Point-to-Multipoint Pseudowire Signaling and Auto-Discovery in Layer 2 VPNs

draft-raggarwa-l2vpn-p2mp-pw-01.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

A Point-to-Multipoint (P2MP) Pseudo Wire (PW) is a mechanism that emulates the essential attributes of a unidirectional P2MP Telecommunications service such as P2MP ATM over a Packet Switched Network (PSN). [RFC4664] describes a number of different ways in which sets of PWs may be combined together into "Provider Provisioned Layer 2 VPNs" (L2 PPVPNs, or L2VPNs), resulting in a number of different kinds of L2VPN. P2MP PWs enable a L2VPN to provide a Virtual Private P2MP unidirectional service (VPMS), which may be in addition to the Virtual Private Wire Service (VPWS) offered by the
This document describes how procedures outlined in [VPLS-MCAST] can be used to signal P2MP PWs and enable a P2MP unidirectional service in a L2VPN.

Table of Contents

1 Specification of requirements .......................... 3
2 Introduction ........................................ 3
3 Layer 2 Multicast VPN ................................ 4
4 Overview ............................................ 5
4.1 P2MP PW Semantics .................................. 5
4.2 Auto-Discovery and Signaling ......................... 6
5 P2MP PW Demultiplexor Exchange ....................... 7
6 P2MP PW Encapsulation Type .......................... 7
7 Data Forwarding ...................................... 8
7.1 MPLS Tree Encapsulation ............................ 8
7.2 Layer 2 MTU ........................................ 9
8 Inter-AS and Multi-Segment P2MP PWs ................. 9
9 Security Considerations ............................... 9
10 IANA Considerations ................................ 9
11 Acknowledgments ..................................... 10
12 References ......................................... 10
12.1 Normative References .............................. 10
12.2 Informative References ............................. 10
13 Author’s Address .................................... 11
14 Intellectual Property Statement ....................... 12
15 Full Copyright Statement ............................ 12
1. Specification of requirements

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 [RFC2119].

2. Introduction

A Point-to-Multipoint (P2MP) Pseudo Wire (PW) is a mechanism that emulates the essential attributes of a unidirectional P2MP Telecommunications service such as P2MP ATM over a Packet Switched Network (PSN). One of the applicabilities of a P2MP PW is to deliver a Layer 2 multicast service, that carries multicast frames (encoded using Layer 2 or IP mechanisms) from a multicast source to one or more multicast receivers.

The required functions of P2MP PWs include encapsulating service-specific PDUs arriving at an ingress Attachment Circuit (AC), and carrying them across a tunnel to one or more egress ACs, managing their timing and order, and any other operations required to emulate the behavior and characteristics of the service as faithfully as possible.

P2MP PWs extend the PWE3 architecture [RFC3985] to offer a P2MP Telecommunications service. They follow the PWE3 architecture as described in [RFC3985] with modifications as outlined in this document. One notable difference between point-to-point (P2P) PWs as outlined in [RFC3985] and P2MP PWs is that the former emulate a bidirectional service whereas the latter emulate a unidirectional service.

[RFC4664] describes a number of different ways in which sets of PWs may be combined together into "Provider Provisioned Layer 2 VPNs" (L2 PPVPNs, or L2VPNs), resulting in a number of different kinds of L2VPN. P2MP PWs enable a L2VPN to provide a Virtual Private Multicast Service (VPMS), which may be in addition to the Virtual Private Wire Service (VPWS) offered by the L2VPN. A VPMS is a L2VPN service that provides point-to-multipoint connectivity traffic to customers.

This document describes how procedures outlined in [VPLS-MCAST] can be used to signal P2MP PWs and enable a P2MP unidirectional service in a L2VPN.
Layer 2 Multicast VPN

A VPMS "customer" is a customer of a Service Provider seeking to provide P2MP connectivity between its 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 PE (provider edge) router is termed the CE (customer edge) device; this device may be a router or a switch. A CE that is the 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.

Within each physical connection from a CE to a PE, there may be multiple ACs circuits.

A P2MP connection is rooted at a single CE, called the root CE (or ingress CE) and has one or more other CEs, called the leaf CEs (or egress CEs), as the leaves. The P2MP PW emulates the connectivity between the root CE and leaf CEs over the PSN.

A L2VPN that offers VPMS is referred to as a L2 Multicast VPN (L2 MVPN) in this document. Such a L2VPN is defined by two sets of sites, Sender Sites set and Receiver Sites set, with the following properties:

- CEs within the Sender Sites set could originate traffic for CEs in the Receiver Sites set. A PE delivers traffic received from a CE in the Sender Sites set to the CEs in the Receiver Sites set using a P2MP PW.

- CEs not in the Receiver Sites set should not be able to receive this traffic.

- CEs within the Receiver Sites set could receive traffic originated by any CEs in the Sender Sites set.

- CEs within the Receiver Sites set should not be able to receive traffic originated by any CE that is not in the Sender Sites set.

A site could be both in the Sender Sites set and Receiver Sites set,
which implies that CEs within such a site could both originate and receive multicast traffic. An extreme case is when the Sender Sites set is the same as the Receiver Sites set, in which case all sites could originate and receive multicast traffic from each other.

Sites within a given L2 MVPN may be either within the same, or in different organizations, which implies than an L2 MVPN can be either an Intranet or an Extranet.

A given site may be in more than one L2 MVPN, which implies that L2 MVPNs may overlap.

Not all sites of a given L2 MVPN have to be connected to the same service provider, which implies that an L2 MVPN can span multiple service providers.

Another way to look at a L2 MVPN is to say that an L2 MVPN is defined by a set of administrative policies. Such policies determine both Sender Sites set and Receiver Site set. Such policies are established by L2 MVPN customers, but implemented/realized by L2 MVPN Service Providers using the existing mechanisms, such as Route Targets, with extensions, as necessary.

Note that there may be local algorithms on the PEs to map a particular sender CE to a set of receiver CEs. Hence there is it is not necessary to configure on each receiver CE which CE it wishes to receive traffic from, since one receiver CE can receive traffic only from one sender CE.

4. Overview

4.1. P2MP PW Semantics

A P2MP PW provides a mechanism for the root CE to send traffic to one or more leaf CEs over a PSN.

A root CE in a sender site sends VPMS traffic on one or more ACs to the root PE. The root PE delivers this traffic over a P2MP PW to one or more leaf PEs. Each leaf PE in turn delivers this traffic to one or more leaf CEs in a receiver site. A particular leaf CE MUST receive this traffic over a single AC.

A particular leaf CE may receive traffic from multiple sender CEs. Traffic from different sender CEs is received by a leaf PE over unique P2MP PWS. The leaf PE must use unique ACs to send traffic received over unique P2MP PW, to the same leaf CE or different leaf CEs. This AC is determined by the leaf PE using local procedures.
which rely on the root CE identifier. For instance an AC may be configured with the root CE identifier it is expecting to receive traffic from. Or there may be an algorithmic mapping between the root CE identifier and the leaf AC.

### 4.2. Auto-Discovery and Signaling

A L2 MVPN requires auto-discovery procedures for the PEs in the Receiver Sites set to discover the PEs (and CEs) in the Sender Sites Set. Depending on the PSN Tunneling technology used the PEs in the Sender Sites set also may require discovering the PEs in the Receiver Sites set. Further a L2 MVPN requires signaling procedures for the root PE to signal P2MP PWs to leaf PEs.

Procedures outlined in [VPLS-MCAST] include the use of BGP for auto-discovery and the concepts of Route Distinguishers (RD) to make VPN advertisements unique, and Route Targets to control VPN topology. [VPLS-MCAST] builds on the mechanisms outlined in [L2VPN-DISC] and [RFC4761] to provide auto-discovery based on BGP. This document uses the procedures described in [VPLS-MCAST] for auto-discovery. The PE that advertises a locally attached CE generates a BGP NLRI that includes the RD and the local CE ID. Note that in [VPLS-MCAST] an equivalent advertisement carries the VE ID in the NLRI.

In addition, the documents mentioned above share the idea that routers not directly connected to VPN customers should carry no VPN state, restricting the provisioning of individual connections to just the edge devices. This is achieved by using P2MP PSN tunnels to carry the traffic, with a demultiplexor that identifies individual root CEs. This document defines a P2MP PW demultiplexor that can be used to identify individual root CEs and enables carrying traffic for multiple P2MP PWs (which may belong to different L2VPNs) over the same P2MP PSN tunnel. The specific approach taken here is to use an upstream assigned MPLS label [MPLS-UPSTREAM] as the P2MP PW demultiplexor. Section 5 describes how this demultiplexor is signaled using mechanisms outlined in [VPLS-MCAST].

The PSN tunnels could be based on P2MP MPLS LSPs signaled using RSVP-TE [RFC4875] or LDP [mLDP] or IP/GRE multicast trees signaled using PIM-SM or PIM-SSM. The signaling of these tunnels is outside the scope of this document.
5. P2MP PW Demultiplexor Exchange

Traffic belonging to different P2MP PWs, which may be in different L2VPNs, may be carried over the same P2MP PSN tunnel. Thus there is a need to identify at the leaf PE the P2MP PW the packet belongs to. This is done by using an inner label that determines to the P2MP PW for which the packet is intended. The ingress PE uses this label as the inner label while encapsulating a customer data packet. Each of the egress PEs must be able to associate this inner label with the same P2MP PW and use it to demultiplex the traffic received over the P2MP PSN tunnel. If downstream label assignment were used this would require all the egress PEs that are leaves of the P2MP PW to agree on a common label.

This problem is similar to the problem of identifying traffic for different VPLSs when Aggregate Trees are used in [VPLS-MCAST]. In that case, the inner label must identify the VPLS, while in the case of P2MP PWs, the inner label must identify the P2MP PW. This document reuses the procedures of [VPLS-MCAST] and requires the use of upstream label assignment by the ingress PE [MPLS-UPSTREAM]. Hence the inner label is assigned by the ingress PE. For details on the procedures, please refer to [VPLS-MCAST].

The ingress PE informs the egress PEs about the inner label as part of the tree binding procedures described in section 12 of [VPLS-MCAST] using the PMSI Tunnel Attribute. As described above the BGP NLRI carries the root CE ID.

6. P2MP PW Encapsulation Type

The set of encapsulation types carried in the L2-info extended community [RFC4761] has been expanded to include the following set. The encapsulation type identifies the Layer 1 or Layer 2 encapsulation, e.g., ATM, Frame Relay etc.

<table>
<thead>
<tr>
<th>Value</th>
<th>Encapsulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Reserved</td>
</tr>
<tr>
<td>1</td>
<td>Frame Relay</td>
</tr>
<tr>
<td>2</td>
<td>ATM AAL5 VCC transport</td>
</tr>
<tr>
<td>3</td>
<td>ATM transparent cell transport</td>
</tr>
<tr>
<td>4</td>
<td>Ethernet VLAN</td>
</tr>
<tr>
<td>5</td>
<td>Ethernet</td>
</tr>
<tr>
<td>6</td>
<td>Cisco-HDLC</td>
</tr>
<tr>
<td>7</td>
<td>PPP</td>
</tr>
<tr>
<td>8</td>
<td>CEM</td>
</tr>
<tr>
<td>9</td>
<td>ATM VCC cell transport</td>
</tr>
<tr>
<td>10</td>
<td>ATM VFC cell transport</td>
</tr>
</tbody>
</table>
7. Data Forwarding

7.1. MPLS Tree Encapsulation

The following diagram shows the progression of a L2 multicast packet as it enters and leaves the SP network when MPLS trees are being used. RSVP-TE P2MP LSPs are examples of such trees. The modification to the Layer 2 frame, by the ingress PE, depends on the Layer 2 encapsulation type. This document requires that the encapsulation methods used in transporting of Layer 2 frames over tunnels be the same as described in [RFC4448], [RFC4618], [RFC4619], and [RFC4717].

```
Packets received at ingress PE   Packets in transit in the service provider network   Packets forwarded by egress PEs
++==================++       ++==================++       ++==================++
||C-L2Hdr     ||       ||C-L2Hdr     ||       ||C-L2Hdr     ||
||C-Payload   ||       ||C-Payload   ||       ||C-Payload   ||
++==================++       ++==================++       ++==================++
```

The receiver PE does a lookup on the outer MPLS tree label and determines the MPLS forwarding table in which to lookup the inner MPLS label. This table is specific to the tree label space. The inner label is unique within the context of the root of the tree (as it is assigned by the root of the tree, without any coordination with any other nodes). Thus it is not unique across multiple roots. So, to unambiguously identify a particular P2MP PW one has to know the label, and the context within which that label is unique. The context is provided by the outer MPLS label [MPLS-UPSTREAM].

The outer MPLS label is stripped. The lookup of the resulting MPLS label determines the set of egress CE ACs that the packet needs to be forwarded to. The egress PE then strips the inner MPLS label and sends the packet to this set of egress CEs.
7.2. Layer 2 MTU

This document requires that the Layer 2 MTU configured on all the access circuits connecting CEs to PEs in a L2VPN be the same. This can be ensured by passing the configured Layer 2 MTU in the Layer2-info extended community [RFC4761]. On receiving the BGP NRLI advertisement 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 L2VPN. If they don’t match, then the BGP NLRI advertisement SHOULD be ignored.

The MTU on the Layer 2 access links MUST be chosen such that the size of the L2 frames plus the P2MP PW header does not exceed the MTU of the SP network. Layer 2 frames that exceed the MTU after encapsulation MUST be dropped.

8. Inter-AS and Multi-Segment P2MP PWs

This document supports all of the inter-AS methodologies described in [VPLS-MCAST] using the procedures of [VPLS-MCAST]. A Multi-Segment P2MP PW is equivalent to a segmented inter-AS tree that is described in [VPLS-MCAST], in the case of inter-AS option (b). A segment of an inter-AS segmented tree is equivalent to a segment of a Multi-Segment P2MP PW. A segmented inter-AS tree for a particular VPLS instance is formed by dynamically stitching intra-AS segments. The same procedures can be used to dynamically stitch segments of a Multi-Segment P2MP PW. Inter-AS segmented tree procedures of [VPLS-MCAST] MUST be used to build Multi-Segment P2MP PWs, by replacing the VE ID with the root CE-ID in the NLRI.

9. Security Considerations

TBD

10. IANA Considerations

IANA is requested to maintain a registry for the encaps type field of the Layer 2 Info Extended Community [RFC4761]. This document defines the following encapsulation types in addition to those defined in [RFC4761]. IANA is requested to add these values in the new registry:

<table>
<thead>
<tr>
<th>Value</th>
<th>Encapsulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Reserved</td>
</tr>
<tr>
<td>1</td>
<td>Frame Relay</td>
</tr>
<tr>
<td>2</td>
<td>ATM AAL5 VCC transport</td>
</tr>
</tbody>
</table>
Acknowledgments

Thanks to Yakov Rekhter and Kireeti Kompella for the discussions that lead to this document.

References

Normative References


Informative References

[VPMS-REQ] Y. Kamite, F. Jounay, "Framework and Requirements for Virtual Private Multicast Service (VPMS)", draft-kamite-l2vpn-vpms-frmwk-requirements-00.txt


13. Author’s Address

Rahul Aggarwal
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089
Email: rahul@juniper.net

Yuji Kamite NTT Communications Corporation Tokyo Opera City Tower
3-20-2 Nishi Shinjuku, Shinjuku-ku, Tokyo 163-1421, Japan
Email: y.kamite@ntt.com

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 at ietf-ipr@ietf.org.

15. Full Copyright Statement

Copyright (C) The IETF Trust (2008). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE,