function supports service traffic transmission over SR or VXLAN tunnels. SR tunnels are classified as segmented SR tunnels or E2E SR tunnels. In E2E SR tunnel scenarios, PEs use BGP VPNv4/VPNv6 or BGP EVPN to connect to a DCN. The control-plane implementation varies according to the protocol used. This section uses BGP VPNv4/VPNv6 as an example.

Networking Introduction

[Figure 1](#) shows the networking of an NFVI distributed gateway (BGP VPNv4/VPNv6 over E2E SR tunnels). DC-GWs, which are the border gateways of the DCN, exchange Internet routes with external devices over PEs. L2GW/L3GW1 and L2GW/L3GW2 are connected to VNFs. VNF1 and VNF2 that function as virtualized NEs are deployed to implement the vUGW functions and vMSE functions, respectively. VNF1 and VNF2 are each connected to L2GW/L3GW1 and L2GW/L3GW2 through IPUs.

Figure 1 NFVI distributed gateway networking



Function Deployment

On the network shown in [Figure 1](#), the number of BDs needs to be planned based on the number of network segments corresponding to each IPU. An example assumes that the four IP addresses planned for four IPUs belong to four network segments. In this case, four BDs need to be planned. You need to configure the BDs and the corresponding VBDIF interfaces on all L2GW/L3GWs and bind all the VBDIF interfaces to the same L3VPN instance. In addition, the following functions need to be deployed on the DCN:

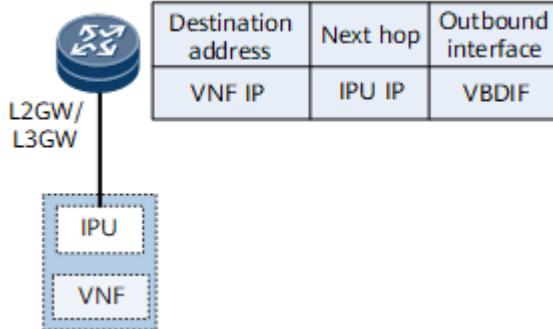
- Establish BGP VPN peer relationships between VNFs and DC-GWs so that the VNFs can advertise mobile phone routes (UE IP) to DC-GWs.
- On L2GW/L3GW1 and L2GW/L3GW2, configure static VPN routes with the IP addresses of VNFs as the destination addresses and the IP addresses of IPUs as next-hop addresses.
- Establish BGP EVPN peer relationships between any DC-GW and L2GW/L3GW. The DC-GW can then advertise local loopback routes and default routes to the L2GW/L3GW. A route-policy needs to be configured on DC-GWs so that the routes sent by DC-GWs to L2GW/L3GWs carry gateway addresses and the next hops of the mobile phone routes received by L2GW/L3GWs from DC-GWs are the VNF addresses. In addition, the BGP EVPN peer relationships established between DC-GWs and L2GW/L3GWs can be used to advertise the routes carrying the IP addresses used for establishing BGP VPN peer relationships, and the BGP EVPN peer relationships established between L2GW/L3GWs can be used to synchronize the MAC or ARP routes and the IP prefix routes carrying gateway addresses with IPUs.
- Deploy EVPN RRs which can be either a standalone device or a DC-GW. In this section, DC-GWs function as EVPN RRs, and L2GW/L3GWs function as RR clients. L2GW/L3GWs can use EVPN RRs to synchronize MAC or ARP routes as well as the IP prefix routes carrying a VNF address as the destination address with IPUs.
- Establish BGP VPNV4/v6 peer relationships between L2GW/L3GWs and PEs. L2GW/L3GWs advertise mobile phone routes to PEs based on BGP VPNV4/v6 peer relationships. DC-GWs send mobile phone routes to L2GW/L3GWs based on BGP EVPN peer relationships. Therefore, mobile phone routes need to be re-encapsulated as BGP VPNV4/v6 routes on L2GW/L3GWs before being advertised to PEs.
- Configure static default routes on PEs and configure the PEs to send static default routes to L2GW/L3GWs based on BGP VPNV4/v6 peer relationships.
- Deploy SR tunnels between PEs and L2GW/L3GWs and between DC-GWs and L2GW/L3GWs to carry service traffic.
- The traffic transmitted between mobile phones and the Internet over VNFs is north-south traffic. The traffic transmitted between VNF1 and VNF2 is east-west traffic. To achieve load balancing of east-west traffic and north-south traffic, deploy the load balancing function on DC-GWs and L2GW/L3GWs.

Establishment of Forwarding Entries

In the networking of an NFVI distributed gateway (E2E SR tunnels), the procedures for establishing forwarding entries on each device are as follows:

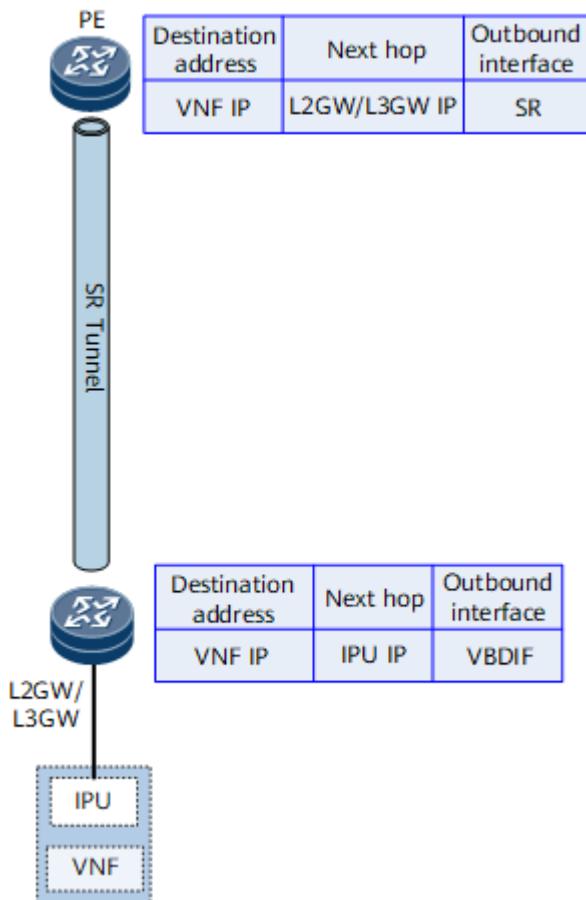
1. On L2GW/L3GWs, the number of BDs is planned based on the number of network segments corresponding to the IPUs, the BDs are bound to the links connecting to the corresponding IPUs, and VBDIF interfaces are configured as the gateways of IPUs. Static VPN routes are configured on L2GW/L3GWs so that the forwarding entries with the destination address as a VNF's address, the next-hop address as an IPU's IP address, and outbound interface as the VBDIF interface can be established on L2GW/L3GWs.

Figure 2 Forwarding entries of static routes on L2GW/L3GWs



- After static VPN routes are configured on L2GW/L3GWs, these static VPN routes are imported to the BGP EVPN routing table and sent to DC-GWs as IP prefix routes based on BGP EVPN peer relationships. A route-policy needs to be configured on L2GW/L3GWs so that L2GE/L3GWs send PEs only the routes destined for VNFs.

Figure 3 Forwarding entries on PEs and L2GW/L3GWs



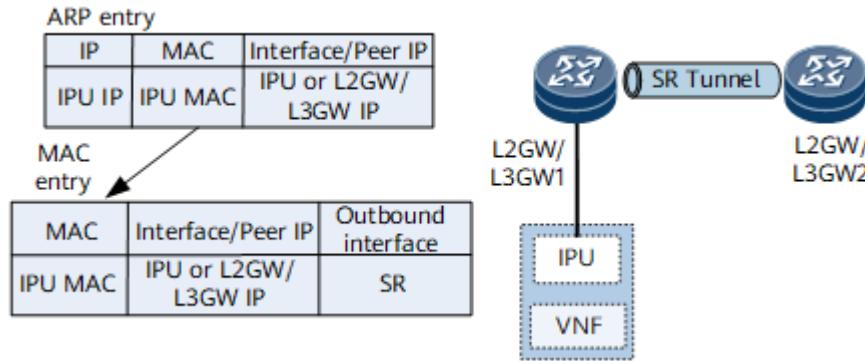
- An L2GW/L3GW learns the MAC addresses and ARP information of IPUs through the data plane. Such information is advertised to another L2GW/L3GW through EVPN routes and can be used for establishment of ARP entries and MAC entries for Layer 2 forwarding. Taking L2GW/L3GW1 as an example, the destination MAC address of the MAC entries on L2GW/L3GW1 is an IPU's MAC address. For the IPU directly connected to L2GW/L3GW1, the IPU's interface is used as the outbound interface in MAC entries. For the IPU connected to another L2GW/L3GW, the MAC entries contain the outbound interface for SR tunnels and the IP address of the BGP EVPN peer of this L2GW/L3GW as the next-hop address.



L2GW/L3GWs exchange routes with a VNF's IP address as the destination IP address. If two VNFs are connected to different L2GW/L3GWs, traffic is forwarded over routes. Otherwise, route loops may occur. Therefore, a route-policy needs to be configured on L2GW/L3GWs, so that the Gateway IP attribute is added to the routes exchanged between L2GW/L3GWs. The Gateway IP attribute is still the next-hop address, which is an IPU's IP address, and the outbound interface is the VBDIF interface. In this manner, L2GW/L3GWs forward traffic based on the MAC forwarding table to prevent route loops.

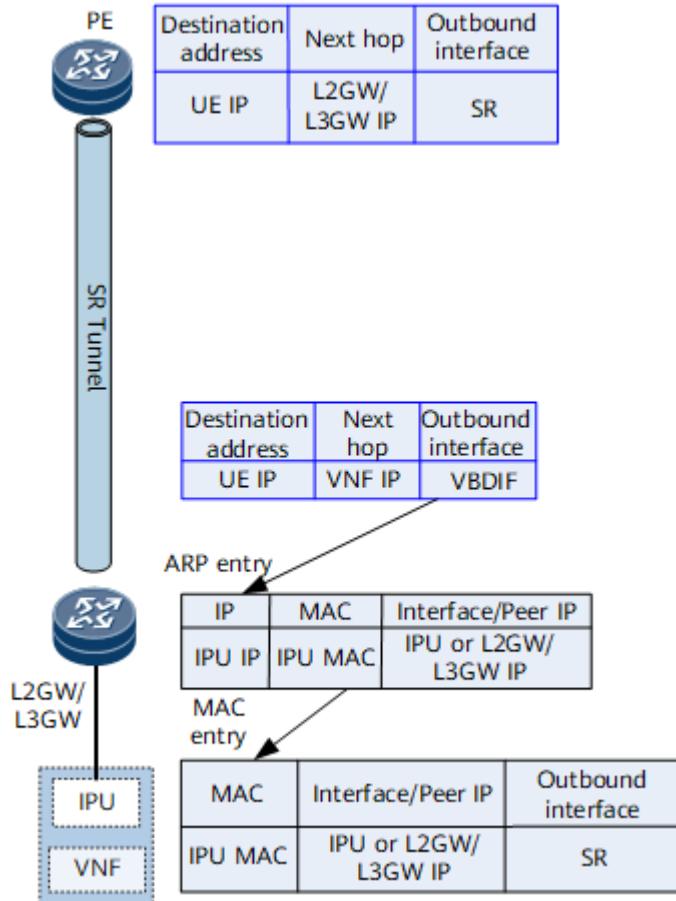
Because multiple links and static routes exist between L2GW/L3GWs and VNFs, to achieve load balancing, the Add-Path function needs to be enabled during configuration of importing static routes to the BGP EVPN routing table.

Figure 4 MAC entries on L2GW/L3GWs



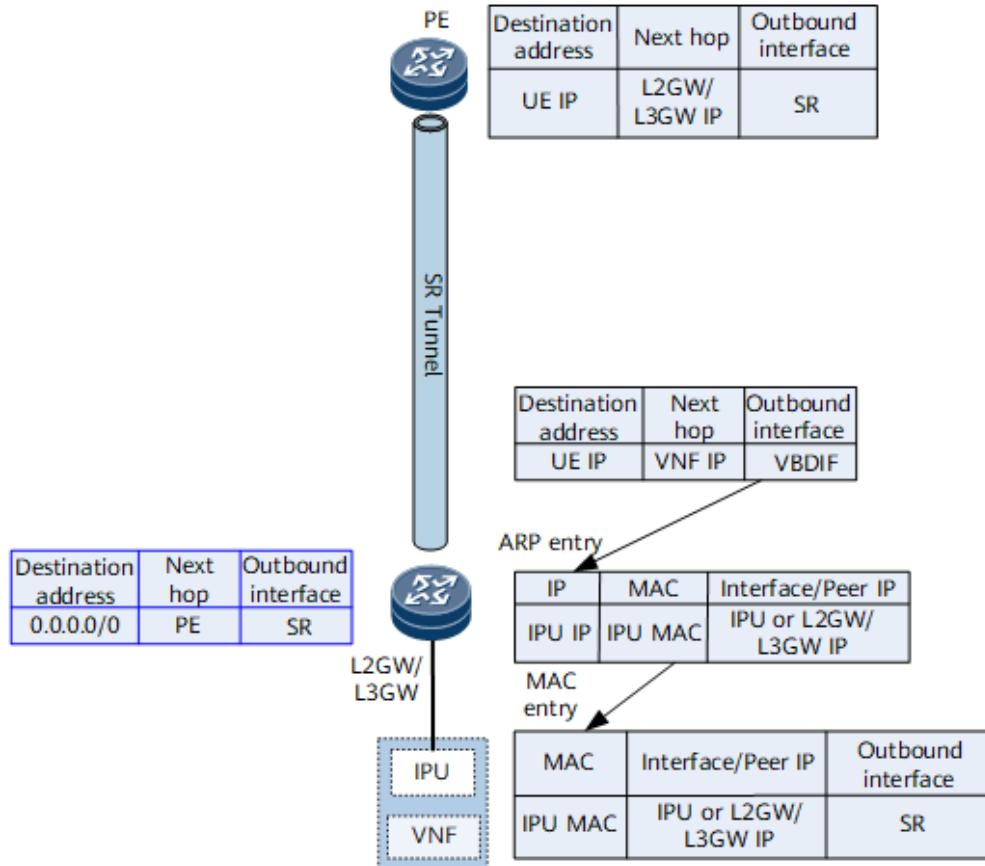
4. DC-GWs can receive the IP prefix routes with a VNF's IP address as the destination IP address from L2GW/L3GWs. However, DC-GWs do not have VBDIF interfaces and therefore cannot recurse routes to the VBDIF interface based on the Gateway IP attribute. DC-GWs need to recurse routes to SR tunnels based on IP prefix routes. Therefore, DC-GWs must be enabled with the function to ignore the Gateway IP attribute to send BGP packets to VNFs over SR tunnels based on next-hop addresses.
5. To establish BGP VPN peer relationships between DC-GWs and VNFs, DC-GWs need to advertise the routes destined for loopback addresses to L2GW/L3GWs. After BGP VPN peer relationships are established between VNFs and DC-GWs, VNFs send mobile phone routes to DC-GWs, and DC-GWs send mobile phone routes to L2GW/L3GWs based on the BGP EVPN peer relationships. A route-policy needs to be configured on DC-GWs to add the Gateway IP attribute to the routes. The Gateway IP attribute is the IP address of a VNF. A route-policy also needs to be configured on L2GW/L3GWs to allow L2GW/L3GWs to receive only the mobile phone routes generated by the directly connected VNFs. This prevents L2GW/L3GWs from sending a large number of repetitive routes to PEs. Upon receipt of mobile phone routes from DC-GWs, L2GW/L3GWs generate VPN forwarding entries and re-encapsulate EVPN routes as BGP VPvN4/v6 routes before sending the routes to PEs. PEs then generate VPN forwarding entries with the IP address of mobile phone routes as the destination address, an L2GW/L3GW's IP address as the next-hop address, and the interface for SR tunnels as the outbound interface.

Figure 5 Forwarding entries on PEs and L2GW/L3GWs



6. Devices on the DCN do not need to get aware of external routes. Therefore, route-policies need to be configured on PEs to allow PEs to send only default routes to L2GW/L3GWs.

Figure 6 Forwarding entries on PEs and L2GW/L3GWs



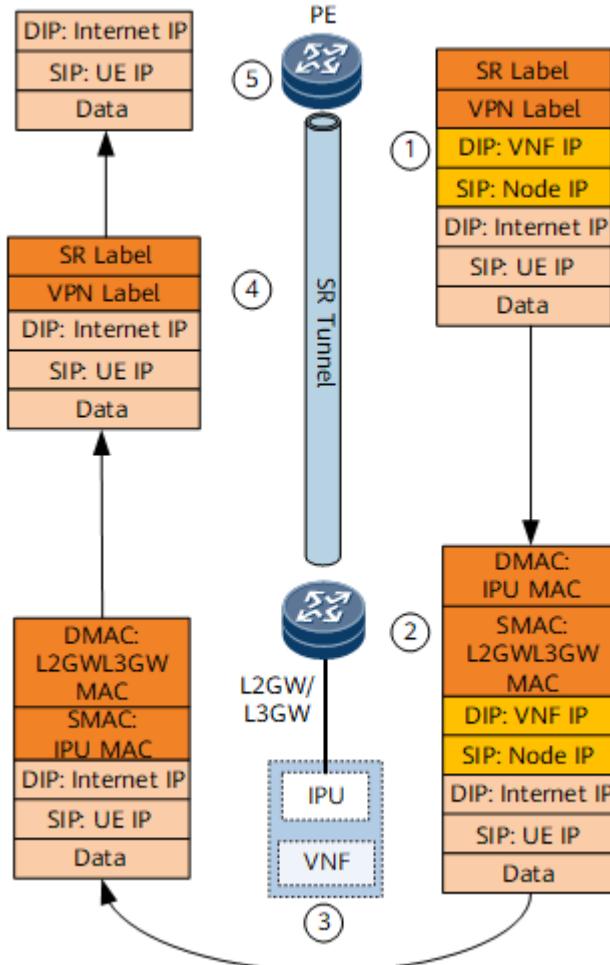
7. PEs can exchange information about Internet routes, such as the Internet server address, with external devices.
8. To achieve load balancing of east-west traffic and north-south traffic, the load balancing function and Add-Path function need to be deployed on PEs and L2GW/L3GWs.
 - Load balancing of north-south traffic: Taking PE1 in [Figure 1](#) as an example, PE1 can receive the VPN routes destined for VNF2 from L2GW/L3GW1 and L2GW/L3GW2. By default, after the load balancing function is configured, PE1 sends half of the traffic destined for VNF2 through L2GW/L3GW1 and the other half of the traffic through L2GW/L3GW2. However, L2GW/L3GW1 connects to VNF2 over one link and L2GW/L3GW2 connects to VNF2 over two links, causing the load balancing function failed to achieve the desired effect. Therefore, the Add-Path function needs to be deployed on L2GW/L3GWs. After the Add-Path function is deployed on L2GW/L3GWs, L2GW/L3GW2 sends two routes with the same destination address to DC-GW1, achieving load balancing.
 - East-west traffic load balancing: Taking L2GW/L3GW1 in [Figure 1](#) as an example, because the Add-Path function is deployed on L2GW/L3GW2, L2GW/L3GW1 receives two EVPN routes from L2GW/L3GW2 and L2GW/L3GW1 has a static route with the next-hop address as IPU3's IP address. The destination addresses of all these routes are VNF2's IP address. Therefore, the load balancing function needs to be configured to balance traffic over static routes and EVPN routes.

Traffic Forwarding Process

As shown in [Figure 7](#), the procedure for forwarding north-south traffic from mobile phones to the Internet is as follows:

1. Mobile phone traffic is sent to base stations (Nodes) and encapsulated with a GPRS tunneling protocol (GTP) header. The destination IP address of the GTP tunnel is a VNF's IP address. PEs send the encapsulated packets to L2GW/L3GWs over SR tunnels based on the VPN routing table.
2. Upon receipt of the encapsulated packets, L2GW/L3GWs look for the VPN routing table and find that the next-hop addresses of the forwarding entries corresponding to VNFs' IP addresses are IPU's IP addresses and the outbound interface is the VBDIF interface. Therefore, routes destined for the network segment corresponding to the VBDIF interface are hit. L2GW/L3GWs search the MAC address that belongs to the network segment in an ARP table, look for the MAC forwarding table based on the ARP information, and forward traffic to VNFs based on the MAC forwarding table.
3. Upon receipt of packets, VNFs decapsulate the GTP tunnel header, search the routing table based on the destination IP address in the decapsulated packets, and forward the packets to L2GW/L3GWs based on the default gateways of VNFs.
4. L2GW/L3GWs search for the VPN routing table on L2GW/L3GWs. The default routes advertised by PEs to L2GW/L3GWs are recursed over SR tunnels and forwarded to PEs after being encapsulated with a VPN label.
5. PEs use the VPN forwarding table to forward the packets to the Internet based on the VPN label.

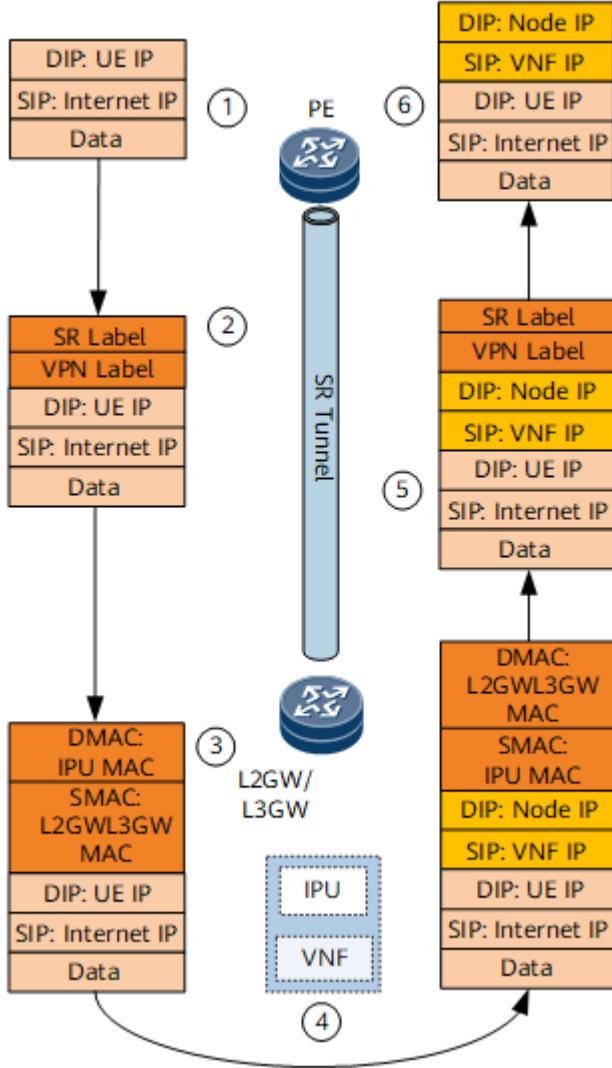
Figure 7 North-south traffic forwarding from mobile phones to the Internet



As shown in [Figure 8](#), the procedure for forwarding east-west traffic from the Internet to mobile phones over VNFs is as follows:

1. Devices on the Internet send mobile phones the reply packets whose destination IP addresses are the IP addresses of mobile phones. This is because mobile phone routes are advertised by L2GW/L3GWs to VNFs based on BGP VPNv4/v6 peer relationships and then advertised to the Internet by PEs. Therefore, reply packets must be first transmitted to L2GW/L3GWs.
2. Upon receipt of reply packets, PEs search the VPN routing table for the forwarding entries corresponding to mobile phone routes whose next-hop addresses are L2GW/L3GWs' IP addresses and the outbound interface is the one for SR tunnels. The reply packets are sent to L2GW/L3GWs after being encapsulated with a VPN label and an SR label.
3. Upon receipt of reply packets, L2GW/L3GWs find the VPN forwarding entries based on the VPN label and the VBDIF interface based on the VPN forwarding entries. The packets are then forwarded to VNFs based on the MAC entries corresponding to the VBDIF interface.
4. Upon receipt of the reply packets, VNFs search for the base stations corresponding to the destination IP addresses of mobile phones and add a tag of tunnel information with the destination IP address as the IP address of a base station. The reply packets are then forwarded to L2GW/L3GWs based on default gateways.
5. Upon receipt of the reply packets, L2GW/L3GWs search the VPN routing table, and the default routes advertised by DC-GWs to L2GW/L3GWs are hit. The reply packets are then forwarded to DC-GWs over SR tunnels after being encapsulated with a VPN label.
6. Upon receipt of reply packets, PEs use the VPN forwarding table to forward the packets to base stations based on the VPN label. The base stations decapsulate the packets before forwarding them to mobile phones.

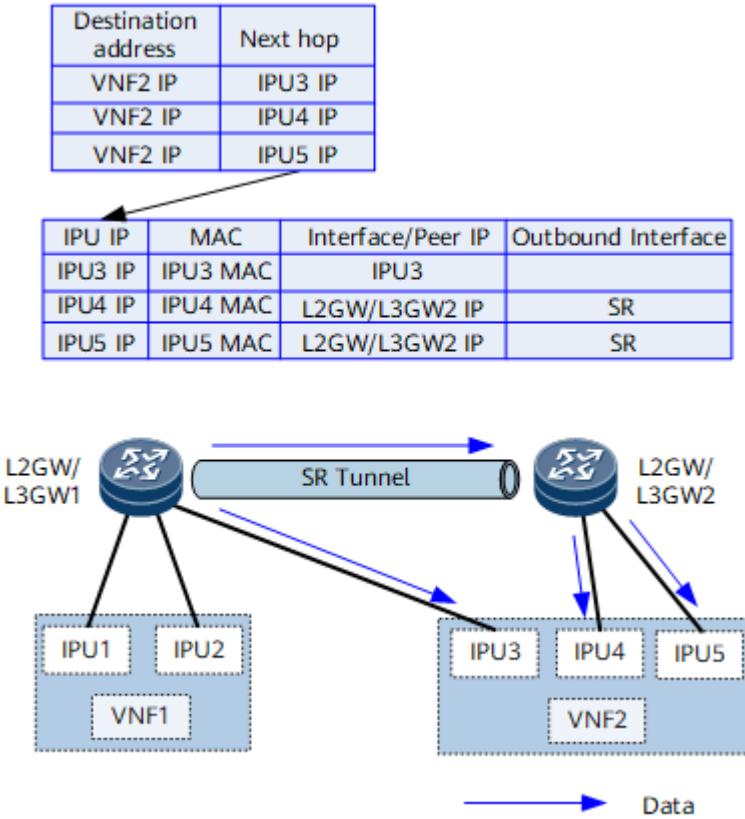
Figure 8 East-west traffic forwarding from the Internet to mobile phones



Upon receipt of user packets, a VNF finds that the packets need to be sent to another VNF for value-added service processing. In this case, east-west traffic occurs. On the network shown in [Figure 9](#), the difference between forwarding east-west traffic and forwarding north-south traffic lies in the processing of packets after they arrive VNF1.

1. Upon receipt of user packets, VNF1 finds that the packets need to be processed by VNF2 and then adds a tunnel label with the destination IP address as VNF2's IP address. The user packets are sent to L2GW/L3GWs based on default routes.
2. Upon receipt of user packets, an L2GW/L3GW searches the VPN forwarding table and finds that multiple load-balancing forwarding entries exist. In some entries, the outbound interface is an IPU or the next-hop address is the IP address of another L2GW/L3GW.
3. If the traffic hits the path of another L2GW/L3GW, an EVPN label is added to the user packets and the routes are recursed to L2GW/L3GW2 over an SR tunnel. L2GW/L3GW2 searches for the BD and destination MAC address based on the EVPN label before forwarding the packets to VNF2.
4. Upon receipt of the user packets, VNF2 processes and forwards the packets to the server. The subsequent forwarding follows the north-south traffic forwarding procedure.

Figure 9 East-west traffic forwarding from VNF1 to VNF2



Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11.9 NFVI Distributed Gateway Function (BGP EVPN over E2E SR Tunnels)

The NFVI telco cloud solution uses the DCI+DCN networking. A large amount of mobile phone traffic is sent to vUGWs and vMSEs on the DCN. After being processed by the vUGWs and vMSEs, the IPv4 or IPv6 mobile phone traffic is forwarded over the DCN to destination devices on the Internet. The destination devices send traffic to mobile phones in similar ways. To achieve these functions and ensure traffic load balancing on the DCN, you need to deploy the NFVI distributed gateway function.

NOTE

A vUGW is a unified gateway developed for Huawei's CloudEdge solution, which can be used for 3GPP access in GPRS, UMTS, and LTE modes. A vUGW can be a GGSN, an S-GW, or a P-GW, which meets the networking requirements of carriers in different stages and operations scenarios.

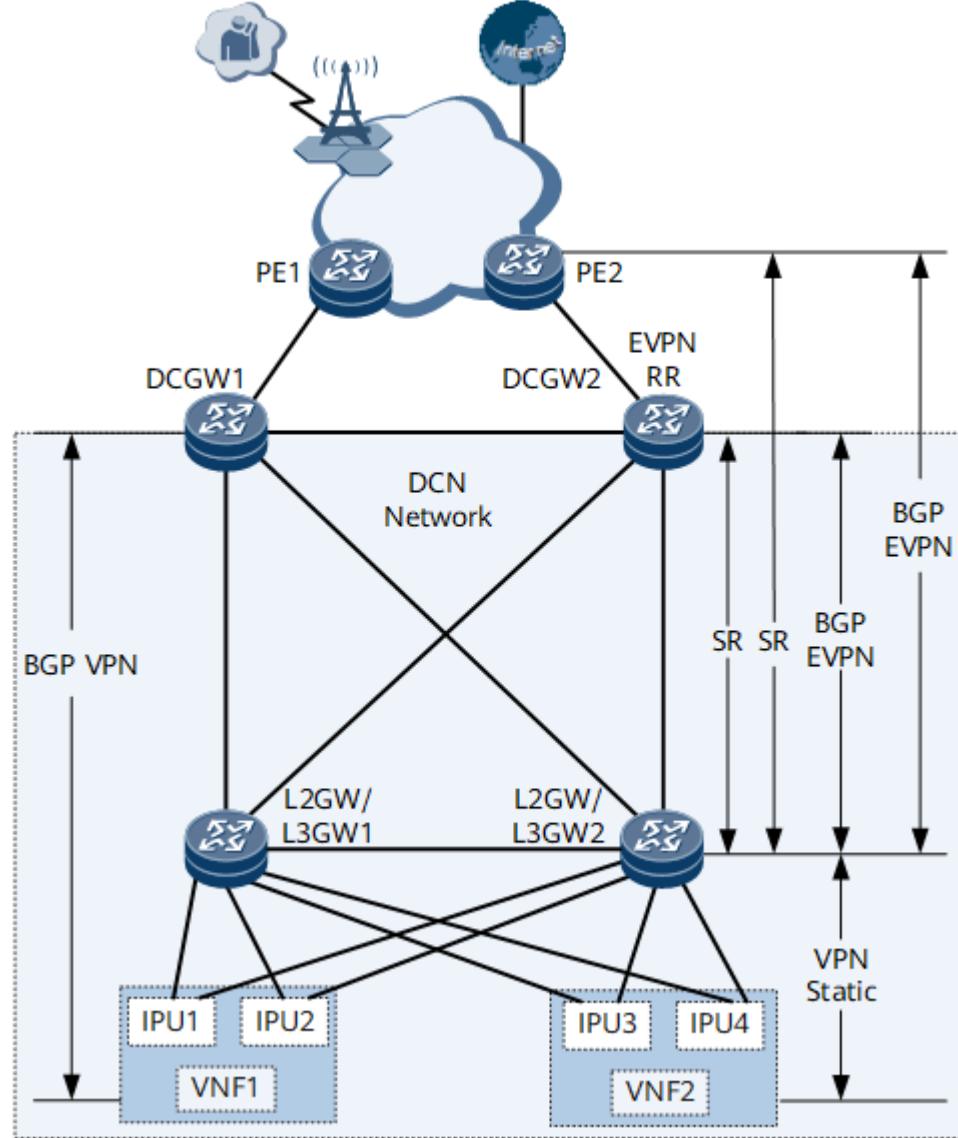
A vMSE is a virtual type of an MSE. The current carrier network includes multiple functional boxes, including the firewall box, video acceleration box, header enhancement box, and URL filter box. All of these functions are enabled through patch installation, causing a more and more complex network and difficult service provisioning and maintenance. To address the problems, vMSEs incorporate the functions of these boxes, uniformly manage these functions, and implement value-added service processing for the data service initiated by users.

The NFVI distributed gateway function supports service traffic transmission over SR or VXLAN tunnels. SR tunnels are classified as segmented SR tunnels or E2E SR tunnels. In E2E SR tunnel scenarios, PEs use BGP VPNv4/VPNv6 or BGP EVPN to connect to a DCN. The control-plane implementation varies according to the protocol used. This section describes the implementation principles in BGP EVPN scenarios.

Networking Introduction

[Figure 1](#) shows the networking of an NFVI distributed gateway (BGP EVPN over E2E SR tunnels). DC-GWs, which are the border gateways of the DCN, exchange Internet routes with external devices over PEs. L2GW/L3GW1 and L2GW/L3GW2 are connected to VNFs. VNF1 and VNF2 that function as virtualized NEs are deployed to implement the vUGW functions and vMSE functions, respectively. VNF1 and VNF2 are each connected to L2GW/L3GW1 and L2GW/L3GW2 through IPUs.

Figure 1 NFVI distributed gateway networking



Function Deployment

On the network shown in [Figure 1](#), the number of BDs needs to be planned based on the number of network segments corresponding to each IPU. An example assumes that the four IP addresses planned for four IPUs belong to four network segments. In this case, four BDs need to be planned. You need to configure the BDs and the corresponding VBDIF interfaces on all L2GW/L3GWs and bind all the

VBDIF interfaces to the same L3VPN instance. In addition, the following functions need to be deployed on the DCN:

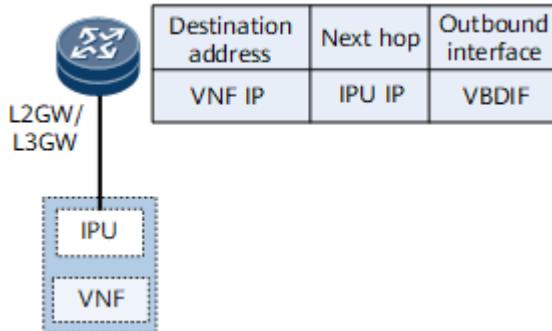
- Establish BGP VPN peer relationships between VNFs and DC-GWs so that the VNFs can advertise mobile phone routes (UE IP) to DC-GWs.
- On L2GW/L3GW1 and L2GW/L3GW2, configure static VPN routes with the IP addresses of VNFs as the destination addresses and the IP addresses of IPUs as next-hop addresses.
- Deploy EVPN RRs which can be either a standalone device or a DC-GW. In this section, BGP EVPN peer relationships are established between all L2GWs/L3GWs, PEs, and DC-GWs, DC-GWs are deployed as RRs to reflect EVPN routes, and L2GW/L3GWs function as RR clients. L2GW/L3GWs can use EVPN RRs to synchronize MAC or ARP routes as well as the IP prefix routes carrying a VNF address as the destination address with IPUs.
- Configure static default routes on PEs and use the EVPN RRs to reflect the static default routes to L2GW/L3GWs.
- Deploy SR tunnels between PEs and L2GW/L3GWs and between DC-GWs and L2GW/L3GWs to carry service traffic.
- The traffic transmitted between mobile phones and the Internet over VNFs is north-south traffic. The traffic transmitted between VNF1 and VNF2 is east-west traffic. To achieve load balancing of east-west traffic and north-south traffic, deploy the load balancing function on DC-GWs and L2GW/L3GWs.

Establishment of Forwarding Entries

In the networking of an NFVI distributed gateway (E2E EVPN over E2E SR tunnels), the procedure for establishing forwarding entries on each device is as follows:

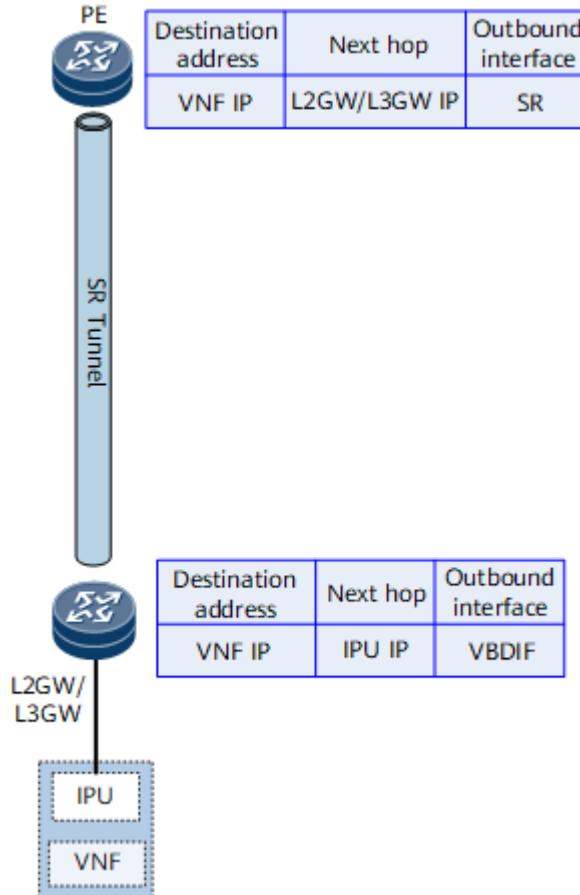
1. On L2GW/L3GWs, the number of BDs is planned based on the number of network segments corresponding to the IPUs, the BDs are bound to the links connecting to the corresponding IPUs, and VBDIF interfaces are configured as the gateways of IPUs. Static VPN routes are configured on L2GW/L3GWs so that the forwarding entries with the destination address as a VNF's address, the next-hop address as an IPU's IP address, and outbound interface as the VBDIF interface can be established on L2GW/L3GWs.

Figure 2 Forwarding entries of static routes on L2GW/L3GWs



2. After static VPN routes destined for VNFs are configured on L2GW/L3GWs, these static VPN routes are imported to the BGP EVPN routing table and IP prefix routes are generated. These routes are sent to DC-GWs and PEs based on BGP EVPN peer relationships. To prevent these routes from being transmitted to DC-GWs or PEs and recursed to VBDIF interfaces (DC-GWs and PEs do not have VBDIF interfaces) because they carry gateway addresses, configure an import route-policy on DC-GWs and PEs to delete gateway addresses from the received routes.

Figure 3 Forwarding entries on PEs and L2GW/L3GWs



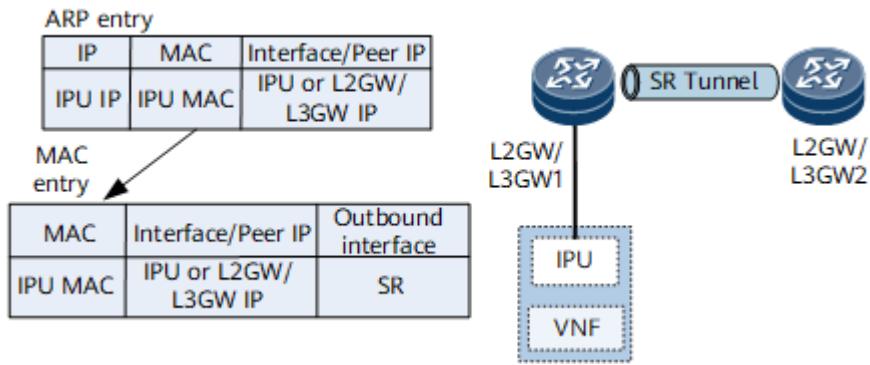
3. An L2GW/L3GW learns the MAC addresses and ARP information of IPUs through the data plane. Such information is advertised to another L2GW/L3GW through EVPN routes and can be used for establishment of ARP entries and MAC entries for Layer 2 forwarding. Taking L2GW/L3GW1 as an example, the destination MAC address of the MAC entries on L2GW/L3GW1 is an IPU's MAC address. For the IPU directly connected to L2GW/L3GW1, the IPU's interface is used as the outbound interface in MAC entries. For the IPU connected to another L2GW/L3GW, the MAC entries contain the outbound interface for SR tunnels and the IP address of the BGP EVPN peer of this L2GW/L3GW as the next-hop address.

NOTE

L2GW/L3GWs exchange routes with a VNF's IP address as the destination IP address. If two VNFs are connected to different L2GW/L3GWs, traffic is forwarded over routes. Otherwise, route loops may occur. Therefore, a route-policy needs to be configured on L2GW/L3GWs, so that the Gateway IP attribute is added to the routes exchanged between L2GW/L3GWs. The Gateway IP attribute is still the next-hop address, which is an IPU's IP address, and the outbound interface is the VBDIF interface. In this manner, L2GW/L3GWs forward traffic based on the MAC forwarding table to prevent route loops.

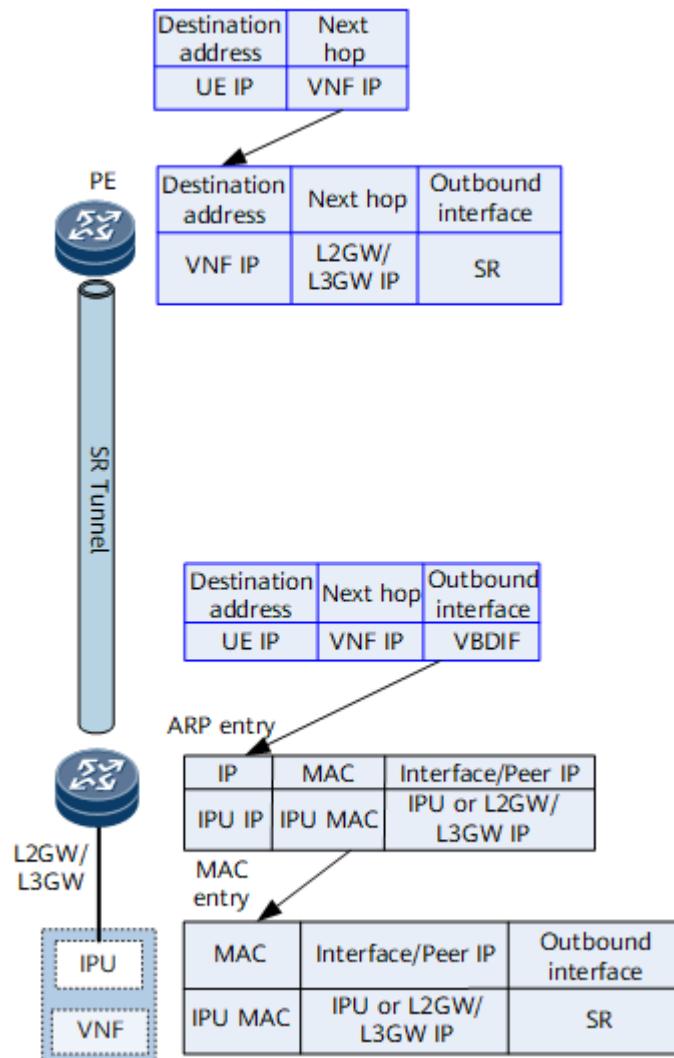
Because multiple links and static routes exist between L2GW/L3GWs and VNFs, to achieve load balancing, the Add-Path function needs to be enabled during configuration of importing static routes to the BGP EVPN routing table.

Figure 4 MAC entries on L2GW/L3GWs



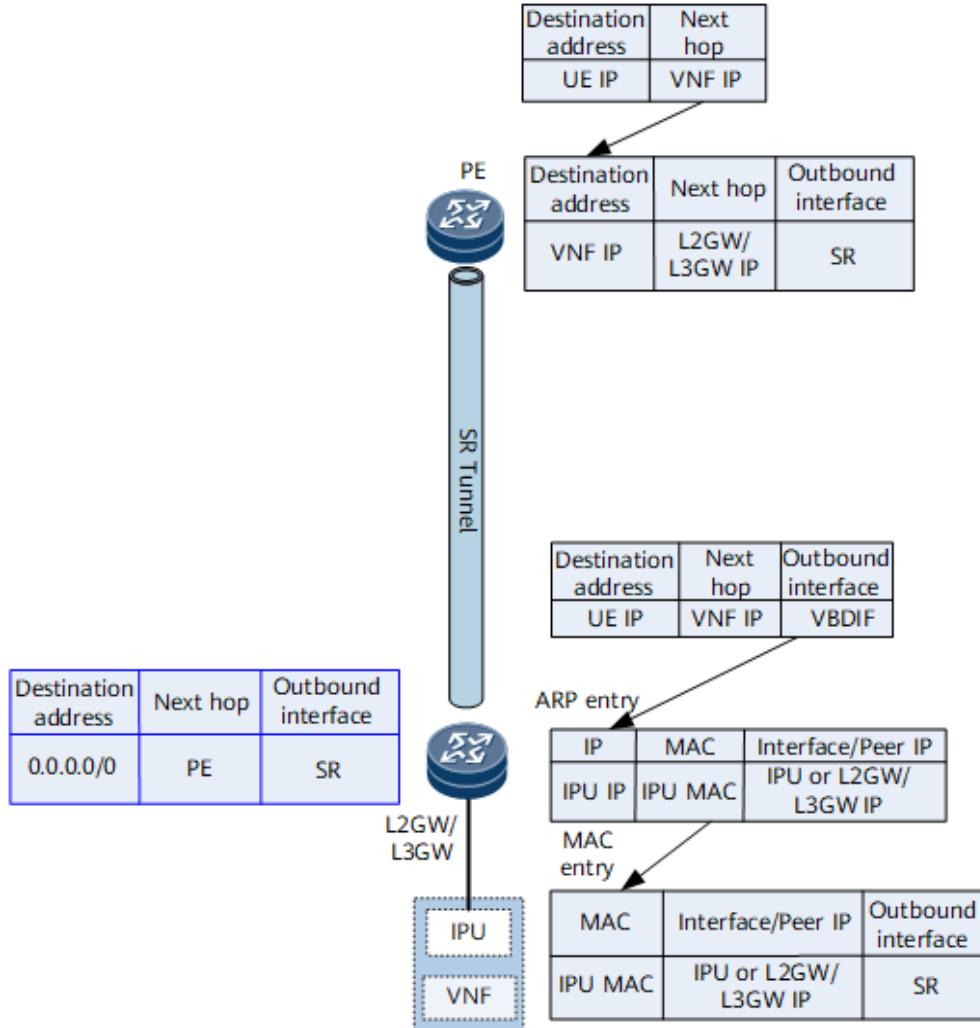
4. To establish BGP VPN peer relationships between DC-GWs and VNFs, DC-GWs need to advertise the routes destined for loopback addresses to L2GW/L3GWs. After BGP VPN peer relationships are established between VNFs and DC-GWs, VNFs can send mobile phone routes to DC-GWs, and DC-GWs send the mobile phone routes as IP prefix routes to L2GW/L3GWs based on the BGP VPN peer relationships. The Gateway IP attribute, which is the IP address of a VNF, is added to the IP prefix routes based on a route-policy. Upon receipt of IP prefix routes, L2GW/L3GWs generate VPN forwarding entries. L2GW/L3GWs then send EVPN routes to PEs based on EBGP EVPN peer relationships so that PEs can generate VPN forwarding entries with the destination address as the IP address of mobile phone routes and the next-hop address as a VNF's IP address. The routes are then recursed to VNFs, and the outbound interface is an SR tunnel.

Figure 5 Forwarding entries on PEs and L2GW/L3GWs



5. Devices on the DCN do not need to get aware of external routes. Therefore, route-policies need to be configured on PEs to allow PEs to send only default routes to L2GW/L3GWs.

Figure 6 Forwarding entries on PEs and L2GW/L3GWs



6. PEs can exchange information about Internet routes, such as the Internet server address, with external devices.
7. To achieve load balancing of east-west traffic and north-south traffic, the load balancing function and Add-Path function need to be deployed on PEs and L2GW/L3GWs.
- Load balancing of north-south traffic: Taking PE1 in [Figure 1](#) as an example, PE1 can receive the VPN routes destined for VNF2 from L2GW/L3GW1 and L2GW/L3GW2. By default, after the load balancing function is configured, PE1 sends half of the traffic destined for VNF2 through L2GW/L3GW1 and the other half of the traffic through L2GW/L3GW2. However, L2GW/L3GW1 connects to VNF2 over one link and L2GW/L3GW2 connects to VNF2 over two links, causing the load balancing function failed to achieve the desired effect. Therefore, the Add-Path function needs to be deployed on L2GW/L3GWs. After the Add-Path function is deployed on L2GW/L3GWs, L2GW/L3GW2 sends two routes with the same destination address to DC-GW1, achieving load balancing.
 - East-west traffic load balancing: Taking L2GW/L3GW1 in [Figure 1](#) as an example, because the Add-Path function is deployed on L2GW/L3GW2, L2GW/L3GW1 receives two EVPN routes from L2GW/L3GW2 and L2GW/L3GW1 has a static route with the next-hop address as IPU3's IP address. The destination addresses of

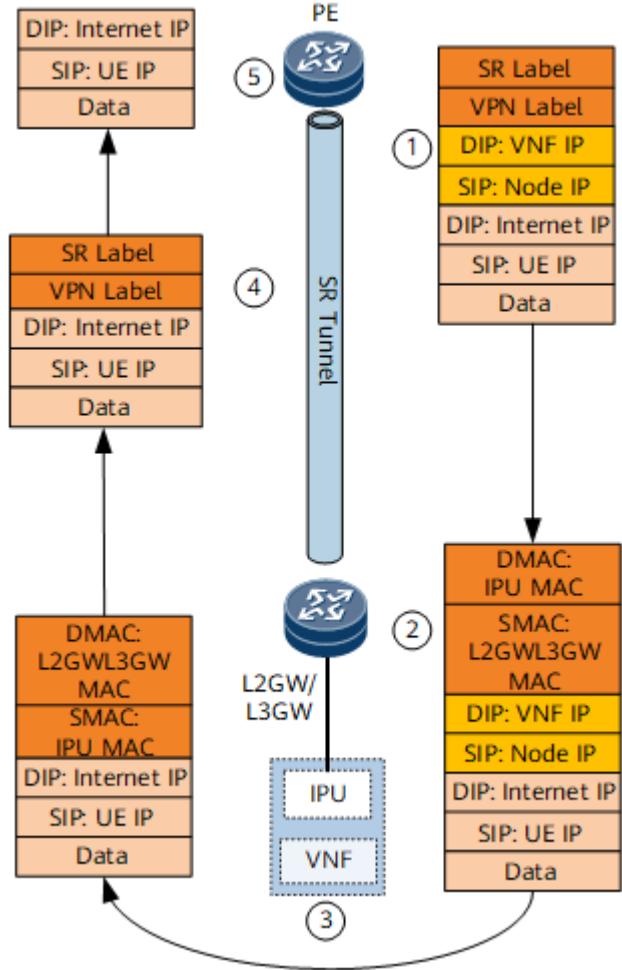
all these routes are VNF2's IP address. Therefore, the load balancing function needs to be configured to balance traffic over static routes and EVPN routes.

