

Network Working Group
Internet Draft
Expiration Date: January 2006

R. Aggarwal
Juniper Networks

Y. Kamite
NTT Communications

L. Fang
AT&T

July 2005

Multicast in VPLS

[draft-raggarwa-l2vpn-vpls-mcast-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

This document describes a solution for overcoming the limitations of existing VPLS multicast solutions. It describes procedures for VPLS multicast that utilize multicast trees in the service provider (SP) network. One such multicast tree can be shared between multiple VPLS instances. Procedures by which a single multicast tree in the backbone can be used to carry traffic belonging only to a specified

set of one or more multicast groups from one or more VPLSs are also described.

Table of Contents

1	Specification of requirements	4
2	Contributors	4
3	Terminology	4
4	Introduction	4
5	Existing Limitation of VPLS Multicast	5
6	Overview	5
7	VPLS Multicast / Broadcast / Unknown Unicast Data Packet Treatment	
7		
8	Propagating Multicast Control Information	7
9	Multicast Tree Leaf Discovery	8
9.1	Inclusive Tree Leaf Discovery	8
9.2	Selective Tree Leaf Discovery	8
10	Demultiplexing Multicast Tree Traffic	8
10.1	One Multicast Tree - One VPLS Mapping	8
10.1.1	One Multicast Tree - Many VPLS Mapping	8
11	Establishing Multicast Trees	9
11.1	RSVP-TE P2MP LSPs	10
11.1.1	P2MP TE LSP - VPLS Mapping	10
11.1.2	Demultiplexing C-Multicast Data Packets	10
11.2	Receiver Initiated MPLS Trees	10
11.2.1	P2MP LSP - VPLS Mapping	11
11.2.2	Demultiplexing C-Multicast Data Packets	11
11.3	PIM Based Trees	11
11.4	Encapsulation of the Aggregate Inclusive Tree and Aggregate Selective Tree	
11		
12	Tree to VPLS / C-Multicast Stream Binding Distribution	12
13	Switching to Aggregate Selective Trees	12
14	BGP Advertisements	13
14.1	Information Elements	13
14.1.1	Inclusive Tree - VPLS Binding Advertisement	13
14.1.2	Selective Tree - C-Multicast Stream Binding Advertisement .14	
14.1.3	Inclusive Tree/Selective Tree Identifier	14
14.2	Encoding	15
15	Aggregation Methodology	15
16	Data Forwarding	16
16.1	MPLS Tree Encapsulation	16
16.2	IP Tree Encapsulation	17
17	Security Considerations	18
18	Acknowledgments	18
19	Normative References	18
20	Informative References	19
21	Author Information	19
22	Intellectual Property Statement	19
23	Full Copyright Statement	20

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. Contributors

Rahul Aggarwal
Yakov Rekhter
Juniper Networks
Yuji Kamite
NTT Communications
Luyuan Fang
AT&T
Chaitanya Kodeboniya
Juniper Networks

3. Terminology

This document uses terminology described in [[VPLS-BGP](#)] and [VPLS-LDP].

4. Introduction

[VPLS-BGP] and [[VPLS-LDP](#)] describe a solution for VPLS multicast that relies on ingress replication. This solution has certain limitations for certain VPLS multicast traffic profiles. This document describes procedures for overcoming the limitations of existing VPLS multicast solutions.

It describes procedures for VPLS multicast that utilize multicast trees in the service provider (SP) network.

It provides mechanisms that allow a single multicast distribution tree in the backbone to carry all the multicast traffic from a specified set of one or more VPLSs. Such a tree is referred to as an "Inclusive Tree" and more specifically as an "Aggregate Inclusive Tree" when the tree is used to carry multicast traffic from more than VPLS.

This document also provides procedures by which a single multicast distribution tree in the backbone can be used to carry traffic belonging only to a specified set of one or more multicast groups,

from one or more VPLSs. Such a tree is referred to as a "Selective Tree" and more specifically as an "Aggregate Selective Tree" when the multicast groups belong to different VPLSs. So traffic from most multicast groups could be carried by an Inclusive Tree, while traffic from, e.g., high bandwidth groups could be carried in one of the "Selective Trees".

5. Existing Limitation of VPLS Multicast

VPLS multicast solutions described in [[VPLS-BGP](#)] and [[VPLS-LDP](#)] rely on ingress replication. Thus the ingress PE replicates the multicast packet for each egress PE and sends it to the egress PE using a unicast tunnel.

This is a reasonable model when the bandwidth of the multicast traffic is low or/and the number of replications performed on an average on each outgoing interface for a particular customer VPLS multicast packet is small. If this is not the case it is desirable to utilize multicast trees in the SP core to transmit VPLS multicast packets [[VPLS-MCAST-REQ](#)]. Note that unicast packets that are flooded to each of the egress PEs, before the ingress PE performs learning for those unicast packets, MAY still use ingress replication.

6. Overview

This document describes procedures for using multicast trees in the SP network to transport VPLS multicast data packets. RSVP-TE P2MP LSPs described in [[RSVP-P2MP](#)] are an example of such multicast trees. The use of multicast trees in the SP network can be beneficial when the bandwidth of the multicast traffic is high or when it is desirable to optimize the number of copies of a multicast packet transmitted by the ingress. This comes at a cost of state in the SP core to build multicast trees and overhead to maintain this state. This document places no restrictions on the protocols used to build SP multicast trees.

Multicast trees used for VPLS can be of two types:

1. Inclusive Trees. A single multicast distribution tree in the SP backbone is used to carry all the multicast traffic from a specified set of one or more VPLSs. These multicast distribution trees can be set up to carry the traffic of a single VPLS, or to carry the traffic of multiple VPLSs. The ability to carry the traffic of more than one VPLS on the same tree is termed 'Aggregation'. The tree will include every PE that is a member of any of the VPLSs that are using the tree. This enables the SP to place a bound on the amount of multicast routing state which the P routers must have. This implies that

a PE may receive multicast traffic for a multicast stream even if it doesn't have any receivers on the path of that stream.

2. Selective Trees. A Selective Tree is used by a PE to send multicast traffic for one or more multicast streams, that belong to the same or different VPLSs, to a subset of the PEs that belong to those VPLSs. Each of the PEs in the subset are on the path to a receiver of one or more multicast streams that are mapped onto the tree. The ability to use the same tree for multicast streams that belong to different VPLSs is termed 'Aggregation'. The reason for having Selective Trees is to provide a PE to have the ability to create separate SP multicast trees for high bandwidth multicast groups. This allows traffic for these multicast groups to reach only those PE routers that have receivers in these groups. This avoids flooding other PE routers in the VPLS.

A SP can use both Inclusive Trees and Selective Trees or either of them for a given VPLS on a PE, based on local configuration. Inclusive Trees can be used for both IP and non-IP data multicast traffic, while Selective Trees can be used only for IP multicast data traffic.

In order to establish Inclusive and Selective multicast trees the root of the tree must be able to discover the VPLS membership of all the PEs and/or the multicast groups that each PE has receivers in. This document describes procedures for doing this for Inclusive multicast trees. For discovering the IP multicast group membership procedures described in [[VPLS-CTRL](#)] are used. Procedures in [[VPLS-CTRL](#)] can also be used with ingress replication to send traffic for a multicast stream to only those PEs that are on the path to receivers for that stream. Aggregation also requires a mechanism for the egresses of the tree to demultiplex the multicast traffic received over the tree. This document describes how upstream label allocation by the root of the tree can be used to perform this demultiplexing. This document also describes procedures based on BGP that are used by the root of an Aggregate Tree to advertise the Inclusive or Selective tree binding and the demultiplexing information to the leaves of the tree

This document uses the prefix 'C' to refer to the customer control or data packets and 'P' to refer to the provider control or data packets.

7. VPLS Multicast / Broadcast / Unknown Unicast Data Packet Treatment

If the destination MAC address of a VPLS packet received by a PE from a VPLS site is a multicast address, a multicast tree SHOULD be used to transport the packet, if possible. If the packet is an IP multicast packet and a Selective tree exists for that multicast stream, the Selective tree SHOULD be used. Else if an Inclusive tree exists for the VPLS, it SHOULD be used.

If the destination MAC address of a VPLS packet is a broadcast address, it is flooded. If Inclusive tree is already established, PE floods over it. If Inclusive Tree cannot be used for some reason, PE MUST flood over multiple unicast PWs, based on [[VPLS-BGP](#)] [[VPLS-LDP](#)].

If the destination MAC address of the packet has not been learned, the flooding of the packet also occurs. Unlike broadcast case, it should be noted that when a PE learns the MAC it might immediately switch to transport over one particular PW. This implies that flooding unknown unicast traffic over Inclusive Tree might lead to packet reordering. This constraint should be taken into consideration if unknown unicast frames are flooded using a Inclusive Tree, instead of multiple unicast PWs based on [[VPLS-BGP](#)] [[VPLS-LDP](#)].

P-multicast trees are intended to be used only for VPLS C-multicast data packets, not for control packets being used by a customer's layer-2 and layer-3 control protocols. For instance, Bridge Protocol Data Units (BPDUs) use an IEEE assigned all bridges multicast MAC address, and OSPF uses OSPF routers multicast MAC address. P-multicast trees SHOULD NOT be used for transporting these control packets.

8. Propagating Multicast Control Information

PEs participating in VPLS need to learn the <C-S, C-G> information for two reasons:

1. With ingress replication, this allows a PE to send the IP multicast packet for a <C-S, C-G> only to other PEs in the VPLS instance, that have receivers interested in that particular <C-S, C-G>. This eliminates flooding.

2. It allows the construction of Aggregate Selective Trees.

Procedures for learning the <C-S, C-G> information are described in [[VPLS-CTRL](#)].

9. Multicast Tree Leaf Discovery

9.1. Inclusive Tree Leaf Discovery

VPLS auto-discovery as described in [[VPLS-BGP](#), [BGP-AUTO](#)] or another VPLS auto-discovery mechanism enables a PE to learn the VPLS membership of other PEs. This is used by the root of the Tree to learn the egresses of the tree.

9.2. Selective Tree Leaf Discovery

This is done using the C-Multicast control information propagation described in [[VPLS-CTRL](#)].

10. Demultiplexing Multicast Tree Traffic

Demultiplexing received VPLS traffic requires the receiving PE to determine the VPLS instance the packet belongs to. The egress PE can then perform a VPLS lookup to further forward the packet.

10.1. One Multicast Tree - One VPLS Mapping

When a multicast tree is mapped to only one VPLS, determining the tree on which the packet is received is sufficient to determine the VPLS instance on which the packet is received. The tree is determined based on the tree encapsulation. If MPLS encapsulation is used, eg: RSVP-TE P2MP LSPs, the outer MPLS label is used to determine the tree. Penultimate-hop-popping MUST be disabled on the RSVP-TE P2MP LSP.

10.1.1. One Multicast Tree - Many VPLS Mapping

As traffic belonging to multiple VPLSs can be carried over the same tree, there is a need to identify the VPLS the packet belongs to. This is done by using an inner label that corresponds to the VPLS for which the packet is intended. The ingress PE uses this label as the inner label while encapsulating a customer multicast data packet. Each of the egress PEs must be able to associate this inner label with the same VPLS and use it to demultiplex the traffic received over the Aggregate Inclusive Tree or the Aggregate Selective Tree. If downstream label assignment were used this would require all the egress PEs in the VPLS to agree on a common label for the VPLS.

We propose a solution that uses upstream label assignment by the

ingress PE [[MPLS-UPSTREAM](#)]. Hence the inner label is allocated by the ingress PE. Each egress PE maintains a separate label space for every other PE. The egress PEs create a forwarding entry for the inner VPN label, allocated by the ingress PE, in this label space. Hence when the egress PE receives a packet over an Aggregate Tree, the Tree identifier specifies the label space to perform the inner label lookup. The same label space may be used for all P-multicast trees rooted at the same ingress PE, or an implementation may decide to use a separate label space for every P-multicast tree.

When PIM based IP/GRE trees are used the root PE source address and the tree P-group address identifies the tree interface. The label space corresponding to the tree interface is the label space to perform the inner label lookup in. A lookup in this label space identifies the VPLS in which the customer multicast lookup needs to be done.

If the tree uses MPLS encapsulation the outer MPLS label and the incoming interface provides the label space of the label beneath it. This assumes that penultimate-hop-popping is disabled. An example of this is RSVP-TE P2MP LSPs. The outer label and incoming interface effectively identifies the Tree interface [[MPLS-UPSTREAM](#), [MPLS-MCAST](#)].

The ingress PE informs the egress PEs about the inner label as part of the tree binding procedures described in [section 12](#).

[11. Establishing Multicast Trees](#)

This document does not place any restrictions on the multicast technology used to setup P-multicast trees. However specific procedures are specified currently only for RSVP-TE P2MP LSPs, LDP P2MP LSPs and PIM-SM and PIM-SSM based trees.

A P-multicast tree can be either a source tree or a shared tree. A source tree is used to carry traffic only for the VPLSs that exist locally on the root of the tree i.e. for which the root has local CEs. A shared tree on the other hand can be used to carry traffic belonging to VPLSs that exist on other PEs as well. For example a RP based PIM-SM Aggregate tree would be a shared tree. Another example of a shared tree is a RSVP-TE P2MP LSP. The shared tree root participates in VPLS auto-discovery. Each of the PEs transport the VPLS traffic to the shared tree root using ingress replication. The shared root splices the traffic onto the shared tree.

11.1. RSVP-TE P2MP LSPs

This section describes procedures that are specific to the usage of RSVP-TE P2MP LSPs for instantiating a tree. The RSVP-TE P2MP LSP can be either a source tree or a shared tree. Procedures in [RSVP-TE-P2MP] are used to signal the LSP. The LSP is signaled after the root of the LSP discovers the leaves. The egress PEs are discovered using the procedures described in [section 9](#). Aggregation as described in this document is supported.

11.1.1. P2MP TE LSP - VPLS Mapping

P2MP TE LSP to VPLS mapping can be learned at the egress PEs using BGP based advertisements of the P2MP TE LSP - VPLS mapping. They require that the root of the tree include the P2MP TE LSP identifier as the tunnel identifier in the BGP advertisements. This identifier contains the following information elements:

- The type of the tunnel is set to RSVP-TE P2MP LSP
- RSVP-TE P2MP LSP's SESSION Object
- Optionally RSVP-TE P2MP LSP's SENDER_TEMPLATE Object

11.1.2. Demultiplexing C-Multicast Data Packets

Demultiplexing the C-multicast data packets at the egress PE require that the PE be able to determine the P2MP TE LSP that the packets are received on. The egress PE needs to determine the P2MP LSP to determine the VPLS that the packet belongs to, as described in [section 10](#). To achieve this the LSP must be signaled with penultimate-hop-popping (PHP) off. This is because the egress PE needs to rely on the MPLS label, that it advertises to its upstream neighbor, to determine the P2MP LSP that a C-multicast data packet is received on.

11.2. Receiver Initiated MPLS Trees

Receiver initiated MPLS trees can also be used. An example of such trees are LDP setup P2MP MPLS Trees [[LDP-P2MP1](#), [LDP-P2MP2](#)].

The LDP P2MP LSP can be either a source tree or a shared tree. Procedures in [[LDP-P2MP1](#), [LDP-P2MP2](#)] are used to signal the LSP. The LSP is signaled after the root of the LSP discovers the leaves and once the leaves receive the LDP FEC for the tree from the root. The egress PEs are discovered using the procedures described in [section 9](#). Aggregation as described in this document is supported.

11.2.1. P2MP LSP - VPLS Mapping

P2MP LSP to VPLS mapping can be learned at the egress PEs using BGP based advertisements of the P2MP LSP - VPLS mapping. They require that the root of the tree include the P2MP LSP identifier as the tunnel identifier in the BGP advertisements. This identifier contains the following information elements:

- The type of the tunnel is set to LDP P2MP LSP
- LDP P2MP FEC which includes an identifier generated by the root.

Each egress PE "joins" the P2MP MPLS tree by sending LDP label mapping messages for the LDP P2MP FEC, that was learned in the BGP advertisement, using procedures described in [[LDP-P2MP1](#), [LDP-P2MP2](#)].

11.2.2. Demultiplexing C-Multicast Data Packets

This follows the same procedures described above for RSVP-TE P2MP LSPs.

11.3. PIM Based Trees

When PIM is used to setup multicast trees in the SP core the Aggregate Inclusive Tree may be a shared tree, rooted at the RP, or a shortest path tree. Aggregate Selective Tree is rooted at the PE that is connected to the multicast traffic source. The root of the Aggregate Tree or the Aggregate Selective Tree has to advertise the P-Group address chosen by it for the tree to the PEs that are leaves of the tree. These other PEs can then Join this tree. The announcement of this address is done as part of the tree binding procedures described in [section 12](#).

11.4. Encapsulation of the Aggregate Inclusive Tree and Aggregate Selective Tree

An Aggregate Inclusive Tree or an Aggregate Selective Tree may use an IP/GRE encapsulation or a MPLS encapsulation. The protocol type in the IP/GRE header in the former case and the protocol type in the data link header in the latter case are as described in [[MPLS-MCAST](#)].

12. Tree to VPLS / C-Multicast Stream Binding Distribution

Once a PE sets up an Aggregate Inclusive Tree or an Aggregate Selective Tree it needs to announce the customer multicast groups being mapped to this tree to other PEs in the network. This procedure is referred to as Inclusive Tree or Selective Tree binding distribution and is performed using BGP. For an Inclusive Tree this discovery implies announcing the mapping of all VPLSs mapped to the Inclusive Tree. The inner label allocated by the ingress PE for each VPLS is included. The Inclusive Tree Identifier is also included. For an Selective Tree this discovery implies announcing all the specific <C-Source, C-Group> entries mapped to this tree along with the Selective Tree Identifier. The inner label allocated for each <C-Source, C-Group> is included. The Selective Tree Identifier is also included.

An Inclusive Tree by definition maps to all the <C-Source, C-Group> entries belonging to all the VPLSs associated with the Inclusive Tree. An Selective Tree maps to the specific <C-Source, C-Group> associated with it.

When PIM or LDP is used to setup SP multicast trees, the egress PE also Joins the P-Group Address or the LDP P2MP FEC corresponding to the Inclusive or Selective tree. This results in setup of the receiver driven multicast tree with IP or MPLS encapsulation.

13. Switching to Aggregate Selective Trees

Selective Trees provide a PE the ability to create separate SP multicast trees for certain <C-S, C-G> entires. The source PE that originates the Selective Tree and the egress PEs have to switch to using the Selective Tree for the <C-S, C-G> entries that are mapped to it.

Once a source PE decides to setup an Selective Tree, it announces the mapping of the <C-S, C-G> entries that are mapped on the tree to the other PEs using BGP. Depending on the SP multicast technology used, this announcement may be done before or after setting up the Selective Tree. After the egress PEs receive the announcement they setup their forwarding path to receive traffic on the Selective Tree if they have one or more receivers interested in the <C-S, C-G> entries mapped to the tree. This implies setting up the demultiplexing forwarding entries based on the inner label as described earlier. The egress PEs may perform this switch to the Selective Tree once the advertisement from the ingress PE is received or wait for a preconfigured timer to do so.

A source PE may use one of two approaches to decide when to start transmitting data on the Selective tree. In the first approach once

the source PE sets up the Selective Tree, it starts sending multicast packets for <C-S, C-G> entries mapped to the tree on both that tree as well as on the Inclusive Tree. After some preconfigured timer the PE stops sending multicast packets for <C-S, C-G> entries mapped on the Selective Tree on the default tree. In the second approach a certain pre-configured delay after advertising the <C-S, C-G> entries mapped to an Selective Tree, the source PE begins to send traffic on the Selective Tree. At this point it stops to send traffic for the <C-S, C-G> entries, that are mapped on the Selective Tree, on the Inclusive Tree. This traffic is instead transmitted on the Selective Tree.

14. BGP Advertisements

The procedures required in this document use BGP for P-Tree - VPLS binding advertisements and P-Tree - Multicast stream binding advertisement. This section first describes the information that needs to be propagated in BGP for achieving the functional requirements. It then describes a suggested encoding.

14.1. Information Elements

14.1.1. Inclusive Tree - VPLS Binding Advertisement

The root of an Aggregate Inclusive Tree maps one or more VPLS instances to the Inclusive Tree. It announces this mapping in BGP. Along with the VPLS instances that are mapped to the Inclusive Tree, the Inclusive Tree identifier is also advertised in BGP.

The following information is required in BGP to advertise the VPLS instance that is mapped to the Inclusive Tree:

1. The address of the router that is the root of the Inclusive Tree.
2. The inner label allocated by the Inclusive Tree root for the VPLS instance. The usage of this label is described in [section 10](#).

When a PE distributes this information via BGP, it must include the following:

1. An identifier of the Inclusive Tree.
2. Route Target Extended Communities attribute. This RT must be an "Import RT" of each VSI in the VPLS. The BGP distribution procedures used by [[VPLS-BGP](#)] or [[BGP-AUTO](#)] will then ensure that the advertised information gets associated with the right VSIs.

14.1.2. Selective Tree - C-Multicast Stream Binding Advertisement

The root of an Aggregate Selective Tree maps one or more <C-Source, C-Group> entries to the tree. These entries are advertised in BGP along with the the Selective Tree identifier to which these entries are mapped.

The following information is required in BGP to advertise the <C-Source, C-Group> entries that are mapped to the Selective Tree:

1. The RD configured for the VPLS instance. This is required to uniquely identify the <C-Source, C-Group> as the addresses could overlap between different VPLS instances.
2. The inner label allocated by the Selective Tree root for the <C-Source, C-Group>. The usage of this label is described in [section 10](#).
3. The C-Source address. This address can be a prefix in order to allow a range of C-Source addresses to be mapped to the Selective Tree.
4. The C-Group address. This address can be a range in order to allow a range of C-Group addresses to be mapped to the Selective Tree.

When a PE distributes this information via BGP, it must include the following:

1. An identifier of the Selective Tree.
2. Route Target Extended Communities attribute. This is used as described in [section 14.1.1](#).

14.1.3. Inclusive Tree/Selective Tree Identifier

Inclusive Tree and Selective Tree advertisements carry the Tree identifier. The following information elements are needed in this identifier.

1. Whether this is a shared Inclusive Tree or not.
2. The type of the tree. For example the tree may use PIM-SM or PIM-SSM.
3. The identifier of the tree. For trees setup using PIM the identifier is a (S, G) value.

14.2. Encoding

Encoding details will be described later.

15. Aggregation Methodology

In general the heuristics used to decide which VPLS instances or <C-S, C-G> entries to aggregate is implementation dependent. It is also conceivable that offline tools can be used for this purpose. This section discusses some tradeoffs with respect to aggregation.

The "congruency" of aggregation is defined by the amount of overlap in the leaves of the client trees that are aggregated on a SP tree. For Aggregate Inclusive Trees the congruency depends on the overlap in the membership of the VPLSs that are aggregated on the Aggregate Inclusive Tree. If there is complete overlap aggregation is perfectly congruent. As the overlap between the VPLSs that are aggregated reduces, the congruency reduces.

If aggregation is done such that it is not perfectly congruent a PE may receive traffic for VPLSs to which it doesn't belong. As the amount of multicast traffic in these unwanted VPLSs increases aggregation becomes less optimal with respect to delivered traffic. Hence there is a tradeoff between reducing state and delivering unwanted traffic.

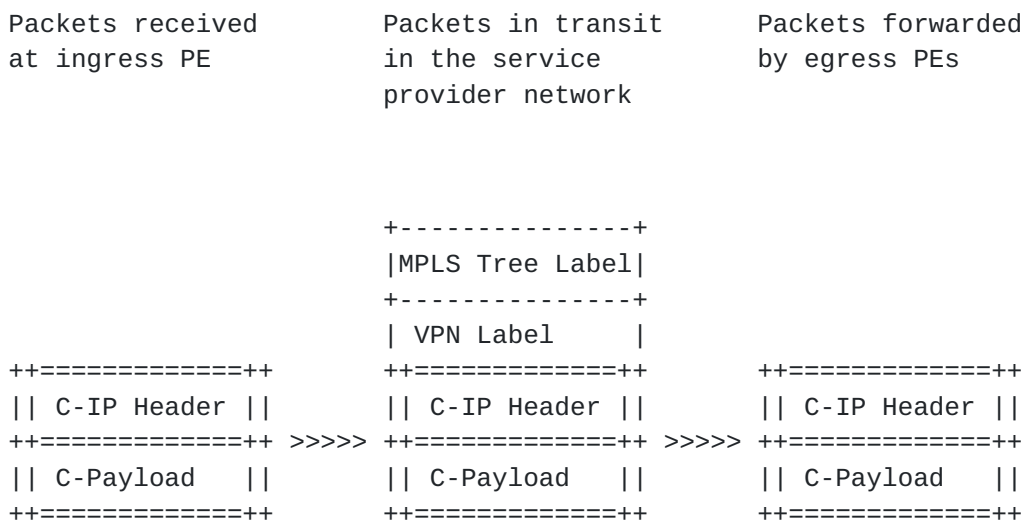
An implementation should provide knobs to control the congruency of aggregation. This will allow a SP to deploy aggregation depending on the VPLS membership and traffic profiles in its network. If different PEs or shared roots' are setting up Aggregate Inclusive Trees this will also allow a SP to engineer the maximum amount of unwanted VPLSs that a particular PE may receive traffic for.

The state/bandwidth optimality trade-off can be further improved by having a versatile many-to-many association between client trees and provider trees. Thus a VPLS can be mapped to multiple Aggregate Trees. The mechanisms for achieving this are for further study. Also it may be possible to use both ingress replication and an Aggregate Tree for a particular VPLS. Mechanisms for achieving this are also for further study.

16. Data Forwarding

16.1. MPLS Tree Encapsulation

The following diagram shows the progression of the VPLS IP multicast packet as it enters and leaves the SP network when MPLS trees are being used for multiple VPLS instances. RSVP-TE P2MP LSPs are examples of such trees.

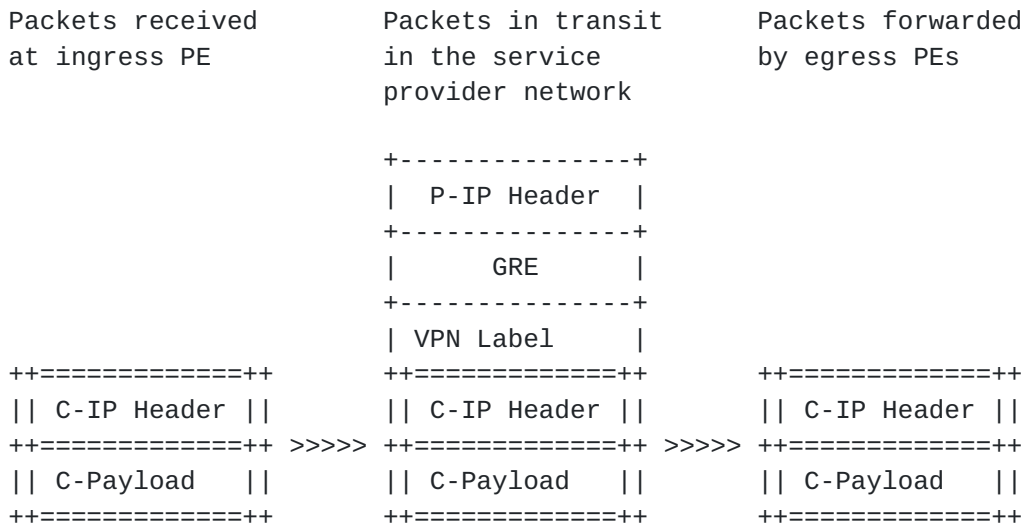


The receiver PE does a lookup on the outer MPLS tree label and determines the MPLS forwarding table in which to lookup the inner MPLS label. This table is specific to the tree label space. The inner label is unique within the context of the root of the tree (as it is assigned by the root of the tree, without any coordination with any other nodes). Thus it is not unique across multiple roots. So, to unambiguously identify a particular VPLS 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 VSI in which the receiver PE needs to do the C-multicast data packet lookup. It then strips the inner MPLS label and sends the packet to the VSI for multicast data forwarding.

16.2. IP Tree Encapsulation

The following diagram shows the progression of the packet as it enters and leaves the SP network when the Aggregate MDT or Aggregate Selective MDTs are being used for multiple VPLS instances. MPLS-in-GRE [[MPLS-IP](#)] encapsulation is used to encapsulate the customer multicast packets.



The P-IP header contains the Aggregate Tree (or Aggregate Selective Tree) P-group address as the destination address and the root PE address as the source address. The receiver PE does a lookup on the P-IP header and determines the MPLS forwarding table in which to lookup the inner MPLS label. This table is specific to the Aggregate Tree (or Aggregate Selective 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 VPLS one has to know the label, and the context within which that label is unique. The context is provided by the P-IP header [[MPLS-UPSTREAM](#)].

The P-IP header and the GRE header is stripped. The lookup of the resulting MPLS label determines the VSI in which the receiver PE needs to do the C-multicast data packet lookup. It then strips the inner MPLS label and sends the packet to the VSI for multicast data forwarding.

17. Security Considerations

Security considerations discussed in [[VPLS-BGP](#)] and [[VPLS-LDP](#)] apply to this document.

18. Acknowledgments

Many thanks to Thomas Morin for his support of this work.

19. Normative References

[RFC2119] "Key words for use in RFCs to Indicate Requirement Levels.", Bradner, March 1997

[RFC3107] Y. Rekhter, E. Rosen, "Carrying Label Information in BGP-4", [RFC3107](#).

[VPLS-BGP] K. Kompella, Y. Rekhter, "Virtual Private LAN Service", [draft-ietf-l2vpn-vpls-bgp-02.txt](#)

[VPLS-LDP] M. Lasserre, V. Kompella, "Virtual Private LAN Services over MPLS", [draft-ietf-l2vpn-vpls-ldp-03.txt](#)

[MPLS-IP] T. Worster, Y. Rekhter, E. Rosen, "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", [draft-ietf-mpls-in-ip-or-gre-08.txt](#)

[BGP-AUTO] H. Ould-Brahim et al., "Using BGP as an Auto-Discovery Mechanism for Layer-3 and Layer-2 VPNs", [draft-ietf-l3vpn-bgpvpn-auto-04.txt](#)

[VPLS-CTRL] R. Aggarwal, Y. Kamite, L. Fang, "Propagation of VPLS IP Multicast Group Membership Information", [draft-raggarwa-l2vpn-vpls-mcast-ctrl-00.txt](#)

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

[MPLS-MCAST] T. Eckert, E. Rosen, R. Aggarwal, Y. Rekhter, "MPLS Multicast Encapsulations", [draft-rosen-mpls-multicast-encaps-00.txt](#)

[VPLS-MCAST-REQ] Y. kamite, et. al., "Requirements for Multicast Support in Virtual Private LAN Services", [draft-kamite-l2vpn-vpls-mcast-reqts-00.txt](#)

20. Informative References

[MVPN] E. Rosen, R. Aggarwal, "Multicast in 2547 VPNs", [draft-ietf-l3vpn-2547bis-mcast-00.txt](#)"

[RSVP-P2MP] R. Aggarwal et. al, "Extensions to RSVP-TE for Point to Multipoint TE LSPs", [draft-ietf-mpls-rsvp-te-p2mp-02.txt](#)

[LDP-P2MP1] I. Minei et. al, "Label Distribution Protocol Extensions for Point-to-Multipoint Label Switched Paths", [draft-minei-mpls-ldp-p2mp-00.txt](#)

[LDP-P2MP2] I. Wijnands et. al., "Multicast Extensions for LDP", [draft-wijnands-mpls-ldp-mcast-ext-00.txt](#)

21. Author Information

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

Yakov Rekhter Juniper Networks 1194 North Mathilda Ave. Sunnyvale, CA 94089 Email: yakov@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

Luyuan Fang AT&T 200 Laurel Avenue, Room C2-3B35 Middletown, NJ 07748 Phone: 732-420-1921 Email: luyuanfang@att.com

Chaitanya Kodeboniya Juniper Networks 1194 North Mathilda Ave. Sunnyvale, CA 94089 Email: ck@juniper.net

22. Intellectual Property Statement

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.

[23. Full Copyright Statement](#)

Copyright (C) The Internet Society (2005). 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 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.

