Internet Working Group
Internet Draft

Ali Sajassi
Cisco Systems

Yetik Serbest
SBC Communications

Frank Brockners
Cisco Systems

Dinesh Mohan
Nortel

Expires: December 2006

VPLS Interoperability with CE Bridges
draft-sajassi-l2vpn-vpls-bridge-interop-03.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Abstract

One of the main motivations behind VPLS is its ability to provide connectivity not only among customer routers and servers/hosts but also among customer bridges. If only connectivity among customer IP routers/hosts was desired, then IPLS solution [IPLS] could have been

used. The strength of the VPLS solution is that it can provide connectivity to both bridge and non-bridge types of CE devices. VPLS is expected to deliver the same level of service that current enterprise users are accustomed to from their own enterprise bridged networks today or the same level of service that they receive from their Ethernet Service Providers using IEEE 802.1ad-based networks [P802.1ad] (or its predecessor QinQ-based network).

When CE devices are IEEE bridges, then there are certain issues and challenges that need to be accounted for in a VPLS network. Majority of these issues have currently been addressed in IEEE 802.1ad standard for provider bridges and they need to be addressed for VPLS networks. This draft discusses these issues and wherever possible, the recommended solutions to these issues.

Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119

Table of Contents

1. Overview

Virtual Private LAN Service is a LAN emulation service intended for providing connectivity between geographically dispersed customer sites across MAN/WAN (over MPLS/IP) network(s), as if they were connected using a LAN. One of the main motivations behind VPLS is its ability to provide connectivity not only among customer routers and servers/hosts but also among customer bridges. If only connectivity among customer IP routers/hosts was desired, then IPLS solution [IPLS] could have been used. The strength of the VPLS solution is that it can provide connectivity to both bridge and non-bridge types of CE devices. VPLS is expected to deliver the same level of service that current enterprise users are accustomed to from their own enterprise bridged networks [802.1D/802.1Q] today or the same level of service that they receive from their Ethernet Service Providers using IEEE 802.1ad-based networks [P802.1ad] (or its predecessor QinQ-based network).

When CE devices are IEEE bridges, then there are certain issues and challenges that need to be accounted for in a VPLS network. Majority of these issues have currently been addressed in IEEE 802.1ad standard for provider bridges and they need to be addressed for VPLS networks. This draft discusses these issues and wherever possible, the recommended solutions to these issues. It also discusses interoperability issues between VPLS and IEEE 802.1ad networks when the end-to-end service spans across both types of networks, as outlined in [VPLS-LDP].

This draft categorizes the CE-bridge issues into two groups: 1) Mandatory and 2) Optional. The issues in group (1) need to be addressed in order to ensure the proper operation of CE bridges. The issues in group (2) would provide additional operational improvement and efficiency and may not be required for interoperability with CE bridges. Sections four and five discuss the mandatory and optional issues respectively.

2. Ethernet Service Instance

Before starting the discussion of bridging issues, it is important to first clarify the Ethernet Service definition. The term VPLS has different meanings in different contexts. In general, VPLS is used in the following contexts [Eth-OAM]: a) as an end-to-end bridged-LAN service over one or more network (one of which being MPLS/IP network), b) as an MPLS/IP network supporting these bridged LAN services, and c) as (V)LAN emulation. For better clarity, we differentiate between its usage as network versus service by using the terms VPLS network and VPLS instance respectively. Furthermore,

we confine VPLS (both network and service) to only the portion of the end-to-end network that spans across an MPLS/IP network. For an

end-to-end service (among different sites of a given customer), we use the term Ethernet Service Instance or ESI.

[MFA-Ether] defines the Ethernet Service Instance (ESI) as an association of two or more Attachment Circuits (ACs) over which an Ethernet service is offered to a given customer. An AC can be either a UNI or a NNI; furthermore, it can be an Ethernet interface or a VLAN, it can be an ATM or FR VC, or it can be a PPP/HDLC interface. If an ESI is associated with more than two ACs, then it is a multipoint ESI. In this document, where ever the keyword ESI is used, it means multipoint ESI, unless it is stated otherwise.

An ESI can correspond to a VPLS instance if its associated ACs are only connected to a VPLS network or an ESI can correspond to a Service VLAN if its associated ACs are only connected to a Provider-Bridged network [P802.1ad]. Furthermore, an ESI can be associated with both a VPLS instance and a Service VLAN when considering an end-to-end service that spans across both VPLS and Provider-Bridged networks. An ESI can span across different networks (e.g., IEEE 802.1ad and VPLS) belonging to the same or different administrative domains.

An ESI (either for a point-to-point or multipoint connectivity) is associated with a forwarding path within the service provider s network and that is different from an Ethernet Class of Service (CoS) which is associated with the frame Quality-of-Service treatment by each node along the path defined by the ESI. An ESI can have one or more CoS associated with it.