Traffic Forwarding Process

As shown in [Figure 7](#), the procedure for forwarding north-south traffic from mobile phones to the Internet is as follows:

1. Mobile phone traffic is sent to base stations (Nodes) and encapsulated with a GPRS tunneling protocol (GTP) header. The destination IP address of the GTP tunnel is a VNF's IP address. PEs send the encapsulated packets to L2GW/L3GWs over SR tunnels based on the VPN routing table.
2. Upon receipt of the encapsulated packets, L2GW/L3GWs look for the VPN routing table and find that the next-hop addresses of the forwarding entries corresponding to VNFs' IP addresses are IPUs' IP addresses and the outbound interface is the VBDIF interface. Therefore, routes destined for the network segment corresponding to the VBDIF interface are hit. L2GW/L3GWs search the MAC address that belongs to the network segment in an ARP table, look for the MAC forwarding table based on the ARP information, and forward traffic to VNFs based on the MAC forwarding table.
3. Upon receipt of packets, VNFs decapsulate the GTP tunnel header, search the routing table based on the destination IP address in the decapsulated packets, and forward the packets to L2GW/L3GWs based on the default gateways of VNFs.
4. L2GW/L3GWs search for the VPN routing table on L2GW/L3GWs. The default routes advertised by PEs to L2GW/L3GWs are recursed over SR tunnels and forwarded to PEs after being encapsulated with a VPN label.
5. PEs use the VPN forwarding table to forward the packets to the Internet based on the VPN label.

Figure 7 North-south traffic forwarding from mobile phones to the Internet

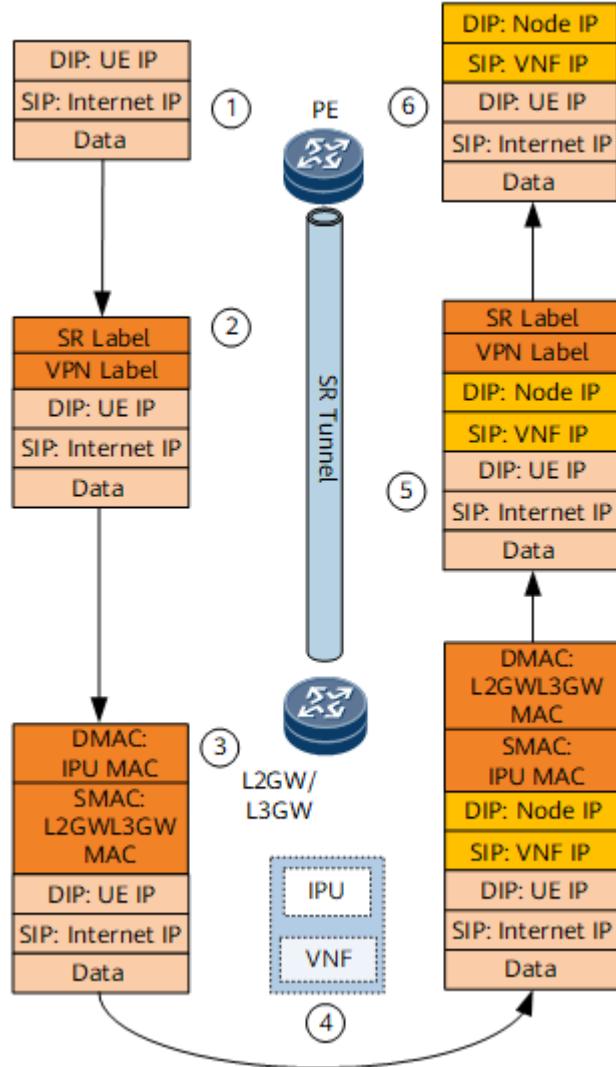


As shown in [Figure 8](#), the procedure for forwarding east-west traffic from the Internet to mobile phones over VNFs is as follows:

1. Devices on the Internet send mobile phones the reply packets whose destination IP addresses are the IP addresses of mobile phones. This is because mobile phone routes are advertised by L2GW/L3GWs to VNFs based on BGP VPNV4/v6 peer relationships and then advertised to the Internet by PEs. Therefore, reply packets must be first transmitted to L2GW/L3GWs.
2. Upon receipt of reply packets, PEs search the VPN routing table for the forwarding entries corresponding to mobile phone routes whose next-hop addresses are L2GW/L3GWs' IP addresses and the outbound interface is the one for SR tunnels. The reply packets are sent to L2GW/L3GWs after being encapsulated with a VPN label and an SR label.
3. Upon receipt of reply packets, L2GW/L3GWs find the VPN forwarding entries based on the VPN label and the VBDIF interface based on the VPN forwarding entries. The packets are then forwarded to VNFs based on the MAC entries corresponding to the VBDIF interface.
4. Upon receipt of the reply packets, VNFs search for the base stations corresponding to the destination IP addresses of mobile phones and add a tag of tunnel information with the destination IP address as the IP address of a base station. The reply packets are then forwarded to L2GW/L3GWs based on default gateways.
5. Upon receipt of the reply packets, L2GW/L3GWs search the VPN routing table, and the default routes advertised by DC-GWs to L2GW/L3GWs are hit. The reply packets are then forwarded to DC-GWs over SR tunnels after being encapsulated with a VPN label.

- Upon receipt of reply packets, PEs use the VPN forwarding table to forward the packets to base stations based on the VPN label. The base stations decapsulate the packets before forwarding them to mobile phones.

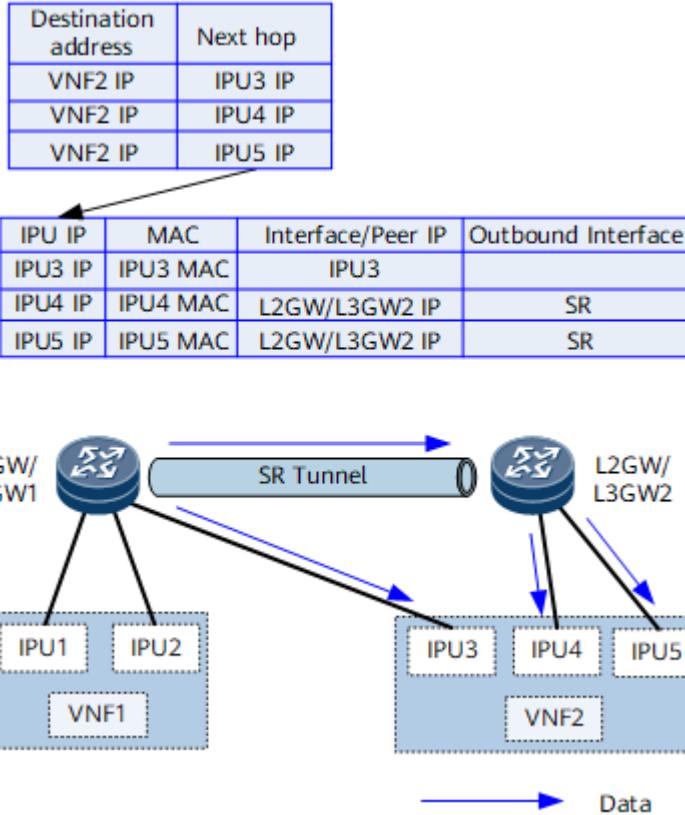
Figure 8 East-west traffic forwarding from the Internet to mobile phones



Upon receipt of user packets, a VNF finds that the packets need to be sent to another VNF for value-added service processing. In this case, east-west traffic occurs. On the network shown in [Figure 9](#), the difference between forwarding east-west traffic and forwarding north-south traffic lies in the processing of packets after they arrive VNF1.

- Upon receipt of user packets, VNF1 finds that the packets need to be processed by VNF2 and then adds a tunnel label with the destination IP address as VNF2's IP address. The user packets are sent to L2GW/L3GWs based on default routes.
- Upon receipt of user packets, an L2GW/L3GW searches the VPN forwarding table and finds that multiple load-balancing forwarding entries exist. In some entries, the outbound interface is an IPU or the next-hop address is the IP address of another L2GW/L3GW.
- If the traffic hits the path of another L2GW/L3GW, an EVPN label is added to the user packets and the routes are recursed to L2GW/L3GW2 over an SR tunnel. L2GW/L3GW2 searches for the BD and destination MAC address based on the EVPN label before forwarding the packets to VNF2.
- Upon receipt of the user packets, VNF2 processes and forwards the packets to the server. The subsequent forwarding follows the north-south traffic forwarding procedure.

Figure 9 East-west traffic forwarding from VNF1 to VNF2



Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.12.11.10 Application Scenarios for EVPN E-LAN Accessing L3VPN

Service Overview

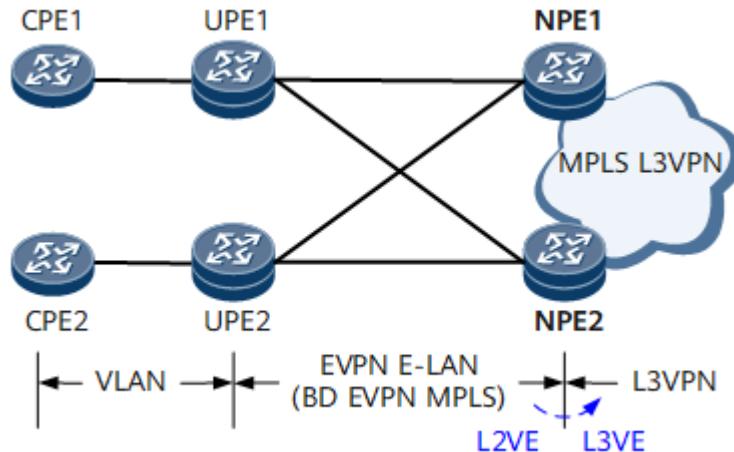
If point-to-multipoint Layer 2 services are carried using EVPN E-LAN and Layer 3 services are carried using MPLS L3VPN, EVPN E-LAN accessing L3VPN must be configured on the devices where Layer 2 and Layer 3 services overlap.

Networking Description

On the network shown in [Figure 1](#), the first layer is the user-side access network where Layer 2 services are carried between CPEs and UPEs to implement VLAN forwarding. The second layer is the aggregation network where services are carried using EVPN E-LAN and BD-based EVPN MPLS is deployed between UPEs and NPEs. The third layer is the core network where services are carried using MPLS L3VPN. In this case, NPEs must be deployed with EVPN E-LAN accessing L3VPN. Specifically, configure an L2VE sub-interface to access BD EVPN on an NPE and then configure an L3VE sub-interface to access L3VPN. In this manner, when the public network packets of EVPN E-LAN reach the L2VE sub-interface on the NPE, the corresponding tag is encapsulated into the packets based on the tag processing behavior configured on the L2VE sub-interface. After the packets

are loopbacked to the L3VE interface, the corresponding L3VE sub-interface is matched based on the tag carried in the packets, and the packets are forwarded using the corresponding L3VPN instance.

Figure 1 Networking of EVPN E-LAN accessing L3VPN



Feature Deployment

Before deploying EVPN E-LAN accessing L3VPN, complete the following tasks on an NPE:

- Create an L2VE interface and an L3VE interface and bind them to the same VE-Group.
- Create an L2VE sub-interface and add it to the BD to which the EVPN instance bound.
- Create an L3VE sub-interface, specify the associated VLAN, set the VLAN encapsulation mode (configure the Tag), and bind it to the corresponding L3VPN instance.

Parent Topic: [Application Scenarios for EVPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.13 PBB VPLS Description

[Overview of PBB VPLS](#)

[Understanding PBB VPLS](#)

[Application Scenarios for PBB VPLS](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#) [Next topic >](#)

1.13.1 Overview of PBB VPLS

Definition

Provider backbone bridge (PBB), a technique defined in IEEE 802.1ah, precedes customer MAC (C-MAC) addresses with backbone MAC (B-MAC) addresses in user packets to completely separate the

user network from the carrier network. Unlike traditional Ethernet local area networks (LANs), a PBB network uses both public and user MAC addresses. This implementation ensures network stability and reduces the number of MAC entries required on public network devices. In addition, the VLAN tag field defined in IEEE 802.1Q has only 12 bits and can identify only a maximum of 4096 VLANs, which cannot meet the user requirements for sufficient tunnels. PBB, which uses instance-virtual service instances (I-VSIs) identified by 24-bit IDs, allows you to establish sufficient tunnels to transmit traffic over an Ethernet transport network.

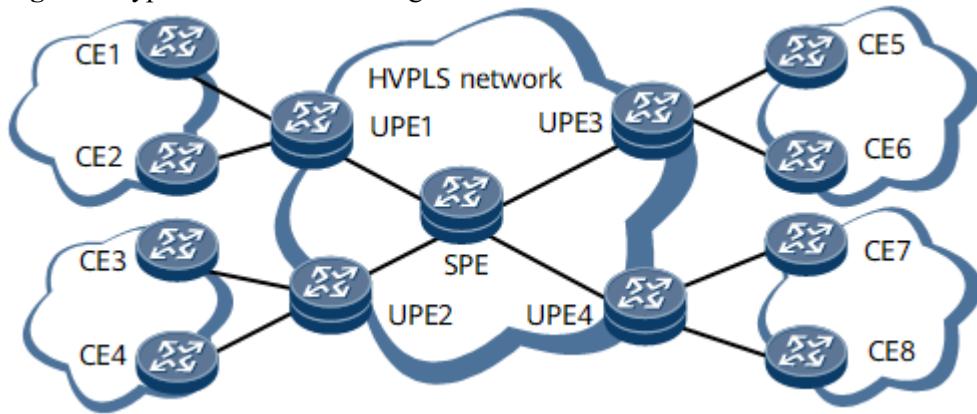
Virtual private LAN service (VPLS) is an L2VPN technology implemented based on Multiprotocol Label Switching (MPLS) and Ethernet technologies.

PBB VPLS uses MAC-in-MAC instead of QinQ to transmit packets over a VPLS network, reducing the number of MAC entries that provider edges (PEs) must learn.

Purpose

Carriers generally use hierarchical virtual private LAN service (HVPLS) to carry metropolitan area network (MAN) services. As shown in [Figure 1](#), by dividing the VPLS network into multiple layers, HVPLS decreases the number of required pseudo wires (PWs), reduces signaling exchange and packet replication frequency, and solves the expansibility problem confronted by full-mesh VPLS networks. HVPLS enables VPLS networks to be applied on a large scale. On the HVPLS network shown in [Figure 1](#), user-end provider edges (UPEs) only need to learn the MAC addresses of local and remote users; superstratum provider edges (SPEs) serving as service aggregation points, however, must learn the MAC addresses of all users on the metro Ethernet. As the metro Ethernet expands in scale, SPEs have to learn increasingly more MAC addresses. Expanding the capacity of MAC forwarding tables on SPEs then becomes a challenging task.

Figure 1 Typical VPLS networking



To address the problem faced by HVPLS networks, PBB VPLS is introduced. When being used with HVPLS, PBB precedes C-SMAC addresses with B-MAC addresses in user packets and enables SPEs on the HVPLS network to transmit these packets based only on B-MAC addresses. By reducing the MAC address learning pressure of SPEs, PBB further improves the expansibility of VPLS networks.

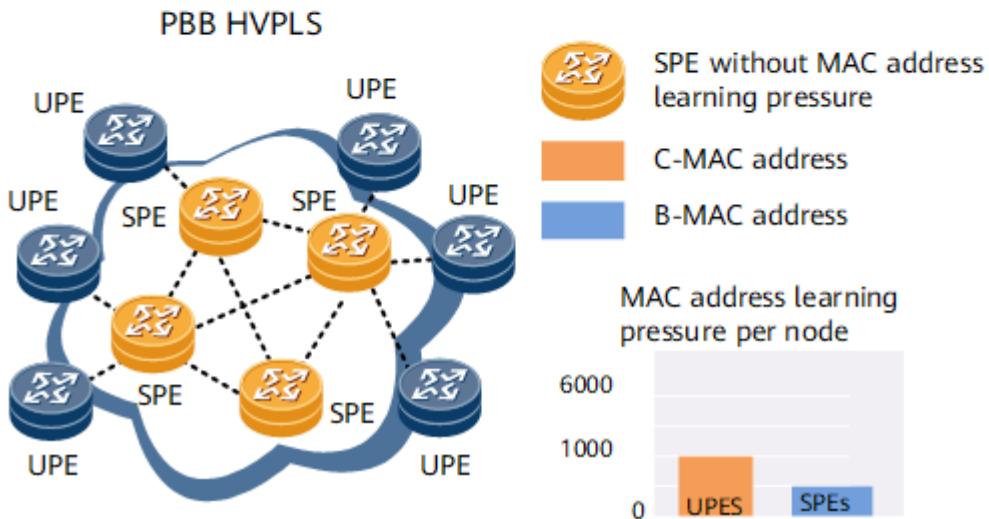
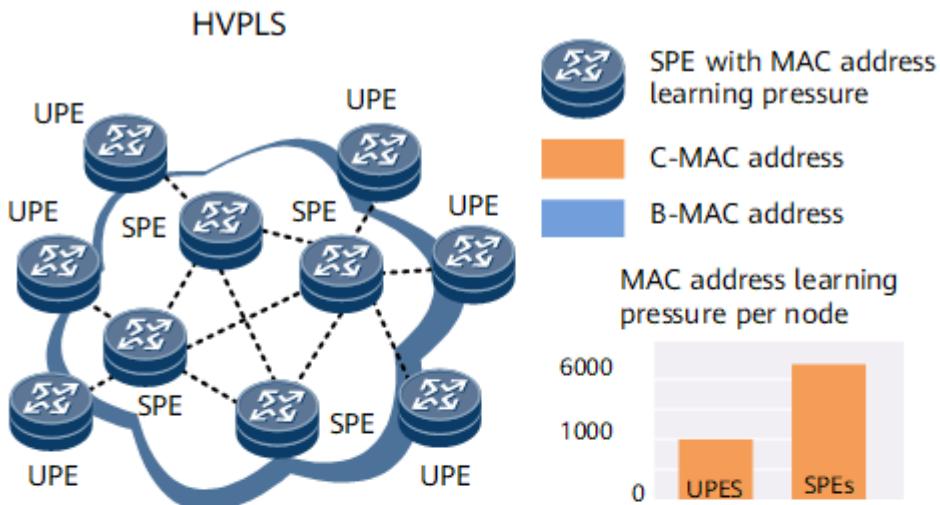
Benefits

Compared to HVPLS, PBB VPLS offers the following benefits:

- Flexible network expansion and reduced network costs

As shown in [Figure 2](#), SPEs on an HVPLS network have to learn all user MAC addresses. Because the MAC address learning capabilities of SPEs are limited, you may have to add additional SPEs when expanding the HVPLS network by adding more UPEs.

Figure 2 Comparison of MAC address learning pressure faced by SPEs on HVPLS and PBB VPLS networks



PBB VPLS enables SPEs to learn only the MAC addresses of UPEs, reducing the MAC address learning pressure of SPEs. The MAC address learning pressure of SPEs does not need to be considered during the expansion of a PBB VPLS network implemented by adding more UPEs. A UPE on an HVPLS network provides access services for a maximum of 4K users, whereas a UPE on a PBB VPLS network provides access services for a maximum of 16K users. In this sense, PBB VPLS helps carriers reduce investments on UPEs.

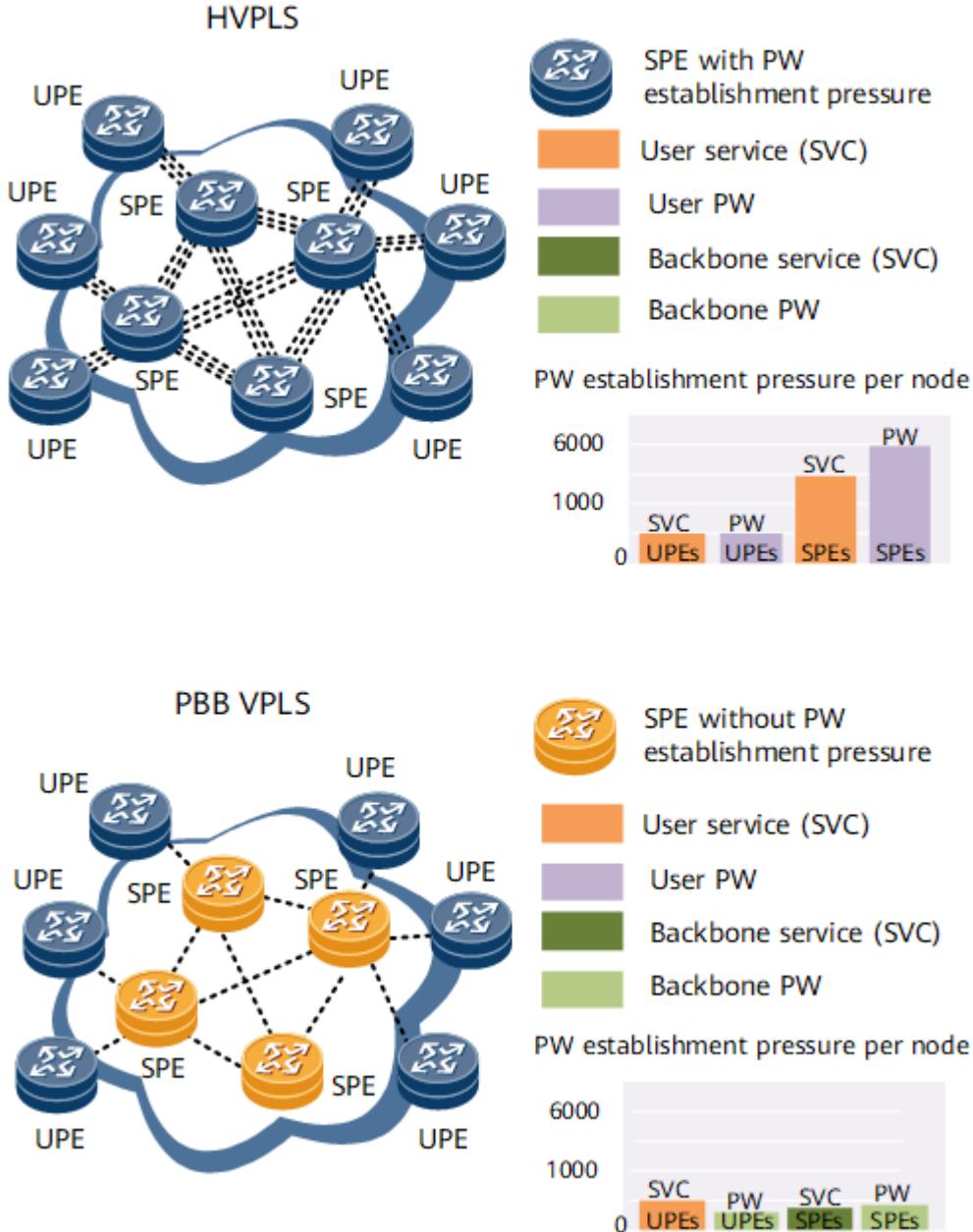
- Improved network security

After a user packet arrives at the public network, B-MAC addresses are inserted before the C-MAC addresses in the packet. This implementation shields detailed packet information and separates the user network from the carrier network, improving network security.

- Reduced maintenance costs

On the HVPLS network shown in [Figure 3](#), the SPE has to establish a PW for each pair of UPEs that need to communicate, incurring high maintenance costs.

Figure 3 Comparison of PW quantities required by HVPLS and PBB HVPLS networks



On a PBB VPLS network, packets sent by different UPEs can share the same PW. The configuration workload required to expand a PBB VPLS network is much less than that required to expand an HVPLS network.

Parent Topic: [PBB VPLS Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.13.2 Understanding PBB VPLS

[PBB VPLS Fundamentals](#)

Parent Topic: [PBB VPLS Description](#)

Copyright © Huawei Technologies Co., Ltd.

1.13.2.1 PBB VPLS Fundamentals

Related Concepts

PBB VPLS divides an HVPLS network into an access VPLS domain and a backbone VPLS domain by deploying I-VSIs and backbone-virtual service instances (B-VSIs) on access-, aggregation-, and core-layer devices. In the access VPLS domain, I-VSIs provide VPLS services; in the backbone VPLS domain, B-VSIs provide VPLS services.

The following table describes basic concepts about PBB VPLS.

Table 1 Basic concepts about PBB VPLS

Concept	Description
PBB	PBB, a technique defined in IEEE 802.1ah, precedes C-MAC addresses with B-MAC addresses in user packets to separate the user network from the carrier network. This implementation enhances network stability and eases the capacity pressure faced by SPEs' MAC forwarding tables.
I-VSI	An I-VSI processes user-side MAC address information, encapsulates packets into PBB frames, and is bound to an attachment circuit (AC) interface for local customer edge (CE) identification.
B-VSI	A B-VSI processes B-MAC addresses and uses the configured peer address to determine a PW's remote PE.
I-VSI-B-VSI binding	PBB VPLS uses both I-VSIs and B-VSIs. Multiple I-VSIs can be bound to the same B-VSI. SPEs are aware of only B-VSIs.
UPE	A UPE is an aggregation device that directly connects to CEs. A UPE has to connect to only one SPE on an HVPLS network. A UPE supports routing and MPLS encapsulation. If a UPE connects to multiple CEs and possesses the basic bridge function, the UPE can perform frame forwarding to reduce the burden on the SPE.
SPE	An SPE connects to all UPEs and all the other SPEs on an HVPLS network. From the perspective of an SPE, its connected UPE acts as a CE, and the PW set up between the UPE and itself serves as an AC. An SPE has to learn the MAC addresses of connected UPE interfaces as well as the MAC addresses of sites connecting to these UPEs.
PBB PE (PPE)	On a PBB VPLS network, a PPE can connect to UPEs, CEs, and SPEs.

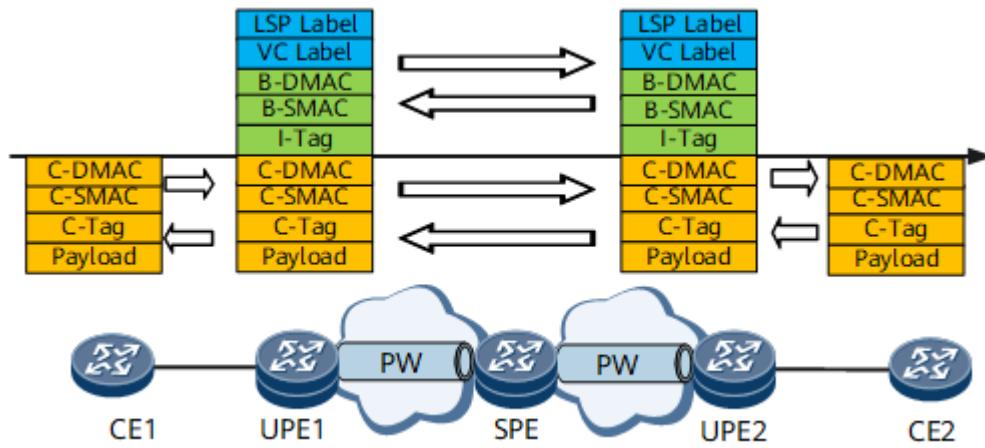
Implementation

[Figure 1](#) describes each field in a PBB packet, and [Table 2](#) shows the transmission of a user packet on a PBB VPLS network.

Table 2 Description of each field in a PBB packet

Packet	Description
Payload	User packet. On a PBB VPLS network, the payload is transparent to UPEs and SPEs.
C-Tag	Used to identify a user VLAN. On a PBB VPLS network, UPEs and SPEs do not need to learn C-tags.
C-SMAC	Source C-MAC address. On a PBB VPLS network, only UPEs need to learn C-SMAC addresses.
C-DMAC	Destination C-MAC address.
I-Tag	An instance tag (I-Tag) is a field defined in 802.1ah for the MAC header. It is used to transmit frame-related service instance information. The I-tag is inserted before user MAC addresses. It is a 4-byte field and includes a 2-byte Ethtype and a 2-byte I-SID. Ethtype 0x88e7 identifies a PBB packet. The I-SID defines a frame-related service instance and uniquely identifies an I-VSI. Upon receipt of a user packet, the UPE adds the I-tag to the packet to identify the corresponding I-VSI. On a PBB VPLS network, SPEs do not need to learn I-tags.
B-SMAC	Source B-MAC address. On a PBB VPLS network, SPEs have to learn B-SMAC addresses.
B-DMAC	Destination B-MAC address. On a PBB VPLS network, SPEs have to learn B-DMAC addresses.
VC Label	Label used to identify a VC.
LSP Label	Label used to identify an LSP.

Figure 1 Transmission of a user packet on a PBB VPLS network



1. After a user packet arrives at the AC interface of UPE1, UPE1 obtains I-tag, B-SMAC address, and B-DMAC address information from the I-VSI, and adds a PBB header containing the B-DMAC address, B-SMAC address, and I-tag to the packet in addition to adding the VC and LSP labels to the packet. Then, UPE1 searches the I-VSI forwarding table for a matching entry to forward the packet. If a matching entry can be found, UPE1 forwards the packet to the SPE. If a matching entry cannot be found, UPE1 broadcasts the packet in the corresponding B-VSI, which may have two PWs working in primary/secondary mode or multiple PWs.
If the packet is destined for another CE connecting to UPE1, UPE1 directly forwards the packet to target AC interface based on I-VSI information.
2. Upon receipt of the packet, the SPE determines the VSI to which the packet belongs and searches the VSI MAC address table for a matching B-DMAC address. After finding a matching B-DMAC address, the SPE unicasts the packet over a PW to UPE2. If the SPE cannot find a matching B-DMAC address, the SPE broadcasts the packet to UPE2. Because the SPE learns only the B-SMAC address of UPE1, the MAC address learning pressure of the SPE is not heavy.
3. Upon receipt of the packet, UPE2 determines the B-VSI to which the packet belongs based on the VC label and determines the corresponding I-VSI based on obtained I-tag information. UPE2 then compares the B-DMAC address carried in the packet with the B-SMAC address of itself. If the two addresses are the same, UPE2 removes the outer label and PBB header from the packet, searches the I-VSI's MAC address table for an AC interface mapped to the C-DMAC address that is carried in the packet, and forwards the packet to CE2 through the AC interface. Otherwise, UPE2 drops the packet. UPE2 learns the B-SMAC address carried in the packet to forward returned traffic.

NOTE

If the B-DMAC address carried in the packet is a multicast MAC address, UPE2 directly forwards the packet to CE2 without comparing this address with the B-SMAC address of itself.

A B-SMAC address can be configured in either an I-VSI or a B-VSI. If an I-VSI and the B-VSI to which the I-VSI is bound have different B-SMAC addresses, the B-SMAC address configured in the I-VSI takes effect.

Parent Topic: [Understanding PBB VPLS](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

1.13.3 Application Scenarios for PBB VPLS

[PBB VPLS Application](#)

Parent Topic: [PBB VPLS Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.13.3.1 PBB VPLS Application

The deployment roadmap is as follows:

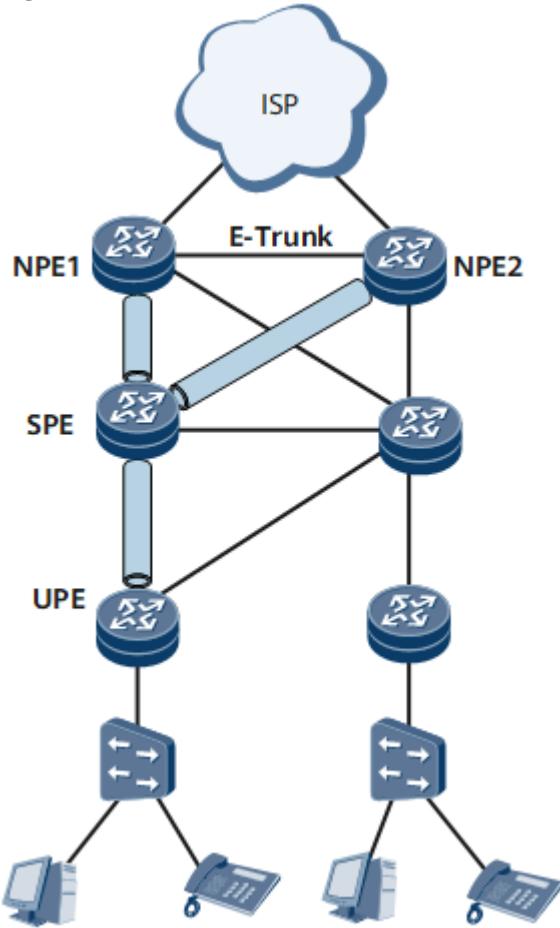
- Before deploying PBB VPLS, complete the following tasks:
 1. Configure an IGP on the backbone network to ensure IP connectivity.
 2. Configure MPLS and MPLS LDP and establish LDP sessions.
 3. Enable MPLS L2VPN to ensure that the MPLS network can transparently transmit Layer 2 user data.
- Deploy PBB VPLS to solve the scalability problem of the VPLS network:
 1. Configure B-VSIs, specify peers, and configure B-SMAC addresses.
 2. Configure I-VSIs and I-Tags, bind AC interfaces to I-VSIs, and configure B-DMAC addresses.
- To improve the reliability of PBB VPLS networks, you can configure different reliability features for different network layers to implement protection switching for links:
 1. Configure E-Trunk to implement link redundancy. If the primary link fails, traffic switches to the secondary link for transmission. To implement rapid traffic switching, the new master device will send LDP MAC Withdraw packets carrying PBB TLV, B-SMAC, and I-Tag information during a master/backup switchover.
 2. Configure VRRP, BFD, and PW redundancy to implement rapid traffic switching when a link fails.

PBB VPLS in E-Trunk Scenarios

- Residential service

On the network shown in [Figure 1](#), an Internet service provider (ISP) network is dual-homed to two network provider edges (NPEs), and an E-Trunk is configured between the two NPEs to determine the master/backup status of NPEs. After the E-Trunk detects a fault on the primary link, the E-Trunk sends an LACP packet to the ISP network, instructing the ISP network to switch traffic to the secondary link. Meanwhile, the new master NPE sends MAC Withdraw messages to SPEs, instructing SPEs to delete B-MAC addresses.

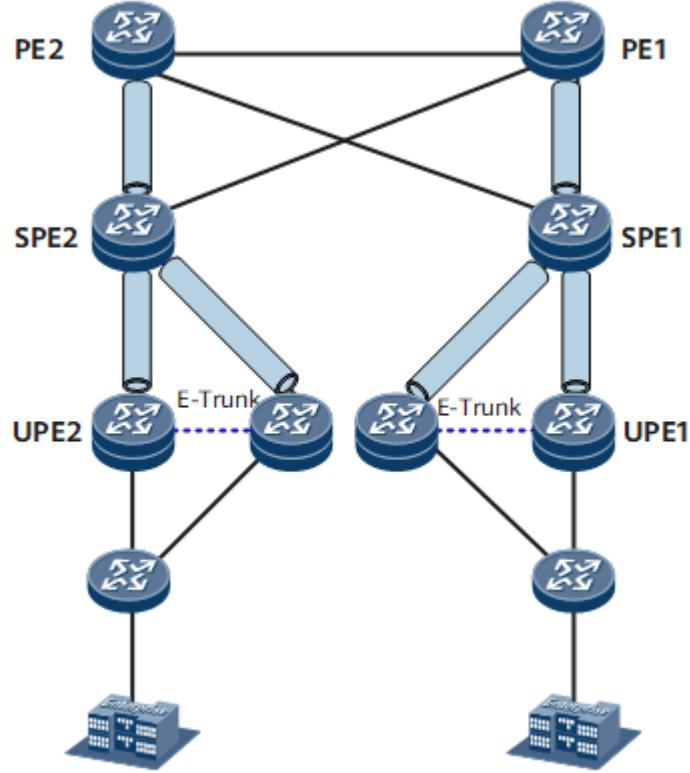
Figure 1 PBB VPLS for a residential service



- Enterprise service

On the network shown in [Figure 2](#), enterprise devices attached to switches are dual-homed to UPEs. An E-Trunk is configured between each pair of UPEs to determine the master/backup status of UPEs. After an E-Trunk detects a fault on the primary link, the E-Trunk sends an LACP packet to the enterprise devices, instructing these devices to switch traffic to the secondary link. Meanwhile, the master UPE sends MAC Withdraw messages to SPEs, instructing SPEs to delete B-MAC addresses.

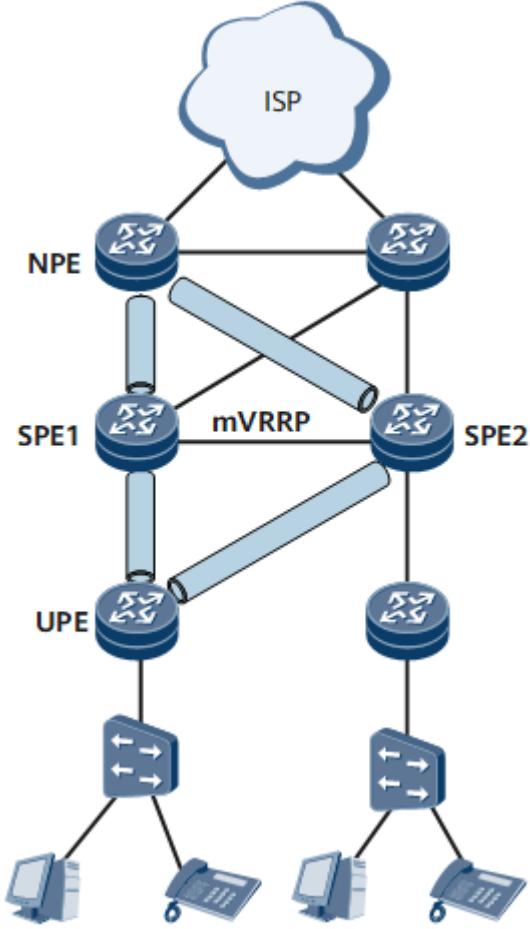
Figure 2 PBB VPLS for an enterprise service



PBB VPLS in VRRP Scenarios

On the network shown in [Figure 3](#), a UPE is dual-homed to SPEs over primary and secondary VPLS PWs. A Management Virtual Router Redundancy Protocol (mVRRP) backup group is configured on the SPEs and a management VSI (mVSI) is configured on the UPE. The mVSI is associated with multiple service VSIs. Link and peer BFD sessions are configured to track the mVRRP group. If a master/backup mVRRP switchover occurs, the new master SPE sends gratuitous Address Resolution Protocol (ARP) packets to the UPE. After receiving the packets, the UPE converts unicast traffic in all service VSIs associated with the mVSI to multicast traffic. The return traffic carrying a label mapped to the previous secondary PW is unicast traffic and traverses the new primary PW.

Figure 3 PBB VPLS for an enterprise service



Parent Topic: [Application Scenarios for PBB VPLS](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.

1.14 Proactive Loop Detection Description

This chapter describes the basic concepts, principles, and applications of proactive loop detection.

[Overview of Proactive Loop Detection](#)

[Understanding Proactive Loop Detection](#)

[Application Scenarios for Proactive Loop Detection](#)

Parent Topic: [VPN](#)

Copyright © Huawei Technologies Co., Ltd.
Copyright © Huawei Technologies Co., Ltd.
[< Previous topic](#)

1.14.1 Overview of Proactive Loop Detection

Definition

Proactive loop detection detects and eliminates Layer 2 network loops. When a device's Ethernet or Eth-Trunk interface physically goes Up or an interface is bound to a VSI, the device proactively detects and eliminates loops, if any.

Purpose

If a device's Ethernet or Eth-Trunk interface goes Up by misoperation or an interface is bound to a VSI and the interface incurs a loop, the device may have services interrupted or even get out of the NMS control. To resolve the loop problem and ensure normal device running, Huawei developed proactive loop detection upon interface Up. This feature allows a device's interface to proactively send loop detection packets. If the interface detects a loop, the device blocks the interface.

Parent Topic: [Proactive Loop Detection Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.14.2 Understanding Proactive Loop Detection

[Proactive Loop Detection](#)

[Loop Detection Packet Format](#)

Parent Topic: [Proactive Loop Detection Description](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

< Previous topic [Next topic >](#)

1.14.2.1 Proactive Loop Detection

Triggering Condition

- Interface going Up

If an Ethernet interface, Ethernet trunk interface, or a specified Ethernet trunk member interface physically goes Up, the proactive loop detection function is triggered to detect whether the Ethernet interface, all members of the Ethernet trunk interface, or the specified Ethernet trunk member has a loop. If they have a loop, this function sets them to Down. Note that if an Ethernet interface goes Down, its associated sub-interfaces also go Down.

- Interface bound to a VSI

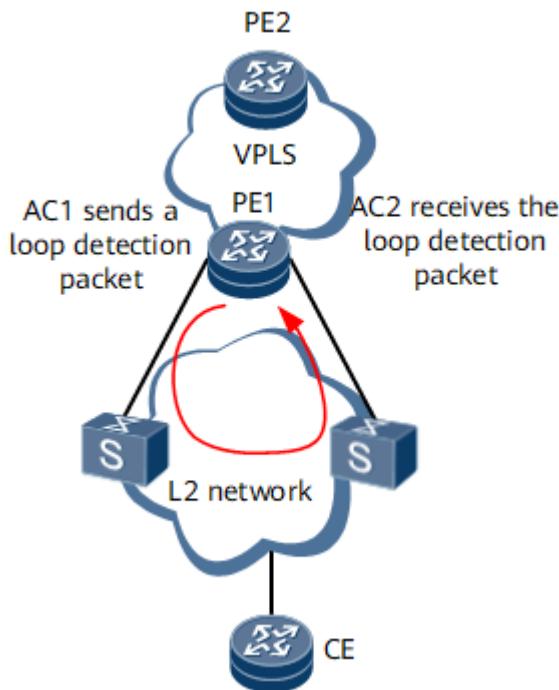
If an Ethernet interface, Ethernet sub-interface, Ethernet trunk interface, or Ethernet trunk sub-interface is bound to a VSI, proactive loop detection is triggered on the Ethernet interface, Ethernet sub-interface, or trunk member interfaces. If they have a loop, this function sets them to Down. An Ethernet trunk sub-interface can be a dot1q sub-interface, dot1q VLAN tag termination sub-interface, or QinQ VLAN tag termination sub-interface.

Detection Principles

When a device's Ethernet or Eth-Trunk interface goes Up or an interface is bound to a VSI, the interface proactively sends a loop detection packet. If the device receives the loop detection packet

sent through a VPLS domain within the configured period, a loop occurs on the network. In this case, the device blocks the interface sending the loop detection packet and reports an alarm.

Figure 1 Proactive loop detection upon interface Up



On the network shown in [Figure 1](#), AC1 sends a loop detection packet. If AC2 receives this packet within a loop detection period, a loop occurs on the network.

NOTE

Proactive loop detection applies only to VPLS scenarios, not VLAN scenarios.

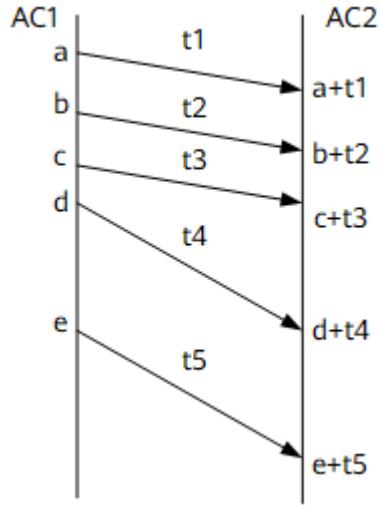
NOTICE

It is recommended that you disable this function on properly running devices. If you have to use this function to detect whether links operate properly during site deployment, be sure to disable this function after this stage.

Loop Detection Period

An interface can send a loop detection packet for a maximum of five times: the sending interval is 3s for the first three times and 10s for the latter two times. The maximum loop detection period is therefore 29s.

Figure 2 Loop detection packet sending



In [Figure 2](#), $t_1 = t_2 = t_3 = 3\text{s}$, and $t_4 = t_5 = 10\text{s}$. AC2 processes only the most recently sent loop detection packet. For example, if AC1 sends a loop detection packet at the **a** second, AC2 determines whether the packet was sent at the **a** second by AC1 upon receiving the packet.

- If so, a loop occurs. The device then sets the link layer protocol of AC1 to Down and reports an alarm to the NMS.
- If not, the device simply waits for the packet to be sent.

Parent Topic: [Understanding Proactive Loop Detection](#)

Copyright © Huawei Technologies Co., Ltd.

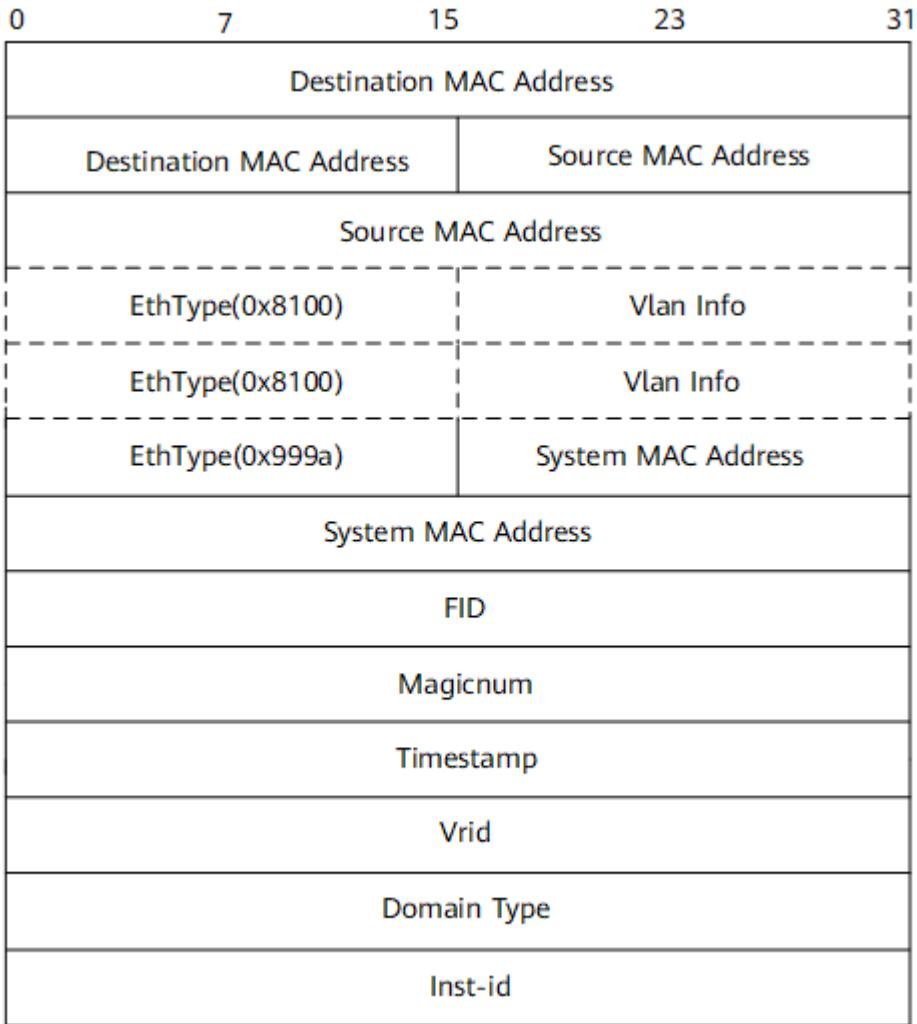
Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.14.2.2 Loop Detection Packet Format

[Figure 1](#) shows the loop detection packet format.

Figure 1 Loop detection packet format



The meanings of the fields are as follows:

- Destination MAC Address: destination MAC address
- Source MAC Address: source MAC address
- EthType: Ethernet type
- VLAN Info: VLAN information
- System MAC Address: system MAC address
- FID: FID of a VLAN or VSI
- Magicnum: index of an AC interface
- Timestamp: timestamp for sending loop detection packets
- Vrid: ID of a virtual router
- Domain Type: domain type
- Inst-id: ID of an instance

NOTE

A loop detection packet carries at most two VLAN tags (for example, in a QinQ VLAN tag termination scenario).

Parent Topic: [Understanding Proactive Loop Detection](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.14.3 Application Scenarios for Proactive Loop Detection

[AC Interface Receiving a Loop Detection Packet](#)

[PW Side Receiving a Loop Detection Packet](#)

Parent Topic: [Proactive Loop Detection Description](#)

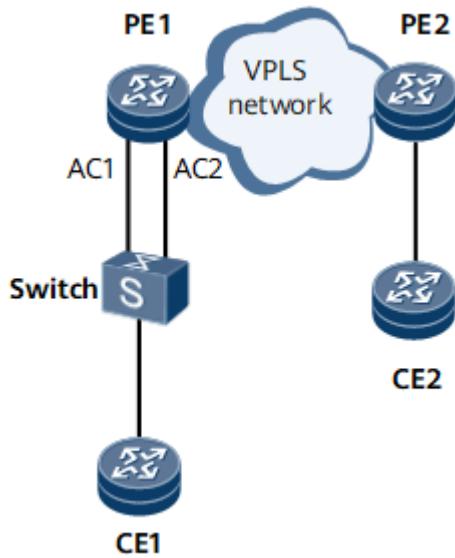
Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)

1.14.3.1 AC Interface Receiving a Loop Detection Packet

Figure 1 AC interface receiving a loop detection packet



In [Figure 1](#), PE1's AC1 is an Ethernet interface. After AC1 goes physically Up, it proactively sends a loop detection packet. If AC2 receives this packet, a loop occurs on the network. PE1 then sets the link layer protocol of AC1 to Down and reports an alarm to the NMS. This mechanism prevents AC1 from sending or receiving any packets.

Parent Topic: [Application Scenarios for Proactive Loop Detection](#)

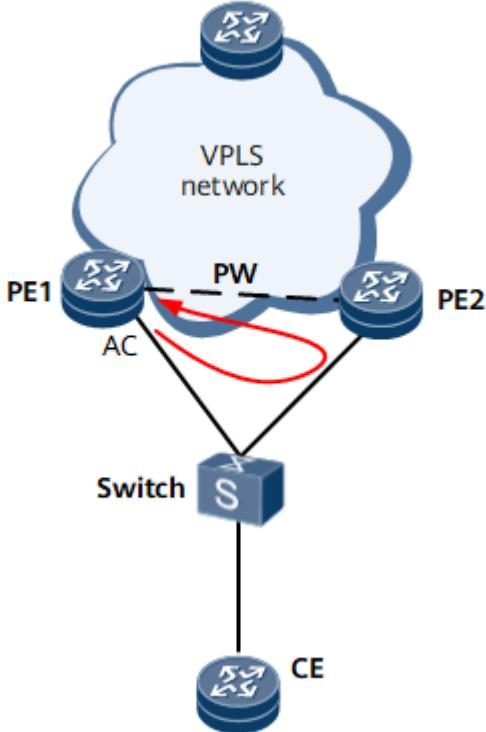
Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[Next topic >](#)

1.14.3.2 PW Side Receiving a Loop Detection Packet

Figure 1 PW side receiving a loop detection packet



In Figure 1, PE1's AC interface is an Ethernet interface. After the AC interface goes physically Up, it proactively sends a loop detection packet. If this packet loops back to PE1 through Switch, PE2, and the PW between PE1 and PE2, a loop occurs on the network. PE1 then sets the link layer protocol of the AC interface to Down and reports an alarm to the NMS. This mechanism prevents the AC interface from sending or receiving any packets.

Parent Topic: [Application Scenarios for Proactive Loop Detection](#)

Copyright © Huawei Technologies Co., Ltd.

Copyright © Huawei Technologies Co., Ltd.

[< Previous topic](#)