IP-MPLS Forum

BGP Autodiscovery and Signaling for VPWS-Based VPN Services

IP-MPLSF 22.0.0

IP-MPLS Forum Technical Committee
April 2009

Note: The user's attention is called to the possibility that implementation of the IP-MPLS Forum implementation agreement contained herein may require the use of inventions covered by patent rights held by third parties. By publication of this IP-MPLS Forum implementation agreement the IP-MPLS Forum makes no representation that the implementation of the specification will not infringe on any third party rights. The IP-MPLS Forum take no position with respect to any claim that has been or may be asserted by any third party, the validity of any patent rights related to any such claims, or the extent to which a license to use any such rights may not be available.

In most team endeavors, the efforts of several individuals deserve special recognition. This specification is no exception. The IP-MPLS Forum gratefully thanks the following individuals and respective employers for their contributions to this document.

Bhupesh Kothari Juniper Networks **(Editor)**
Ed Sierecki AT&T **(Editor)**

For more information contact:

The IP-MPLS Forum
Suite 117
48377 Fremont Blvd.
Fremont, CA 94538 USA

Phone: +1 (510) 492-4000
FAX: +1 (510) 492-4001
E-Mail: info@mfaforum.org
WWW: http://www.mfaforum.org/

Full Notice

Table of Contents

1 Introduction

1.1 Purpose

Layer 2 Virtual Private Networks (L2VPNs) based on Frame Relay (FR) or ATM circuits have been around a long time; more recently, Ethernet VPNs, including Virtual Private LAN Service (VPLS), have become popular. L2VPNs based on FR or ATM often require a separate Service Provider infrastructure for each type, and yet another for the Internet and IP VPNs. In addition, L2VPN provisioning was cumbersome.

Three types of L2VPNs are described in [RFC 4026]: Virtual Private Wire Service (VPWS), Virtual Private LAN Service (VPLS), and IP-only LAN-like Service (IPLS). This specification presents a new approach to the problem of offering VPWS services where the VPWS customer's experience is virtually identical to that offered by traditional Layer 2 VPNs, but such that a Service Provider can maintain a single network for different L2VPNs, IP VPNs and the Internet, as well as a common provisioning methodology for all services.

This document provides specification for VPWS-based VPN services that uses BGP as the control plane protocol. BGP is used to auto-discover the end-points of a VPWS and is also used to signal the VPWS pseudowire.

1.2 Overview

VPNs based on Ethernet Virtual Local Area Networks (VLANs) and Virtual Private LAN Service (VPLS) ([RFC 4761] and [RFC 4762]) have become quite popular. All of these come under the classification of Layer 2 VPNs (L2VPNs), as the customer to Service Provider (SP) hand-off is at Layer 2.

Two major L2 VPN models are distinguished in [RFC4026]: (1) Virtual Private LAN Service (VPLS) which provides for the connection of several LAN segments over a packet network and (2) Virtual Private Wire Service (VPWS) which provides for the connection of two Customer Edge devices (point-to-point). As specified in [RFC 4665], VPWS is not tied to a particular type of L2 service, but applies to all services such as Ethernet, ATM and Frame Relay. It is important to distinguish between a single Layer 2 circuit or a VPWS, which connects two customer sites, and a Layer 2 VPN, which is a set of circuits that connect sites belonging to the same customer. This document addresses VPN services that will be referred to as VPWS-based VPNs.

There has been much progress in network "convergence", whereby Layer 2 traffic, Internet traffic and IP VPN traffic can be carried over a single, consolidated network infrastructure based on IP/MPLS tunnels; this is made possible by techniques such as those described in [RFC 4448], [RFC 4618], [RFC 4619], and [RFC 4717] for Layer 2 traffic, and [RFC 4364] for IP VPN traffic. These developments go a long way towards addressing the problem of network technology proliferation.

The IETF developed technology to support both of the L2 VPN models, however, only one signaling technology (LDP) supports both models, whereas the other VPLS signaling technology (BGP) does not support VPWS. Service providers derive significant operational savings by delivering multiple services using a common technology platform. It is highly desirable to support both of the L2VPN service models using BGP when BGP is the preferred platform by a service provider. For service providers using BGP, this will provide significant operational savings.

1.3 Scope

This specification describes the use of BGP auto-discovery and signaling to offer VPWS-based VPN service. RFC 4761 describes procedures to offer a VPLS service using BGP auto-discovery and signaling. The auto-discovery and signaling procedures used to offer a VPWS-based VPN service in this document are based on those defined in RFC 4761 [RFC 4761]. Advantages of using BGP as the control plane protocol for VPWS-based VPN service, such as separation of administrative responsibilities between a service provider and a customer, are described in Section 5. Procedures for auto-discovery and signaling are described in Section 6. VPWS encapsulations supported are described in Section 7. OAM support is described in Section 9. Operation when a VPWS-based VPN service spans multiple ASes is described in Section 10.

The solution described in this specification has no impact on the mechanism defined in RFC 4448 for setting up pseudowires for L2VPN service using Label Distribution Protocol.

The following functionalites are for further study and are considered outside the scope of this document.

1. Interworking between L2 VPNs using LDP signaling and L2 VPNs using BGP signaling.
2. Multi-segment Pseudowires using LDP.
3. Supporting both LDP and BGP signaling for L2 VPN in the same MPLS network.
4. Services crossing an MPLS Inter-Carrier Interconnect (ICI).

2 Definitions

Must, Shall or Mandatory — the item is an absolute requirement of this specification.

Should — the item is desirable.

May or Optional — the item is not compulsory, and may be followed or ignored according to the needs of the implementer.

2.1 Acronyms

Acronym	Description
AC	Attachment Circuit
ATM	Asynchronous Transfer Mode
AS	Autonomous System
BFD	Bidirectional Forwarding Detection
BGP	Border Gateway Protocol
CAS	Channel-Associated Signaling
CE	Customer Edge
CoS	Class of Service
DLCI	Data Link Connection Identifier
FR	Frame Relay

GRE	Generic Router Encapsulation
HDLC	High-Level Data Link Connection
IETF	Internet Engineering Task Force
IP	Internet Protocol
L2VPN	Layer 2 VPN
LDP	Label Distribution Protocol
LSP	Label Switched Path
LSR	Label Switching Router
MPLS	Multi Protocol Label Switching
MTU	Maximal Transfer Unit
NLRI	Network Layer Reachability Information
OAM	Operations, Administration and Management
P	Provider
PDU	Protocol Data Unit
PE	Provider Edge
PSN	Packet Switched Network
PW	Pseudowire
PWE3	Pseudo Wire Emulation Edge to Edge
RFC	Request for Comments
RSVP-TE	Resource Reservation Protocol with Traffic Engineering Extensions
RTP	Real-time Transport Protocol (IETF RFC3550)
SP	Service Provider
SSRC	Synchronization Source
TDM	Time Division Multiplexing
TLV	Type Length Value
VCCV	Virtual Circuit Connectivity Verification
VE	VPLS Edge
VLAN	Virtual Local Area Network
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
VPWS	Virtual Private Wire Service

3 Normative References

[BFD-Base] Katz & Ward, “Bidirectional Forwarding Detection”, draft-ietf-bfd-base-08.txt, IETF work in progress.

[RFC 3032] Rosen, E. et al., “MPLS Label Stack Encoding”, RFC 3032, January 2001.

[RFC 3985] Bryant, S. and Pate, P. “Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture”, RFC 3985, March 2005.

[RFC 4026] Andersson, L. and Madsen, T. “Provider Provisioned Virtual Private Network (VPN) Terminology”

[RFC 4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[RFC 4364] Rosen, E and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC4379] Kompella, K. and Swallow, G. “Detecting MPLS Data Plane Failures”, IETF, RFC 4379, February 2006.

[RFC 4448] Martini, L., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks", RFC 4448, April 2006.

[RFC 4553] Vainshtein, A. and Stein, YJ. "Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)”, RFC 4553, June 2006.

[RFC 4618] Martini, L., Rosen, E., Heron, G., and A. Malis, “Encapsulation Methods for Transport of PPP/High-Level Data Link Control (HDLC) over MPLS Networks”, RFC 4618, September 2006.

[RFC 4619] Martini, L., Kawa, C., and A. Malis, "Encapsulation Methods for Transport of Frame Relay over Multiprotocol Label Switching (MPLS) Networks", RFC 4619, September 2006.

[RFC 4717] Martini, L., Jayakumar, J., Bocci, M., El-Aawar, N., Brayley, J., and G. Koleyni, "Encapsulation Methods for Transport of Asynchronous Transfer Mode (ATM) over MPLS Networks", RFC 4717, December 2006.

[RFC 4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC 4816] Malis, A, et al., "Pseudowire Emulation Edge-to-Edge (PWE3) Asynchronous Transfer Mode (ATM) Transparent Cell Transport Service", RFC 4816, February 2007.

[RFC 5085] Nadeau, T. and Pignataro, C. "Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires”, RFC 5085, December 2007.

[RFC5086] Vainshtein, A. et al.,“Structure-Aware Time Division Multiplexed (TDM) Circuit Emulation Service over Packet Switched Network (CESoPSN)”, RFC 5086, December 2007.

[RFC6624] Kompella, K et al.,“Layer 2 Virtual Private Networks Using BGP for Auto-discovery and Signaling”, RFC 6624, February 2012.

Note: RFC 6624 allocated the required code points for the document completed in 2009.

[802.1ag] IEEE 802.1 Connectivity Fault Management,December 2007.

3.1 Informative References

[RFC 2796] Bates, T. et al.,“ BGP Route Reflection - An Alternative to Full Mesh IBGP”, RFC 2796, April 2000.

[RFC 4447] Martini, L, El-Aawar, N., Smith, T., and Heron, G. “Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)”, RFC 4447, April 2006.

[RFC 4665] Augustyn, W. and Serbest, Y. "Service Requirements for Layer 2 Provider-Provisioned Virtual Private Networks", RFC 4665, September 2006.

[RFC 4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

4 Terminogy

The terminology used is from [RFC 4761] and [RFC 4364], and is briefly repeated here. A "customer" is a customer of a service provider seeking to interconnect their various "sites" (each an independent network) at Layer 2 through the service provider's network, while maintaining privacy of communication and address space. The device in a customer site that connects to a service provider router is termed the CE (customer edge) device; this device may be a router or a switch. The service provider router to which a CE connects is termed a PE. A VPLS PE is also known as a VE (VPLS Edge). A router in the service provider's network which doesn't connect directly to any CE is termed P. Every pair of PEs is connected by a logical PSN "tunnel"; within a tunnel, VPN data is distinguished by a "demultiplexor", which in this document is an MPLS label.

Each CE within a VPN is assigned a CE ID, a number that uniquely identifies a CE within an L2 VPN. More accurately, the CE ID identifies a physical connection from the CE device to the PE, since a CE may be connected to multiple PEs (or multiply connected to a PE); in such a case, the CE would have a CE ID for each connection. A CE may also be part of many L2 VPNs; it would need one (or more) CE ID(s) for each L2 VPN of which it is a member. The number space for CE IDs is scoped to a given VPN.

In the case of inter-Provider L2 VPNs, there needs to be some coordination of allocation of unique CE IDs across multiple ASes. How CE IDs can be coordinated across multiple ASes is outside the scope of this specification.

Within each physical connection from a CE to a PE, there may be multiple virtual circuits. These will be referred to as Attachment Circuits (ACs), following [RFC 4447]. Similarly, the entity that connects two attachment circuits across the service provider network is called a pseudowire (PW). The solution presented in this specification conforms to the PWE3 architecture described in RFC 3985 [RFC 3985].

For the purpose of this document VPWS is always used to refer to a VPWS-based VPN.

4.1 Assumptions

The service provider network is a packet switched network. The PEs are assumed to be logically connected with tunnels over which packets that belong to a service (such as VPWS) are encapsulated and forwarded. These tunnels can be IP tunnels, such as Generic Routing Encapsulation (GRE), or MPLS tunnels, established by Resource Reservation Protocol – Traffic Engineering (RSVP-TE) or Label Distribution Protocol (LDP). These tunnels are established independently of the services offered over them; the signaling used for the establishment of these tunnels is outside the scope of this document.

4.2 Functional Model

See section 2 of RFC 4761.

5 Layer 2 VPNs

A VPWS is a provider provisioned VPN where the service provider (SP) participates in management and provisioning of the VPN. The following sections describes advantages of BGP based VPWS and addresses some of the requirements.

5.1 VPN Provisioning

Provisioning is an important factor for service providers when deploying a new service. Use of a common protocol to offer multiple services is very advantageous to service providers as it reduces operational complexity involved in deploying and managing multiple protocols. Some service providers prefer the BGP protocol to offer both L2 and L3 VPN services. For example, to service providers who already offer BGP-based VPLS service, BGP based VPWS could be the preferred choice as both use BGP for auto-discovery and signaling and thus, both services could be deployed using the same provisioning methodologies.

5.2 Configuration

The configuration required to deploy a service is another important factor for service providers as it impacts the manageability of the service. BGP VPWS uses auto-discovery to discover the service endpoints (ACs from CEs terminating on PE routers) where VPWS pseudowires terminate. This reduces the configuration required on each PE as it eliminates the need to manually configure the list of remote PEs and the identifiers of the ACs terminating on the remote PEs. In addition, when additional CEs are added in the network to participate in the VPWS, additional changes on existing CEs and PEs already participating in the VPWS service can be avoided by over-provisioning. Consider a case where 20 DLCIs are associated with a VPWS when only 10 are needed. With this over-provisioning, adding a new CE to a VPWS requires configuring just the new CE and its associated PE; existing CEs and their PEs require no change in configuration. Note that if DLCIs at the CE edge are expensive, e.g. if these DLCIs are provisioned across a switched network, one could provision them as and when needed

5.3 Separation of Administrative Responsibilities

BGP VPWS provides control plane separation in the following three cases.

1. Customer and service provider domain

2. Multiple domains within a single service provider

3. Multiple domains across multiple service providers

In VPWS, the service provider is responsible for Layer 2 connectivity and the customer is responsible for Layer 3 connectivity, which includes routing. If the customer says that host x in site A cannot reach host y in site B, the service provider need only demonstrate that site A is connected to site B. The details of how routes for host y reach host x are the customer's responsibility. Any instability in the customer’s control plane will have no affect on the service providers’ network due to lack of customer’s control plane state on service provider PE routers.

Use of BGP to administer services across domains belonging to a single service provider or to multiple service providers is well established. Section 11 describes inter-domain VPWS operations.

5.4 PE Scaling

In the BGP VPWS, as described in this specification, each PE only needs to transmit information about CEs that are connected to that PE to every other PE. This means that both the Forwarding Information Base and the Routing Information Base of the PE scale well with the number of sites and number of VPWS. Furthermore, the scaling properties are independent of the customer’s Layer 3 routing: because it is a layer 2 service and the only factors impacting PSN routing are the total number of VPWS and the number of sites within each VPWS.

Use of BGP for carrying L2VPN NLRIs to set up pseudowires is expected to have insignificant impact on other services that use BGP in the control plane, such as L3VPN. The number of VPWS prefixes carried in BGP is expected to be a few tens of thousand for large scale deployment, which is still relatively low compared to the number of prefixes BGP carries for the Internet or for large scale L3VPN deployments. Note that the VPWS prefixes carry information about the endpoints of pseudowires only, and thus are far lower in total number when compared to L3VPNs, where customer IPv4 or IPv6 prefixes are carried in BGP.

A service provider might want to prioritize BGP updates that belong to a particular service, such as L3VPN, over updates that belong to other services, such as VPWS. Such operation and others that relate to how BGP updates are processed on a PE are outside the scope of this document.

5.5 Class of service

Providing different Classes of Service is an important offering to service providers. Depending on the attachment circuit type (Ethernet, ATM, etc), it may or may not be possible to provide Class of Service. As long as the attachment circuit type has the capability to carry CoS mappings, it is possible to map customer traffic to different Classes of Service that a provider offers. For example, for Ethernet attachment circuits, mapping the 802.1p bits in a customer’s Ethernet frame can be used to map to a provider VLAN tag and to an appropriate setting of EXP bits in the PSN tunnel label. If the customer traffic is known to be IP only, then the class of service carried in the IP header can be used to map to provider service mappings.

6 Control plane

There are two primary functions of the VPWS control plane: auto-discovery, and setup and teardown of the pseudowires that constitute the VPWS, often called signaling. Section 3 of RFC 4761 describes these two functions for VPLS. In general, the same procedures defined in Section 3 of RFC 4761 for VPLS are applicable to BGP VPWS. Note that the references to RFC 4761 should replace VPLS with VPWS.

Section 6.1 and Section 6.2 describe the two control plane functions: auto-discovery and signaling. The two functions use Layer 2 Info Extended Community and L2VPN NLRI defined in RFC 4761. Details on how to use these fields are provided in Section 8. In particular, it should be noted that Section 8.5 extends the Encaps Type code values from the signal value defined in Section 3.2.4 of RFC 4761.

6.1 Auto-Discovery

Discovery refers to the process of finding all the PEs that participates in a given VPWS domain. A PE either can be configured with the identities of all the other PEs in a given VPWS domain or can use some protocol to discover the other PEs. The latter is called auto-discovery.

For more details on auto-discovery, refer to Section 3.1 in RFC 4761.

6.2 Signaling

Once discovery is completed, each pair of PEs in a VPWS must be able to establish (and tear down) pseudowires to each other, i.e., exchange (and withdraw) demultiplexors. This process is known as signaling. Signaling is also used to transmit certain characteristics of the pseudowires that a PE sets up for a given VPWS.

The following sections are required to implement the functions needed for signaling operations on VPWS pseudowires.

6.2.1 Label Blocks

Details are provided in Section 3.2.1 in RFC 4761.

6.2.2 L2VPN BGP NLRI

Details are provided in Section 3.2.2 in RFC 4761.

6.2.3 PW Setup and Teardown

Details are provided in Section 3.2.3 in RFC 4761.

6.2.4 Signaling PE Capabilities

Details are provided in Section 3.2.4 in RFC 4761.

7 VPWS Encapsulation

PEs set up PWs on behalf of the CEs to enable the CEs to communicate with each other. A PSN tunnel is required to carry the PW traffic. PEs provide the necessary encapsulation and decapsulation functionality to handle the customer traffic. The PW emulation provided in this specification follows the Psuedo Wire Emulation Edge-to-Edge (PWE3) architecture defined in RFC 3985.

This section describes the encapsulation supported for a BGP-based VPWS service. The VPWS encapsulation must comply with RFC 3031 and RFC 3032 [RFC 3032] for MPLS tunnels, and RFC 3985 for pseudowire emulation

7.1 TDM encapsulation

For TDM PWs, the encapsulation can either be structure-aware or structure-agnostic.

Structure-aware emulation is the transport of structured TDM taking at least some level of the structure into account. If structure-aware encapsulation is required, the PE must use the encapsulation specified in RFC 5086 [RFC5086].

Structure-agnostic emulation is the transport of unstructured TDM, or of structured TDM when the structure is completely disregarded by the transport mechanism. It maintains the precise bit sequence of data and any structure overhead that may present. For structure-agnostic emulation, PEs must support the encapsulation specified in RFC 4553 [RFC4553]. The encapsulation supports the following TDM services: DS1, E1, DS3 and E3. Support of octet-aligned payload for structure-agnostic emulation of DS1 circuits is optional.

7.1.1 ATM encapsulation

For ATM encapsulation, PEs must use the N-to-1 mode of RFC 4717 [RFC4717]. This capability allows the encapsulation of several VCs or VPs on one PW which share an ATM class of service, which minimizes the number of PWs. PEs may also support the N-to-1 mode of RFC 4717 with a mapping of a complete ATM port to each PW, as specified in RFC 4816 [RFC4816].

7.1.2 Ethernet encapsulation

For Ethernet encapsulation, PEs must use RFC 4448 [RFC4448] and support Ethernet Raw and Tagged Mode.

7.1.3 Frame Relay encapsulation

For Frame Relay encapsulation, PEs must use RFC 4619 [RFC4619].

7.1.4 HDLC/PPP encapsulation

For HDLC/PPP encapsulation, PEs must use RFC 4618 [RFC4618].

8 PE Information Exchange

This section describes the parameters that are carried in the BGP message between the PEs for PW operations.

8.1 VE ID

A VE ID uniquely identifies a particular customer site. PEs are generally configured with unique VE IDs. The exception is multi-homing where a customer site is connected to multiple PEs for redundancy. In the case of multi-homing, each PE that is connected to the same customer site must configure the same VE ID. The VE ID is carried in the L2VPN NLRLI as described in Section 3.2.2 in RFC 4761.

8.2 VE ID Range

A VE ID range provides a means to overprovision a VPWS service. Typically, a VE ID will map to a connection identifier. Depending on the native Layer 2 connectivity, the connection identifier can be a VLAN in the case of Ethernet or a DLCI in the case of Frame Relay. To overprovision, a PE can configure more connection identifiers than the service being provided for so that later when the service is required to expand to more connection identifiers, no more provisioning will be required.

The VE ID range is carried in the L2VPN NLRI as described in Section 3.2.2 in RFC 4761. Each PE is required to allocate and advertise label blocks for all remote VE IDs that it receives from remote network peers. Thus, a PE can configure more VE IDs than required in order to overprovision.

8.3 Route Target

Route Target identifies a particular VPWS domain that consists of all the sites of a customer. Route Target is carried in Layer 2 Info Extended Community, as described in Section 3.2.4 of RFC 4761. PEs connected to sites belonging to the same customer must configure the same Route Target.

8.4 Route Distinguisher

Route Distinguisher (RD) is used to uniquely identify routes belonging to a particular VPWS domain on a PE. It is recommended that each PE configure a unique RD for each of its customer VPWS domains.

Route Distinguisher is carried in L2VPN NLRI as described in Section 3.2.2 in RFC 4761.

8.5 Encapsulation Type

The set of encapsulation types carried in the L2-info extended community has been expanded to include the following encapsulation types. The Encaps Type is single-octet. RFC 4761 defines value 19 for VPLS. The following Encaps Types are defined for VPWS.

Encaps Type is carried in Layer 2 Info Extended Community as described in Section 3.2.4 of RFC 4761. For Encaps Type values see Table 1 in section 3 of RFC 6624 [RFC6624].

Encaps Type	Description	Reference
TDM		
17	Structure-agnostic E1 over packet	RFC 4553
18	Structure-agnostic T1 (DS1) over packet	RFC 4553
40	Structure-agnostic E3 over packet	RFC 4553
20	Structure-agnostic T3 (DS3) over packet	RFC 4553
41 (Note 1)	Octet-aligned payload for Structure-agnostic DS1 circuits	RFC 4553
21	Nx64kbit/s Basic Service using Structure-aware	RFC 5086
42 (Note 2)	E1 Nx64kbit/s with CAS using Structure-aware	RFC 5086
43	DS1 (ESF) Nx64kbit/s with CAS using Structure-aware	RFC 5086
44	DS1 (ESF) Nx64kbit/s with CAS using Structure-aware	RFC 5086
ATM		
3	ATM transparent cell transport	RFC 4816
9	ATM n-to-one VCC cell transport	RFC 4717
10	ATM n-to-one VPC cell transport	RFC 4717
Ethernet		
4	Ethernet Tagged Mode	RFC 4448
5	Ethernet Raw Mode	RFC 4448
Frame Relay		
15	Frame Relay Port mode	RFC 4619
25	Frame Relay DLCI	RFC 4619
Other		
6	HDLC	RFC 4618
7	PPP	RFC 4618
11	IP Layer2 Transport	RFC3032

Note 1: Allocation of separate code point for Encaps Type will eliminate the need for TDM payload size.
Note 2: Allocated separate code points for Encaps Type to specify the trunk framing (i.e, E1, T1 ESF or T1 SF) with CAS.

8.6 Layer 2 MTU

This specification requires that the Layer 2 MTU configured on all the access circuits connecting CEs to PEs in a VPWS domain be the same. This can be ensured by passing the configured Layer 2 MTU in the Layer2 info extended community when advertising label-blocks. On receiving label-blocks from remote PEs in a VPN, the MTU value carried in the Layer2-info extended community should be compared against the configured value for the VPN. If they do not match, then the label-block should be ignored.

The MTU on the Layer 2 access links must be chosen such that the size of the L2 frames plus the VPWS PW header does not exceed the MTU of the SP network. Layer 2 frames that exceed the MTU after encapsulation must be dropped.

MTU is carried in Layer 2 Info Extended Community as described in Section 3.2.4 of RFC 4761.

8.7 Interface Parameters

[RFC4447] defines extensions to LDP that are required to exchange service parameters for various Layer2 services (Ethernet, FR, ATM, TDM, HDLC etc.). This section describes how these parameters are used with pseudowires for BGP-based VPWS

8.7.1 TDM Interface Parameters

Control Protocol extensions for set up of TDM pseudowires in MPLS networks is described in IETF RFC 5287 [RFC5287]. This section specifies how the interface parameters for TDM pseudowires are provisioned.

1. TDM Payload Bytes

   The default payload size defined for the corresponding service (see [RFC 4553], [RFC 5086] must be used. It is the same for each direction of the emulated circuit.

2. RTP

   This parameter specifies whether the RTP header is to be used or not. RTP will be used only if both endpoints are configured to receive it. The default is not to use RTP.

   If RTP is used, PEs at either end of the pseudowire must be configured with the following parameters.

   - Differential timestamping Mode – If it is set, indicates that the PW endpoint use Differential timestamping mode in the packets sent.
   - Frequency – Frequency of the timestamping clock in units of 8 khz (e.g., a bit rate clock for an E1 circuit would be encoded as 256).
   - SSRC – indicates the value of the SSRC ID in the RTP header. Value 0 means that SSRC ID value check will not be used for detecting misconnections.

8.7.2 ATM Interface Parameters

The ATM PW-specific interface parameter defined in section 14 of RFC 4717 is "Maximum Number of concatenated ATM cells". This parameter specifies the maximum number of concatenated ATM cells that can be processed as a single PDU. An ingress PE transmitting concatenated cells on the PW can concatenate a number of cells up to the value of this parameter, but must not exceed it.

This parameter must be configured on both ends of the connection, with the same value for both directions of a specific PW.

8.7.3 Frame Relay Interface Parameters

The FR PW-specific interface parameter defined in section 7.9.1 of RFC 4619 is "Frame Relay Header Length". This parameter indicates the length of the FR header expressed in octets.

This parameter is not configured, the default value of 2 is assumed.

8.7.4 Ethernet Interface Parameters

The Ethernet PW-specific interface parameter defined in section 4.3 of RFC 4448 is “Requested VLAN ID”. This parameter is not required for Ethernet Raw mode.

9 Operation, Administration and Maintenance (OAM)

A VPWS pseudowire is bidirectional, and can be modeled as composed of two simplex connections going in opposite directions. A simplex connection in one direction consists of 3 segments: 1) the local access circuit between the source CE and the ingress PE, 2) the tunnel LSP between the ingress and egress PEs, and 3) the access circuit between the egress PE and the destination CE.

To monitor the status of a VPWS pseudowire, a PE needs to monitor the status of both simplex connections. Since it knows the status of its access circuit, and the status of the tunnel towards the remote PE, it can inform the remote PE of these two. Similarly, the remote PE can inform the local PE of the status of its access circuit to its local CE and the status of the tunnel to the local PE. Combining the local and the remote information, a PE can determine the status of a pseudowire.

Emulated services that have native OAM (e.g. ATM, Ethernet, etc) must be supported. Native service OAM is transported transparently over the corresponding PW as user data. For Ethernet, as an example, IEEE 802.1ag [802.1ag] continuity check messages between two maintenance end-points can be transported transparently as user data over corresponding PW.

For TDM PWs, indication of status of the TDM attachment circuits is carried in-band in the data plane.

For the MPLS LSP, PEs should support LSP Ping [RFC 4379], and may support BFD [BFD-Base] for the detection of defects. For the pseudowire, PEs should support VCCV-Ping as per RFC 5085 [RFC 5085] and VCC-BFD. For defect notification, PEs should support the mapping of attachment circuit OAM messages to the pseudowire. Note: These mechanisms were IETF work in progress at the time of publication of this specification.

9.1 Circuit status vector

A new sub-TLV, called circuit status vector, carries the status of the VPWS pseudowire between a pair of PEs.

The basic unit of advertisement in VPWS-based VPN for a given CE is a label-block. Each label within a label-block corresponds to a pseudowire. Each pseudowire corresponds to an AC. The local status information for all pseudowires corresponding to a label-block is advertised along with the NLRI for the label-block, using the status vector TLV.

The Type field of this TLV is value 1. The Length field of the TLV specifies the length of the value field in bits. The Value field of this TLV is a bit-vector, each bit of which indicates the status of the pseudowire associated with the corresponding label in the label-block. Bit position 0 corresponds to the pseudowire associated with the first label in the label block. A bit value of 0 indicates that the corresponding local circuit and the tunnel LSP to the remote PE are up, while a value of 1 indicates that either or both of them are down. The Value field is padded to the nearest octet boundary.

Any change in VPWS PW state is advertised to all BGP speakers participating in the same VPN. If PE A receives an L2VPN NLRI, it can determine the status of the corresponding pseudowire between its local CE ‘n’ and remote CE ‘m’ by looking at the appropriate bit in the received curcuit status vector.

10 Inter-domain operation

The auto-discovery and signaling functions are typically announced via I-BGP. This assumes that all customer sites are connected to PEs that are in a single Autonomous System (AS).

However, if PEs providing VPWS services are in different ASes, some mechanism is needed to connect customer sites connected to those PEs. Section 3.4 in [RFC4761] describes three methods (a, b and c) to exchange L2VPN advertisements among PEs that are across multiple AS. In method (a), VPWS advertisements do not cross AS boundaries, and thus, all operations described in this document are applicable as is for method (a). However, for both method (b) and (c), VPWS advertisements do cross AS boundaries. This section describes the VPWS operation in inter-AS method (b) and method (c).

10.1 Method (b): EBGP Redistribution of VPWS Information between ASes

Details are provided in Section 3.4.2 in RFC 4761.

10.2 Method (c): EBGP Redistribution of VPWS Information between ASes

Details are provided in Section 3.4.3 in RFC 4761.

Appendix I

Operation of a Layer 2 VPN
(Informative)

This appendix describes the operation of a layer 2 VPN through a simple example of a customer with 4 sites connected to 3 PE routers in a Service Provider network to illustrate the various aspects of the operation of a VPWS or Layer 2 VPN. For simplicity, it is assumed that a full-mesh topology is used.

In what follows, Frame Relay serves as the Layer 2 medium, and each CE has multiple DLCIs to its PE, each to connect to another CE in the VPN. If the Layer 2 medium were ATM, then each CE would have multiple VPI/VCIs to connect to other CEs. For PPP and Cisco HDLC, each CE would have multiple physical interfaces to connect to other CEs. In the case of IP-only Layer 2 interworking, each CE could have a mix of one or more of the above Layer 2 mediums to connect to another

I.1 Network Topology

Consider a Service Provider network with edge routers PE0, PE1, and PE2. Assume that PE0 and PE1 are IGP neighbors, and PE2 is more than one hop away from PE0.

Suppose that a customer C has 4 sites S0, S1, S2 and S3 that C want to connect via the service provider's network using Frame Relay. Site S0 has CE0 and CE1 both connected to PE0. Site S1 has CE2 connected to PE0. Site S2 has CE3 connected to PE1, and CE4 connected to PE2. Site S3 has CE5 connected to PE2 (See Figure 1 below.). Suppose further that C wants to "over-provision" each current site, in expectation that the number of sites will grow to at least 10 in the near future. However, CE4 is only provisioned with 9 DLCIs. (Note that the signaling mechanism discussed in Section 6.2 will allow a site to grow in terms of connectivity to other sites at a later point in time at the cost of additional signaling, i.e., over-provisioning is not a must but a recommendation).

Suppose finally that CE0 and CE2 have DLCIs 100 through 109 provisioned; CE1 and CE3 have DLCIs 200 through 209 provisioned; CE4 has DLCIs 107, 209, 265, 301, 414, 555, 654, 777 and 888 provisioned; and CE5 has DLCIs 417-426.

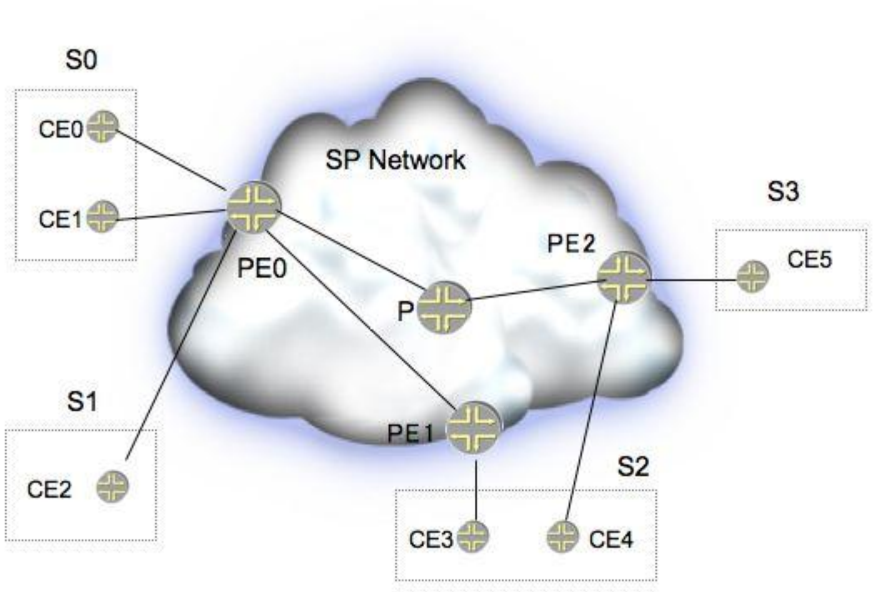

Figure 1: Example Network Topology

I.2 Configuration

The following sub-sections detail the configuration that is needed to provision the above VPN. For the purpose of exposition, assume that the customer will connect to the SP with Frame Relay circuits.

While focusing primarily on the configuration that an SP has to do, these sub-sections also touch upon the configuration requirements of CEs. Most importantly, the PEs and CEs must agree on the DLCIs that will be used on the interface connecting them.

If the PE-CE connection is Frame Relay, it is recommended that LMI be used between the PE and CE. For the case of ATM VCs, OAM cells may be used. For PPP and Cisco HDLC, keepalives may be used directly between CEs; however, in this case, PEs would not have visibility as to the state of the customer's circuits.

In the case of IP-only Layer 2 interworking, if CE1, attached to PE0, connects to CE3, attached to PE1, via a L2VPN circuit, the Layer 2 medium between CE1 and PE0 is independent of the Layer 2 medium between CE3 and PE1. Each side will run its own Layer 2-specific link management protocol, e.g., LMI, LCP, etc. PE0 will inform PE1 about the status of its local circuit to CE1 via the circuit status vector TLV defined in Section 9. Similarly PE1 will inform PE0 about the status of its local circuit to CE3.

I.2.1 CE Configuration

Each CE that belongs to a VPN is given a "CE ID". CE IDs must be unique in the context of a VPN. For the example, we assume that the CE ID for CE-k is k.

Each CE is configured to communicate with its corresponding PE with the set of DLCIs given above; for example, CE0 is configured with DLCIs 100 through 109. In general, a CE is configured with a list of circuits, all with the same Layer 2 technoly, e.g., VPWS, DLCIs, VCIs, physical PPP interface etc. (IP-only Layer 2 interworking allows a mix of Layer 2 encapsulation types). The size of this list/set determines the number of remote CEs a given CE can communicate with. Denote the size of this list/set as the CE's range. A CE's range must be at least the number of remote CEs that the CE will connect to in a given VPN; if the range exceeds this, then the CE is over-provisioned, in anticipation of growth of the VPN.

Each CE also "knows" which DLCI connects it to each other CE. The methodology followed in this example is to use the CE ID of the other CE as an index into the DLCI list this CE has (with zero-based indexing, i.e., 0 is the first index). For example, CE0 is connected to CE3 through its fourth DLCI, 103; CE4 is connected to CE2 by the third DLCI in its list, namely 265. This is just the methodology used in the example the actual methodology used to pick the DLCI to be used is a local matter; the key factor is that CE-k may communicate with CE-m using a different DLCI from the DLCI that CE-m uses to communicate with CE-k, i.e., the SP network effectively acts as a giant Frame Relay switch. This is very important, as it decouples the DLCIs used at each CE site, making for much simpler provisioning.

I.2.2 PE Configuration

Each PE is configured with the VPNs in which it participates. Each VPN is associated with one or more Route Target communities [RFC 4360] which serve to define the topology of the VPN. For each VPN, the PE must determine a Route Distinguisher (RD) to use; this may either be configured or chosen by the PE. RDs do not have to be unique across the VPN. For each CE attached to the PE in a given VPN, the PE must know the set of virtual circuits (DLCI, VCI/VPI or VPWS) connecting it to the CE, and a CE ID identifying the CE within the VPN. CE IDs must be unique in the context of a given VPN.

I.2.3 Adding a New Site

The first step in adding a new site to a VPN is to pick a new CE ID. If all current members of the VPN are over-provisioned, i.e., their range includes the new CE ID, adding the new site is a purely local task. Otherwise, the sites whose range doesn't include the new CE ID and that wish to communicate directly with the new CE must have their ranges increased by allocating additional local circuits to incorporate the new CE ID.

The next step is ensuring that the new site has the required connectivity. This usually requires the addition of a new virtual circuit between the PE and CE; in most cases, this configuration is limited to the PE in question.

The rest of the configuration is a local matter between the new CE and the PE to which it is attached.

It bears repeating that the key to making additions easy is over-provisioning and the algorithm for mapping a CE-id to a DLCI which is used for connecting to the corresponding CE. However, what is being over-provisioned is the number of DLCIs/VCIs that connect the CE to the PE. This is a local matter between the PE and CE, and does not affect other PEs or CEs.

END OF DOCUMENT