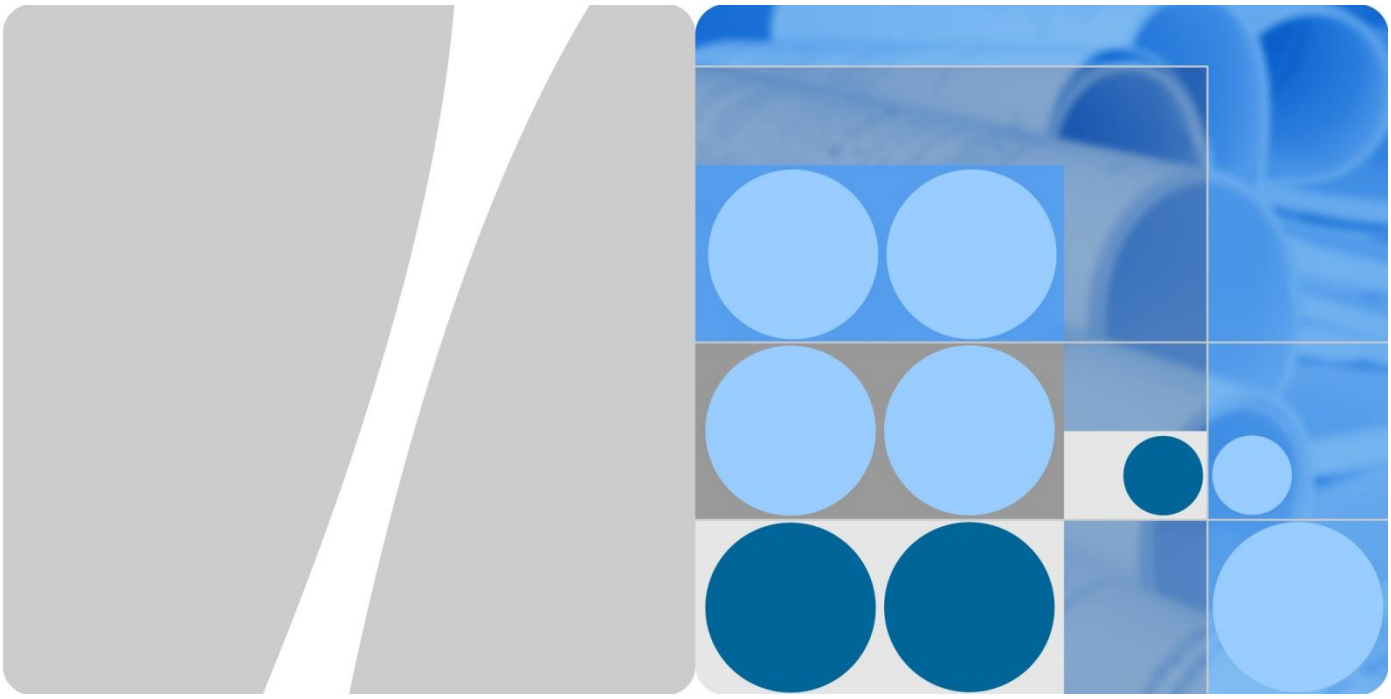


VPLS Technology White Paper

Issue 01

Date 2012-10-30

HUAWEI TECHNOLOGIES CO., LTD.

HUAWEI

Trademarks and Permissions

and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: http://enterprise.huawei.com

About This Document

Keywords

AC, MPLS, NPE, PW, UPE, VPLS, VSI

Abstract

Virtual Private LAN Service (VPLS) is a point-to-multipoint L2VPN service provided on public networks, and combines the advantages of Ethernet and MPLS. VPLS connects geographically dispersed user sites through a MAN or WAN by forming a single bridging domain, or VPN. This document describes how to implement VPLS in general and how Huawei specifically implements VPLS.

List of Acronyms

Acronym	Full Name
AC	Attachment Circuit
ATM	Asynchronous Transfer Mode
BFD	Bidirectional Forwarding Detection
CE	Custom Edge
FR	Frame Relay
LAN	Local Area Network
MAN	Metropolitan Area Network
MPLS	Multi-protocol Label Switching
MTU	Multi-Tenant Unit
NPE	Network Provider Edge
PE	Provider Edge
PW	Pseudo Wire
STP	Spanning Tree Protocol

Acronym	Full Name
UPE	User-facing Provider Edge
VC	Virtual Circuit
VE	VPLS Edge
VLL	Virtual Leased Line
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
VPWS	Virtual Private Wire Service
VSI	Virtual Switch Instance
WAN	Wide Area Network

Contents

1 Overview

1.1 Background

As the world economy develops, increasing enterprises have to span greater distances to provide quality services to an extensive clientele base. The employees of these enterprises also have to travel more frequently. As a result, enterprises seek out services that enable them to interconnect their branches, so that their employees can easily access enterprise networks from anywhere.

Originally, service providers filled this need by providing leased lines, but leased lines have significant disadvantages. For example, leased lines are not applicable when there are a large number of branches or when the number of branches grows quickly. Furthermore, this method is relatively expensive and a network based on leased lines is difficult to manage.

Then, Asynchronous Transfer Mode (ATM) and Frame Relay (FR) emerged, and service providers turned to them to provide virtual circuits. With these new methods, enterprises could establish their own Layer 3 networks for IP or IPX traffic based on the virtual circuits. However, the virtual links are point-to-point Layer 2 links, which make networks difficult to configure and maintain, especially when a new site is deployed.

Later, after IP networks had become present almost everywhere in the world, service providers began to focus on how to provide enterprises with low-cost private network services using the existing IP networks. The Multi-protocol Label Switching (MPLS) VPN technology was developed to address this demand. This technology is easy to configure and allows service providers to change bandwidth settings easily.

MPLS VPNs fall into two categories: MPLS L3VPN and MPLS L2VPN. MPLS L3VPN services require that service providers manage internal routes on user networks. The traditional MPLS L2VPN technology (VLL) provides point-to-point L2VPN services on public networks. The VLL virtual links allow sites to communicate as if they were directly connected by a link, but VLL supports only point-to-point exchange.

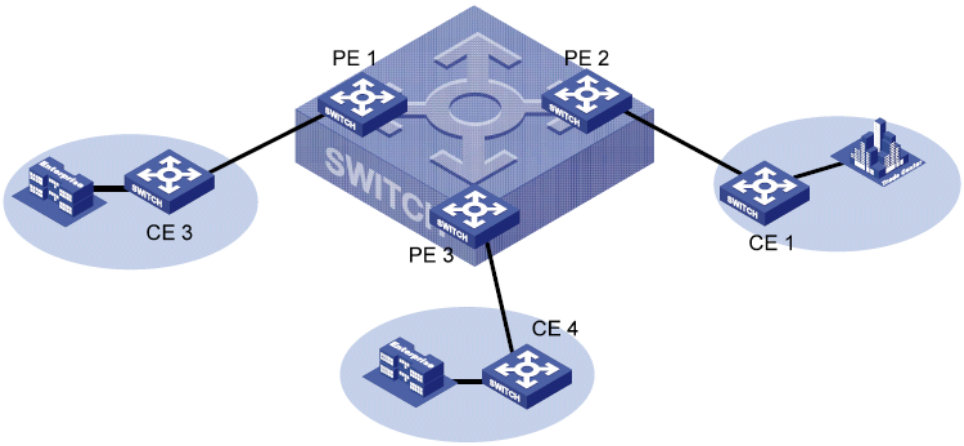
Virtual Private LAN Service (VPLS) is based on the traditional VLL technology. It supports multipoint-to-multipoint communication and has proven to be a better solution for service providers.

1.2 Benefits

VPLS combines the advantages of Ethernet and MPLS and also provides all the functions of a traditional LAN. It can connect geographically dispersed Ethernets through IP/MPLS networks so that the Ethernets can work as a single LAN.

In Figure 1-1, a service provider simulates an Ethernet bridge on the MPLS backbone network by using VPLS. The bridge forwards frames based on MAC addresses or MAC addresses and VLAN tags. In the simplest case, all sites connected to the PEs belong to a single VPLS instance, and each CE needs to communicate with the other CEs in the VPLS instance. For the CEs, the MPLS backbone functions just like an Ethernet bridge.

Figure 1-1 Ethernet bridge emulated by VPLS



VPLS provides the following advantages:

- VPLS uses Ethernet interfaces at the user side, supporting quick and flexible service deployment at the border between the LAN and the WAN.
- With VPLS, users control and maintain routing policies on the user networks. This simplifies management of the service provider network.
- All CEs of a VSI belong to the same subnet, which simplifies IP address planning.
- The VPLS service is invisible to users, and it is not involved in IP addressing and routing.

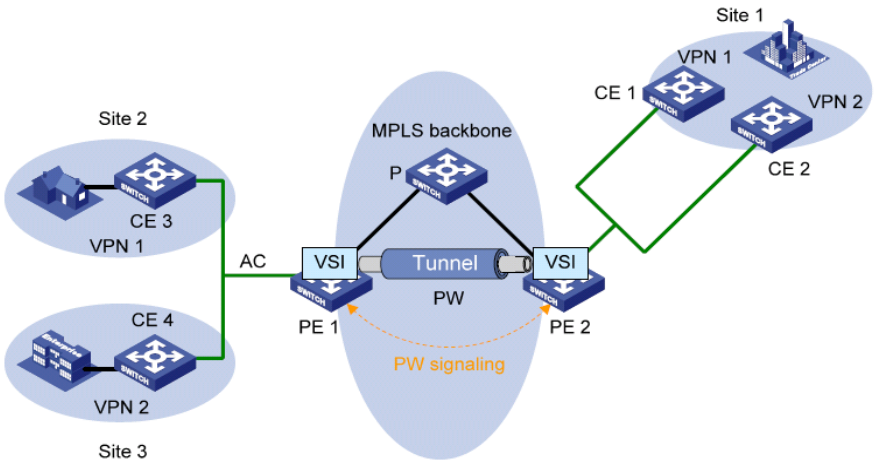
2 VPLS Implementation

2.1 Concepts

- Customer edge device (CE): is directly connected to the service provider network.
- Provider edge device (PE): connects one or more CEs to the service provider network for VPN service access. A PE converts packet formats and forwards packets between private networks and public network tunnels. In Hierarchical VPLS (H-VPLS) networking, a PE can be a UPE or NPE.
- User-facing provider edge device (UPE): functions as the user access convergence device.
- Network provider edge device (NPE): functions as the network core PE. An NPE resides at the edge of a VPLS network core domain and provides transparent VPLS transport services between core networks.
- Service delimiter: is an identifier added before a user data frame to identify which VPN the packet belongs to. A service delimiter is effective only on the local device. An example of a service delimiter is the outer tag used in QinQ.
- 802.1Q in 802.1Q (QinQ): is a tunneling protocol based on 802.1Q. It offers point-to-multipoint L2VPN services. QinQ encapsulates private network VLAN tags in public network VLAN tags, so packets travel across the service provider network with two layers of VLAN tags. QinQ is a simple Layer 2 VPN tunneling service.