An ESI most often represents a customer or a specific service requested by a customer. Since traffic isolation among different customers (or their associated services) is of paramount importance in service provider networks, its realization shall be done such that it provides a separate MAC address domain and broadcast domain per ESI. A separate MAC address domain is provided by using a separate filtering database per ESI (for both VPLS and IEEE 802.1ad networks) and separate broadcast domain is provided by using a full-mesh of PWs per ESI over the IP/MPLS core in a VPLS network and/or a dedicated Service VLAN per ESI in an IEEE 802.1ad network.

3. VPLS-Capable PE Model with Bridge Module

[L2VPN-FRWK] defines three models for VPLS-capable PE (VPLS-PE) based on the bridging functionality that needs to be supported by the PE. If the CE devices can include both routers/hosts and IEEE bridges, then the second model is the most suitable and adequate one and it is consistent with IEEE standards for Provider Bridges [P802.1ad]. We briefly describe the second model and then elaborate

upon this model to show its sub-components based on [P802.1ad] Provider Bridge model.

As described in [L2VPN-FRWK], the second model for VPLS-PE contains a single bridge module supporting all the VPLS instances on that PE where each VPLS instance is represented by a unique VLAN inside that bridge module (also known as Service VLAN or S-VLAN). The bridge module has at least a single Emulated LAN interface over which each VPLS instance is represented by a unique S-VLAN tag. Each VPLS instance can consist of a set of PWs and its associated forwarder corresponding to a single Virtual LAN (VLAN) as depicted in the following figure. Thus, sometimes it is referred to as V-LAN emulation.

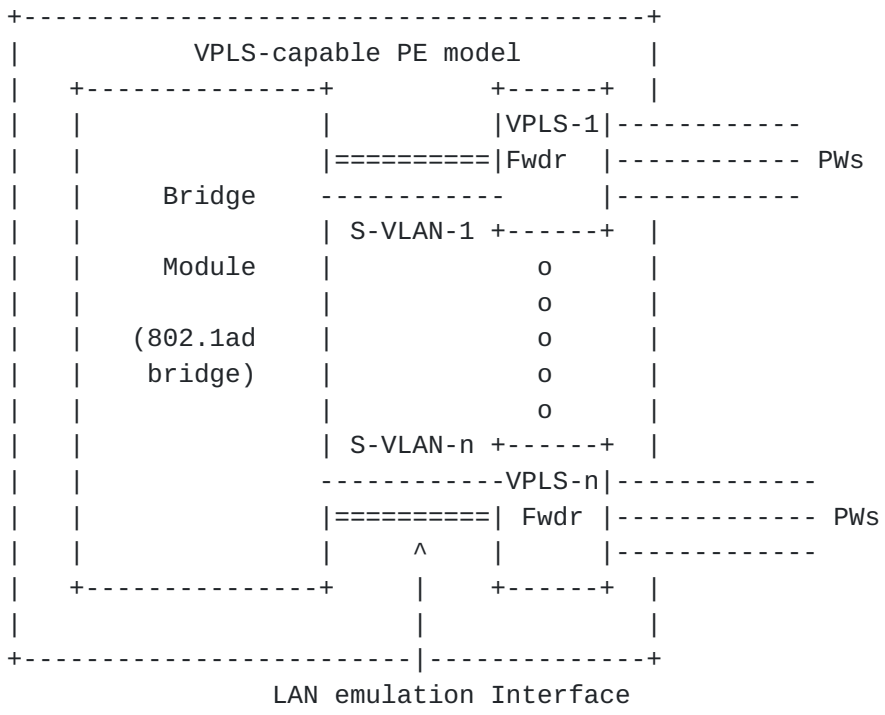

Figure 1. VPLS-capable PE Model

Customer frames associated with a given ESI, carry the S-VLAN ID for that ESI over the LAN emulation interface. The S-VLAN ID is stripped before transmitting the frames over the set of PWs associated with that VPLS instance (assuming raw mode PW is used as specified in [PWE3-Ethernet]).

The bridge module can itself consist of one or two sub-components depending on the functionality that it needs to perform. The following figure depicts the model for bridge module based on [P802.1ad].

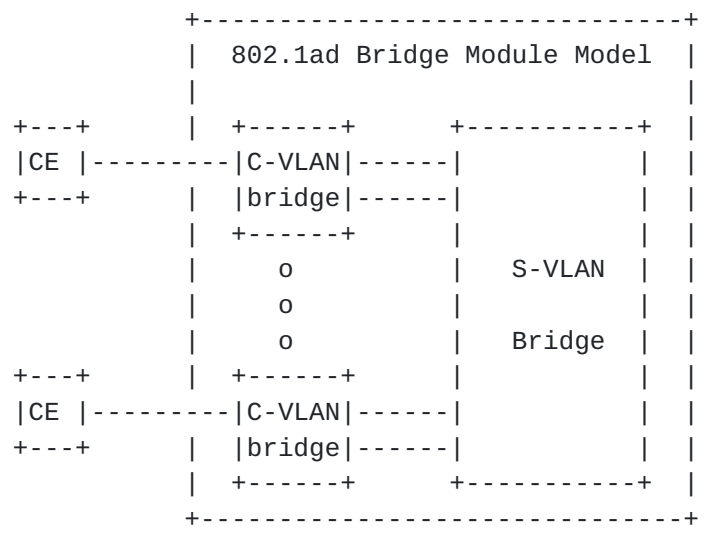

Figure 2. The Model of 802.1ad Bridge Module

The S-VLAN bridge component is always required and it is responsible for tagging customer frames with S-VLAN tags in the ingress direction (from customer UNIs) and removing S-VLAN tags in the egress direction (toward customer UNIs). It is also responsible for running the provider s bridge protocol such as RSTP, MSTP, GVRP, GMRP, etc. among provider bridges within a single administrative domain.

The C-VLAN bridge component is required when the customer Attachment Circuits are VLANs (aka C-VLANs). In such cases, the VPLS-capable PE needs to participate in some of the customer s bridging protocol such as RSTP and MSTP. The reason that such participation is required is because a customer VLAN (C-VLAN) at one site can be mapped into a different C-VLAN at a different site or in case of asymmetric mapping (as describe in the previous section), a customer Ethernet port at one site can be mapped into a customer VLAN (or group of C-VLANs) at a different site.

In scenarios where C-VLAN bridge component is required, then there will be one such component per customer UNI port in order to avoid local switching within the C-VLAN bridge component and thus limiting local switching among different UNIs for the same customer to S-VLAN bridge component.

The C-VLAN bridge component does service selection and identification based on C-VLAN tags. Each frame from the customer device is assigned to a C-VLAN and presented at one or more internal port-based interfaces, each supporting a single service instance that the customer desires to carry that C-VLAN. Similarly frames from the provider network are assigned to an internal interface or LAN (e.g, between C-VLAN and S-VLAN components) on the basis of

the S-VLAN tag. Since each internal interface supports a single service instance, the S-VLAN tag can be, and is, removed at this

interface by the S-VLAN bridge component. If multiple C-VLANs are supported by this service instance (e.g., VLAN bundling or port-based), then the frames will have already been tagged with C-VLAN tags. If a single C-VLAN is supported by this service instance (e.g., VLAN-based), then the frames shall not have been tagged with C-VLAN tag since C-VLAN can be derived from the S-VLAN (e.g., one to one mapping). The C-VLAN aware bridge component applies a port VLAN ID (PVID) to untagged frames received on each internal LAN , allowing full control over the delivery of frames for each C-VLAN through the Customer UNI Port.

4. Mandatory Issues

4.1. Service Mapping

Different Ethernet AC types can be associated with a single Ethernet Service Instance (ESI). For example, an ESI can be associated with only physical Ethernet ports, VLANs, or a combination of two (e.g., one end of the service be associated with physical Ethernet ports and the other end be associated with VLANs). In [VPLS-LDP], unqualified and qualified learning is used to refer to port-based and VLAN-based operation respectively and it does not describe the possible mappings between different types of Ethernet ACs (e.g., 802.1D, 802.1Q or 802.1ad frames). In general, the mapping of a customer port or VLAN to a given service instance is a local function performed by the local PE and the service provisioning shall accommodate it. In other words, there is no reason to restrict and limit an ESI to have only port-based ACs or to have only VLAN-based ACs. [P802.1ad] allows for each customer AC (either a physical port, or a VLAN, or a group of VLANs) to be mapped independently to an ESI which provides better service offering to Enterprise customers. For better and more flexible service offerings and for interoperability purposes between VPLS and 802.1ad networks, it is imperative that both networks offer the same capabilities in terms of customer ACs mapping to the customer service instance.

The following table lists possible mappings that can exist between customer ACs and its associated ESI this table is extracted from [MFA-Ether]. As it can be seen, there are several possible ways to perform such mapping. In the first scenario, it is assumed that an Ethernet physical port only carries untagged traffic and all the traffic is mapped to the corresponding service instance or ESI. This is referred to as port-based w/ untagged traffic . In the second scenario, it is assumed that an Ethernet physical port carries both tagged and untagged traffic and all that traffic is mapped to the corresponding service instance or ESI. This is referred to as port-based w/ tagged and untagged traffic . In the third scenario, it is assumed that only a single VLAN is mapped to the corresponding

service instance or ESI (referred to as VLAN-based). Finally, in the fourth scenario, it is assumed that a group of VLANs from the

Ethernet physical interface is mapped to the corresponding service instance or ESI.

Ethernet I/F & Associated Service Instance(s)				
	Port-based Untagged	port-based tagged & untagged	VLAN-based	VLAN bundling
Port-based untagged	Y	N	Y(Note-1)	N
Port-based tagged & untagged	N	Y	Y(Note-2)	Y
VLAN-based	Y(Note-1)	Y(Note-2)	Y	Y(Note-3)
VLAN Bundling	N	Y	Y(Note-3)	Y

Note-1: In this asymmetric mapping scenario, it is assumed that the CE device with VLAN-based AC is a device capable of supporting [802.1Q] frame format.

Note-2: In this asymmetric mapping scenario, it is assumed that the CE device with VLAN-based AC is a device that can support [P802.1ad] frame format because it will receive Ethernet frames with two tags; where the outer tag is S-VLAN and the inner tag is C-VLAN received from port-based AC. One application example for such CE device is in feature server for DSL aggregation over Metro Ethernet network.

Note 3: In this asymmetric mapping scenario, it is assumed that the CE device with VLAN-based AC can support the [P802.1ad] frame format because it will receive Ethernet frames with two tags; where the outer tag is S-VLAN and the inner tag is C-VLAN received from VLAN bundling AC.

If a PE uses an S-VLAN tag for a given ESI (either by adding an S-VLAN tag to customer traffic or by replacing a C-VLAN tag with a S-VLAN tag), then the frame format and EtherType for S-VLAN shall adhere to [P802.1ad].

As mentioned before, the mapping function between the customer AC and its associated ESI is a local function and thus when the AC is a

single customer VLAN, it is possible to map different customer VLANs

at different sites to a single ESI without coordination among those sites.

When a port-based or a VLAN-bundling is used, then the may PE use an additional S-VLAN tag to mark the customer traffic received over that AC as belonging to a given ESI. If the PE uses the additional S-VLAN tag, then in the opposite direction the PE shall strip the S-VLAN tag before sending the customer frames over the same AC. However, when VLAN-mapping mode is used at an AC and if the PE uses S-VLAN tag locally, then if the Ethernet interface is a UNI, the tagged frames over this interface shall have a frame format based on [802.1Q] and the PE shall translate the customer tag (C-VLAN) into the provider tag (S-VLAN) upon receiving a frame from the customer and in the opposite direction, the PE shall translate from provider frame format (802.1ad) back to customer frame format (802.1Q).

All the above asymmetric services can be supported via the PE model with the bridge module depicted in figure-2 (based on [802.1ad]).

4.2. CE Bridge Protocol Handling

When a VPLS-capable PE is connected to a CE bridge, then depending on the type of Attachment Circuit, different protocol handling may be required by the bridge module of the PE. [P802.1ad] states that when a PE is connected to a CE bridge, then the service offered by the PE may appear to specific customer protocols running on the CE in one of the four ways:

i) Transparent to the operation of the protocol among CEs of different sites using the service provided, appearing as an individual LAN without bridges; or,
ii) Discarding frames, acting as a non-participating barrier to the operation of the protocol; or,
iii) Peering, with a local protocol entity at the point of provider ingress and egress, participating in and terminating the operation of the protocol; or,
iv) Participation in individual instances of customer protocols.

All the above CE bridge protocol handling can be supported via the PE model with the bridge module depicted in figure-2 (based on [802.1ad]). For example, when an Attachment Circuit is port-based, then the bridge module of the PE can operate transparently with respect to the CE s RSTP/MSTP protocols (and thus no C-VLAN component is required for that customer UNI). However, when an Attachment Circuit is VLAN-based (either VLAN-based or VLAN bundling), then the bridge module of the PE needs to peer with the RSTP/MSTP protocols running on the CE (and thus the C-VLAN bridge component is required). There are also protocols that require peering but are independent from the type of Attachment Circuit. An

example of such protocol is link aggregation protocol [802.3ad]; however, this is a media-dependent protocol as its name implies. Therefore, the peering requirement can be generalized such that the media-independent protocols (RSTP/MSTP, CFM, etc) that require peering are for VLAN-based or VLAN-bundling Attachment Circuit.

[P802.1ad] reserves a block of 16 MAC addresses for the operation of C-VLAN and S-VLAN bridge components and it shows which of these reserved MAC addresses are only for C-VLAN bridge component and which ones are only for S-VLAN bridge component and which ones apply to both C-VLAN and S-VLAN components.

4.3. Partial-mesh of Pseudowires

The effect of a PW failure (resulting in creation of partial-mesh of PWs) on the CE devices and their supported services should be well known. If the CE devices belonging to an ESI are routers running link state routing protocols that use LAN procedures over that ESI, then a partial-mesh of PWs can result in black holing traffic among the selected set of routers. And if the CE devices belonging to an ESI are IEEE bridges, then a partial-mesh of PWs can cause broadcast storms in the customer and provider networks. Furthermore, it can cause multiple copies of a single frame to be received by the CE and/or PE devices. Therefore, it is of paramount importance to be able to detect PW failure and to take corrective action to prevent creation of partial-mesh of PWs.

[Rosen-Mesh] defines a procedure for detection of partial mesh in which each PE keeps track of the status of PW Endpoint Entities (EEs - e.g., VPLS forwarders) for itself as wells the ones reported by other PEs. Therefore, upon a PW failure, the PE that detects the failure not only takes notice locally but it notifies other PEs belonging to that service instance of such failure so that all the participants PEs have a consistent view of the PW mesh. The procedure defined in [Rosen-Mesh] is for detection of partial mesh per service instance and in turn it relies on additional procedure for PW failure detection such as BFD or VCCV. Given that there can be ten (or even hundreds) of thousands of PWs in a PE, the scalability aspects of this procedure needs to be worked out. Also [Rosen-Mesh] acknowledges that many of the details regarding operational aspects of such procedure are missing and need to be worked out.

If the PE model, with bridge module as depicted in figure-2, is used, then [P802.1ag] procedures could be used for detection of partial-mesh of PWs. [p802.1ag] defines a set of procedures for fault detection, verification, isolation, and notification per ESI.

Fault detection mechanism of [p8021.ag] can be used to perform connectivity check among PEs belonging to a given VPLS instance. It

checks the integrity of a service instance end-to-end within an administrative domain e.g., from one AC at one end of the network to another AC at the other end of the network. Therefore, its path coverage includes bridge module within a PE and it is not limited to just PWs. Furthermore, [P802.1ag] operates transparently over the full-mesh of PWs for a given service instance since it operates at the Ethernet level (and not at PW level). It should be noted that [P802.1ag] assumes that the Ethernet links or LAN segments connecting provider bridges are full-duplex and the failure in one direction results in the failure of the whole link or LAN segment. However, that is not the case for VPLS instance since a PW consists of two uni-directional LSPs and one direction can fail independent from the other causing inconsistent view of the full-mesh by the participating PEs till the detected failure in one side is propagated to the other side.

4.4. Multicast Traffic

VPLS follows a centralized model for multicast replication within an ESI. VPLS relies on ingress replication. The ingress PE replicates the multicast packet for each egress PE and sends it to the egress PE using PtP PW over a unicast tunnel. VPLS operates on an overlay topology formed by the full mesh of pseudo-wires. Thus, depending on the underlying topology, the same datagram can be sent multiple times down the same physical link. VPLS currently does not offer any mechanisms to restrict the distribution of multicast or broadcast traffic of an ESI throughout the network causing additional burden on the ingress PE for unnecessary packet replication, causing additional load on the MPLS core network, and causing additional processing at the receiving PE where multicast packet is discarded.

One possible approach, to deliver multicast more efficiently over VPLS network, is to include the use of IGMP snooping in order to send the packet only to the PEs that have receivers for that traffic, rather than to all the PEs in the VPLS instance. If the customer bridge or its network has dual-home connectivity as described in section 7, then for proper operation of IGMP snooping, the PE must generate a General Query over that customer UNIs upon receiving a customer topology change notification as described in [IGMP-SNOOP]. A General Query by the PE results in proper registration of the customer multicast MAC address(s) at the PE when there is customer topology change. It should be noted that IGMP snooping provides a solution for IP multicast packets and is not applicable to general multicast data.

Using the IGMP-snooping as described, the ingress PE can select a sub-set of PWs for packet replication; therefore, avoiding sending multicast packets to the egress PEs that don t need them. However,

the replication is still performed by the ingress PE. In order to avoid, replication at ingress PE, one may want to use multicast

distribution trees (MDTs) in the provider core network; however, this comes with its potential pitfalls. If the MDT is used for all multicast traffic of a given customer, then this results in customer multicast and unicast traffic to be forwarded on different PWs and even on a different physical topology within the provider network. This is a serious issue for customer bridges because customer BPDUs, which are multicast data, can take a different path through the network than the unicast data. Situations might arise where either unicast OR multicast connectivity is lost. If unicast connectivity is lost, but multicast forwarding continues to work, the customer spanning tree would not take notice which results in loss of its unicast traffic. Similarly, if multicast connectivity is lost, but unicast is working, then the customer spanning tree will activate the blocked port which may result in loop within the customer network. Therefore, the MDT cannot be used for both customer multicast control and data traffic. If it is used, it should only be limited to customer data traffic. However, there can be potential issue even when it is used for customer data traffic since the MDT doesn t fit the PE model described in Figure-1 (it operates independently from the full-mesh of PWs that correspond to an S-VLAN). It is also not clear how CFM procedures (802.1ag) used for ESI integrity check (e.g., per service instance) can be applied to check the integrity of the customer multicast traffic over the provider MDT. Because of these potential issues, the specific applications of the provider MDT to customer multicast traffic shall be documented and its limitations be clearly specified.

5. Optional Issues

5.1. Customer Network Topology Changes

A single CE or a customer network can be connected to a provider network using more than one User-Network Interface (UNI). Furthermore, a single CE or a customer network can be connected to more than one provider network. [L2VPN-REQ] provides some examples of such customer network connectivity that are depicted in the figure below. Such network topologies are designed to protect against the failure or removal of network components with the customer network and it is assumed that the customer leverages the spanning tree protocol to protect against these cases. Therefore, in such scenarios, it is important to flush customer MAC addresses in the provider network upon the customer topology change to avoid black holing of customer frames.

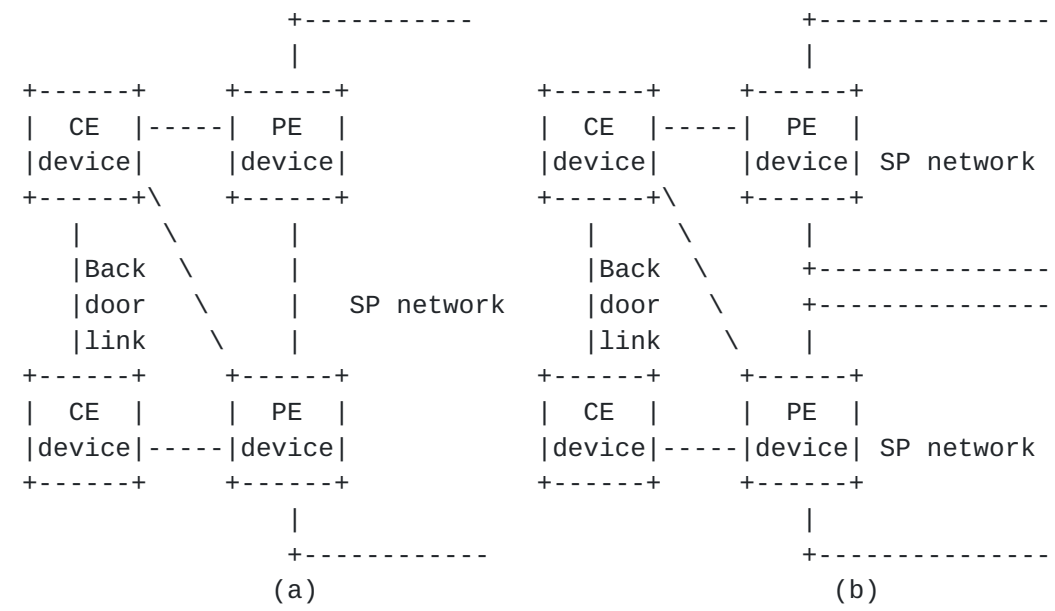

Figure 3. Combination of Dual-Homing and Backdoor Links for CE Devices

The customer networks use their own instances of spanning tree protocol to configure and partition their active topology, so that the provider connectivity doesn t result in data loop. Reconfiguration of a customer s active topology can result in the apparent movement of customer end stations from the point of view of the PEs. However, the requirement for mutual independence of the distinct ESIs that can be supported by a single provider spanning tree active topology does not permit either the direct receipt of provider topology change notifications from the CEs or the use of received customer spanning tree protocol topology change notifications to stimulate topology change signaling on a provider spanning tree.

To address this issue, [P802.1ad] requires that customer topology change notification to be detected at the ingress of the S-VLAN bridge component and the S-VLAN bridge transmits a Customer Change Notification (CCN) BPDU tagged with the S-VLAN ID associated with that service instance and a destination MAC address as specified in the block of 16 reserved multicast MAC addresses. Upon receiving the CCN, the provider bridge will flush all the customer MAC addresses associated with that S-VLAN ID on all the provider bridge interfaces except the one that the CCN message is received from.

Based on the provider bridge model depicted in figure (1), there are two methods of propagating the CCN message over the VPLS network. The first method is to translate the in-band CCN message into an out-of-band MAC Address Withdrawal message as specified in [VPLS-LDP] and the second method is to treat the CCN message as customer

data and pass it transparently over the set of PWs associated with that VPLS instance. The second method is recommended because of ease

of interoperability between the bridge and the LAN emulation modules of the PE.

5.2. Redundancy

[VPLS-LDP] talks about dual-homing of a given u-PE to two n-PEs over a provider MPLS access network to provide protection against link and node failure e.g., in case the primary n-PE fails or the connection to it fails, then the u-PE uses the backup PWs to reroute the traffic to the backup n-PE. Furthermore, it discusses the provision of redundancy when a provider Ethernet access network is used and how any arbitrary access network topology (not just limited to hub-and-spoke) can be supported using the provider s MSTP protocol and how the provider MSTP for a given access network can be confined to that access network and operate independently from MSTP protocols running in other access networks.

In both types of redundancy mechanisms (Ethernet versus MPLS access networks), only one n-PE is active for a given VPLS instance at any time. In case of an Ethernet access network, core-facing PWs (for a VPLS instance) at the n-PE are blocked by the MSTP protocol; whereas, in case of a MPLS access network, the access-facing PW is blocked at the u-PE for a given VPLS instance.

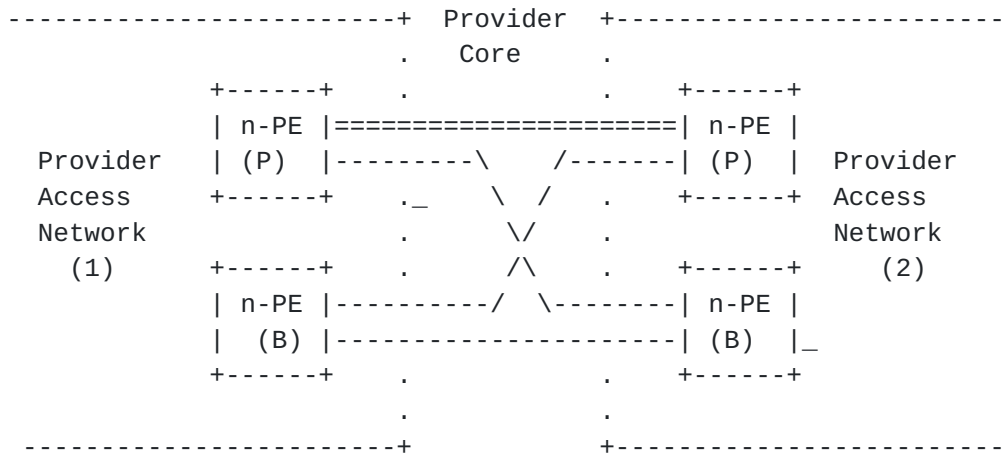

Figure 4. Bridge Module Model

Figure-4 shows two provider access networks each with two n-PEs where the n-PEs are connected via a full mesh of PWs for a given VPLS instance. As shown in the figure, only one n-PE in each access network is serving as a Primary PE (P) for that VPLS instance and the other n-PE is serving as the backup PE (B). In this figure, each primary PE has two active PWs originating from it. Therefore, when a multicast, broadcast, and unknown unicast frame arrives at the primary n-PE from the access network side, the n-PE replicates the

frame over both PWs in the core even though it only needs to send the frames over a single PW (shown with == in the figure) to the

primary n-PE on the other side. This is an unnecessary replication of the customer frames that consumes core-network bandwidth (half of the frames get discarded at the receiving n-PE). This issue gets aggravated when there are more than two n-PEs per provider access network e.g., if there are three n-PEs or four n-PEs per access network, then 67% or 75% of core-BW for multicast, broadcast and unknown unicast are respectively wasted.

Therefore, it is recommended to have a protocol among n-PEs that can disseminate the status of PWs (active or blocked) among themselves and furthermore to have it tied up with the redundancy mechanism such that per VPLS instance the status of active/backup n-PE gets reflected on the corresponding PWs emanating from that n-PE.

The above discussion was centered on the lack of efficiency with regards to packet replication over MPLS core network for current VPLS redundancy mechanism. Another important issue to consider is the interaction between customer and service provider redundancy mechanisms especially when customer devices are IEEE bridges. If CEs are IEEE bridges, then they can run RSTP/MSTP protocols, RSTP convergence and detection time is much faster than its predecessor (IEEE 802.1D STP which is obsolete). Therefore, if the provider network offers VPLS redundancy mechanism, then it should provide transparency to the customer s network during a failure within its network e.g., the failure detection and recovery time within the service provider network to be less than the one in the customer network. If this is not the case, then a failure within the provider network can result in unnecessary switch-over and temporary flooding/loop within the customer s network that is dual homed.

5.3. MAC Address Learning

When customer devices are routers, servers, or hosts, then the number of MAC addresses per customer sites is very limited (most often one MAC address per CE). However, when CEs are bridges, then there can be many customer MAC addresses (e.g., hundreds of MAC addresses) associated with each CE.

[P802.1ad] has devised a mechanism to alleviate MAC address learning within provider Ethernet networks that can equally be applied to VPLS networks. This mechanism calls for disabling MAC address learning for an S-VLAN (or a service instance) within a provider bridge (or PE) when there is only one ingress and one egress port associated with that service instance on that PE. In such cases, there is no need to learn customer MAC addresses on that PE since the path through that PE for that service instance is fixed. For example, if a service instance is associated with four CEs at four different sites, then the maximum number of provider bridges (or

PEs), that need to participate in that customer MAC address learning, is only three regardless of how many PEs are in the path

of that service instance. This mechanism can reduce the number of MAC addressed learned in a H-VPLS with QinQ access configuration.

If the provider access network is of type Ethernet (e.g., IEEE 802.1ad-based network), then the MSTP protocol can be used to partition access network into several loop-free spanning tree topologies where Ethernet service instances (S-VLANs) are distributed among these tree topologies. Furthermore, GVRP can be used to limit the scope of each service instance to a subset of its associated tree topology (and thus limiting the scope of customer MAC address learning to that sub-tree). Finally, the MAC address disabling mechanism (described above) can be applied to that sub-tree, to further limit the number of nodes (PEs) on that sub-tree that need to learn customer MAC addresses for that service instance.

Furthermore, [p802.1ah] provides the capability of encapsulating customers MAC addresses within the provider MAC header. A u-PE capable of this functionality can reduce the number of MAC addressed learned significantly within the provider network for H-VPLS with QinQ access as well as H-VPLS with MPLS access.

6. Interoperability with 802.1ad Networks

[VPLS-LDP] discusses H-VPLS provider-network topologies with both Ethernet [P802.1ad] as well as MPLS access networks. Therefore, it is of paramount importance to ensure seamless interoperability between these two types of networks.

Provider bridges as specified in [P802.1ad] are intended to operate seamlessly with customer bridges and provide the required services. Therefore, if a PE is modeled based on Figures 1&2 which includes a [802.1ad] bridge module, then it should operate seamlessly with Provider Bridges given that the issues discussed in this draft have been taken into account.

7. Acknowledgments

The authors would like to thank Norm Finn for his comments and feedbacks.

8. Security Considerations

There are no additional security aspects beyond that of VPLS/H-VPLS that needs to be discussed here.

9. Normative References

[L2VPN-REQ] Agustyn, W. et al, "Service Requirements for Layer-2 Provider Provisioned Virtual Provider Networks", work in progress

[L2VPN-FRWK] Andersson, L. and et al, "Framework for Layer 2 Virtual Private Networks (L2VPNs)", Work in Progress

[VPLS-LDP] Lasserre, M. and et al, "Virtual Private LAN Services over MPLS", work in progress

[P802.1ad] IEEE Draft P802.1ad/D2.4 Virtual Bridged Local Area Networks: Provider Bridges , Work in progress, September 2004

[P802.1ag] IEEE Draft P802.1ag/D0.1 Virtual Bridge Local Area Networks: Connectivity Fault Management , Work in Progress, October 2004

10. Informative References

[IPLS] Shah, H. and et al, "IP-Only LAN Service (IPLS) ", work in progress, October 2004

[MFA-Ether] Sajassi, A. and et al, Ethernet Service Interworking Over MPLS , Work in Progress, September 2004

[Rosen-Mesh] Detecting and Reacting to Failures of the Full Mesh in IPLS and VPLS ,draft-rosen-l2vpn-mesh-failure-01.txt, March 2004

[PWE3-Ethernet] "Encapsulation Methods for Transport of Ethernet Frames Over IP/MPLS Networks", draft-ietf-pwe3-ethernet-encap-01.txt, Work in progress, November 2002.

[802.1D-REV] IEEE Std. 802.1D-2003 Media Access Control (MAC) Bridges .

[802.1Q] IEEE Std. 802.1Q-2003 "Virtual Bridged Local Area Networks".

[IGMP-SNOOP] Christensen, M. and et al, Considerations for IGMP and MLD Snooping Switches , Work in progress, May 2004

[Eth-OAM] Dinesh Mohan, Ali Sajassi, and et al, L2VPN OAM Requirements and Framework , Work in progress, July 2006

11. Authors' Addresses

Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
Email: sajassi@cisco.com

Yetik Serbest
SBC Labs
9505 Arboretum Blvd.
Austin, TX 78759
Email: yetik_serbest@labs.sbc.com

Frank Brockners
Cisco Systems, Inc.
Hansaallee 249
40549 Duesseldorf
Germany
Email: fbrockne@cisco.com

Dinesh Mohan
Nortel Networks
3500 Carling Ave
Ottawa, ON K2H8E9
Email: mohand@nortel.com

12. Intellectual Property Considerations

This document is being submitted for use in IETF standards discussions.

13. Full Copyright Statement

14. IPR Notice

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF ietf ipr@ietf.org.