Network Working Group
Internet Draft
Category: Standards Track
Expiration Date: January 2011

R. Aggarwal
Juniper Networks

Y. Kamite
NTT Communications

F. Jounay
France Telecom

July 12, 2010

BGP based Virtual Private Multicast Service Auto-Discovery and Signaling

draft-raggarwa-l2vpn-p2mp-pw-03.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and 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.

Copyright and License Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Abstract

A Point-to-Multipoint (P2MP) Pseudowire (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.

[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.

VPMS framework and requirements are described in [VPLS-REQ]. One of the VPMS requirements is auto-discovery. This document describes how procedures outlined in [VPLS-MCAST] can be used for auto-discovery (A-D) in VPMS using BGP. This document also describes BGP based procedures for P2MP PW signaling for VPMS that may be used when BGP is used for VPMS auto-discovery.

Table of Contents

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) Pseudowire (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. Encapsulation details and procedures of P2MP PWs are described in [P2MP-PW].

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 [P2MP-PW-REQ] and [P2MP-PW-ENCAP].

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.

VPMS framework and requirements are described in [VPLS-REQ]. One of the VPMS requirements is auto-discovery.

This document describes how procedures outlined in [VPLS-MCAST] can be used for auto-discovery (A-D) in VPMS using BGP. The BGP based A-D procedures also allow meeting other VPMS requirements such as redundancy.

This document also describes BGP based procedures for P2MP PW signaling for VPMS that may be used when BGP is used for VPMS auto-discovery.

The BGP based auto-discovery procedures that are specified in this document are meant to allow VPMS edge devices to discover each other even when the P2MP PW signaling protocol is a protocol other than BGP such as LDP [LDP-P2MP-PW]. However this version of the document does not provide all the details for that particular case.

3. 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.

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 have multiple connections 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 following the definition of sender site and receiver site in [VPMS-REQ]. These sites have the following properties:

- CEs within the Sender Sites set could originate traffic for CEs in the Receiver Sites set. A PE MUST deliver 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.

There may be multiple sender sites in a given L2 MVPN. On each PE that has a L2 MVPN instance, there may be multiple receiver sites in that instance. One possible policy may be for each receiver site to receive traffic from all the sender site. Another policy might be for a given receiver site to receive traffic only from a given sender site. To accomplish this there may be local algorithms on the PEs to map a particular sender CE to a set of receiver CEs. It is not necessary in this case to configure on each receiver CE which CE it wishes to receive traffic from.

4. Mapping Sender Attachment ACs to Receiver ACs

A P2MP PW provides a mechanism for the root CE to send traffic to one or more leaf CEs over a PSN. P2MP PW semantics are covered in [P2MP-PW-REQ] and P2MP PW encapsulation is described in [P2MP-PW-ENCAP].

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 may use unique ACs or the same AC to send traffic received over unique P2MP PW, to the same leaf CE. This AC is determined by the leaf PE using local procedures which rely on the policy in the L2 MVPN and may 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. Or the policy might be to send all the traffic that is received by a receiver PE in a L2 MVPN to all ACs that are in the receiver site set.

5. VPMS Auto-Discovery

As specified in [VPMS-REQ] a VPMS instance requires 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.

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 reuses the procedures described in [VPLS-MCAST] for auto-discovery with modifications described in this document.

The PE that advertises a locally attached VPMS CE MUST generate a BGP NLRI that includes the RD and the local CE ID <RD, CE ID>. Note that in [VPLS-MCAST] an equivalent advertisement carries the VE ID in the NLRI. The BGP A-D route MUST carry the set of Route Targets being exported by the VPMS instance.

The BGP based auto-discovery procedures that are specified in this document are meant to allow VPMS edge devices to discover each other even when the P2MP PW signaling protocol is a protocol other than BGP such as LDP [LDP-P2MP-PW]. However this version of the document does not provide all the details for that particular case. The details will be provided once [LDP-P2MP-PW] procedures mature.

As described in the section "Layer 2 Multicast VPN" the information about whether a CE belongs to a sender site or a receiver site is determined from the Route Targets (RT) that are configured to enforce the administrative policies of a L2 MVPN. These RTs are advertised in the corresponding BGP A-D routes. For instance if some of the sites in a VPMS are only in sender site set while others are only in receiver sites set, then CEs that are in the receiver site set are

configured to import only sender site set RTs. While CEs that are in the sender site set are configured to import only the receiver site set RTs. In this case two RTs are required to provision the VPMS instance.

5.1. Redundancy

[VPMS-REQ] describes requirements for redundancy that rely on multi-homing a sender CE to multiple PEs. The goal is to allow redundancy of the ingress PE. BGP based auto-discovery procedures allow each ingress PE that is part of the multi-homed PE set for a given sender CE to advertise a BGP NLRI for the CE. If the CE ID is configured to the be the same on all the ingress PEs, BGP path selection procedures ensure that only a given PE is chosen as the primary PE at a given time. In other words egress PEs receive traffic only from a given PE at a time for a multi-homed sender CE. It is a matter of local policy as to whether a) the other ingress PEs transmit traffic on the P2MP PW and the egress PEs drop this traffic or b) the other ingress PEs drop traffic that they receive from the sender CE.

6. VPMS P2MP PW Signaling

Documents mentioned above [VPLS-MCAST], [L2VPN-DISC], [RFC4761], 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 PWs to carry the traffic using the encapsulation described in [P2MP-PW-ENCAP]. A L2 MVPN requires signaling procedures for the root PE to signal P2MP PWs to leaf PEs.

As described in [P2MP-PW-ENCAP], upstream assigned MPLS labels are used as P2MP PW demultiplexors. This section describes how this demultiplexor is signaled using BGP based mechanisms outlined in [VPLS-MCAST]. Note that procedures in this section are not required if another mechanism or procedures are used for P2MP PW signaling. LDP based P2MP PW signaling [LDP-P2MP-PW] is one such mechanism. Even if that is the case BGP based A-D procedures as specified in this document MUST be used for VPMS auto-discovery.

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. As described in [P2MP-PW-ENCAP] this is done by using an upstream assigned MPLS label that determines the P2MP PW for which the packet is intended. The ingress PE MUST use this label as the bottom-most label while encapsulating a customer data packet.

The P2MP PW signaling problem is similar to the problem of identifying traffic fordifferent 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] to signal this label and the binding of the P2MP PW to the PSN Tunnel. For details on the procedures, please refer to [VPLS-MCAST].

The ingress PE MUST inform 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.

6.1. 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.

Value	Encapsulation
0	Reserved
1	Frame Relay
2	ATM AAL5 VCC transport
3	ATM transparent cell transport
4	Ethernet VLAN
5	Ethernet
6	Cisco-HDLC
7	PPP
8	CEM
9	ATM VCC cell transport
10	ATM VPC cell transport

7. Data Forwarding

Data forwarding follows the procedures specified in [P2MP-PW-ENCAP].

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] when the signaling procedures of this document are used along with the auto-discovery procedures of this document.

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:

Value	Encapsulation
0	Reserved
1	Frame Relay
2	ATM AAL5 VCC transport
3	ATM transparent cell transport
4	Ethernet VLAN
5	Ethernet
6	Cisco-HDLC
7	PPP
8	CEM
9	ATM VCC cell transport
10	ATM VPC cell transport

11. Acknowledgments

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

12. References

12.1. Normative References

[VPLS-MCAST] R. Aggarwal et. al., "Multicast in VPLS", draft-ietf-l2vpn-vpls-mcast-03.txt", November 2007

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[MPLS-UPSTREAM] R. Aggarwal, Y. Rekhter, E. Rosen, "MPLS Upstream Label Assignment and Context Specific Label Space", draft-ietf-mpls-upstream-label-00.txt

[P2MP-PW-ENCAP] R. Aggarwal., "Point-to-Multipoint PW Encapsulation", draft-raggarwa-pwe3-p2mp-pw-00.txt, work in progress

12.2. 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

[L2VPN-DISC] E. Rosen et. al., "Provisioning, Autodiscovery, and Signaling in L2VPNs", draft-ietf-l2vpn-signaling-08.txt

[RFC3985] S. Bryant et. al., "Pseudowire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, March 2005.

[RFC4664] L. Andersson etl. al,, "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, September 2006.

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC4875] R. Aggarwal et. al, "Extensions to RSVP-TE for Point to Multipoint TE LSPs", RFC4875

[LDP-P2MP-PW] F. Jounay et. al, "LDP Extensions for Source-initiated Point-to-Multipoint Pseudowire", draft-jounay-niger-pwe3-source-

initiated-p2mp-pw-03.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

Frederic Jounay
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex
FRANCE
Email: frederic.jounay@orange-ftgroup.com