2.2 VPLS Network Architecture

Figure 2-1 VPLS network architecture



As shown in Figure 2-1, a VPLS network consists of these primary components:

- Attachment circuit (AC): is a link between a CE and a PE that connects a user with the service provider. The ends of an AC can only be Ethernet interfaces.
- Pseudo wire (PW): is a bidirectional virtual connection between Virtual Switch Instances (VSIs) on two PEs. A PW, also called an emulated circuit, consists of two unidirectional MPLS virtual circuits (VCs).
- Tunnel: is a direct channel between a local PE and the peer PE for transparent data transmission. Tunnels are used to carry PWs. There are two types of tunnels: MPLS tunnels and GRE tunnels.
- PW Signaling: is the protocol that VPLS is based on. It is used for creating and maintaining PWs and automatically discovering VSI peer PEs. Currently, there are two PW signaling protocols: Label Distribution Protocol (LDP) and border gate protocol (BGP).
- Virtual switch instance (VSI): is an Ethernet bridge function entity of a VPLS instance on a PE. It forwards Layer 2 frames based on MAC addresses and VLAN tags.

2.3 Establishing a PW

A PW is a communication tunnel on the public network. It can be established on an MPLS tunnel (a common LSP or a CR-LSP) or a GRE tunnel. To establish a PW, perform the following configurations:

1. Establish an MPLS or GRE tunnel between the local end and the peer PE.
2. Determine the address of the peer PE. If the peer PE is in the same VSI as the local end, you can specify the address of the peer PE manually or let the signaling protocol find the peer PE automatically.
3. Use the LDP or BGP signaling protocol to assign multiplex distinguishing flags (that is, VC labels) and advertise the assigned VC flags to the peer PE to establish unidirectional

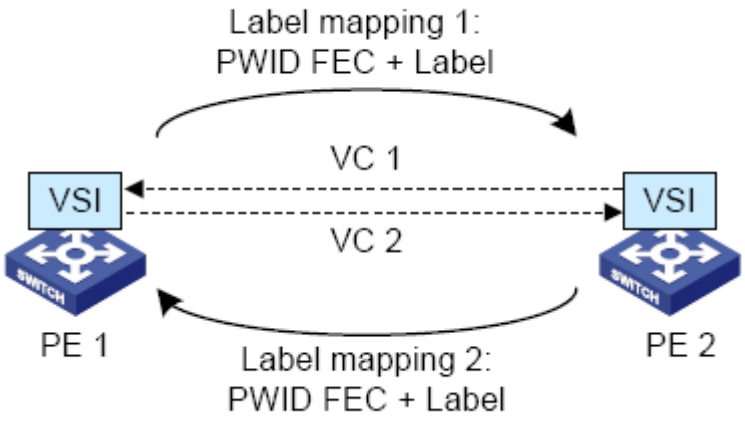
VCs. If a PW is established on an MPLS tunnel, a packet transported over the PW will contain two levels of labels. The inner label, or VC label, identifies the VC to which the packet belongs so that the packet is forwarded to the correct CE. The outer label, or public network MPLS tunnel label, is for guaranteeing that the packet is correctly transmitted on the MPLS tunnel.

The following describes how to use the two signaling protocols (LDP and BGP) to establish a PW.

2.3.1 LDP

When VPLS uses extended LDP (remote LDP sessions) as the PW signaling protocol, it is called Martini VPLS.

Figure 2-2 Establish a PW by using LDP



As shown in Figure 2-2, the process of establishing a PW using LDP is as follows:

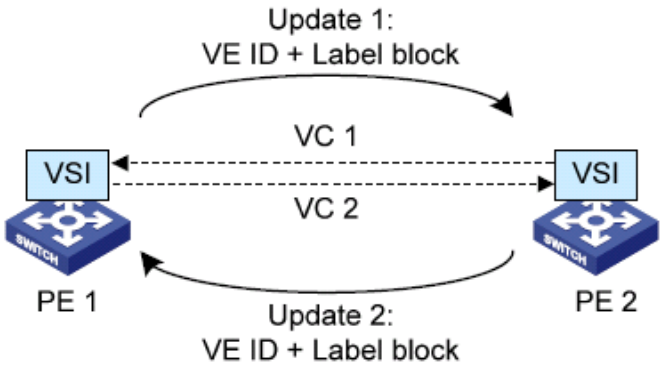
1. After being associated with a VSI, each PE uses LDP in downstream unsolicited (DU) mode to send a label mapping message to its peer PE without first being solicited. The message contains the PWID FEC, the VC label bound with the PWID FEC, and the interface settings, such as the maximum transmission unit.
2. Upon receiving a label mapping message, each PE determines whether it is associated with the PWID. If so, the PE accepts the label mapping message and responds with its own label mapping message.
3. After a unidirectional VC is established in each direction, the PW is formed. A PW can be considered a virtual Ethernet interface of a VSI.

Martini VPLS is easy to implement. However, as LDP does not provide automatic VPLS member discovery, PE peers need to be manually configured. In addition, when a new PE is deployed on the network, configurations of all PEs on the network need to be modified.

2.3.2 BGP

When VPLS uses extended BGP as the PW signaling protocol, it is called Kompella VPLS.

Figure 2-3 Establish a PW by using BGP



As shown in Figure 2-3, the process of establishing a PW using BGP is as follows:

1. A PE (PE1) sends BGP update messages to all peer PEs to notify the peers of its VE ID and label block information. A VE ID is the unique identifier of a site connected to the PE in the VPN and is assigned by the service provider. A label block is a group of consecutive labels.
2. Upon receiving a BGP update message, a peer PE ("PE 2") figures out a unique label value based on its own VE ID information and the label block in the update message, and uses the label value as the VC label. At the same time, PE2 determines the VC label of the peer PE (PE1) based on the VE ID and local label block in the message.
3. Once two PE peers receive update messages from each other and figure out the VC labels, a PW is established between them.

Kompella VPLS implements automatic VPLS member discovery using VPN target configurations and requires no manual configuration when a PE is added to or removed from the network; therefore, Kompella VPLS is highly scalable. However, the BGP protocol is complex.

2.4 VPLS Packet Encapsulation

2.4.1 Packet Encapsulation on an AC

On an AC, two packet encapsulation types are available: VLAN access and Ethernet access.

- VLAN access: The Ethernet header of a packet sent from a CE to a PE or from a PE to a CE includes a VLAN tag, which is added in the header as a service delimiter. Then, the service provider network can use the service delimiter to identify the user. The VLAN tag is also called a P-Tag.
- Ethernet access: The Ethernet header of a packet sent from a CE to a PE or from a PE to a CE does not contain any service delimiter. Conversely, if the Ethernet header contains a VLAN tag, it is the internal VLAN tag of the user, or "U-Tag," which means nothing to the PE.

2.4.2 Packet Encapsulation on a PW

A PW is uniquely identified by its PWID and PW encapsulation type. Two PE peers can establish a PW only when they have the same PWID and PW encapsulation type.

Two packet encapsulation types are available on a PW: raw and tagged.

- In raw mode, packets transmitted on a PW do not contain P-tags. If a packet from a CE contains a service delimiter, the PE removes the service delimiter and adds the PW label and tunnel label into the packet before sending the packet out. If the packet contains no delimiter, the PE directly adds the PW label and tunnel label into the packet and then sends the packet out. When receiving a packet destined for the CE, PE determines whether to add the service delimiter to the packet depending on your configuration. However, the PE cannot rewrite or remove existing tags of the packet.
- In tagged mode, every packet on a PW must carry a P-Tag. If a packet from a CE contains a service delimiter, the PE retains the P-Tag or changes it into the VLAN tag specified by the peer PE or an empty tag (with a value of 0). Then, the PE adds the PW label and tunnel label into the packet. If the packet contains no service delimiter, the PE directly adds the VLAN tag specified by the peer PE or an empty tag, and then adds the PW label and tunnel label into the packet before sending the packet out. When receiving a packet destined for the CE, the PE rewrites, removes, or retains the service delimiter depending on your configuration.

2.5 MAC Address Management

For user networks, a VPLS network emulates an Ethernet bridge, which forwards packets based on MAC addresses or based on MAC addresses and VLAN tags. Each PE associated with a particular VPLS service establishes a VSI for the VPLS instance. Each VSI maintains a MAC address table and supports packet flooding and forwarding, as well as MAC address learning and aging.

2.5.1 Flooding and Forwarding

PEs on a VPLS network forward data packets by searching the MAC address table in VSIs.

Upon receiving a unicast packet, multicast packet, or broadcast packet with an unknown destination MAC address from an AC in a VSI, a PE sends the packet to all the local ACs and PWs in the VSI.

Upon receiving a unicast packet, multicast packet, broadcast packet with an unknown destination MAC address from a PW in a VSI, the PE sends the packet to all the local ACs in the VSI and does not send the packet to any PWs in the VSI.

2.5.2 MAC Address Learning

There are two types of MAC address learning: remote and local.

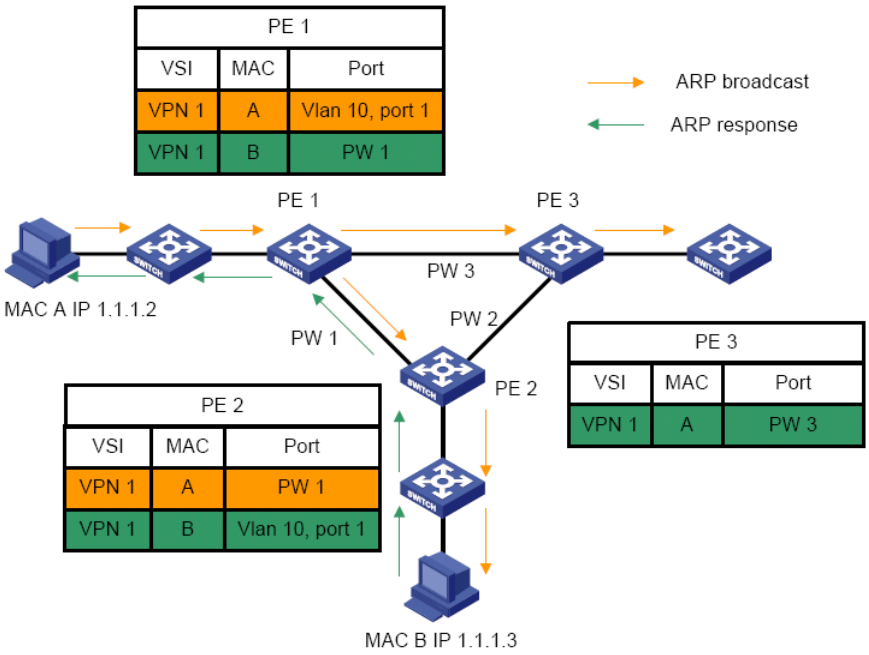
- Remote MAC address learning associated with PWs

  A PW consists of two unidirectional VC LSPs. A PW is in Up state only when both of the VC LSPs are up. When the inbound VC LSP learns a new MAC address, the PE associates the learned MAC address with the PW (namely, the virtual Ethernet interface). In other words, the PE maps the MAC address to the outbound VC LSP.
- Local MAC address learning of interfaces directly connected to users

A PE learns source MAC addresses from Layer 2 packets originated by CEs. This occurs on the Ethernet interfaces directly connected to CEs.

Figure 2-4 illustrates the MAC address learning process and flooding process.

Figure 2-4 MAC address learning



2.5.3 MAC Address Aging

Learned MAC address entries that are no longer in use need to be aged out. The aging mechanism is based on the source MAC addresses of received packets. Whenever a PE receives a packet, it sets an activated or effective flag for the corresponding MAC address entry. If a MAC address entry is not tagged activated or effect within a specified period of time, the MAC address is removed from the MAC address table.

2.6 Loop Avoidance

In general, Layer 2 networks use STP to avoid loops. However, STP is not applicable for VPLS networks because VPLS users are not aware of the service provider network topology. In VPLS, PW full mesh and split horizon forwarding are used to avoid loops.

- PW full mesh

  PEs are logically fully meshed. In other words, each PE must create for each VPLS forwarding instance a tree to all the other PEs that share the same instance.

- Split horizon forwarding

  Each PE must support split horizon to avoid loops. Upon receiving a packet from a PW, a PE does not forward the packet to the other PWs of the VSI. Instead, any two PEs can communicate directly through the PW connecting them. This is why PW full mesh is required for each VSI instance.

2.7 Hierarchical VPLS Implementation

VPLS requires that the PEs be fully meshed. Therefore, the following formula can be used for a VPLS instance:

Number of PWs = Number of PEs x (Number of PEs - 1)/2

When the VPLS network is large, there is a large number of PWs. As a result, the PW signaling cost is very high and the network is difficult to manage and expand. Hierarchical VPLS (H-VPLS) simplifies network management and improves network scalability.

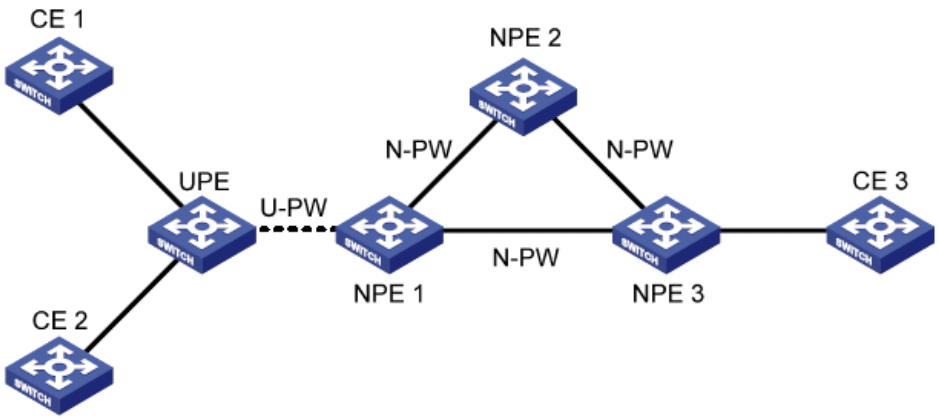
In H-VPLS, a PE can be a UPE or NPE. A UPE can function as an MTU and connect CEs to the service provider network. Meanwhile, an NPE resides at the edge of a VPLS network core domain and can provide transparent VPLS transport services between core networks. NPEs must be fully meshed, but each UPE does not need to be connected to all NPEs. Due to its hierarchical structure, H-VPLS requires fewer PWs and eases the PW signaling burden.

2.7.1 H-VPLS Access Modes

Depending on the connection modes between UPEs and NPEs, there are two H-VPLS access modes: LSP access and QinQ.

- LSP access

Figure 2-5 H-VPLS LSP access



As shown in Figure 2-5, a UPE functions as an MTU and establishes a U-PW only with NPE 1. To establish a U-PW, the NPE and UPE must have a VSI and peer configured, and they must have the same PWID.

In this scenario, a packet is forwarded as follows:

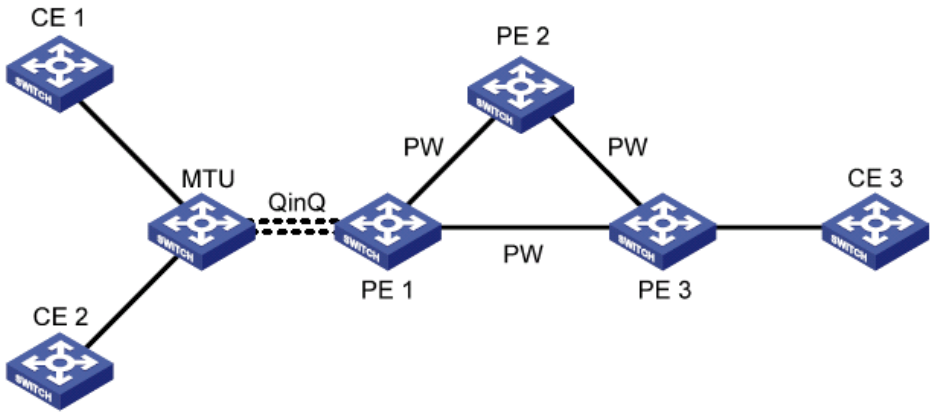
1. For each packet from CE 1 or CE 2, the UPE adds the VC label of U-PW (VC label that NPE 1 assigns for PW multiplexing/demultiplexing) to the packet, and then forwards the packet to NPE 1.
2. NPE 1 uses the packet's VC label to figure out the VSI of the packet, adds the VC label of an N-PW to the packet based on the destination MAC address of the packet, and then sends the packet out.

3. When NPE 1 receives a packet from an N-PW, it adds the VC label of an U-PW to the packet and forwards the packet to the UPE, which in turn forwards the packet to the destination CE.

If the UPE supports bridging, it can forward packets between CE 1 and CE 2 directly without NPE 1. However, the situation is different for packets with unknown destination MAC addresses and broadcast packets. Upon receiving such a packet from a local CE, the UPE not only broadcasts the packet to the other local CEs using the bridging function, but also forwards it through the U-PW to NPE 1, which replicates the packet and sends a copy to each peer CE.

- QinQ access

Figure 2-6 H-VPLS QinQ access



As shown in Figure 2-6, an Ethernet QinQ connection is established between the MTU (UPE) and PE 1. To establish a QinQ connection, enable QinQ on the MTU interfaces connected to CEs and configure VLAN access mode on PE 1.

Upon receiving a packet from CE 1 or CE 2, the MTU adds an outer VLAN tag to the packet and forwards the packet to PE 1. Then PE 1 interprets the outer VLAN tag as the service provider VLAN tag based on the VLAN access mode. The VLAN tag is the service delimiter assigned to the user by the service provider. With this delimiter, PE 1 maps the packet to the corresponding VSI instance and unicasts or broadcasts the packet according to the VSI instance.

The following details the forwarding process:

1. With QinQ enabled on the interfaces connecting CEs, the MTU adds a VLAN tag to each packet from the CEs and transparently forwards the packet through the QinQ tunnel to PE 1.
2. PE 1 uses the packet's VLAN tag to determine the VSI of the packet, adds the PW label of the PW to the packet based on the destination MAC address of the packet, and then sends the packet out.
3. PE 1 uses the PW label to determine the VSI of the packet and labels the packet with the VLAN tag based on the destination MAC address of the packet. Then, PE 1 forwards the packet through the QinQ tunnel to the MTU, which in turn forwards the packet to the destination CE.

Since the MTU supports bridging, it can forward packets between CE 1 and CE 2 directly without PE 1. However, the situation is different for packets with unknown destination MAC addresses and broadcast packets. Upon receiving such a packet from a local CE, the MTU not only broadcasts it to the other local CEs using the bridging function, but also forwards it through the QinQ tunnel to PE 1, which replicates the packet and sends a copy to each peer CE.

2.7.2 Link Redundancy in H-VPLS

If there is only a single link between a UPE and an NPE or between an MTU and a PE, all VPNs supported by the aggregation device are disconnected in the event of a link failure. Therefore, link redundancy is required for both H-VPLS access modes. Normally, the aggregation device uses only one link for access. This is called the primary link. When the primary link fails, the backup link is used instead.

In H-VPLS LSP access mode, since LDP runs between a UPE and a NPE, the status of the primary PW can be determined by checking the status of the LDP session. In HVPLS QinQ access mode, you need to configure STP between the MTU and the PE connected to it, so that a backup link is used when the primary link fails.

Figure 2-7 Link redundancy in H-VPLS LSP access mode

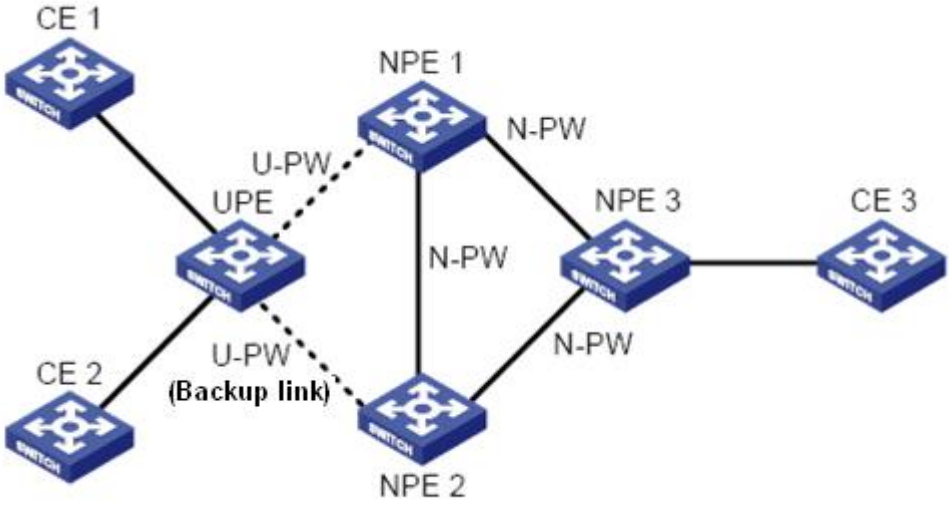
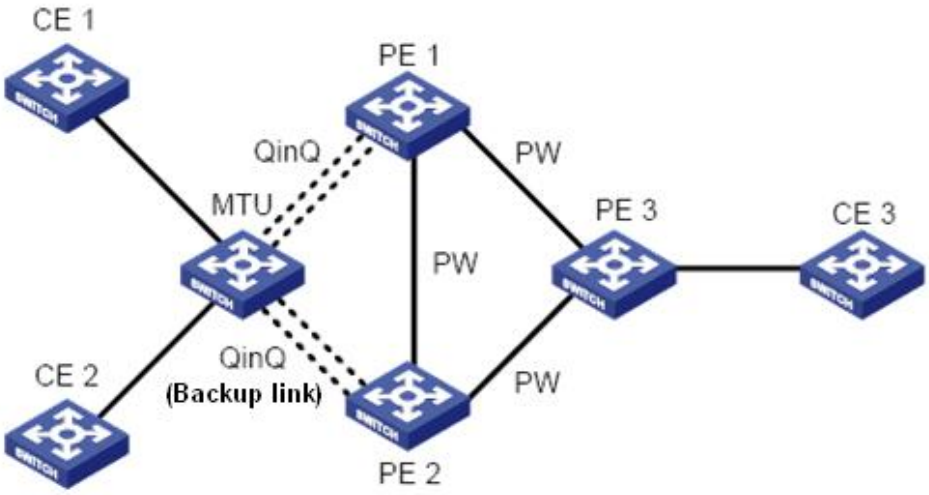


Figure 2-8 Link redundancy in H-VPLS QinQ access mode



2.7.3 Loop Avoidance in H-VPLS

Loop avoidance in H-VPLS is different from that in common VPLS in the following aspects:

- Only NPEs need to be fully meshed and a UPE does not need to be connected to all NPEs.
- Upon receiving a packet from a PW connected to another NPE, an NPE does not forward the packet to the other PWs of the VSI that are connected to NPEs, but it forwards the packet to the PWs connected to UPEs.
- When an NPE receives a packet from a PW connected to a UPE, it forwards the packet to all PWs of the VSI that are connected to other NPEs.

2.8 Restrictions

2.8.1 QinQ Configuration and Packet Encapsulation on PWs

Some products do not resolve the outer tags of received packets according to Packet Encapsulation on an AC. Instead, they process tags based on whether QinQ is enabled on the private network interfaces and the packet encapsulation types of PWs. When using these products, note that:

- When a VPLS service board processes VPLS traffic, it always treats outer tags as P-Tags (service delimiters). If QinQ is not enabled on the private network interfaces and PWs are working in raw mode, the board removes the outer tag, even if the tag is a U-Tag. Therefore, if you do not want the U-tags to be removed, enable QinQ on the private network interfaces, so that the board removes only the tags added during QinQ access.

- The value of the Requested VLAN ID field advertised to the peer depends on the number of ACs supported by VPLS, that is, the number of interfaces that can be bound with a VSI. If a VSI is bound with a single AC, the value of the Requested VLAN ID field is the VLAN ID of the local private network interface; otherwise, the value is 0. Therefore, if a VSI can be bound with more than one AC and the encapsulation type of the PW is tagged, the P-Tag of a packet on the PW is a null tag.

2.8.2 H-VPLS QinQ Access

When configuring H-VPLS QinQ access mode, note that:

- On an MTU, you need to enable STP or MSTP on the Ethernet interfaces connected to NPEs and CEs, so that BPDU messages are exchanged between NPEs to avoid loops.
- On an NPE, ensure that MSTP is not enabled on the Ethernet interfaces connected to an MTU. Otherwise, BPDU messages will not be able to be transferred properly.
- To prevent BPDU messages from being transferred to other PEs in the VPLS network domain, map BPDU messages to a VPLS instance other than the one for user data packets.

3 Huawei Implementation Characteristics

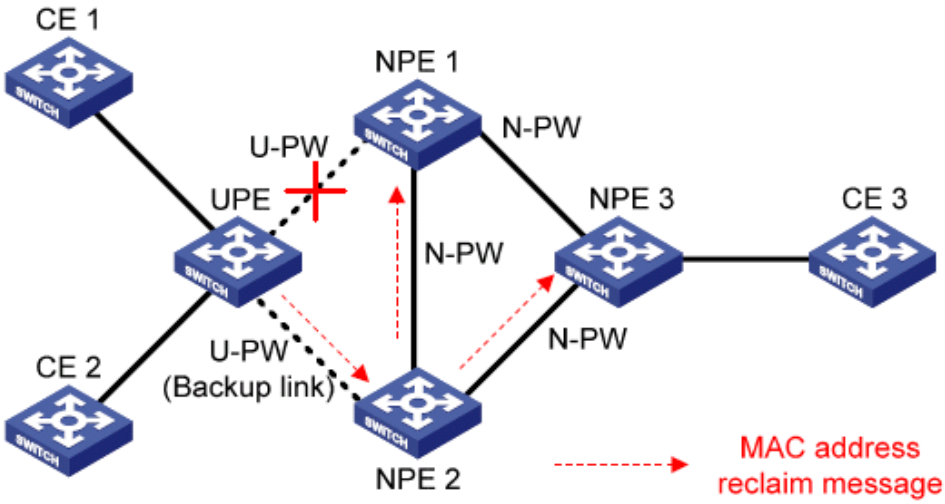
3.1 H-VPLS Networking

3.1.1 MAC Address Reclaiming

A UPE belongs to two NPEs. When the PW in use (for example, the primary PW) fails, the UPE initiates a PW switchover to use the backup PW. However, for a short period of time after the original PW fails, the NPEs of the other sites continue to send packets over the original PW, which the NPE cannot forward. To improve the convergence speed, a mechanism is introduced to inform other NPEs of the PW switchover, so that they remove their MAC entries for the VSI, initiate MAC address relearning, and reestablish MAC forwarding paths. This mechanism uses LDP address reclaim messages.

In VRP implementation of VRP, the UPE sends MAC address reclaim messages. As shown in Figure 3-1, the UPE sends a MAC address reclaim message to the NPE connected by the newly activated PW, which in turn forwards the message to other NPEs.

Figure 3-1 MAC address reclaiming in VRP implementation



An address reclaim message contains a MAC TLV. When a device receives an address reclaim message, it removes or relearns MAC addresses according to the parameters specified in the TLV. When the quantity of MAC addresses is large, a null MAC address list can be sent to improve convergence performance. When an NPE receives an address reclaim message containing an empty MAC address list, it removes all the MAC addresses of the specified VSI except the one learned from the PE sending the message. VRP can reclaim MAC addresses only by sending an empty MAC address list.

3.1.2 BFD Detection and Redundancy

In an H-VPLS network, two PWs are established from one UPE to two NPEs for redundancy: one PW is the primary and the other is the secondary. When the primary PW fails, traffic is switched to the secondary PW.

BFD is a detection mechanism used throughout the network for quick detection and monitoring of link connectivity. It sends detection packets at an interval as short as 10 ms to detect the route reachability from a UPE to a NPE. When BFD detects that a connection from a UPE to an NPE fails, the UPE initiates a PW switchover and uses the backup PW to forward traffic.

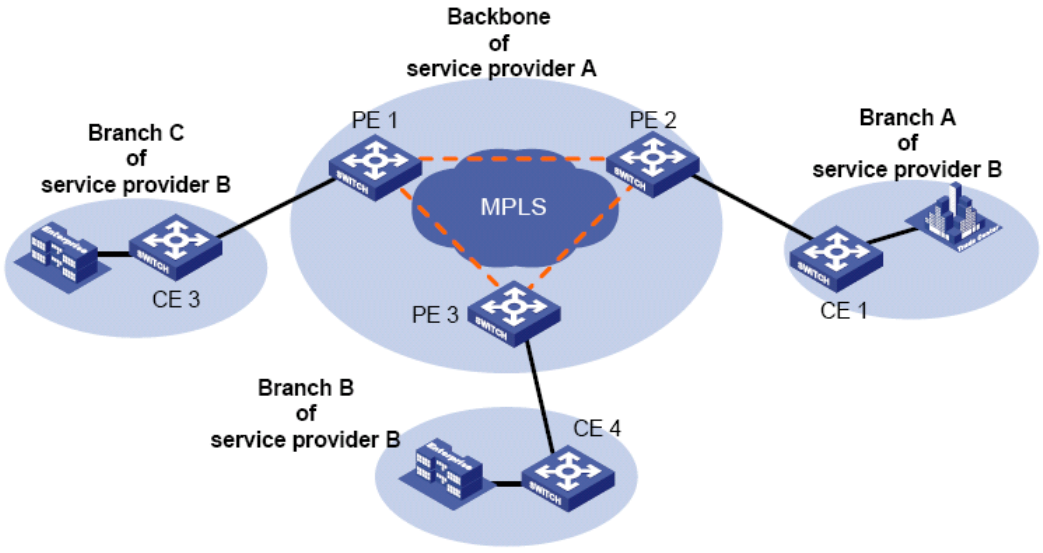
4 Application Scenarios

VPLS applies to large enterprises with the following characteristics:

- They have their own maintenance staff.
- Their networks are large and have many routes.
- They have many routes and geographically dispersed branches,
- They have a high requirement for service quality.

Figure 4-1 illustrates a typical application scenario where service provider A has a backbone network that covers an entire country. Service provider B wants to rent bandwidth from service provider A to connect branches in several cities. Service provider B has enough network management and maintenance capability. To provide high-quality services and ensure privacy, service provider B adopts VPLS for its network.

Figure 4-1 Typical VPLS networking



5 References

- RFC 4447: Pseudo-Wire Setup and Maintenance Using the Label Distribution Protocol (LDP)
- RFC 4761: Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling
- RFC 4762: Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling