MVPN: Optimized use of PIM, Wild Card Selectors, Bidirectional Tunnels, Extranets

draft-rosen-l3vpn-mvpn-mspmsi-02.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

Specifications for a number of important topics were arbitrarily omitted from the initial MVPN specifications, so that those specifications could be "frozen" and advanced. The current document provides some of the missing specifications. The topics covered are: (a) using Wild Card selectors to bind multicast data streams to tunnels, (b) using Multipoint-to-Multipoint Label Switched Paths as tunnels, (c) binding bidirectional customer multicast data streams to specific tunnels, (d) running PIM (i.e., sending and receiving multicast control traffic) over a set of tunnels that are created...
only if needed to carry multicast data traffic, and (e) extranets.

Table of Contents

1 Specification of requirements .......................... 3
2 Introduction ........................................ 3
2.1 Topics Covered ....................................... 3
2.2 Terminology .......................................... 4
3 Wild Card Selectors in S-PMSI A-D Routes ............... 5
4 Binding C-(*,G) to a Unidirectional P-Tunnel ........... 6
5 S-PMSI Procedures for Using MP2MP LSPs as P-tunnels ... 6
  5.1 General Procedures: MS-PMSIs .......................... 7
  5.2 Use of Multiple MP2MP LSPs .............................. 8
    5.2.1 Binding C-(S,G) ..................................... 8
    5.2.2 Binding C-(*,G) Flows from Unidirectional C-trees ... 9
    5.2.3 Binding C-(*,G) Flows from Bidirectional C-trees ..... 9
    5.2.4 Binding C-(*,*) ...................................... 10
    5.2.5 Default Tunnel Identifier ............................ 11
  5.3 Single MP2MP LSP ....................................... 12
6 PIM over MS-PMSI ....................................... 13
7 S-PMSI Join Extensions .................................. 14
  7.1 mLDP P2MP P-Tunnels .................................. 14
  7.2 IPv6 (S,G) with GRE P-tunnels .......................... 16
  7.3 Multiple S-PMSI Joins per Datagram ..................... 16
8 Extranets using PIM as the MVPN Control Plane ........... 16
  8.1 Default PMSI ........................................ 17
  8.2 Red method ........................................... 17
    8.2.1 Control Plane RPF Check ............................ 18
    8.2.2 Data Plane RPF Check ............................. 18
  8.3 Blue method .......................................... 18
  8.4 Binding Specific Extranet C-Flows to S-PMSIs .......... 19
9 IANA Considerations .................................... 19
10 Security Considerations .................................. 20
11 Authors’ Addresses ..................................... 20
12 Normative References ................................... 21
13 Informative References .................................. 21
14 Full Copyright Statement ................................ 21
15 Intellectual Property ................................... 22
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

The documents [MVPN] and [MVPN-BGP] contain specifications for a large number of MVPN topics. However, a number of important topics have been declared to be "out of scope" of those documents. This document provides the specifications for some of those topics. This document is not expected to be read as a stand-alone document; terminology from [MVPN] is used freely and knowledge of [MVPN] and [MVPN-BGP] is presupposed.

Any necessary procedures not explicitly specified here are as in [MVPN] and/or [MVPN-BGP].

2.1. Topics Covered

The topics covered in this document are the following:


  In [MVPN] and [MVPN-BGP], one can use an S-PMSI A-D route to assign a particular C-multicast flow, identified as C-(S,G), to a particular S-PMSI. The Wild Card Selectors specified in this document provide additional functionality:

  * One can send an S-PMSI A-D route whose semantics are "assign all the traffic traveling the C-(*,G) tree to this S-PMSI".

  * One can send an S-PMSI A-D route whose semantics are "use this S-PMSI as the default method for carrying any C-(S,G) traffic that isn’t assigned to a different S-PMSI". That is, it allows for the use of S-PMSIs as the default PMSIs for carrying data traffic.

- The use of Multipoint-to-Multipoint Label Switched Paths (MP2MP LSPs) as P-tunnels.

  A new kind of PMSI is defined, the MS-PMSI. An S-PMSI is defined in [MVPN] to have a single PE as its transmitter. An MS-PMSI is a set of S-PMSIs which together are instantiated by a single MP2MP LSP. This allows one to create P-tunnels which contain
only a subset of the PEs attached to a given VPN, but which can be used by any member of that subset to transmit to the other members of the subset. MS-PMSIs are advertised using the S-PMSI A-D routes of [MVPN] and [MVPN-BGP].

- PIM over MS-PMSI.

[MVPN] specifies how to run PIM [PIM] as the multicast routing protocol of a particular MVPN, by running it over an MI-PMSI for that MVPN. In this specification, we show how to run PIM over an MS-PMSI. A potential disadvantage of running PIM over MI-PMSI is that it can result in the creation of P-tunnels that only carry PIM messages, but do not carry multicast data. However, when PIM is run over an MS-PMSI, there is never any need to create a P-tunnel just for control messages; the only P-tunnels needed are those which carry multicast data.

- MVPN Extranets with PIM Control Plane.

In an MVPN "extranet", the transmitter of a multicast traffic flow is in a different VPN than the receivers. Additional procedures are defined to determine how the traffic is associated with a particular MI-PMSI or MS-PMSI, and how the RPF checks are done.

2.2. Terminology

In the following, we will sometimes talk of a PE receiving traffic from a PMSI and then discarding it. If PIM is being used as the multicast control protocol between PEs, this always implies that the discarded traffic will not be seen by PIM on the receiving PE.

In the following, we will sometimes speak of an S-PMSI A-D route being "ignored". When we say the route is "ignored", we do not mean that it’s normal BGP processing is not done, but that the route is not considered when determining which P-tunnel to use when sending multicast data, and that the MPLS label values it conveys are not used. We will generally use "ignore" in quotes to indicate this meaning.
3. Wild Card Selectors in S-PMSI A-D Routes

As specified in [MVPN-BGP], an S-PMSI A-D route can be used to bind a specified C-multicast flow to a specified P-tunnel. The P-tunnel is identified in the PMSI Tunnel Attribute (PTA) of the A-D route. The C-multicast flow is identified as (C-S,C-G), and specified as part of the route’s NLRI. The NLRI is defined as follows:

```
+-----------------------------------+
|      RD   (8 octets)              |
+-----------------------------------+
| Multicast Source Length (1 octet) |
+-----------------------------------+
|  Multicast Source (Variable)      |
+-----------------------------------+
|  Multicast Group Length (1 octet) |
+-----------------------------------+
|  Multicast Group   (Variable)     |
+-----------------------------------+
|   Originating Router’s IP Addr    |
+-----------------------------------+
```

The Multicast Source field contains the C-S address, and the Multicast Group field contains the C-G address.

However, [MVPN-BGP] does not specify any means of encoding wild cards ("*", in multicast terminology) in the Source or Group fields.

This omission makes it difficult to provide optimized multicast routing for customers that use ASM ("Any Source Multicast") multcasts, in which flows may be traveling along "shared" C-trees. By "shared C-trees", we mean both the unidirectional "RPT trees" used in sparse mode, and the bidirectional trees used in BIDIR-PIM [BIDIR-PIM].

When a customer is using ASM multicast, it is useful to be able to select the set of flows that are traveling along a shared C-tree, and to bind that entire set of flows to a specified P-tunnel. Conceptually, we would like to have a way to express that we want (C-* , C-G) traffic bound to the specified P-tunnel.

Another useful feature would be a way of using an S-PMSI A-D route to say "by default, all multicast traffic (within a given VPN) that has not been bound to any other P-tunnel is bound to the specified P-tunnel". To do this, we need to have a way to express that we want (C-*, C-*) traffic bound to the P-tunnel.
This specification therefore establishes the following convention. The use of a zero length source or group field is to be interpreted as specifying a wild card value for the respective field. The following two combinations MUST BE supported:

- C-(*,*): Source Wildcard, Group Wildcard.

This specification does not provide support for the combination of a specified source and a group wildcard. A received S-PMSI A-D route specifying this combination will be "ignored".

4. Binding C-(*,G) to a Unidirectional P-Tunnel

Consider an S-PMSI A-D Route whose NLRI specifies C-(*,G), and which contains a PTA that specifies a unidirectional P-tunnel. The P-tunnel may be a P2MP LSP, or it may be a unidirectional PIM-created multicast distribution tree specified either as P-(*,G) or as P-(S,G).

If C-G is known to be an SSM group address, the S-PMSI A-D route is "ignored".

Otherwise, the semantics of this S-PMSI A-D route are the following: the originator of the S-PMSI A-D route is saying that if it receives, over a VRF interface, any traffic that is traveling on the C-(*,G) shared tree, it will transmit such traffic on the specified P-tunnel. Any PE interested in receiving such traffic from the originator MUST join that P-tunnel.

(A PE receiving C-(S,G) multicast traffic can always tell whether that traffic is traveling on a C-(*,G) shared tree by consulting its C-PIM state. Similarly, each PE in an MVPN, by virtue of running C-PIM, knows whether it is interested in receiving traffic from the C-(*,G) tree.)

5. S-PMSI Procedures for Using MP2MP LSPs as P-tunnels
5.1. General Procedures: MS-PMSIs

There are two methods for using MP2MP LSPs as P-tunnels. In one method, a single MP2MP LSP is used to connect n PE routers. In another method, multiple MP2MP LSPs are used. These two methods are considered separately. Which method is in use is a matter of provisioning.

In both cases, an S-PMSI A-D route whose PTA specifies a particular MP2MP LSP MUST be originated by the PE that is the root of the LSP. The Tunnel Identifier for the MP2MP LSP consists of an MLDP MP2MP FEC element [MLDP], which itself consists of an IP address of the originating PE router, followed by an "opaque value" identifying the MP2MP LSP in the context of that PE router. This opaque value may be configured or autogenerated, and there is no need for different PEs attached to a given MVPN to use the same opaque value. The IP address which appears in the tunnel identifier field of the PTA MUST be the same IP address that the PE uses for sending and receiving PIM control messages.

According to the definition of S-PMSI in [MVPN], only a single PE can transmit onto a given S-PMSI. However, an MP2MP LSP that contains n PEs can therefore be used to instantiate n S-PMSIs, each of which has a different PE as its transmitter. Each PE can use the tunnel to transmit data to the other n-1 PEs. Therefore when a MP2MP LSP is specified in the PTA of an S-PMSI A-D route, we consider that it implicitly advertises a number of S-PMSIs: one for the root PE, and one for each PE that receives and processes the route. We will call the latter S-PMSIs the "implicitly advertised reverse S-PMSIs" (or just "reverse S-PMSIs").

When an MP2MP LSP is specified in the PTA of an S-PMSI A-D route, we will use the term "MS-PMSI" to refer the set of S-PMSIs that (including the reverse S-PMSIs) that are advertised (explicitly or implicitly) in that A-D route. The PE that originated the S-PMSI A-D route is known as the "root" of the MS-PMSI. When PE1 is the root of an MS-PMSI, we will sometimes refer to the MS-PMSI as "PE1's MS-PMSI". (Of course, a given PE may be the root for more than one MS-PMSI, for the same or different MVPNs. Rules governing the association of an S-PMSI A-D route with a given MVPN are as specified in [MVPN] and [MVPN-BGP].)

If the PTA in the S-PMSI A-D route contains an MPLS label, then any PE that, as a result of having received that route, transmits a packet onto the MS-PMSI will first push that label onto the packet’s label stack. The interpretation of that label when the packet is received is as specified in [MVPN] and [MVPN-BGP]. The use of this label allows multiple VPNs to share a single MP2MP LSP.
An S-PMSI A-D route whose PTA specifies a MP2MP LSP MUST be "ignored" UNLESS it is originated by the root of the LSP. Any MPLS label specified in the PTA of an "ignored" route MUST be ignored. Any PE Distinguisher Labels specified in the "ignored" route MUST be ignored.

5.2. Use of Multiple MP2MP LSPs

In this method, each PE attached to a given MVPN is potentially the root of a distinct MP2MP LSP. Each such PE may originate an S-PMSI A-D route whose PTA specifies an MP2MP LSP for which the originating PE is the root. In effect, each such PE advertises an MS-PMSI. We will sometimes refer to the MS-PMSIs as "partitions", and to the PE that advertised it as the root of the MS-PMSI or the root of the partition. This notion is useful both in support for BIDIR-PIM C-multicast traffic and for running PIM over MS-PMSI. Details are given in later sections.

The procedures that follow presuppose when a packet is received from a MP2MP LSP, it can be associated with one or more VRFs, and processed in the context of that VRF or VRFs. If the PTA that specified the MP2MP LSP has no MPLS label, then all packets received from the LSP are associated with the same set of VRFs. If the PTA did specify a label, then received packets must will carry a label (beneath the label that identifies the LSP itself), and the label must be processed in order to determine the context.

5.2.1. Binding C-(S,G)

When PE1 advertises an S-PMSI A-D route that binds a C-(S,G) flow to a MP2MP LSP, the semantics are as follows. PE1 is stating that any C-(S,G) traffic that it needs to transmit to other PEs will be transmitted on the specified LSP. Any other PE that needs to receive such traffic from PE1 (i.e., any other PE that needs to receive C-(S,G) traffic and which has selected PE1 as the upstream PE for C-S) MUST join that LSP.

If a PE has joined the LSP, but does not need to receive the C-(S,G) traffic, or if it needs to receive C-(S,G) traffic but has not selected PE1 as the upstream PE for C-S, then the PE MUST discard any such received traffic. Please note that if PIM is being used as the multicast control protocol, traffic that is discarded will not be seen by PIM.
5.2.2. Binding C-(*,G) Flows from Unidirectional C-trees

When PE1 advertises an S-PMSI A-D route that binds C-(*,G) to a MP2MP LSP, where C-G is not an SSM group, and the C-(*,G) traffic is traveling on a unidirectional shared C-tree, the semantics are as follows. PE1 is stating that any traffic to C-G that is traveling the shared C-tree and which PE1 needs to transmit to other PEs will be transmitted on the specified LSP. Any other PE that needs to receive such traffic from PE1 (i.e., any other PE that needs to receive C-(*,G) traffic and which has selected PE1 as the upstream PE for the C-RP corresponding to the C-G group) MUST join that LSP.

If a PE has joined the LSP, but does not need to receive the C-(*,G) traffic, or if it needs to receive C-(*,G) traffic but has not selected PE1 as the upstream PE for the C-RP that corresponds to C-G, then the PE MUST discard any such received traffic. Please note that if PIM is being used as the multicast control protocol, traffic that is discarded will not be seen by PIM.

5.2.3. Binding C-(*,G) Flows from Bidirectional C-trees

When PE1 advertises an S-PMSI A-D route that binds C-(*,G) to a MP2MP LSP, where C-G is not an SSM group, and the C-(*,G) traffic is traveling on a bidirectional shared C-tree, the semantics are as follows:

- PE1 is stating that any traffic to C-G that it (PE1) needs to send downstream will be sent on the specified LSP (or equivalently, on the MS-PMSI that it instantiates by virtue of the A-D route having been sent).

- Any other PE that is interested in receiving C-(*,G) traffic MUST join the specified LSP (or equivalently, must become a member of the MS-PMSI).

- Any other PE, say PE2, that (a) has traffic to C-G to send upstream and (b) has selected PE1 as its upstream PE for the C-RPA corresponding to C-G, MUST join the specified LSP (become a member of the MS-PMSI), and MUST send such traffic on the specified LSP. (i.e., such traffic is bound to the MS-PMSI instantiated by the MP2MP LSP that is rooted at PE2.)

- If a PE, say PE3, has joined the specified LSP, but does not need to receive the C-(*,G) traffic, or has not selected PE1 as the upstream PE for the C-RPA corresponding to C-G, then PE3 MUST NOT send any C-(*,G) traffic on that LSP, and MUST discard any C-(*,G) traffic it received on that LSP.
These procedures implement, for S-PMSIs, the "partitioning" scheme described in section 11.2 of [MVPN], with each MS-PMSI being a "partition".

The specification given so far requires an S-PMSI A-D route to be sent for each C-(*,G) that is using a bidirectional C-tree. A more efficient method is given in the next section.

5.2.4. Binding C-(*,*)

When PE1 advertises an S-PMSI A-D route that binds C-(*,*) to a specified MP2MP LSP of which PE1 is the root, the semantics are as that the MP2MP LSP is to be used to carry C-multicast traffic in the following sets of cases:

1. If PE1 has C-(S,G) traffic that is traveling on a source-specific C-tree, and PE1 needs to transmit that data to one or more other PEs, and PE1 has not bound C-(S,G) or C-(*,G) to a different P-tunnel, then the C-(S,G) traffic is sent by PE1 on the specified MP2MP LSP.

2. If PE1 has C-(*,G) traffic that is traveling on a unidirectional shared C-tree, and PE1 needs to transmit that data to one or more other PEs, and PE1 has not bound C-(*,G) to a different P-tunnel, then the C-(*,G) traffic is sent by PE1 on the specified MP2MP LSP.

3. If PE1 has C-(*,G) traffic that is traveling on a bidirectional shared C-tree, and PE1 needs to transmit that data to one or more other PEs, and PE1 has not bound C-(*,G) to a different P-tunnel, then the C-(*,G) traffic is sent by PE1 on the specified MP2MP LSP.

4. Consider some other PE, PE2, that has received the S-PMSI A-D route from PE1. If PE2 has C-(*,G) traffic that is traveling on a bidirectional shared C-tree, and PE2 needs to transmit that traffic UPSTREAM, and PE2 has selected PE1 as the upstream PE for the C-RPA corresponding to C-G, and PE1 has not bound C-(*,G) to any other P-tunnel, then the C-(*,G) traffic is sent by by PE2 on the specified MP2MP LSP.

5. If a PE receives traffic from a particular MS-PMSI, and the traffic is traveling a unidirectional C-(*,G) or C-(S,G) tree, and the root of the MS-PMSI is not the PE’s selected upstream PE for the C-(*,G) or C-(S,G), the PE MUST discard the traffic.
6. If a PE receives traffic from a particular MS-PMSI, and the traffic is traveling a bidirectional C-(*,G) tree, and the PE’s selected upstream PE for the C-RPA corresponding to C-G is not the root of the MS-PMSI, then the PE MUST discard the traffic.

With respect to traffic traveling a bidirectional C-tree, these procedures implement, for S-PMSIs, the "partitioning" scheme described in section 11.2 of [MVPN], without the need to send an S-PMSI A-D route for each C-(*,G) that is using a bidirectional C-tree. Each PE becomes the root of an MS-PMSI, and binds the double wildcard selector to it. The MS-PMSIs serve as the "partitions". The MS-PMSI rooted at PE1 becomes the default MS-PMSI for all traffic that PE1 needs to send downstream to other PEs. It also becomes the default MS-PMSI for all traffic that other PEs need to send upstream, as long as those other PEs have selected PE1 as the upstream PE for the C-RPA corresponding to that traffic.

Note that other PEs SHOULD NOT join the specified MP2MP LSP unless they have a need to send or receive data over it. A PE knows when it needs to receive data by virtue of having certain multicast state in its C-PIM instance. With regard to multicast data traveling on a bidirectional C-(*,G) tree, a PE may not know whether it has to send data until such data actually arrives over a VRF interface; the PE may be on a "sender-only" branch. However, the PE in this case would have to know, through provisioning or some automatic procedure such as "Bootstrap Routing Protocol for PIM" (BSR) [BSR], the set of C-RPAs that are being used to support C-(*,G) traffic. For each C-RPA, the PE could join the MP2MP LSP advertised by its selected upstream PE for that C-RPA. Alternatively the PE could defer joining the LSP until it actually has data to send.

5.2.5. Default Tunnel Identifier

To identify a MP2MP LSP, the PMSI Tunnel Attribute of an S-PMSI A-D route contains an MP2MP FEC Element [mLDP] in its "Tunnel Identifier" field. This contains the IP address of the PE at the root of the LSP, as well as an "opaque value" which is unique at that PE. Each PMSI Tunnel is associated at its root PE with a particular VRF, and each VRF in a given PE has a unique default RD. Therefore one way to uniquely identify a MP2MP LSP is to use an MP2MP FEC Element whose Opaque Value length is 8 and whose Opaque Value value is the default RD of the associated VRF. This method of assigning a Tunnel Identifier MUST be the default method for any PMSI Tunnel which is bound to C-(*,*) traffic. Other methods MAY be available as well.

Note that if aggregation of multiple VPNs onto a single default MS-PMSI is not being supported, this method of assigning the Tunnel

Rosen, et al.                                                   [Page 11]
Identifier allows each PE to algorithmically determine the Tunnel Identifier that has been assigned by a particular upstream PE. A PE decides to join a particular MS-PMSI because it has chosen that MS-PMSI’s root as the upstream PE for a particular VPN-IP address. The RD of that VPN-IP address is the contents of the Opaque Value field of the corresponding MS-PMSI.

5.3. Single MP2MP LSP

When a single MP2MP LSP is used for a given VPN (rather than multiple MP2MP LSPs), the PE at the root of the LSP MUST advertise it in the PTA of an S-PMSI A-D root. The PE that is at the root of the LSP MUST include a "PE Distinguisher Labels" attribute in either in its I-PMSI A-D route, or in the S-PMSI A-D route containing the PTA that identifies the LSP. The PE MUST use the attribute to bind an upstream-assigned MPLS label to the IP address of each other PE that attaches to the same MVPN (as determined by the RTs of the A-D route). That is, the PE as the root of the LSP assigns a distinct label to each of the other PEs attaching to the same MVPN. This set of PEs is learned via the reception of I-PMSI A-D routes.

The procedures for using the single MP2MP LSP differ from the procedures for using a mesh of MP2MP LSPs only in the following way. Let PE1 be the root of the LSP. When a packet that is traveling on a unidirectional C-tree is transmitted on the LSP by a particular PE, say PE2, PE2 must push on the packet’s label stack the label that PE1 assigned to PE2 via the procedure above. When a packet that is traveling on a bidirectional C-tree is transmitted on the LSP by PE2, it must push on the packet’s label stack the label that PE1 assigned to PE3, where PE3 is the upstream PE that PE2 has selected for the C-RPA corresponding to C-G.

For unidirectional flows, this allows the transmitter to be identified, and for bidirectional flows, this allows the partition to be identified. Packets received from the wrong upstream PE or from the wrong partition MUST be discarded. (In effect, this is a case of LSP hierarchy, where the PE Distinguisher Labels represent the set of MP2MP LSPs described in previous sections, but those LSPs are all tunneled through a single MP2MP LSP.)

If the PTA identifying the MP2MP LSP contains an MPLS label, then that label shall appear in the label stack immediately preceding the label specified in the PE Distinguisher Labels attribute.
6. PIM over MS-PMSI

[MVPN] provides two alternative means of distributing C-multicast routing information: PIM or BGP. Procedures for running PIM over MI-PMSI are specified in that document. However, a number of efficiencies can be obtained by running PIM instead over an MS-PMSI, instantiated as a set of MP2MP LSPs. The procedures for this are as follows.

Each PE that attaches to a given MVPN MUST originate an Intra-AS I-PMSI A-D route that does NOT contain a PTA. Each such PE MUST also originate an S-PMSI A-D route whose PTA is a MP2MP LSP rooted at the originating PE. This S-PMSI A-D MUST bind the LSP to the "double wildcard" (*, *). The use of these LSPs for sending and receiving data traffic is as specified in the previous section. In effect, each PE in the MVPN has advertised an MS-PMSI for which it is the root.

If PE1 needs to direct a PIM Join/Prune message to PE2, PE1 MUST join the PE2’s MS-PMSI by joining the LSP advertised in PE2’s corresponding S-PMSI A-D route. The PIM J/P messages MUST be sent over that MS-PMSI.

If PE1 does not need to direct a PIM Join/Prune message to PE2, then PE1 SHOULD not join the LSP advertised in PE2’s S-PMSI A-D route, as PE1 will not be receiving any multicast data on that LSP.

Any PE that sends a PIM Join/Prune message on a given LSP is automatically considered to be a PIM adjacency of every PE that receives the message on that LSP. This implies that any PE receiving the LSP MUST accept a PIM Join/Prune message on that LSP from any other PE, even if the PE that transmitted the Join/Prune messages has not previously transmitted a PIM Hello. That is, the "adjacency relationship" does not depend on the reception of PIM Hellos.

PIM Hellos may still be useful for OAM purposes. Any PIM Hellos that PE1 sends MUST be sent on the LSP advertised in PE1’s S-PMSI A-D route above.

Standard PIM procedures are used, except for:

- The above change in the adjacency maintenance procedures.

- Changes in the "RPF determination" or "RPF checking" procedures as may be defined in [MVPN] or in subsequent sections of this document (such as section 8.2).

Note that the data handling procedures of the previous section will
prevent PIM from ever seeing any packets that come from the wrong transmitter or that are in the wrong partition; when such packets are received they are discarded, rather than being passed to PIM’s state machinery. As a result, such packets do not cause Asserts to be generated. Other standard PIM procedures, such as Join Suppression and Prune Override may come into play, however.

By running PIM over MS-PMSI instead of over MI-PMSI, one completely avoids the need to have PEs join P-tunnels that would carry only control messages. A PE need not ever join a particular a P-tunnel unless it either has data to send on it, or needs to receive data on it.

It is also possible to run PIM over MS-PMSI when a single MP2MP LSP is used. In that case, the PE at the root of the LSP MUST include a PE Distinguisher Labels attribute in its S-PMSI A-D route, and must assign a label to each of the other PEs that attach to the same MVPN. (This set is auto-discovered through the I-PMSI A-D routes.) When sending a PIM J/P packet, one must push onto its label stack the label identifying the PE to which the J/P packet is being directed. When receiving a PIM J/P packet, a PE discards any that are not carrying the PE distinguisher label that has been bound to its own IP address.

All other MVPN-specific PIM procedures are as specified in [MVPN].

7. S-PMSI Join Extensions

7.1. mLDP P2MP P-Tunnels

The S-PMSI Join message is defined in section 7.4.2.2 of [MVPN]. In this specification, we define the "type 2" and "type 3" S-PMSI Joins, which are used when the S-PMSI tunnel is a P2MP LSP created by mLDP, and the tunnel is to carry C-flows of, respectively, IPv4 or IPv6 multicast traffic.
Type (8 bits):
- 2 if C-Source and C-Group are IPv4 addresses,
- 3 if C-Source and C-Group are IPv6 addresses.

Length (16 bits): the total number of octets in the Type, Length, Reserved and Value fields combined, rounded up to the next multiple of 4, encoded as an unsigned binary integer.

Reserved (8 bits): This field SHOULD be zero when transmitted, and MUST be ignored when received.

C-Source: address of the traffic source in the VPN
- for type 2, a 32-bit IPv4 address
- for type 3, a 128-bit IPv6 address

C-Group: address of the traffic destination in the VPN
- for type 2, a 32-bit IPv4 address
- for type 3, a 128-bit IPv6 address

FEC Element: this variable length field is a P2MP FEC element, encoded as a TLV as specified in [MLDP].

Padding: 0–3 bytes, as needed for 32-bit alignment. The padding bytes SHOULD be zero on transmission and MUST be ignored on reception.
7.2. IPv6 (S,G) with GRE P-tunnels

MVPN defines the S-PMSI Join type used when assigning IPv4 (S,G) to a GRE P-tunnel. When assigning IPv6 (S,G) to a GRE P-tunnel, S-PMSI type 4 is used, and the C-Source and C-Group are IPv6 addresses.

7.3. Multiple S-PMSI Joins per Datagram

A single UDP datagram MAY carry multiple S-PMSI Joins, as many as can fit entirely within it. If there are multiple S-PMSI Joins in a UDP datagram, they MUST be of the same S-PMSI Join type. The end of the last S-PMSI Join (as determined by the S-PMSI Join length field) MUST coincide with the end of the UDP datagram, as determined by the UDP length field. When processing a received UDP datagram that contains one or more S-PMSI Joins, a router MUST be able to process all the S-PMSI Joins that fit into the datagram.

8. Extranets using PIM as the MVPN Control Plane

Suppose there are two VPNs. VPN1 consists of a set of VRFs, each of which has been configured with RT1 as its export and import Route Target. VPN2 consists of a set of VRFs, each of which has been configured with RT2 as its export and import Route Target. For convenience, we will use the term "blue" instead of "RT1" and the term "red" instead of "RT2". Thus we will call VPN1 the "blue VPN" and VPN2 the "red VPN". Similarly, the blue VPN consists of a number of "blue sites" containing "blue systems"; these sites are attached to PEs via VRF interfaces that are associated with "blue VRFs".

We want to create an MVPN extranet in which blue receivers can join multicast groups whose sources and/or RPs are red.

The first step is to ensure that the blue VRFs (or the subset of blue VRFs whose attached sites are allowed to receive multicasts from red sources) import routes to the red sources. This is done as follows:

- The red VRFs are configured so that the subset of red routes that are to be part of the extranet are exported with a seconds RT value (call it RT3), as well as with RT2. For convenience, we will call RT3 "violet".

- The blue VRFs are configured so that they import violet routes as well as blue routes.

There are two different methods of providing the extranets, which will shall call the "red method" and the "blue method". (Remember
that the red VPN contains the transmitter, and the blue VPN contains
the receivers.)

This document assumes that in the case of non-SSM extranet multicast
groups, the mapping between a group address and an RP is pre-
configured in the PEs.

This document does not provide support for bidirectional C-trees in
extranets.

8.1. Default PMSI

Some of the procedures subsequently specified in this section are
largely independent of whether PIM is used with (a) an MI-PMSI or (b)
with an MS-PMSI that has been bound to the double wildcard. We will
use the term "default PMSI" as a general term to mean either (a) or
(b), depending upon which technique is actually being used in a given
network.

8.2. Red method

In the "red method", extranet multicasts are carried by default in
the default PMSI of the red VPN, which we will of course call the
"red PMSI".

To use this method, blue VRFs must be configured to import "red" I-
PMSI A-D routes and red S-PMSI A-D routes. If MI-PMSIs are being
used, the blue VRFs must immediately join the P-tunnels specified in
the red I-PMSI A-D routes. If MS-PMSIs are being used, a blue VRF
need not join the MS-PMSI P-tunnel rooted at a particular PE unless a
PIM Join needs to be sent to that PE.

The PIM C-instance associated with a blue VRF will treat the red and
blue default PMSIs as two different PIM interfaces.

The blue VRFs must also be configured to "associate" violet unicast
routes with the red default PMSI. What this means is that the red
default PMSI will be considered to be the RPF interface for the
violet unicast routes. The RPF interface for the blue unicast routes
remains, as usual, the blue default PMSI.

All that remains to be specified is how the control plane and data
plane RPF checks are done. Apart from these MVPN-specific procedures
for the RPF check, ordinary PIM procedures are used.
8.2.1. Control Plane RPF Check

Suppose a PE receives a PIM Join(S,G) from a CE, over a VRF interface that is associated with a blue VRF. The PE does the RPF check for S by looking up S in the blue VRF. If the route matching S is a blue route (i.e., carries the blue RT but not the violet RT), then a Join is sent over the blue default PMSI. However, if the route matching S is a violet route (i.e., carries the violet RT), a Join is sent over the red default PMSI.

If the PE receives a PIM Join(*,G) from a CE, the RPF check is done against the address of the corresponding RP; otherwise the procedure is the same.

8.2.2. Data Plane RPF Check

Suppose a red default PMSI has been associated with a blue VRF, as specified above, and an (S,G) multicast data packet is received from the red default PMSI. Then S is looked up in the (blue) VRF. If it matches a violet route, the packet is forwarded normally. However, if it matches a blue route, the packet is discarded as having failed the RPF check.

This prevents the blue sites from receiving packets from red transmitters, except in the case where routes to the red receivers have been explicitly imported into the blue VRF.

8.3. Blue method

In the "blue method", extranet multicasts are carried by default in the default PMSI of the blue VPN.

In the blue method, the red VRFs must be configured to import "blue" I-PMSI and S-PMSI A-D routes. If MI-PMSIs are being used the P-tunnels specified therein must be joined immediately. If MS-PMSIs are being used, the P-tunnels need not be joined unless and until it is necessary to send a PIM Join to the root of the P-tunnel.

The PIM C-instance associated with a red VRF will treat the red default PMSI and the blue default PMSI as two different PIM interfaces.

PIM Joins from blue receivers are then received at the red VRF over the blue PMSI, whereas PIM Joins from red receivers are received at the red VRF over the red PMSI. As a result, PIM may add one or the other or both PMSIs to a particular multicast tree’s olist.
In this method, the blue VRFs are associated with only one default PMSI, so the RPF check for both blue and violet sources (and RPs) always resolves to that PMSI. Hence the special RPF check procedures of the red method are not necessary. However, a PE with a red VRF may need to transmit multicast traffic on more than one MI-PMSI.

Note that since the data plane RPF check of section 8.2.2 is not needed, one does not really need a "violet" RT value. Rather, one may simply configure certain routes from the red VRF to be exported with both the red and the blue RTs.

8.4. Binding Specific Extranet C-Flows to S-PMSIs

If the procedure of [MVPN] section 7.4.2 is used, the S-PMSI Join message MUST be sent on whatever default PMSI or default PMSIs are used to carry the C-flow identified in the message.

If the procedure of [MVPN] section 7.4.1 is used, then procedures differ slightly depending upon whether the red method or the blue method is in use.

If the red method is in use, and if a C-flow whose target source is exported from a red VRF is bound to an S-PMSI, then the S-PMSI A-D route that specifies the binding must carry both the red RT and the violet RT. Blue VRFs must be configured to import the violet S-PMSI A-D routes.

If the blue method is in use, and if a C-flow whose target source is exported from a red VRF is bound to an S-PMSI, then the S-PMSI A-D route that specifies the binding:

- must carry the red RT if the C-flow has any receivers on the red default PMSI, and

- must carry the blue RT if the C-flow has any receivers on the blue default PMSI.

9. IANA Considerations

[MVPN] creates an IANA registry for the "S-PMSI Join Message Type Field". This document requires three new values:

- The value 2 should be registered, and its description should read "mLDP P2MP S-PMSI for IPv4 traffic (unaggregated)".
- The value 3 should be registered, and its description should read "mLDP P2MP S-PMSI for IPv6 traffic (unaggregated)".
- The value 4 should be registered, and its description should read "GRE S-PMSI for IPv6 traffic (unaggregated)".

10. Security Considerations

There are no additional security considerations beyond those of [MVPN] and [MVPN-BGP].

11. Authors’ Addresses

Arjen Boers
Cisco Systems, Inc.
170 Tasman Drive
San Jose, CA, 95134
E-mail: aboers@cisco.com

Yiqun Cai
Cisco Systems, Inc.
170 Tasman Drive
San Jose, CA, 95134
E-mail: ycai@cisco.com

Eric C. Rosen
Cisco Systems, Inc.
1414 Massachusetts Avenue
Boxborough, MA, 01719
E-mail: erosen@cisco.com

IJsbrand Wijnands
Cisco Systems, Inc.
De kleetlaan 6a Diegem 1831
Belgium
E-mail: ice@cisco.com
12. Normative References


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

13. Informative References


14. 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.
15. Intellectual Property

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.