Internet Engineering Task Force (IETF)
Request for Comments: 6514
Category: Standards Track
ISSN: 2070-1721

R. Aggarwal
Juniper Networks
E. Rosen
Cisco Systems, Inc.
T. Morin
France Telecom - Orange
Y. Rekhter
Juniper Networks

February 2012

BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs

Abstract

This document describes the BGP encodings and procedures for exchanging the information elements required by Multicast in MPLS/BGP IP VPNs, as specified in RFC 6513.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6514.

Table of Contents

1. Introduction

This document describes the BGP encodings and procedures for exchanging the information elements required by Multicast in MPLS/BGP IP VPNs, as specified in [MVPN]. This document assumes a thorough familiarity with the procedures, concepts, and terms described in [MVPN].

This document defines a new Network Layer Reachability Information (NLRI), MCAST-VPN NLRI. The MCAST-VPN NLRI is used for MVPN auto-discovery, advertising MVPN to Inclusive P-Multicast Service Interface (I-PMSI) tunnel binding, advertising (C-S,C-G) to Selective PMSI (S-PMSI) tunnel binding, VPN customer multicast routing information exchange among Provider Edge routers (PEs), choosing a single forwarder PE, and for procedures in support of co-locating a Customer Rendezvous Point (C-RP) on a PE.

This document specifies two new BGP attributes: the P-Multicast Service Interface Tunnel (PMSI Tunnel) attribute and the PE Distinguisher Label attribute.

This document also defines two new BGP Extended Communities: the Source Autonomous System (AS) Extended Community and the VPN Routing and Forwarding (VRF) Route Import Extended Community.

2. 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].

3. Terminology

In the context of this document, we will refer to the MVPN auto-discovery/binding information carried in BGP as "auto-discovery routes" ("A-D routes"). For a given MVPN, there are the following types of A-D routes:

+ Intra-AS I-PMSI A-D route;

+ Inter-AS I-PMSI A-D route;

+ S-PMSI A-D route;

+ Leaf A-D route;

+ Source Active A-D route.

In the context of this document, we will refer to the MVPN customers' multicast routing information carried in BGP as "C-multicast routes". For a given MVPN, there are the following types of C-multicast routes:

+ Shared Tree Join route;

+ Source Tree Join route;

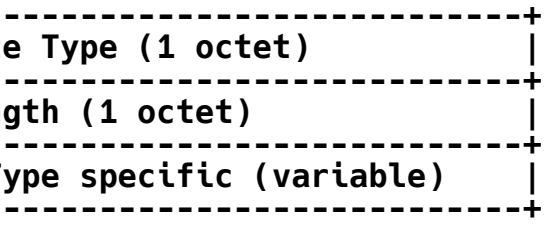
For each MVPN present on a PE, the PE maintains a Tree Information Base (MVPN-TIB). This is the same as TIB defined in [RFC4601], except that instead of a single TIB, a PE maintains multiple MVPN-TIBs: one per each MVPN.

Throughout this document, we will use the term "VPN-IP route" to mean a route that is either in the VPN-IPv4 address family [RFC4364] or in the VPN-IPv6 address family [RFC4659].

4. MCAST-VPN NLRI

This document defines a new BGP NLRI, called the MCAST-VPN NLRI.

The following is the format of the MCAST-VPN NLRI:



The Route Type field defines the encoding of the rest of MCAST-VPN NLRI (Route Type specific MCAST-VPN NLRI).

The Length field indicates the length in octets of the Route Type specific field of the MCAST-VPN NLRI.

This document defines the following Route Types for A-D routes:

+ 1 - Intra-AS I-PMSI A-D route;
+ 2 - Inter-AS I-PMSI A-D route;
+ 3 - S-PMSI A-D route;
+ 4 - Leaf A-D route;
+ 5 - Source Active A-D route.

This document defines the following Route Types for C-multicast routes:

+ 6 - Shared Tree Join route;
+ 7 - Source Tree Join route;

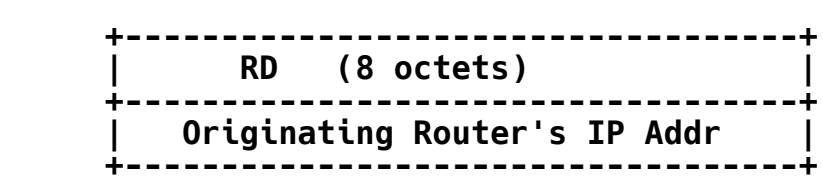
The MCAST-VPN NLRI is carried in BGP [RFC4271] using BGP Multiprotocol Extensions [RFC4760] with an Address Family Identifier (AFI) of 1 or 2 and a Subsequent AFI (SAFI) of MCAST-VPN. The NLRI field in the MP_REACH_NLRI/MP_UNREACH_NLRI attribute contains the MCAST-VPN NLRI (encoded as specified above). The value of the AFI field in the MP_REACH_NLRI/MP_UNREACH_NLRI attribute that carries the MCAST-VPN NLRI determines whether the multicast source and multicast group addresses carried in the S-PMSI A-D routes, Source Active A-D routes, and C-multicast routes are IPv4 or IPv6 addresses (AFI 1 indicates IPv4 addresses, AFI 2 indicates IPv6 addresses).

In order for two BGP speakers to exchange labeled MCAST-VPN NLRIs, they must use a BGP Capabilities Advertisement to ensure that they both are capable of properly processing such an NLRI. This is done as specified in [RFC4760], by using capability code 1 (multiprotocol BGP) with an AFI of 1 or 2 and an SAFI of MCAST-VPN.

The following describes the format of the Route Type specific MCAST-VPN NLRI for various Route Types defined in this document.

4.1. Intra-AS I-PMSI A-D Route

An Intra-AS I-PMSI A-D Route Type specific MCAST-VPN NLRI consists of the following:

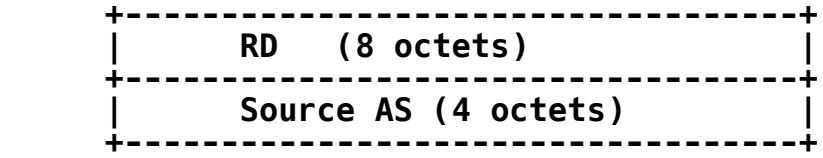


The Route Distinguisher (RD) is encoded as described in [RFC4364].

Usage of Intra-AS I-PMSI A-D routes is described in Section 9.2.

4.2. Inter-AS I-PMSI A-D Route

An Inter-AS I-PMSI A-D Route Type specific MCAST-VPN NLRI consists of the following:



The RD is encoded as described in [RFC4364].

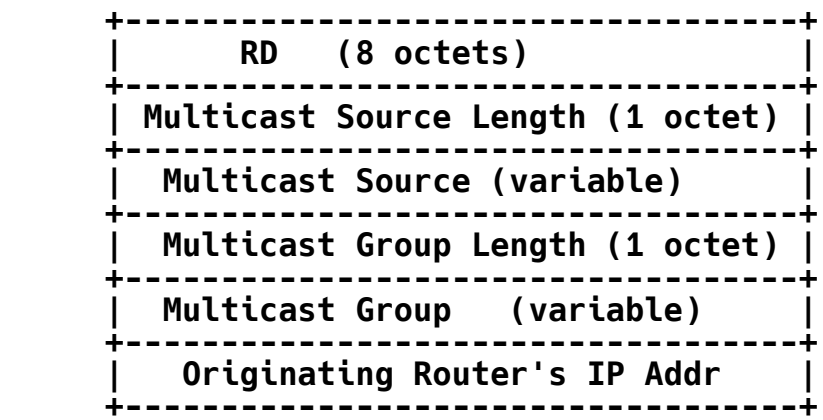
The Source AS contains an Autonomous System Number (ASN).

Two-octet ASNs are encoded in the two low-order octets of the Source AS field, with the two high-order octets set to zero.

Usage of Inter-AS I-PMSI A-D routes is described in Section 9.1.

4.3. S-PMSI A-D Route

An S-PMSI A-D Route Type specific MCAST-VPN NLRI consists of the following:



The RD is encoded as described in [RFC4364].

The Multicast Source field contains the C-S address. If the Multicast Source field contains an IPv4 address, then the value of the Multicast Source Length field is 32. If the Multicast Source field contains an IPv6 address, then the value of the Multicast Source Length field is 128.

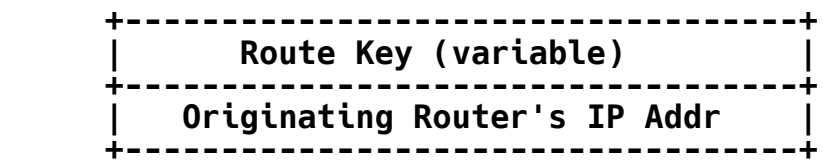
The Multicast Group field contains the C-G address or C-LDP (Label Distribution Protocol) MP Opaque Value Element (use of C-LDP MP Opaque Value Element is described in the Section 11.3.2. If the Multicast Group field contains an IPv4 address, then the value of the Multicast Group Length field is 32. If the Multicast Group field contains an IPv6 address, then the value of the Multicast Group Length field is 128.

Usage of other values of the Multicast Source Length and Multicast Group Length fields is outside the scope of this document.

Usage of S-PMSI A-D routes is described in Section 12.

4.4. Leaf A-D Route

A Leaf A-D Route Type specific MCAST-VPN NLRI consists of the following:



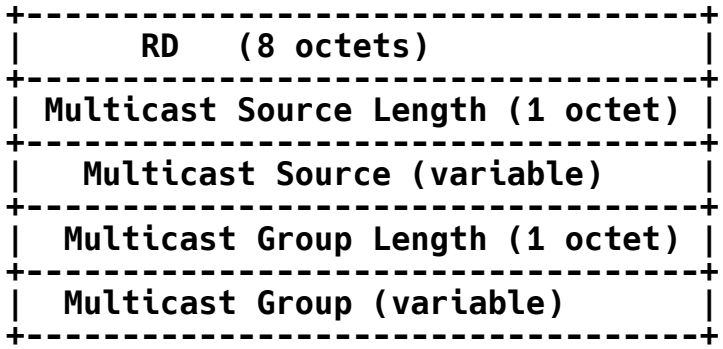
Leaf A-D routes may be originated as a result of processing a received Inter-AS I-PMSI A-D route or S-PMSI A-D route. A Leaf A-D route is originated in these situations only if the received route has a PMSI Tunnel attribute whose "Leaf Information Required" bit is set to 1.

If a Leaf A-D route is originated as a result of processing one of the received routes specified in the previous paragraph, the Route Key of the Leaf A-D route is set to the NLRI of the received route.

Details of the use of the Leaf A-D route may be found in Sections 9.2.3.2.1, 9.2.3.3, 9.2.3.4.1, 9.2.3.5, and 12.3.

4.5. Source Active A-D Route

A Source Active A-D Route Type specific MCAST-VPN NLRI consists of the following:



The RD is encoded as described in [RFC4364].

The Multicast Source field contains the C-S address. If the Multicast Source field contains an IPv4 address, then the value of the Multicast Source Length field is 32. If the Multicast Source field contains an IPv6 address, then the value of the Multicast Source Length field is 128.

Use of the Source Active A-D routes with the Multicast Source Length field of 0 is outside the scope of this document.

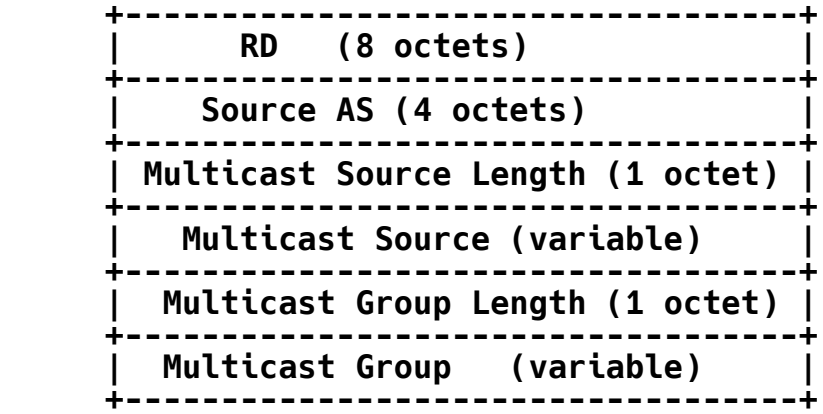
The Multicast Group field contains the C-G address. If the Multicast Group field contains an IPv4 address, then the value of the Multicast Group Length field is 32. If the Multicast Group field contains an IPv6 address, then the value of the Multicast Group Length field is 128.

Source Active A-D routes with a Multicast group belonging to the Source Specific Multicast (SSM) range (as defined in [RFC4607], and potentially extended locally on a router) MUST NOT be advertised by a router and MUST be discarded if received.

Usage of Source Active A-D routes is described in Sections 13 and 14.

4.6. C-Multicast Route

A Shared Tree Join route and a Source Tree Join Route Type specific MCAST-VPN NLRI consists of the following:



The RD is encoded as described in [RFC4364].

The Source AS contains an ASN. Two-octet ASNs are encoded in the low-order two octets of the Source AS field.

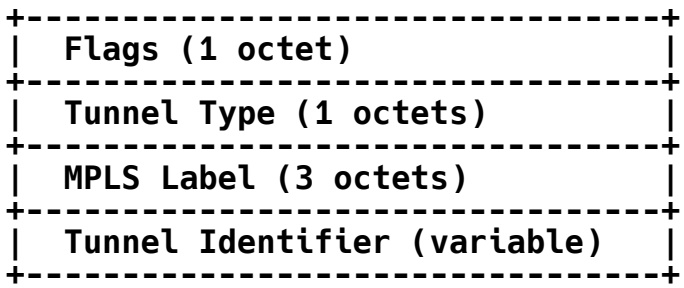
For a Shared Tree Join route, the Multicast Source field contains the C-RP address; for a Source Tree Join route, the Multicast Source field contains the C-S address. If the Multicast Source field contains an IPv4 address, then the value of the Multicast Source Length field is 32. If the Multicast Source field contains an IPv6 address, then the value of the Multicast Source Length field is 128.

The Multicast Group field contains the C-G address or C-MP Opaque Value Element. If the Multicast Group field contains an IPv4 address, then the value of the Multicast Group Length field is 32. If the Multicast Group field contains an IPv6 address, then the value of the Multicast Group Length field is 128.

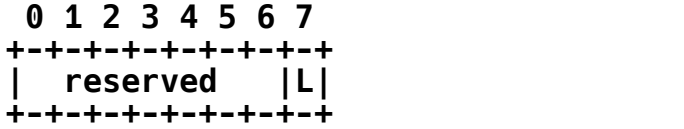
Usage of C-multicast routes is described in Section 11.

5. PMSI Tunnel Attribute

This document defines and uses a new BGP attribute called the "P-Multicast Service Interface Tunnel (PMSI Tunnel) attribute". This is an optional transitive BGP attribute. The format of this attribute is defined as follows:



The Flags field has the following format:



This document defines the following flags:

+ Leaf Information Required (L)

The Tunnel Type identifies the type of the tunneling technology used to establish the PMSI tunnel. The type determines the syntax and semantics of the Tunnel Identifier field. This document defines the following Tunnel Types:

+ 0 - No tunnel information present
+ 1 - RSVP-TE P2MP LSP
+ 2 - mLDP P2MP LSP
+ 3 - PIM-SSM Tree
+ 4 - PIM-SM Tree
+ 5 - BIDIR-PIM Tree
+ 6 - Ingress Replication
+ 7 - mLDP MP2MP LSP

If the MPLS Label field is non-zero, then it contains an MPLS label encoded as 3 octets, where the high-order 20 bits contain the label value. Absence of an MPLS Label is indicated by setting the MPLS Label field to zero.

When the Tunnel Type is set to "No tunnel information present", the PMSI Tunnel attribute carries no tunnel information (no Tunnel Identifier). This type is to be used only in the following case: to enable explicit tracking for a particular customer multicast flow (by setting the Leaf Information Required flag to 1), but without binding this flow to a particular provider tunnel (by omitting any tunnel information).

When the Tunnel Type is set to RSVP - Traffic Engineering (RSVP-TE) Point-to-Multipoint (P2MP) Label Switched Path (LSP), the Tunnel Identifier is <Extended Tunnel ID, Reserved, Tunnel ID, P2MP ID> as carried in the RSVP-TE P2MP LSP SESSION Object [RFC4875].

When the Tunnel Type is set to multipoint Label Distribution Protocol (mLDP) P2MP LSP, the Tunnel Identifier is a P2MP Forwarding Equivalence Class (FEC) Element [mLDP].

When the Tunnel Type is set to Protocol Independent Multicast - Sparse Mode (PIM-SM) tree, the Tunnel Identifier is <Sender Address, P-Multicast Group>. The node that originated the attribute MUST use the address carried in the Sender Address as the source IP address for the IP/GRE (Generic Routing Encapsulation) encapsulation of the MVPN data.

When the Tunnel Type is set to PIM-SSM tree, the Tunnel Identifier is <P-Root Node Address, P-Multicast Group>. The node that originates the attribute MUST use the address carried in the P-Root Node Address as the source IP address for the IP/GRE encapsulation of the MVPN data. The P-Multicast Group in the Tunnel Identifier of the Tunnel attribute MUST NOT be expected to be the same group for all Intra-AS A-D routes for the same MVPN. According to [RFC4607], the group address can be locally allocated by the originating PE without any consideration for the group address used by other PE on the same MVPN.

When the Tunnel Type is set to BIDIR-PIM tree, the Tunnel Identifier is <Sender Address, P-Multicast Group>. The node that originated the attribute MUST use the address carried in the Sender Address as the source IP address for the IP/GRE encapsulation of the MVPN data.

When the Tunnel Type is set to PIM-SM or BIDIR-PIM tree, then the P-Multicast Group in the Tunnel Identifier of the Tunnel attribute SHOULD contain the same multicast group address for all Intra-AS I-PMSI A-D routes for the same MVPN originated by PEs within a given AS. How this multicast group address is chosen is outside the scope of this specification.

When the Tunnel Type is set to Ingress Replication, the Tunnel Identifier carries the unicast tunnel endpoint IP address of the local PE that is to be this PE's receiving endpoint address for the tunnel.

When the Tunnel Type is set to mLDP Multipoint-to-Multipoint (MP2MP) LSP, the Tunnel Identifier is an MP2MP FEC Element [mLDP].

The use of mLDP MP2MP LSPs as Provider tunnels (P-tunnels) requires procedures that are outside the scope of this document.

A router that supports the PMSI Tunnel attribute considers this attribute to be malformed if either (a) it contains an undefined tunnel type in the Tunnel Type field of the attribute, or (b) the router cannot parse the Tunnel Identifier field of the attribute as a tunnel identifier of the tunnel types specified in the Tunnel Type field of the attribute.

When a router that receives a BGP Update that contains the PMSI Tunnel attribute with its Partial bit set determines that the attribute is malformed, the router SHOULD treat this Update as though all the routes contained in this Update had been withdrawn.

An implementation MUST provide debugging facilities to permit issues caused by a malformed PMSI Tunnel attribute to be diagnosed. At a minimum, such facilities MUST include logging an error when such an attribute is detected.

The PMSI Tunnel attribute is used in conjunction with Intra-AS I-PMSI A-D routes, Inter-AS I-PMSI A-D routes, S-PMSI A-D routes, and Leaf A-D routes.

6. Source AS Extended Community

This document defines a new BGP Extended Community called "Source AS".

The Source AS is an AS-specific Extended Community, of an extended type, and is transitive across AS boundaries [RFC4360].

The Global Administrator field of this Community MUST be set to the ASN of the PE. The Local Administrator field of this Community MUST be set to 0.

Consider a given MVPN that uses BGP for exchanging C-multicast routes, and/or uses segmented inter-AS tunnels. A PE that has sites of that MVPN connected to it, and originates a (unicast) route to VPN-IP addresses associated with the destinations within these sites, MUST include in the BGP Update message that carries this route the Source AS Extended Community.

The usage of a received Source AS Extended Community is described in Section 11.1.3.

7. VRF Route Import Extended Community

This document defines a new BGP Extended Community called "VRF Route Import".

The VRF Route Import is an IP-address-specific Extended Community, of an extended type, and is transitive across AS boundaries [RFC4360].

To support MVPN in addition to the import/export Route Target(s) Extended Communities used by the unicast routing, each VRF on a PE MUST have an import Route Target Extended Community, except if it is known a priori that none of the (local) MVPN sites associated with the VRF contain multicast source(s) and/or C-RP; in which case, the VRF need not have this import Route Target.

We refer to this Route Target as the "C-multicast Import RT", as this Route Target controls imports of C-multicast routes into a particular VRF.

A PE constructs C-multicast Import RT as follows:

+ The Global Administrator field of the C-multicast Import RT MUST be set to an IP address of the PE. This address SHOULD be common for all the VRFs on the PE (e.g., this address may be the PE's loopback address).

+ The Local Administrator field of the C-multicast Import RT associated with a given VRF contains a 2-octet number that uniquely identifies that VRF within the PE that contains the VRF (procedures for assigning such numbers are purely local to the PE and are outside the scope of this document).

The way C-multicast Import RT is constructed allows it to uniquely identify a VRF.

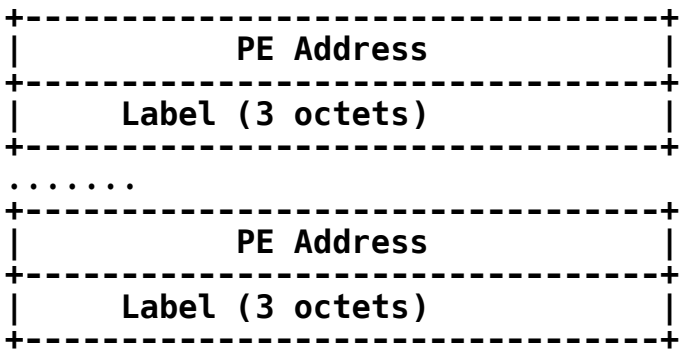
A PE that has site(s) of a given MVPN connected to it needs to communicate the value of the C-multicast Import RT associated with the VRF of that MVPN on the PE to all other PEs that have sites of that MVPN. To accomplish this, a PE that originates a (unicast) route to VPN-IP addresses MUST include in the BGP Updates message that carries this route the VRF Route Import Extended Community that has the value of the C-multicast Import RT of the VRF associated with the route, except if it is known a priori (e.g., via provisioning) that none of these addresses could act as multicast sources and/or RP; in which case, the (unicast) route MUST NOT carry the VRF Route Import Extended Community.

If a PE uses Route Target Constraint [RT-CONSTRAIN], the PE SHOULD advertise all such C-multicast Import RTs using Route Target Constraints (note that doing this requires just a single Route Target Constraint advertisement by the PE). This allows each C-multicast route to reach only the relevant PE. To constrain distribution of the Route Target Constraint routes to the AS of the advertising PE, these routes SHOULD carry the NO_EXPORT Community [RFC1997].

Usage of VRF Route Import Extended Community is described in Section 11.1.3.

8. PE Distinguisher Labels Attribute

This document defines a new BGP attribute, called the "PE Distinguisher Labels" attribute. This is an optional transitive BGP attribute. The format of this attribute is defined as follows:



The Label field contains an MPLS label encoded as 3 octets, where the high-order 20 bits contain the label value.

A router that supports the PE Distinguisher Labels attribute considers this attribute to be malformed if the PE Address field does not contain a unicast address. The attribute is also considered to be malformed if: (a) the PE Address field is expected to be an IPv4 address, and the length of the attribute is not a multiple of 7 or (b) the PE Address field is expected to be an IPv6 address, and the length of the attribute is not a multiple of 19. The length of the Route Type field of MCAST-VPN NLRI of the route that carries the PE Distinguisher Labels attribute provides the information on whether the PE Address field contains an IPv4 or IPv6 address. Each of the PE addresses in the PE Distinguisher Labels attribute MUST be of the same address family as the "Originating Router's IP Address" of the route that is carrying the attribute.

When a router that receives a BGP Update that contains the PE Distinguisher Labels attribute with its Partial bit set determines that the attribute is malformed, the router SHOULD treat this Update as though all the routes contained in this Update had been withdrawn.

An implementation MUST provide debugging facilities to permit issues caused by malformed PE Distinguisher Label attribute to be diagnosed. At a minimum, such facilities MUST include logging an error when such an attribute is detected.

Usage of this attribute is described in [MVPN].

9. MVPN Auto-Discovery/Binding

This section specifies procedures for the auto-discovery of MVPN memberships and the distribution of information used to instantiate I-PMSIs.

There are two MVPN auto-discovery/binding mechanisms, dubbed "intra-AS" and "inter-AS" respectively.

The intra-AS mechanisms provide auto-discovery/binding within a single AS.

The intra-AS mechanisms also provide auto-discovery/binding across multiple ASes when non-segmented inter-AS tunnels are being used.

The inter-AS mechanisms provide auto-discovery/binding across multiple ASes when segmented inter-AS tunnels are being used.

Note that if a multi-AS system uses option (a) of section 10 of [RFC4364], the notion of inter-AS tunnels does not apply, and so it needs only the intra-AS mechanisms.

9.1. MVPN Auto-Discovery/Binding - Intra-AS Operations

This section describes exchanges of Intra-AS I-PMSI A-D routes originated/received by PEs within the same AS, or if non-segmented inter-AS tunnels are used, then by all PEs.

9.1.1. Originating Intra-AS I-PMSI A-D Routes

To participate in the MVPN auto-discovery/binding, a PE router that has a given VRF of a given MVPN MUST, except for the cases specified in this section, originate an Intra-AS I-PMSI A-D route and advertises this route in IBGP. The route is constructed as follows.

The route carries a single MCAST-VPN NLRI with the RD set to the RD of the VRF, and the Originating Router's IP Address field set to the IP address that the PE places in the Global Administrator field of the VRF Route Import Extended Community of the VPN-IP routes advertised by the PE. Note that the <RD, Originating Router's IP Address> tuple uniquely identifies a given multicast VRF.

The route carries the PMSI Tunnel attribute if and only if an I-PMSI is used for the MVPN (the conditions under which an I-PMSI is used can be found in [MVPN]). Depending on the technology used for the P-tunnel for the MVPN on the PE, the PMSI Tunnel attribute of the Intra-AS I-PMSI A-D route is constructed as follows.

+ If the PE that originates the advertisement uses a P-multicast tree for the P-tunnel for the MVPN, the PMSI Tunnel attribute MUST contain the identity of the tree (note that the PE could create the identity of the tree prior to the actual instantiation of the tree).

+ A PE that uses a P-multicast tree for the P-tunnel MAY aggregate two or more MVPNs present on the PE onto the same tree. In this case, in addition to carrying the identity of the tree, the PMSI Tunnel attribute of the Intra-AS I-PMSI A-D route MUST carry an MPLS upstream-assigned label that the PE has bound uniquely to the MVPN associated with this route (as determined by its RTs).

  If the PE has already advertised Intra-AS I-PMSI A-D routes for two or more MVPNs that it now desires to aggregate, then the PE MUST re-advertise those routes. The re-advertised routes MUST be the same as the original ones, except for the PMSI Tunnel attribute and the label carried in that attribute.

+ If the PE that originates the advertisement uses ingress replication for the P-tunnel for the MVPN, the route MUST include the PMSI Tunnel attribute with the Tunnel Type set to Ingress Replication and Tunnel Identifier set to a routable address of the PE. The PMSI Tunnel attribute MUST carry a downstream-assigned MPLS label. This label is used to demultiplex the MVPN traffic received over a unicast tunnel by the PE.

+ The Leaf Information Required flag of the PMSI Tunnel attribute MUST be set to zero and MUST be ignored on receipt.

Discovery of PE capabilities in terms of what tunnel types they support is outside the scope of this document. Within a given AS, PEs participating in an MVPN are expected to advertise tunnel bindings whose tunnel types are supported by all other PEs that are

participating in this MVPN and are part of the same AS. In addition, in the inter-AS scenario with non-segmented inter-AS tunnels, the tunnel types have to be supported by all PEs that are participating in this MVPN, irrespective of whether or not these PEs are in the same AS.

The Next Hop field of the MP_REACH_NLRI attribute of the route MUST be set to the same IP address as the one carried in the Originating Router's IP Address field.

By default, the distribution of the Intra-AS I-PMSI A-D routes is controlled by the same Route Targets as the ones used for the distribution of VPN-IP unicast routes. That is, by default, the Intra-AS I-PMSI A-D route MUST carry the export Route Target used by the unicast routing. If any other PE has one of these Route Targets configured as an import Route Target for a VRF present on the PE, it treats the advertising PE as a member in the MVPN to which the VRF belongs. The default could be modified via configuration by having a set of Route Targets used for the Intra-AS I-PMSI A-D routes being distinct from the ones used for the VPN-IP unicast routes (see also Section 10).

To constrain distribution of the intra-AS membership/binding information to the AS of the advertising PE, the BGP Update message originated by the advertising PE SHOULD carry the NO_EXPORT Community [RFC1997].

Note that if non-segmented inter-AS P-tunnels are being used, then the Intra-AS I-PMSI routes need to be distributed to other ASes and MUST NOT carry the NO_EXPORT Community.

When BGP is used to exchange C-multicast routes, if (a) it is known a priori that, as a matter of policy, none of the MVPN sites connected to a given PE are allowed to send multicast traffic to other sites of that MVPN (in other words, all these sites are only in the Receiver Sites set), (b) the PE does not use ingress replication for the incoming traffic of that MVPN, and (c) none of the other PEs that have VRFs of that MVPN use RSVP-TE P2MP LSP for that MVPN, then the local PE SHOULD NOT originate an Intra-AS I-PMSI A-D route.

When BGP is used to exchange C-multicast routes, if it is known a priori that, as a matter of policy, none of the MVPN sites connected to a given PE can receive multicast traffic from other sites of that MVPN (in other words, all these sites are only in the Sender Sites set), and the PE uses ingress replication for that MVPN, then the PE SHOULD NOT originate an Intra-AS I-PMSI A-D route for that MVPN.

9.1.2. Receiving Intra-AS I-PMSI A-D Routes

When a PE receives a BGP Update message that carries an Intra-AS I-PMSI A-D route such that (a) at least one of the Route Targets of the route matches one of the import Route Targets configured for a particular VRF on the local PE, (b) either the route was originated by some other PE within the same AS as the local PE, or the MVPN associated with the VRF uses non-segmented inter-AS tunnels, and (c) the BGP route selection determines that this is the best route with respect to the NLRI carried by the route, the PE performs the following.

If the route does not carry the PMSI Tunnel attribute and ingress replication is not used, either a) the PE that originated the route will be using only S-PMSIs to send traffic to remote PEs, or b) as a matter of policy, the PE that originated the route cannot send multicast traffic from the MVPN sites connected to it to other sites of that MVPN (in other words, the sites connected to the PE are only in the Receiver Sites set).

When BGP is used to exchange C-multicast routes, to distinguish between cases (a) and (b), we use the presence/absence of the VRF Route Import Extended Community in the unicast VPN routes, as follows. As specified in Section 7, if it is know a priori that none of the addresses carried in the NLRI of a given (unicast) VPN route could act as multicast sources and/or C-RP, then such a route does not carry the VRF Route Import Extended Community. Hence, based on the Upstream Multicast Hop (UMH) selection algorithm specified in [MVPN], such a route will be ineligible for the UMH selection. This implies that if a given VPN route is selected by the UMH selection procedures, and the PE that originates this VPN route also originates an Intra-AS I-PMSI A-D route, but this route does not carry the PMSI Tunnel attribute, then this PE will be using only S-PMSIs for sending (multicast) data.

If the route carries the PMSI Tunnel attribute, then:

+ If the Tunnel Type in the PMSI Tunnel attribute is set to Ingress Replication, then the MPLS label and the address carried in the Tunnel Identifier field of the PMSI Tunnel attribute should be used when the local PE sends multicast traffic to the PE that originated the route.

+ If the Tunnel Type in the PMSI Tunnel attribute is set to mLDP P2MP LSP, mLDP MP2MP LSP, PIM-SSM tree, PIM-SM tree, or BIDIR-PIM tree, the PE SHOULD join as soon as possible the P-multicast tree whose identity is carried in the Tunnel Identifier.

+ If the Tunnel Type in the PMSI Tunnel attribute is set to RSVP-TE P2MP LSP, then the PE that originated the route MUST establish an RSVP-TE P2MP LSP with the local PE as a leaf. This LSP may have been established before the local PE receives the route, or it may be established after the local PE receives the route.

+ The receiving PE has to establish the appropriate state to properly handle the traffic received on the P-multicast tree.

+ If the PMSI Tunnel attribute does not carry a label, then all packets that are received on the P-multicast tree, as identified by the PMSI Tunnel attribute, are forwarded using the VRF that has at least one of its import Route Targets that matches one of the Route Targets of the received Intra-AS I-PMSI A-D route.

+ If the PMSI Tunnel attribute has the Tunnel Type set to mLDP P2MP LSP, PIM-SSM tree, PIM-SM tree, BIDIR-PIM tree, or RSVP-TE P2MP LSP, and the attribute also carries an MPLS label, then this is an upstream-assigned label, and all packets that are received on the P-multicast tree, as identified by the PMSI Tunnel attribute, with that upstream-assigned label are forwarded using the VRF that has at least one of its import Route Targets that matches one of the Route Targets of the received Intra-AS I-PMSI A-D route.

Irrespective of whether the route carries the PMSI Tunnel attribute, if the local PE uses RSVP-TE P2MP LSP for sending (multicast) traffic from the VRF to the sites attached to other PEs, then the local PE uses the Originating Router's IP address information carried in the route to add the PE that originated the route as a leaf node to the LSP.

9.2. MVPN Auto-Discovery/Binding - Inter-AS Operations

This section applies only to the case where segmented inter-AS tunnels are used.

An Autonomous System Border Router (ASBR) may be configured to support a particular MVPN as follows:

+ An ASBR MUST be configured with a set of (import) Route Targets (RTs) that specifies the set of MVPNs supported by the ASBR. These Route Targets control acceptance of Intra-AS/Inter-AS I-PMSI A-D routes by the ASBR. As long as unicast and multicast connectivity are congruent, this could be the same set of Route Targets as the one used for supporting unicast (and therefore would not require any additional configuration above and beyond of

what is required for unicast). Note that instead of being configured, the ASBR MAY obtain this set of (import) Route Targets (RTs) by using Route Target Constraint [RT-CONSTRAIN].

+ The ASBR MUST be (auto-)configured with an import Route Target called "ASBR Import RT". ASBR Import RT controls acceptance of Leaf A-D routes and C-multicast routes by the ASBR, and is used to constrain distribution of both Leaf A-D routes and C-multicast routes (see Section 11).

  ASBR Import RT is an IP-address-specific Route Target. The Global Administrator field of the ASBR Import RT MUST be set to the IP address carried in the Next Hop of all the Inter-AS I-PMSI A-D routes and S-PMSI A-D routes advertised by this ASBR (if the ASBR uses different Next Hops, then the ASBR MUST be (auto-)configured with multiple ASBR Import RTs, one per each such Next Hop). The Local Administrator field of the ASBR Import RT MUST be set to 0.

  If the ASBR supports Route Target Constraint [RT-CONSTRAIN], the ASBR SHOULD advertise its ASBR Import RT within its own AS using Route Target Constraints. To constrain distribution of the Route Target Constraint routes to the AS of the advertising ASBR, these routes SHOULD carry the NO_EXPORT Community [RFC1997].

+ The ASBR MUST be configured with the tunnel types for the intra-AS segments of the MVPNs supported by the ASBR, as well as (depending on the tunnel type) the information needed to create the PMSI attribute for these tunnel types. Note that instead of being configured, the ASBR MAY derive the tunnel types from the Intra-AS I-PMSI A-D routes received by the ASBR.

+ If the ASBR originates an Inter-AS I-PMSI A-D route for a particular MVPN present on some of the PEs within its own AS, the ASBR MUST be (auto-)configured with an RD for that MVPN. It is RECOMMENDED that one of the following two options be used:

(1) To allow more aggregation of Inter-AS I-PMSI A-D routes, it is recommended that all the ASBRs within an AS that are configured to originate an Inter-AS I-PMSI A-D route for a particular MVPN be configured with the same RD (although for a given MVPN each AS may assign this RD on its own, without coordination with other ASes).

(2) To allow more control over spreading MVPN traffic among multiple ASBRs within a given AS, it is recommended that each ASBR have a distinct RD per each MVPN; in which case, such an RD SHOULD be auto-configured.

If an ASBR is configured to support a particular MVPN, the ASBR MUST participate in the intra-AS MVPN auto-discovery/binding procedures for that MVPN within the ASBR's own AS, as specified in Section 9.1.

Moreover, in addition to the above, the ASBR performs procedures described in Sections 9.2.1, 9.2.2, and 9.2.3.

9.2.1. Originating Inter-AS I-PMSI A-D Routes

For a given MVPN configured on an ASBR when the ASBR determines (using the intra-AS auto-discovery procedures) that at least one of the PEs of its own AS has (directly) connected site(s) of the MVPN, the ASBR originates an Inter-AS I-PMSI A-D route and advertises it in External BGP (EBGP). The route is constructed as follows:

+ The route carries a single MCAST-VPN NLRI with the RD set to the RD configured for that MVPN on the ASBR, and the Source AS set to the ASN of the ASBR.

+ The route carries the PMSI Tunnel attribute if and only if an I-PMSI is used for the MVPN. The Tunnel Type in the attribute is set to Ingress Replication; the Leaf Information Required flag is set to 1; the attribute carries no MPLS labels.

+ The Next Hop field of the MP_REACH_NLRI attribute is set to a routable IP address of the ASBR.

+ The default policy for aggregation of Intra-AS I-PMSI A-D routes into an Inter-AS I-PMSI A-D route is that a given Inter-AS I-PMSI A-D route aggregates only the Intra-AS I-PMSI A-D routes that carry exactly the same set of RTs (note that this set may have just one RT). In this case, an Inter-AS I-PMSI A-D route originated by an ASBR carries exactly the same RT(s) as the RT(s) carried by the Intra-AS I-PMSI A-D routes that the ASBR aggregates into that Inter-AS I-PMSI A-D route. An implementation MUST support the default policy for aggregation of Intra-AS I-PMSI A-D routes into an Inter-AS I-PMSI A-D route.

+ The default policy for aggregation could be modified via configuration on the ASBR. An implementation MAY support such functionality. Modified policy MUST include rules for constructing RTs carried by the Inter-AS I-PMSI A-D routes originated by the ASBR.

An Inter-AS I-PMSI A-D route for a given <AS, MVPN> indicates the presence of the MVPN sites connected to one or more PEs of the AS.

An Inter-AS I-PMSI A-D route originated by an ASBR aggregates Intra-AS I-PMSI A-D routes originated within the ASBR's own AS. Thus, while the Intra-AS I-PMSI A-D routes originated within an AS are at the granularity of <PE, MVPN> within that AS, outside of that AS the (aggregated) Inter-AS I-PMSI A-D routes could be at the granularity of <AS, MVPN>.

9.2.2. When Not to Originate Inter-AS I-PMSI A-D Routes

If, for a given MVPN and a given AS, all of the sites connected to the PEs within the AS are known a priori to have no multicast sources, then ASBRs of that AS MAY refrain from originating an Inter-AS I-PMSI A-D route for that MVPN at all.

9.2.3. Propagating Inter-AS I-PMSI A-D Routes

An Inter-AS I-PMSI A-D route for a given MVPN originated by an ASBR within a given AS is propagated via BGP to other ASes.

9.2.3.1. Propagating Inter-AS I-PMSI A-D Routes - Overview

Suppose that ASBR A installs an Inter-AS I-PMSI A-D route for MVPN V that originated at a particular AS, AS1. The BGP Next Hop of that route becomes A's "upstream multicast hop" on a multicast distribution tree for V that is rooted at AS1. When the Inter-AS I-PMSI A-D routes have been distributed to all the necessary ASes, they define a "reverse path" from any AS that supports MVPN V back to AS1. For instance, if AS2 supports MVPN V, then there will be a reverse path for MVPN V from AS2 to AS1. This path is a sequence of ASBRs, the first of which is in AS2, and the last of which is in AS1. Each ASBR in the sequence is the BGP Next Hop of the previous ASBR in the sequence on the given Inter-AS I-PMSI A-D route.

This reverse path information can be used to construct a unidirectional multicast distribution tree for MVPN V, containing all the ASes that support V, and having AS1 at the root. We call such a tree an "inter-AS tree". Multicast data originating in MVPN sites connected to PEs within a given AS will travel downstream along the tree, which is rooted at that AS.

The path along an inter-AS tree is a sequence of ASBRs; it is still necessary to specify how the multicast data gets from a given ASBR to the set of ASBRs that are immediately downstream of the given ASBR along the tree. This is done by creating "segments": ASBRs in adjacent ASes will be connected by inter-AS segments, ASBRs in the same AS will be connected by "intra-AS segments".

An ASBR initiates creation of an intra-AS segment when the ASBR receives an Inter-AS I-PMSI A-D route from an EBGP neighbor. Creation of the segment is completed as a result of distributing, via IBGP, this route within the ASBR's own AS.

For a given inter-AS tunnel, each of its intra-AS segments could be constructed by its own independent mechanism. Moreover, by using upstream-assigned labels within a given AS multiple intra-AS segments of different inter-AS tunnels of either the same or different MVPNs may share the same P-multicast tree.

If the P-multicast tree that serves as a particular intra-AS segment of an inter-AS tunnel is created by a multicast control protocol that uses receiver-initiated joins (e.g., mLDP, any PIM variant), and this P-multicast tree does not aggregate multiple segments, then all the information needed to create that segment is present in the PMSI Tunnel attribute of the Inter-AS I-PMSI A-D routes. However, if the P-multicast tree that serves as the segment is created by a protocol that does not use receiver-initiated joins (e.g., RSVP-TE, ingress unicast replication), or if this P-multicast tree aggregates multiple segments (irrespective of the multicast control protocol used to create the tree), then it is also necessary to use Leaf A-D routes. The precise conditions under which Leaf A-D routes need to be used are described in subsequent sections.

Since (aggregated) Inter-AS I-PMSI A-D routes could have granularity of <AS, MVPN>, an MVPN that is present in N ASes could have a total of N inter-AS tunnels. Thus, for a given MVPN, the number of inter-AS tunnels constituting the I-PMSIs is independent of the number of PEs that have this MVPN.

The precise rules for distributing and processing the Inter-AS I-PMSI A-D routes across ASes are given in the following sections.

9.2.3.2. Inter-AS I-PMSI A-D Route Received via EBGP

When an ASBR receives, from one of its EBGP neighbors, a BGP Update message that carries an Inter-AS I-PMSI A-D route, if (a) at least one of the Route Targets carried in the message matches one of the import Route Targets configured on the ASBR, and (b) the ASBR determines that the received route is the best route for its NLRI, the ASBR re-advertises this route to other PEs and ASBRs within its own AS (handling of this route by other PEs and ASBRs is described in Section 9.2.3.4).

When re-advertising an Inter-AS I-PMSI A-D route, the ASBR MUST set the Next Hop field of the MP_REACH_NLRI attribute to a routable IP address of the ASBR.

If the received Inter-AS I-PMSI A-D route carries the PMSI Tunnel attribute, then, depending on the technology used to instantiate the intra-AS segment of the inter-AS tunnel, the ASBR constructs the PMSI Tunnel attribute of the re-advertised Inter-AS I-PMSI A-D route as follows.

+ If the ASBR uses ingress replication for the intra-AS segment of the inter-AS tunnel, the re-advertised route MUST carry the PMSI Tunnel attribute with the Tunnel Type set to Ingress Replication, but no MPLS labels.

+ If the ASBR uses a P-multicast tree for the intra-AS segment of the inter-AS tunnel, the PMSI Tunnel attribute MUST contain the identity of the tree (note that the ASBR could create the identity of the tree prior to the actual instantiation of the tree). If, in order to instantiate the tree, the ASBR needs to know the leaves of the tree, then the ASBR obtains this information from the Leaf A-D routes received from other PEs/ASBRs in the ASBR's own AS (as described in Section 9.2.3.5) by setting the Leaf Information Required flag in the PMSI Tunnel attribute to 1.

+ An ASBR that uses a P-multicast tree as the intra-AS segment of the inter-AS tunnel MAY aggregate two or more MVPNs present on the ASBR onto the same tree. In this case, in addition to the identity of the tree, the PMSI Tunnel attribute of the Inter-AS I-PMSI A-D route MUST carry an MPLS upstream-assigned label that the PE has bound uniquely to the MVPN associated with this route (as determined by its RTs).

  If the ASBR has already advertised Inter-AS I-PMSI A-D routes for two or more MVPNs that it now desires to aggregate, then the ASBR MUST re-advertise those routes. The re-advertised routes MUST be the same as the original ones, except for the PMSI Tunnel attribute and the MVPN label.

9.2.3.2.1. Originating Leaf A-D Route into EBGP

In addition, the ASBR MUST send to the EBGP neighbor from whom it received the Inter-AS I-PMSI A-D route, a BGP Update message that carries a Leaf A-D route constructed as follows.

+ The route carries a single MCAST-VPN NLRI with the Route Key field set to the MCAST-VPN NLRI of the Inter-AS I-PMSI A-D route received from that neighbor and the Originating Router's IP address set to the IP address of the ASBR (this MUST be a routable IP address).

+ The Leaf A-D route MUST include the PMSI Tunnel attribute with the Tunnel Type set to Ingress Replication and the Tunnel Identifier set to a routable address of the advertising router. The PMSI Tunnel attribute MUST carry a downstream-assigned MPLS label that is used by the advertising router to demultiplex the MVPN traffic received over a unicast tunnel from the EBGP neighbor.

+ The ASBR constructs an IP-based Route Target Extended Community by placing the IP address carried in the Next Hop of the received Inter-AS I-PMSI A-D route in the Global Administrator field of the Community, with the Local Administrator field of this Community set to 0 and setting the Extended Communities attribute of the Leaf A-D route to that Community. Note that this Route Target is the same as the ASBR Import RT of the EBGP neighbor from which the ASBR received the Inter-AS I-PMSI A-D route.

+ The Next Hop field of the MP_REACH_NLRI attribute of the route MUST be set to the same IP address as the one carried in the Originating Router's IP Address field of the route.

+ To constrain the distribution scope of this route, the route MUST carry the NO_ADVERTISE BGP Community [RFC1997].

Handling of this Leaf A-D route by the EBGP neighbor is described in Section 9.2.3.3.

The ASBR MUST set up its forwarding state such that packets that arrive on the one-hop ASBR-ASBR LSP, as specified in the PMSI Tunnel attribute of the Leaf A-D route, are transmitted on the intra-AS segment, as specified in the PMSI Tunnel attribute of the Inter-AS I-PMSI A-D route that the ASBR re-advertises in its own AS. However, the packets MAY be filtered before forwarding, as specified in Section 9.2.3.6.

9.2.3.3. Leaf A-D Route Received via EBGP

When an ASBR receives, via EBGP, a Leaf A-D route originated by its neighbor ASBR, if the Route Target carried in the Extended Communities attribute of the route matches one of the ASBR Import RT (auto-)configured on the ASBR, the ASBR performs the following.

+ The ASBR finds an Inter-AS I-PMSI A-D route whose MCAST-VPN NLRI has the same value as the Route Key field of the Leaf A-D route.

+ If the found Inter-AS I-PMSI A-D route was originated by ASBR itself, then the ASBR sets up its forwarding state such that packets received on the intra-AS tunnels originating in the ASBR's own AS are transmitted on the one-hop ASBR-ASBR LSP specified by

the MPLS label carried in the PMSI Tunnel attribute of the received Leaf A-D route. (However, the packets MAY be filtered before transmission as specified in Section 9.2.3.6). The intra-AS tunnels are specified in the PMSI Tunnel attribute of all the Intra-AS I-PMSI A-D routes received by the ASBR that the ASBR aggregated into the Inter-AS I-PMSI A-D route. For each of these intra-AS tunnels, if a non-zero MPLS label is carried in the PMSI Tunnel attribute (i.e., aggregation is used), then only packets received on the inner LSP corresponding to that label MUST be forwarded, not the packets received on the outer LSP, as the outer LSP possibly carries the traffic of other VPNs.

+ If the found Inter-AS I-PMSI A-D route was originated by some other ASBR, then the ASBR sets up its forwarding state such that packets received on the intra-AS tunnel segment, as specified in the PMSI Tunnel attribute of the found Inter-AS I-PMSI A-D route, are transmitted on the one-hop ASBR-ASBR LSP, as specified by the MPLS label carried in the PMSI Tunnel attribute of the Leaf A-D route.

9.2.3.4. Inter-AS I-PMSI A-D Route Received via IBGP

In the context of this section, we use the term "PE/ASBR router" to denote either a PE or an ASBR router.

If a given Inter-AS I-PMSI A-D route is received via IBGP by a BGP route reflector, the BGP route reflector MUST NOT modify the Next Hop field of the MP_REACH_NLRI attribute when re-advertising the route into IBGP (this is because the information carried in the Next Hop is used for controlling flow of C-multicast routes, as specified in Section 11.2).

If a given Inter-AS I-PMSI A-D route is advertised within an AS by multiple ASBRs of that AS, the BGP best route selection performed by other PE/ASBR routers within the AS does not require all these PE/ASBR routers to select the route advertised by the same ASBR -- to the contrary, different PE/ASBR routers may select routes advertised by different ASBRs.

When a PE/ASBR router receives, from one of its IBGP neighbors, a BGP Update message that carries an Inter-AS I-PMSI A-D route, if (a) at least one of the Route Targets carried in the message matches one of the import Route Targets configured on the PE/ASBR, and (b) the PE/ASBR determines that the received route is the best route to the destination carried in the NLRI of the route, the PE/ASBR performs the following operations.

+ If the router is a PE, then the router imports the route into the VRF(s) that have the matching import Route Targets.

+ If the router is an ASBR, then the ASBR propagates the route to its EBGP neighbors. When propagating the route to the EBGP neighbors, the ASBR MUST set the Next Hop field of the MP_REACH_NLRI attribute to a routable IP address of the ASBR. If the received Inter-AS I-PMSI A-D route carries the PMSI Tunnel attribute, then the propagated route MUST carry the PMSI Tunnel attribute with the Tunnel Type set to Ingress Replication; the attribute carries no MPLS labels.

+ If the received Inter-AS I-PMSI A-D route carries the PMSI Tunnel attribute with the Tunnel Type set to mLDP P2MP LSP, PIM-SSM tree, PIM-SM tree, or BIDIR-PIM tree, the PE/ASBR SHOULD join as soon as possible the P-multicast tree whose identity is carried in the Tunnel Identifier.

+ If the received Inter-AS I-PMSI A-D route carries the PMSI Tunnel attribute with the Tunnel Identifier set to RSVP-TE P2MP LSP, then the ASBR that originated the route MUST establish an RSVP-TE P2MP LSP with the local PE/ASBR as a leaf. This LSP MAY have been established before the local PE/ASBR receives the route, or it MAY be established after the local PE receives the route.

+ If the received Inter-AS I-PMSI A-D route carries the PMSI Tunnel attribute with the Tunnel Type set to mLDP P2MP LSP, RSVP-TE P2MP LSP, PIM-SSM, PIM-SM tree, or BIDIR-PIM tree, but the attribute does not carry a label, then the P-multicast tree, as identified by the PMSI Tunnel attribute, is an intra-AS LSP segment that is part of the inter-AS tunnel for the MVPN advertised by the Inter-AS I-PMSI A-D route and rooted at the AS that originated the Inter-AS I-PMSI A-D route. If the PMSI Tunnel attribute carries a (upstream-assigned) label, then a combination of this tree and the label identifies the intra-AS segment. If the receiving router is an ASBR, this intra-AS segment may further be stitched to the ASBR-ASBR inter-AS segment of the inter-AS tunnel. If the PE/ASBR has local receivers in the MVPN, packets received over the intra-AS segment must be forwarded to the local receivers using the local VRF.

9.2.3.4.1. Originating Leaf A-D Route into IBGP

If the Leaf Information Required flag in the PMSI Tunnel attribute of the received Inter-AS I-PMSI A-D route is set to 1, then the PE/ASBR MUST originate a new Leaf A-D route as follows.

+ The route carries a single MCAST-VPN NLRI with the Route Key field set to the MCAST-VPN NLRI of the Inter-AS I-PMSI A-D route received from that neighbor and the Originating Router's IP address set to the IP address of the PE/ASBR (this MUST be a routable IP address).

+ If the received Inter-AS I-PMSI A-D route carries the PMSI Tunnel attribute with the Tunnel Type set to Ingress Replication, then the Leaf A-D route MUST carry the PMSI Tunnel attribute with the Tunnel Type set to Ingress Replication. The Tunnel Identifier MUST carry a routable address of the PE/ASBR. The PMSI Tunnel attribute MUST carry a downstream-assigned MPLS label that is used to demultiplex the MVPN traffic received over a unicast tunnel by the PE/ASBR.

+ The PE/ASBR constructs an IP-based Route Target Extended Community by placing the IP address carried in the Next Hop of the received Inter-AS I-PMSI A-D route in the Global Administrator field of the Community, with the Local Administrator field of this Community set to 0 and setting the Extended Communities attribute of the Leaf A-D route to that Community.

+ The Next Hop field of the MP_REACH_NLRI attribute of the route MUST be set to the same IP address as the one carried in the Originating Router's IP Address field of the route.

+ To constrain the distribution scope of this route, the route MUST carry the NO_EXPORT Community [RFC1997].

+ Once the Leaf A-D route is constructed, the PE/ASBR advertises this route into IBGP.

9.2.3.5. Leaf A-D Route Received via IBGP

When an ASBR receives, via IBGP, a Leaf A-D route, if the Route Target carried in the Extended Communities attribute of the route matches one of the ASBR Import RT (auto-)configured on the ASBR, the ASBR performs the following.

The ASBR finds an Inter-AS I-PMSI A-D route whose MCAST-VPN NLRI has the same value as the Route Key field of the Leaf A-D route.

The received route may carry either (a) no PMSI Tunnel attribute, or (b) the PMSI Tunnel attribute, but only with the Tunnel Type set to Ingress Replication.

If the received route does not carry the PMSI Tunnel attribute, the ASBR uses the information from the received route to determine the leaves of the P-multicast tree rooted at the ASBR that would be used for the intra-AS segment associated with the found Inter-AS I-PMSI A-D route. The IP address of a leaf is the IP address carried in the Originating Router's IP address field of the received Leaf A-D route.

If the received route carries the PMSI Tunnel attribute with the Tunnel Type set to Ingress Replication, the ASBR uses the information carried by the route to construct the intra-AS segment with ingress replication.

9.2.3.6. Optimizing Bandwidth by IP Filtering on ASBRs

An ASBR that has a given Inter-AS I-PMSI A-D route MAY discard some of the traffic carried in the tunnel specified in the PMSI Tunnel attribute of this route, if the ASBR determines that there are no downstream receivers for that traffic.

When BGP is being used to distribute C-multicast routes, an ASBR that has a given Inter-AS I-PMSI A-D route MAY discard traffic from a particular customer multicast source C-S and destined to a particular customer multicast group address C-G that is carried over the tunnel specified in the PMSI Tunnel attribute of the route, if none of the C-multicast routes on the ASBR with RD and Source AS being the same as the RD and Source AS of the Inter-AS I-PMSI A-D route matches the (C-S,C-G) tuple. A C-multicast route is said to match a (C-S,C-G) tuple, if it is a Source Tree Join route with Multicast Source set to C-S and Multicast Group set to C-G or a Shared Tree Join route with Multicast Group set to C-G.

The above procedures MAY also apply to an ASBR that originates a given Inter-AS I-PMSI A-D route. In this case, the ASBR applies them to the traffic carried over the tunnels specified in the PMSI Tunnel attribute of the Intra-AS I-PMSI A-D routes that the ASBR aggregates into the Inter-AS I-PMSI A-D route and whose tails are stitched to the one-hop ASBR-ASBR tunnel specified in the Inter-AS I-PMSI A-D route.

10. Non-Congruent Unicast and Multicast Connectivity

It is possible to deploy MVPN such that the multicast routing and the unicast routing are "non-congruent". For instance, the CEs may be distributing to the PEs a special set of unicast routes that are to be used exclusively for the purpose of upstream multicast hop selection, and not used for unicast routing at all. (For example, when BGP is the CE-PE unicast routing protocol, the CEs may be using SAFI 2 ("Network Layer Reachability Information used for multicast

forwarding" [IANA-SAFI]), and either IPv4 or IPv6 AFI to distribute a special set of routes that are to be used for, and only for, upstream multicast hop selection.) In such a situation, we will speak of the MVPN as having two VRFs on a given PE: one containing the routes that are used for unicast, the other containing the unicast routes that are used for UMH selection. We will call the former the "unicast routing VRF" and the latter the "UMH VRF" (upstream-multicast-hop VRF).

In this document, when we speak without qualification of the MVPN's VRF, then if the MVPN has both a unicast VRF and a UMH VRF, we are speaking of the UMH VRF. (Of course, if there is no separate UMH VRF, then we are speaking of the unicast VRF.)

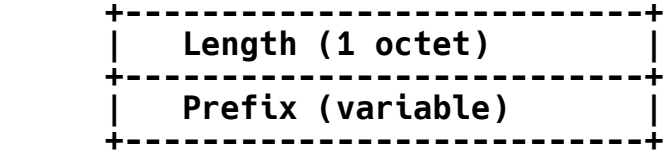
If there is a separate UMH VRF, it MAY have its own import and export Route Targets, different from the ones used by the unicast VRF. These Route Targets MUST be used to control distribution of auto-discovery routes. In addition, the export Route Targets of the UMH VRF are added to the Route Targets used by the unicast VRF when originating (unicast) VPN-IP routes. The import Route Targets associated with a given UMH VRF are used to determine which of the received (unicast) VPN-IP routes should be accepted into the UMH VRF.

If a PE maintains an UMH VRF for that MVPN, then it is RECOMMENDED that the UMH VRF use the same RD as the one used by the unicast VRF of that MVPN.

If an MVPN site is multihomed to several PEs, then to support non-congruent unicast and multicast connectivity, on each of these PEs, the UMH VRF of the MVPN MUST use its own distinct RD (although on a given PE, the RD used by the UMH VRF SHOULD be the same as the one used by the unicast VRF).

If an MVPN has a UMH VRF distinct from its unicast VRF, then one option to support non-congruency is to exchange the routes to/from that UMH VRF by using the same AFI/SAFI as used by the routes from the unicast VRF.

Another option is to exchange the routes to/from the UMH VRF using the IPv4 or IPv6 AFI (as appropriate), but with the SAFI set to SAFI 129 "Multicast for BGP/MPLS IP Virtual Private Networks (VPNs)" [IANA-SAFI]. The NLRI carried by these routes is defined as follows:



The use and the meaning of these fields are as follows:

a) Length:

The Length field indicates the length, in bits, of the address prefix.

b) Prefix:

The Prefix field contains a Route Distinguisher as defined in [RFC4364] prepended to an IPv4 or IPv6 address prefix, followed by enough trailing bits to make the end of the field fall on an octet boundary. Note that the value of trailing bits is irrelevant.

These routes MUST carry the VRF Route Import Extended Community. If, for a given MVPN, BGP is used for exchanging C-multicast routes, or if segmented inter-AS tunnels are used, then these routes MUST also carry the Source AS Extended Community.

The detailed procedures for selecting forwarder PE in the presence of such routes are outside the scope of this document. However, this document requires these procedures to preserve the constraints imposed by the single forwarder PE selection procedures, as specified in [MVPN].

11. Exchange of C-Multicast Routing Information among PEs

VPN C-Multicast Routing Information is exchanged among PEs by using C-multicast routes that are carried using an MCAST-VPN NLRI. These routes are originated and propagated as follows.

11.1. Originating C-Multicast Routes by a PE

Part of the procedures for constructing MCAST-VPN NLRI depends on the multicast routing protocol between CE and PE (C-multicast protocol).

11.1.1. Originating Routes: PIM as the C-Multicast Protocol

The following specifies the construction of MCAST-VPN NLRI of C-multicast routes for the case where the C-multicast protocol is PIM. These C-multicast routes are originated as a result of updates in the (C-S,C-G), or (C-*,C-G) state learned by a PE via the C-multicast protocol.

Note that creation and deletion of (C-S,C-G,rpt) states on a PE when the C-multicast protocol is PIM do not result in any BGP actions.

11.1.1.1. Originating Source Tree Join C-Multicast Route

Whenever (a) a C-PIM instance on a particular PE creates a new (C-S,C-G) state, and (b) the selected upstream PE for C-S (see [MVPN]) is not the local PE, then the local PE MUST originate a C-multicast route of type Source Tree Join. The Multicast Source field in the MCAST-VPN NLRI of the route is set to C-S; the Multicast Group field is set of C-G.

This C-multicast route is said to "correspond" to the C-PIM (C-S,C-G) state.

The semantics of the route are such that the PE has one or more receivers for (C-S,C-G) in the sites connected to the PE (the route has the (C-S,C-G) Join semantics).

Whenever a C-PIM instance on a particular PE deletes a (C-S,C-G) state, the corresponding C-multicast route MUST be withdrawn. (The withdrawal of the route has the (C-S,C-G) Prune semantics). The MCAST-VPN NLRI of the withdrawn route is carried in the MP_UNREACH_NLRI attribute.

11.1.1.2. Originating Shared Tree Join C-Multicast Route

Whenever (a) a C-PIM instance on a particular PE creates a new (C-*,C-G) state, and (b) the selected upstream PE for the C-RP corresponding to the C-G (see [MVPN]) is not the local PE, then the local PE MUST originate a C-multicast route of type Shared Tree Join. The Multicast Source field in the MCAST-VPN NLRI of the route is set to the C-RP address. The Multicast Group field in the MCAST-VPN NLRI is set to the C-G address.

This C-multicast route is said to "correspond" to the C-PIM (C-*,C-G) state.

The semantics of the route are such that the PE has one or more receivers for (C-*,C-G) in the sites connected to the PE (the route has the (C-*,C-G) Join semantics).

Whenever a C-PIM instance on a particular PE deletes a (C-*,C-G) state, the corresponding C-multicast route MUST be withdrawn. (The withdrawal of the route has the (C-S,C-G) Prune semantics). The MCAST-VPN NLRI of the withdrawn route is carried in the MP_UNREACH_NLRI attribute.

11.1.2. Originating Routes: mLDP as the C-Multicast Protocol

The following specifies the construction of the MCAST-VPN NLRI of C-multicast routes for the case where the C-multicast protocol is mLDP [mLDP].

Whenever a PE receives, from one of its CEs, a P2MP Label Map <X, Y, L> over interface I, where X is the Root Node Address, Y is the Opaque Value, and L is an MPLS label, the PE checks whether it already has state for <X, Y> in the VRF associated with the CE. If so, then all the PE needs to do in this case is to update its forwarding state by adding <I, L> to the forwarding state associated with <X, Y>.

If the PE does not have state for <X, Y> in the VRF associated with the CE, then the PE constructs a Source Tree Join C-multicast route whose MCAST-VPN NLRI contains X as the Multicast Source field, and Y as the Multicast Group field.

Whenever a PE deletes a previously created <X, Y> state that had resulted in originating a C-multicast route, the PE withdraws the C-multicast route. The MCAST-VPN NLRI of the withdrawn route is carried in the MP_UNREACH_NLRI attribute.

11.1.3. Constructing the Rest of the C-Multicast Route

The rest of the C-multicast route is constructed as follows (the same procedures apply to both PIM and mLDP as the C-Multicast protocol).

The local PE executes the procedures of [MVPN] to find the selected Upstream Multicast Hop (UMH) route and the selected upstream PE for the address carries in the Multicast Source field of MCAST-VPN NLRI.

From the selected UMH route, the local PE extracts (a) the ASN of the upstream PE (as carried in the Source AS Extended Community of the route), and (b) the C-multicast Import RT of the VRF on the upstream PE (the value of this C-multicast Import RT is the value of the VRF Route Import Extended Community carried by the route). The Source AS field in the C-multicast route is set to that AS. The Route Target Extended Community of the C-multicast route is set to that C-multicast Import RT.

If there is more than one (remote) PE that originates the (unicast) route to the address carried in the Multicast Source field of the MCAST-VPN NLRI, then the procedures for selecting the UMH route and the upstream PE to reach that address are as specified in [MVPN].

If the local and the upstream PEs are in the same AS, then the RD of the advertised MCAST-VPN NLRI is set to the RD of the VPN-IP route that contains the address carried in the Multicast Source field.

The C-multicast route is then advertised into IBGP.

If the local and the upstream PEs are in different ASes, then the local PE finds in its VRF an Inter-AS I-PMSI A-D route whose Source AS field carries the ASN of the upstream PE. The RD of the found Inter-AS I-PMSI A-D route is used as the RD of the advertised C-multicast route. The local PE constructs an IP-based Route Target Extended Community by placing the Next Hop of the found Inter-AS I-PMSI A-D route in the Global Administrator field of this Community, with the Local Administrator field of this Community set to 0; it then adds this Community to the Extended Communities attribute of the C-multicast route. (Note that this Route Target is the same as the ASBR Import RT of the ASBR identified by the Next Hop of the found Inter-AS I-PMSI A-D route.)

Inter-AS I-PMSI A-D routes are not used to support non-segmented inter-AS tunnels. To support non-segmented inter-AS tunnels, if the local and the upstream PEs are in different ASes, the local system finds in its VRF an Intra-AS I-PMSI A-D route from the upstream PE (the Originating Router's IP Address field of that route has the same value as the one carried in the VRF Route Import of the (unicast) route to the address carried in the Multicast Source field). The RD of the found Intra-AS I-PMSI A-D route is used as the RD of the advertised C-multicast route. The Source AS field in the C-multicast route is set to value of the Originating Router's IP Address field of the found Intra-AS I-PMSI A-D route.

The Next Hop field of the MP_REACH_NLRI attribute MUST be set to a routable IP address of the local PE.

If the Next Hop of the found (Inter-AS or Intra-AS) I-PMSI A-D route is an EBGP neighbor of the local PE, then the PE advertises the C-multicast route to that neighbor. If the Next Hop of the found (Inter-AS or Intra-AS) I-PMSI A-D route is within the same AS as the local PE, then the PE advertises the C-multicast route into IBGP.

11.1.4. Unicast Route Changes

The particular UMH route that is selected by a given PE for a given C-S may be influenced by the network's unicast routing. In that case, a change in the unicast routing may invalidate prior choices of the UMH route for some C-S. If this happens, the local PE MUST execute the UMH route selection procedures for C-S again. If the

result is that a different UMH route is selected, then for all C-G, any previously originated C-multicast routes for (C-S,C-G) MUST be re-originated.

Similarly, if a unicast routing change results in a change of the UMH route for a C-RP, then for all C-G such that C-RP is the RP associated with C-G, any previously originated C-multicast routes for (C-*,C-G) MUST be re-originated.

11.2. Propagating C-Multicast Routes by an ASBR

When an ASBR receives a BGP Update message that carries a C-multicast route, if at least one of the Route Targets of the route matches one of the ASBR Import RTs (auto-)configured on the ASBR, the ASBR finds an Inter-AS I-PMSI A-D route whose RD and Source AS matches the RD and Source AS carried in the C-multicast route. If no matching route is found, the ASBR takes no further action. If a matching route is found, the ASBR proceeds as follows.

To support non-segmented inter-AS tunnels, instead of matching the RD and Source AS carried in the C-multicast route against the RD and Source AS of an Inter-AS I-PMSI A-D route, the ASBR should match it against the RD and the Originating Router's IP Address of the Intra-AS I-PMSI A-D routes.

The ASBR first checks if it already has one or more C-multicast routes that have the same MCAST-VPN NLRI as the newly received route. If such a route(s) already exists, the ASBR keeps the newly received route, but SHALL NOT re-advertise the newly received route. Otherwise, the ASBR re-advertises the route, as described in this section.

When an ASBR receives a BGP Update message that carries a withdrawal of a previously advertised C-multicast route, the ASBR first checks if it already has at least one other C-multicast route that has the same MCAST-VPN NLRI. If such a route already exists, the ASBR processes the withdrawn route, but SHALL NOT re-advertise the withdrawal. Otherwise, the ASBR re-advertises the withdrawal of the previously advertised C-multicast route, as described below.

If the ASBR is the ASBR that originated the found Inter-AS I-PMSI A-D route, then before re-advertising the C-multicast route into IBGP, the ASBR removes from the route the Route Target that matches one of the ASBR Import RTs (auto-)configured on the ASBR.

If the ASBR is not the ASBR that originated the found Inter-AS I-PMSI A-D route, then before re-advertising the C-multicast route, the ASBR modifies the Extended Communities attribute of the C-multicast route

by replacing the Route Target of the route that matches one of the ASBR Import RTs (auto-)configured on the ASBR with a new Route Target constructed as follows. The new Route Target is an IP-based Route Target that has the Global Administrator field set to the Next Hop of the found Inter-AS I-PMSI A-D route, and Local Administrator field of this Community set to 0. Note that this newly constructed Route Target is the same as the ASBR Import RT of the ASBR identified by the Next Hop of the found Inter-AS I-PMSI A-D route. The rest of the Extended Communities attribute of the route SHOULD be passed unmodified.

The Next Hop field of the MP_REACH_NLRI attribute SHOULD be set to an IP address of the ASBR.

If the Next Hop field of the MP_REACH_NLRI of the found (Inter-AS or Intra-AS) I-PMSI A-D route is an EBGP neighbor of the ASBR, then the ASBR re-advertises the C-multicast route to that neighbor. If the Next Hop field of the MP_REACH_NLRI of the found (Inter-AS or Intra-AS) I-PMSI A-D route is an IBGP neighbor of the ASBR, the ASBR re-advertises the C-multicast route into IBGP. If it is the ASBR that originated the found Inter-AS I-PMSI A-D route in the first place, then the ASBR just re-advertises the C-multicast route into IBGP.

11.3. Receiving C-Multicast Routes by a PE

When a PE receives a C-multicast route the PE checks if any of the Route Target Extended Communities carried in the Extended Communities attribute of the route match any of the C-multicast Import RTs associated with the VRFs of any MVPN maintained by the PE. If no match is found, the PE SHOULD discard the route. Otherwise, (if a match is found), the PE checks if the address carried in the Multicast Source field of the C-multicast route matches one of the (unicast) VPN-IP routes advertised by PE from the VRF. If no match is found the PE SHOULD discard the route. Otherwise, (if a match is found), the PE proceeds as follows, depending on the multicast routing protocol between CE and PE (C-multicast protocol).

11.3.1. Receiving Routes: PIM as the C-Multicast Protocol

The following describes procedures when PIM is used as the multicast routing protocol between CE and PE (C-multicast protocol).

Since C-multicast routing information is disseminated by BGP, PIM messages are never sent from one PE to another.

11.3.1.1. Receiving Source Tree Join C-Multicast Route

If the received route has the Route Type set to Source Tree Join, then the PE creates a new (C-S,C-G) state in its MVPN-TIB from the Multicast Source and Multicast Group fields in the MCAST-VPN NLRI of the route, if such a state does not already exist.

If the local policy on the PE is to bind (C-S,C-G) to an S-PMSI, then the PE adds the S-PMSI to the outgoing interface list of the (C-S,C-G) state, if it is not already there. Otherwise, the PE adds an I-PMSI to the outgoing interface list of the (C-S,C-G) state, if it is not already there.

When, for a said VRF, the last Source Tree Join C-multicast route for (C-S,C-G) is withdrawn, resulting in the situation where the VRF contains no Source Tree Join C-multicast route for (C-S,C-G), the PE MUST remove the I-PMSI/S-PMSI from the outgoing interface list of the (C-S,C-G) state. Depending on the (C-S,C-G) state of the PE-CE interfaces, this may result in the PE using PIM procedures to prune itself off the (C-S,C-G) tree. If C-G is not in the SSM range for the VRF, then removing the I-PMSI/S-PMSI from the outgoing interface list of the (C-S,C-G) state SHOULD be done after a delay that is controlled by a timer. The value of the timer MUST be configurable.

The purpose of this timer is to ensure that the PE does not stop forwarding (C-S,C-G) onto a PMSI tunnel until all the PEs of the same MVPN have had time to receive the withdrawal of the Source Active A-D route for (C-S,C-G) (see Section 13.1), and the PE connected to C-RP starts forwarding (C-S,C-G) on the C-RPT.

Note that before the PE stops forwarding (C-S,C-G), there is a possibility to have (C-S,C-G) packets being sent at the same time on the PMSI by both the local PE and the PE connected to the site that contains C-RP. This would result in a transient unnecessary traffic on the provider backbone. However, no duplicates will reach customer hosts subscribed to C-G as long as the downstream PEs apply procedures described in Section 9.1 of [MVPN].

11.3.1.2. Receiving Shared Tree Join C-Multicast Route

If the received route has the Route Type set to Shared Tree Join, then the PE creates a new (C-*,C-G) state in its MVPN-TIB with the RP address for that state taken from the Multicast Source, and C-G for that state taken from the Multicast Group fields of the MCAST-VPN NLRI of the route, if such a state does not already exist. If there is no S-PMSI for (C-*,C-G), then the PE adds I-PMSI to the outgoing

interface list of the state if it is not already there. If there is an S-PMSI for (C-*,C-G), then the PE adds S-PMSI to the outgoing interface list of the state if it is not already there.

When, for a said VRF, the last Shared Tree Join C-multicast route for (C-*,C-G) is withdrawn, resulting in the situation where the VRF contains no Shared Tree Join C-multicast route for (C-*,C-G), the PE MUST remove the I-PMSI/S-PMSI from the outgoing interface list of the (C-*,C-G) state. Depending on the (C-*,C-G) state of the PE-CE interfaces, this may result in the PE using PIM procedures to prune itself off the (C-*,C-G) tree.

11.3.2. Receiving Routes: mLDP as the C-Multicast Protocol

The following describes procedures when mLDP is used as the multicast routing protocol between CE and PE (C-multicast protocol).

When mLDP is used as a C-multicast protocol, the only valid type of a C-multicast route that a PE could receive is a Source Tree Join C-multicast route.

When the PE receives a Source Tree Join C-multicast route, the PE applies, in the scope of this VRF, the P2MP mLDP procedures for a transit node using the value carried in the Multicast Source field of the route as the C-Root Node Identifier, and the value carried in the Multicast Group of the route as the C-LDP MP Opaque Value Element.

If there is no S-PMSI for <C-Root Node Identifier, C-LDP MP Opaque Value Element>, then the PE creates and advertises an S-PMSI as described in Section 12 using C-Root Node Identifier as the value for the Multicast Source field of the S-PMSI A-D route and C-LDP MP Opaque Value Element as the value for the Multicast Group field of the route.

To improve scalability when mLDP is used as the C-Multicast protocol for a given MVPN, within each AS that has sites of that MVPN connected to the PEs of that AS, all the S-PMSIs of that MVPN MAY be aggregated into a single P-multicast tree (by using upstream-assigned labels).

11.4. C-Multicast Routes Aggregation

Note that C-multicast routes are "de facto" aggregated by BGP. This is because the MCAST-VPN NLRIs advertised by multiple PEs, for a C-multicast route for a particular C-S and C-G (or a particular C-* and C-G) of a given MVPN are identical.

Hence, a BGP route reflector or ASBR that receives multiple such routes with the same NLRI will re-advertise only one of these routes to other BGP speakers.

This implies that C-multicast routes for a given (S,G) of a given MVPN originated by PEs that are clients of a given route reflector are aggregated by the route reflector. For instance, if multiple PEs that are clients of a route reflector, have receivers for a specific SSM channel of a MVPN, they will all advertise an identical NLRI for the Source Tree Join C-multicast route. However, only one C-multicast route will be advertised by the route reflector for this specific SSM channel of that MVPN, to other PEs and route reflectors that are clients of the route reflector.

This also implies that an ASBR aggregates all the received C-multicast routes for a given (S,G) (or a given (*,G)) of a given MVPN into a single C-multicast route.

To further reduce the routing churn due to C-multicast routes changes, a route reflector that re-advertises a C-multicast route SHOULD set the Next Hop field of the MP_REACH_NLRI attribute of the route to an IP address of the route reflector. Likewise, an ASBR that re-advertises a C-multicast route SHOULD set the Next Hop field of the MP_REACH_NLRI attribute of the route to an IP address of the ASBR.

Further, a BGP receiver, which receives multiple such routes with the same NLRI for the same C-multicast route, will potentially create forwarding state based on a single C-multicast route. Per the procedures described in Section 11.3, this forwarding state will be the same as the state that would have been created based on another route with same NLRI.

12. Using S-PMSI A-D Routes to Bind C-Trees to P-Tunnels

This section describes BGP-based procedures for using S-PMSIs A-D routes to bind (C-S,C-G) trees to P-tunnels.

12.1. Originating S-PMSI A-D Routes

The following describes procedures for originating S-PMSI A-D routes by a PE.

The PE constructs the MCAST-VPN NLRI of an S-PMSI A-D route for a given (C-S,C-G) as follows.

+ The RD in this NLRI is set to the RD of the MVPN's VRF associated with (C-S,C-G).

+ The Multicast Source field MUST contain the source address associated with the C-multicast stream, and the Multicast Source Length field is set appropriately to reflect this.

+ The Multicast Group field MUST contain the group address associated with the C-multicast stream, and the Multicast Group Length field is set appropriately to reflect this.

+ The Originating Router's IP Address field MUST be set to the IP address that the (local) PE places in the Global Administrator field of the VRF Route Import Extended Community of the VPN-IP routes advertised by the PE. Note that the <RD, Originating Router's IP address> tuple uniquely identifies a given multicast VRF.

The PE constructs the rest of the S-PMSI A-D route as follows.

Depending on the type of P-multicast tree used for the P-tunnel, the PMSI Tunnel attribute of the S-PMSI A-D route is constructed as follows:

+ The PMSI Tunnel attribute MUST contain the identity of the P-multicast tree (note that the PE could create the identity of the tree prior to the actual instantiation of the tree).

+ If, in order to establish the P-multicast tree, the PE needs to know the leaves of the tree within its own AS, then the PE obtains this information from the Leaf A-D routes received from other PEs/ASBRs within its own AS (as other PEs/ASBRs originate Leaf A-D routes in response to receiving the S-PMSI A-D route) by setting the Leaf Information Required flag in the PMSI Tunnel attribute to 1.

+ If a PE originates S-PMSI A-D routes with the Leaf Information Required flag in the PMSI Tunnel attribute set to 1, then the PE MUST be (auto-)configured with an import Route Target, which controls acceptance of Leaf A-D routes by the PE. (Procedures for originating Leaf A-D routes by the PEs that receive the S-PMSI A-D route are described in Section 12.3.)

  This Route Target is IP address specific. The Global Administrator field of this Route Target MUST be set to the IP address carried in the Next Hop of all the S-PMSI A-D routes advertised by this PE (if the PE uses different Next Hops, then the PE MUST be (auto-)configured with multiple import RTs, one per each such Next Hop). The Local Administrator field of this Route Target MUST be set to 0.

If the PE supports Route Target Constraint [RT-CONSTRAIN], the PE SHOULD advertise this import Route Target within its own AS using Route Target Constraints. To constrain distribution of the Route Target Constraint routes to the AS of the advertising PE, these routes SHOULD carry the NO_EXPORT Community [RFC1997].

+ A PE MAY aggregate two or more S-PMSIs originated by the PE onto the same P-multicast tree. If the PE already advertises S-PMSI A-D routes for these S-PMSIs, then aggregation requires the PE to re-advertise these routes. The re-advertised routes MUST be the same as the original ones, except for the PMSI Tunnel attribute. If the PE has not previously advertised S-PMSI A-D routes for these S-PMSIs, then the aggregation requires the PE to advertise (new) S-PMSI A-D routes for these S-PMSIs. The PMSI Tunnel attribute in the newly advertised/re-advertised routes MUST carry the identity of the P-multicast tree that aggregates the S-PMSIs. If at least some of the S-PMSIs aggregated onto the same P-multicast tree belong to different MVPNs, then all these routes MUST carry an MPLS upstream-assigned label [RFC5331].

If all these aggregated S-PMSIs belong to the same MVPN, and this MVPN uses PIM as its C-multicast routing protocol, then the corresponding S-PMSI A-D routes MAY carry an MPLS upstream-assigned label [RFC5331]. Moreover, in this case, the labels MUST be distinct on a per-MVPN basis and MAY be distinct on a per-route basis.

If all these aggregated S-PMSIs belong to the MVPN(s) that uses mLDP as its C-multicast routing protocol, then the corresponding S-PMSI A-D routes MUST carry an MPLS upstream-assigned label [RFC5331], and these labels MUST be distinct on a per-route (per-mLDP FEC) basis, irrespective of whether the aggregated S-PMSIs belong to the same or different MVPNs.

The Next Hop field of the MP_REACH_NLRI attribute of the route MUST be set to the same IP address as the one carried in the Originating Router's IP Address field.

The route always carries a set of Route Targets. The default set of Route Targets is determined as follows:

+ If there is a (unicast) VPN-IP route to C-S originated from the VRF, but no (unicast) VPN-IP route to C-RP originated from the VRF, then the set of Route Targets is formed by a set intersection between the set of Route Targets carried in the Intra-AS I-PMSI A-D route originated from the VRF and the set of Route Targets carried by the (unicast) VPN-IP route to C-S.

+ If there is no (unicast) VPN-IP route to C-S originated from the VRF, but there is a (unicast) VPN-IP route to C-RP originated from the VRF, then the set of Route Targets is formed by a set intersection between the set of Route Targets carried in the intra-AS I-PMSI A-D route originated from the VRF and the set of Route Targets carried by the (unicast) VPN-IP route to C-RP.

+ If there is a (unicast) VPN-IP route to C-S originated from the VRF, and a (unicast) VPN-IP route to C-RP originated from the VRF, then the set of Route Targets is formed by a set intersection between the set of Route Targets carried in the Intra-AS I-PMSI A-D route originated from the VRF and the set union of Route Targets carried by the (unicast) VPN-IP route to C-S and the (unicast) VPN-IP route to C-RP.

In each of the above cases, an implementation MUST allow the set of Route Targets carried by the route to be specified by configuration. In the absence of a configured set of Route Targets, the route MUST carry the default set of Route Targets, as specified above.

12.2. Handling S-PMSI A-D Routes by ASBRs

Procedures for handling an S-PMSI A-D route by ASBRs (both within and outside of the AS of the PE that originates the route) are the same as specified in Section 9.2.3, except that instead of Inter-AS I-PMSI A-D routes, the procedures apply to S-PMSI A-D routes.

12.2.1. Merging S-PMSI into an I-PMSI

Consider the situation where:

+ An ASBR is receiving (or expecting to receive) inter-AS (C-S,C-G) data from upstream via an S-PMSI.

+ The ASBR is sending (or expecting to send) the inter-AS (C-S,C-G) data downstream via an I-PMSI.

This situation may occur if the upstream providers have a policy of using S-PMSIs but the downstream providers have a policy of using I-PMSIs. To support this situation, an ASBR MAY, under certain conditions, merge one or more upstream S-PMSIs into a downstream I-PMSI.

An S-PMSI (corresponding to a particular S-PMSI A-D route) MAY be merged by a particular ASBR into an I-PMSI (corresponding to a particular Inter-AS I-PMSI A-D route) if and only if the following conditions all hold:

+ BGP is used to exchange C-multicast routes.

+ The S-PMSI A-D route and the Inter-AS I-PMSI A-D route originate in the same AS. The Inter-AS I-PMSI A-D route carries the originating AS in the Source AS field of the NLRI of the route and in the AS_PATH attribute of the route. The S-PMSI A-D route carries the originating AS in the AS_PATH attribute of the route.

+ The S-PMSI A-D route and the Inter-AS I-PMSI A-D route have exactly the same set of RTs.

+ For each (C-S,C-G) mentioned in the S-PMSI route, if the ASBR has installed a Source Tree Join (C-S,C-G) C-multicast route, then the S-PMSI route was originated by the upstream PE of the C-multicast route. The address of the upstream PE is carried in the RT of the C-multicast route. The address of the PE that originated the S-PMSI route is carried in the Originating Router's IP Address field of the MCAST-VPN NLRI of the route.

+ The ASBR supports the optional capability to discard (C-S,C-G) traffic received on an I-PMSI.

An ASBR performs merging by stitching the tail end of the P-tunnel, as specified in the PMSI Tunnel attribute of the S-PMSI A-D route received by the ASBR, to the head of the P-tunnel, as specified in the PMSI Tunnel attribute of the Inter-AS I-PMSI A-D route re-advertised by the ASBR.

IP processing during merge: If an ASBR merges a (C-S,C-G) S-PMSI A-D route into an Inter-AS I-PMSI A-D route, the ASBR MUST discard all (C-S,C-G) traffic it receives on the tunnel advertised in the I-PMSI A-D route.

An ASBR that merges an S-PMSI A-D route into an Inter-AS I-PMSI A-D route MUST NOT re-advertise the S-PMSI A-D route.

12.3. Receiving S-PMSI A-D Routes by PEs

Consider a PE that receives an S-PMSI A-D route. If one or more of the VRFs on the PE have their import Route Targets that contain one or more of the Route Targets carried by the received S-PMSI A-D route, then for each such VRF (and associated MVPN-TIB) the PE performs the following.

Procedures for receiving an S-PMSI A-D route by a PE (both within and outside of the AS of the PE that originates the route) are the same as specified in Section 9.2.3.4 except that (a) instead of Inter-AS

I-PMSI A-D routes, the procedures apply to S-PMSI A-D routes and (b) a PE performs procedures specified in that section only if, in addition to the criteria there, one of the following is true:

- the PE originates a Source Tree Join (C-S,C-G) C-multicast route, and the upstream PE of that route is the PE that originates the S-PMSI A-D route; or

- the PE does not originate a Source Tree Join (C-S,C-G) C-multicast route, but it originates a Shared Tree Join (C-*,C-G) C-multicast route. The best (as determined by the BGP route selection procedures) Source Active A-D route for (C-S,C-G) selected by the PE is originated by the same PE as the one that originates the S-PMSI A-D route; or

- the PE does not originate a Source Tree Join (C-S,C-G), has not received any Source Active A-D routes for (C-S,C-G), but does originate a Shared Tree Join (C-*,C-G) route. The upstream PE for that route is the PE that originates the received S-PMSI A-D route.

If the received S-PMSI A-D route has a PMSI Tunnel attribute with the Leaf Information Required flag set to 1, then the PE originates a Leaf A-D route. The Route Key of the Leaf A-D route is set to the MCAST-VPN NLRI of the S-PMSI A-D route. The rest of the Leaf A-D route is constructed using the same procedures as specified in section 9.2.3.4.1, except that instead of originating Leaf A-D routes in response to receiving Inter-AS I-PMSI A-D routes, the procedures apply to originating Leaf A-D routes in response to receiving S-PMSI A-D routes.

In addition to the procedures specified in Section 9.2.3.4.1, the PE MUST set up its forwarding path to receive (C-S,C-G) traffic from the tunnel advertised by the S-PMSI A-D route (the PE MUST switch to the S-PMSI).

If a PE that is a leaf node of a particular Selective tunnel determines that it no longer needs to receive any of (C-S,C-G)s carried over that tunnel, the PE SHOULD prune itself off that tunnel. Procedures for pruning are specific to a particular tunneling technology.

13. Switching from Shared a C-Tree to a Source C-Tree

The procedures defined in this section only apply when the C-multicast routing protocol is PIM [RFC4601]; moreover, they only apply for the multicast ASM mode and MUST NOT be applied to multicast

group addresses belonging to the SSM range. The procedures also MUST NOT be applied when the C-multicast routing protocol is BIDIR-PIM [RFC5015].

The procedures of this section are applicable only to MVPNs that use both shared (i.e., rooted at a C-RP) and source (i.e., rooted at a C-S) inter-site C-trees.

These procedures are not applicable to MVPNs that do not use shared inter-site C-trees and rely solely on source inter-site C-trees. See Section 14 for the procedures applicable to that scenario.

Whether or not a given MVPN uses both inter-site shared and source C-trees must be known a priori (e.g., via provisioning).

In the scenario where an MVPN customer switches from a C-RP-based tree (RPT) to the shortest path tree (SPT), in order to avoid packet duplication, choosing of a single consistent upstream PE, as described in [MVPN], may not suffice. To illustrate this, consider a set of PEs {PE2, PE4, PE6} that are on the C-RP tree for (C-*,C-G) and have chosen a consistent upstream PE, as described in [MVPN], for (C-*,C-G) state. Further, this upstream PE, say PE1, is using a Multidirectional Inclusive PMSI (MI-PMSI) for (C-*,C-G). If a site attached to one of these PEs, say PE2, switches to the C-S tree for (C-S,C-G), PE2 generates a Source Tree Join C-multicast route towards the upstream PE that is on the path to C-S, say PE3. PE3 also uses the MI-PMSI for (C-S,C-G), as PE1 uses for (C-*,C-G). This results in {PE2, PE4, PE6} receiving duplicate traffic for (C-S,C-G) -- both on the C-RP tree (from PE1) and C-S tree (from PE3). If it is desirable to suppress receiving duplicate traffic, then it is necessary to choose a single forwarder PE for (C-S,C-G). The following describes how this is achieved.

13.1. Source within a Site - Source Active Advertisement

When, as a result of receiving a Source Tree Join C-multicast route for (C-S,C-G) from some other PE the local PE adds either the S-PMSI or the I-PMSI to the outgoing interface list of the (C-S,C-G) state (see Section 11.3.1.1), the local PE MUST originate a Source Active A-D route if the PE has not originated such route already. The route carries a single MCAST-VPN NLRI constructed as follows:

+ The RD in this NLRI is set to the RD of the VRF of the MVPN on the PE.

+ The Multicast Source field MUST be set to C-S. The Multicast Source Length field is set appropriately to reflect this.

+ The Multicast Group field MUST be set to C-G. The Multicast Group Length field is set appropriately to reflect this.

The Next Hop field of the MP_REACH_NLRI attribute MUST be set to the IP address that the PE places in the Global Administrator field of the VRF Route Import Extended Community of the VPN-IP routes advertised by the PE from the MVPN's VRF.

The route SHOULD carry the same set of Route Targets as the Intra-AS I-PMSI A-D route of the MVPN originated by the PE.

Using the normal BGP procedures, the Source Active A-D route is propagated to all the PEs of the MVPN.

Note that the advertisement of a Source Active A-D route for a given (C-S,C-G) could be combined, if desired, with the advertisement of an S-PMSI A-D route for the same (C-S,C-G). This is accomplished by using the same BGP Update message to carry both the NLRI of the S-PMSI A-D route and the NLRI of the Source Active A-D route.

Note that even if the originating PE advertises both the Source Active A-D route and the S-PMSI A-D route in the same BGP Update message, an implementation cannot assume that all other PEs will receive both of these routes in the same Update message.

When, as a result of receiving a withdrawal of the previously advertised Source Tree Join C-multicast route for (C-S,C-G), the PE is going to remove the S-PMSI/I-PMSI from the outgoing interface list of the (C-S,C-G) state. The local PE MUST also withdraw the Source Active A-D route for (C-S,C-G), if such a route has been advertised.

Note that if the PE is also acting as a C-RP, but inter-site shared trees are being used, the reception of a PIM Register message by the PE does not result in the origination of a Source Active A-D route.

13.2. Receiving Source Active A-D Route

When a PE receives a new Source Active A-D route from some other PE, the PE finds a VRF whose import Route Targets match one or more of the Route Targets carried by the route. If the match is found, then the PE updates the VRF with the received route.

We say that a given (C-S,C-G) Source Active A-D route stored in a given VRF on a PE matches a given (C-*,C-G) entry present in the MVPN-TIB associated with the VRF if C-G carried by the route is the same as C-G of the entry, and the PE originates a Shared Tree Join C-multicast route for the same C-G as the C-G of the entry.

When (as a result of receiving PIM messages from one of its CEs) a PE creates in one of its MVPN-TIBs a (new) (C-*,C-G) entry with a non-empty outgoing interface list that contains one or more PE-CE interfaces, the PE MUST check if it has any matching Source Active A-D routes. If there is one or more such matching route, such that the PE does not have (C-S,C-G) state in its MVPN-TIB for (C-S,C-G) carried in the route, then the PE selects one of them (using the BGP route selection procedures), and sets up its forwarding path to receive (C-S,C-G) traffic from the tunnel that the originator of the selected Source Active A-D route uses for sending (C-S,C-G).

When, as a result of receiving a new Source Active A-D route, a PE updates its VRF with the route, the PE MUST check if the newly received route matches any (C-*,C-G) entries. If (a) there is a matching entry, (b) the PE does not have (C-S,C-G) state in its MVPN-TIB for (C-S,C-G) carried in the route, and (c) the received route is selected as the best (using the BGP route selection procedures), then the PE sets up its forwarding path to receive (C-S,C-G) traffic from the tunnel the originator of the selected Source Active A-D route uses for sending (C-S,C-G).

Note that if the PE is also acting as a C-RP, and inter-site shared trees are being used, the BGP Source Active A-D routes do not replace the Multicast Source Discovery Protocol (MSDP) or PIM-based Anycast RP peerings among C-RPs that would be needed to disseminate source discovery information among C-RPs.

13.2.1. Pruning Sources off the Shared Tree

In addition to the procedures in the previous section, a PE applies the following procedure when importing a Source Active A-D route for (C-S,C-G) into a VRF.

The PE finds a (C-*,C-G) entry in the MVPN-TIB whose C-G is the same as the C-G carried in the Multicast Group field of the Source Active A-D route.

If the outgoing interface list (oif) for the found (C-*,C-G) entry in the MVPN-TIB on the PE contains either I-PMSI or S-PMSI, and the PE does not originate the Source Tree Join C-multicast route for (C-S,C-G) (where C-S is address carried in the Multicast Source field and C-G is the address carried in the Multicast Group field of the received Source Active A-D route), then the PE MUST transition the (C-S,C-G,rpt) downstream state machine on I-PMSI/S-PMSI to the Prune state. (Conceptually, the C-PIM state machine on the PE will act "as if" it had received Prune (C-S,C-G,rpt) on I-PMSI/S-PMSI, without

actually having received one.) Depending on the (C-S,C-G,rpt) state of the PE-CE interfaces, this may result in the PE using PIM procedures to prune the C-S off the (C-*,C-G) tree.

Transitioning the state machine to the Prune state SHOULD be done after a delay that is controlled by a timer. The value of the timer MUST be configurable. The purpose of this timer is to ensure that the C-S is not pruned off the shared tree until all PEs have had time to receive the Source Active A-D route for (C-S,C-G).

Note that before C-S is pruned off the shared tree, there is a possibility to have (C-S,C-G) packets sent at the same time on the PMSI by distinct PEs. This would result in a transient unnecessary traffic on the provider backbone. However, no duplicates will reach customer hosts subscribed to C-G as long as the downstream PEs apply procedures described in Section 9.1 of [MVPN].

The PE MUST keep the (C-S,C-G,rpt) downstream state machine on I-PMSI/S-PMSI in the Prune state for as long as (a) the outgoing interface list (oif) for the found (C-*,C-G) entry in the MVPN-TIB on the PE contains either I-PMSI or S-PMSI, (b) the PE has at least one Source Active A-D route for (C-S,C-G), and (c) the PE does not originate the Source Tree Join C-multicast route for (C-S,C-G). Once any of these conditions become no longer valid, the PE MUST transition the (C-S,C-G,rpt) downstream state machine on I-PMSI/S-PMSI to the NoInfo state.

Note that changing the state on the downstream state machine on I-PMSI/S-PMSI, as described above, does not imply exchanging PIM messages over I-PMSI/S-PMSI.

Also, note that except for the scenario described in the third paragraph of this section, in all other scenarios relying solely on PIM procedures on the PE is sufficient to ensure the correct behavior when pruning sources off the shared tree.

14. Supporting PIM-SM without Inter-Site Shared C-Trees

The procedures defined in this section only apply when the C-multicast routing protocol is PIM [RFC4601]; moreover, only apply for the multicast ASM mode, and MUST NOT be applied to multicast group addresses belonging to the SSM range. The procedures also MUST NOT be applied when the C-multicast routing protocol is BIDIR-PIM [RFC5015].

The procedures of this section are applicable only to MVPNs that do not use inter-site shared (i.e., rooted at a C-RP) C-trees.

These procedures are not applicable to MVPNs that use both shared and shortest path inter-site C-trees. See Section 13 for the procedures applicable to that scenario.

Whether or not a given MVPN uses inter-site shared C-trees must be known a priori (e.g., via provisioning).

14.1. Discovering Active Multicast Sources

A PE can obtain information about active multicast sources within a given MVPN in a variety of ways. One way is for the PE to act as a fully functional customer RP (C-RP) for that MVPN. Another way is to use PIM Anycast RP procedures [PIM-ANYCAST-RP] to convey information about active multicast sources from one or more of the MVPN C-RPs to the PE. Yet another way is to use MSDP [MSDP] to convey information about active multicast sources from the MVPN C-RPs to the PE.

When a PE using any of the above methods first learns of a new (multicast) source within that MVPN, the PE constructs a Source Active A-D route and sends this route to all other PEs that have one or more sites of that MVPN connected to them. The route carries a single MCAST-VPN NLRI constructed as follows:

+ The RD in this NLRI is set to the RD of the VRF of the MVPN on the PE.

+ The Multicast Source field MUST be set to the source IP address of the multicast data packet carried in the PIM Register message (RP/PIM register case) or of the MSDP Source-Active message (MSDP case). The Multicast Source Length field is set appropriately to reflect this.

+ The Multicast Group field MUST be set to the group IP address of the multicast data packet carried in the PIM Register message (RP/PIM register case) or of the MSDP Source-Active message (MSDP case). The Multicast Group Length field is set appropriately to reflect this.

The Next Hop field of the MP_REACH_NLRI attribute MUST be set to the IP address that the PE places in the Global Administrator field of the VRF Route Import Extended Community of the VPN-IP routes advertised by the PE.

The route SHOULD carry the same set of Route Targets as the Intra-AS I-PMSI A-D route of the MVPN originated by the PE.

Using the normal BGP procedures, the Source Active A-D route is propagated to all the PEs of the MVPN.

When a PE that previously advertised a Source Active A-D route for a given (multicast) source learns that the source is no longer active (the PE learns this by using the same mechanism by which the PE learned that the source was active), the PE SHOULD withdraw the previously advertised Source Active route.

14.2. Receiver(s) within a Site

A PE follows the procedures specified in Section 11.1, except that the procedures specified in Section 11.1.1.2 are replaced with the procedures specified in this section.

When a PE receives a new Source Active A-D route, the PE finds a VRF whose import Route Targets match one or more of the Route Targets carried by the route. If the match is found, then the PE updates the VRF with the received route.

We say that a given (C-S,C-G) Source Active A-D route stored in a given VRF matches a given (C-*,C-G) entry present in the MVPN-TIB associated with the VRF if C-G carried by the route is the same as C-G of the entry.

When (as a result of receiving PIM messages from one of its CEs) a PE creates, in one of its MVPN-TIBs, a (new) (C-*,C-G) entry with a non-empty outgoing interface list that contains one or more PE-CE interfaces, the PE MUST check if it has any matching Source Active A-D routes. If there is one or more such matching routes, and the best path to C-S carried in the matching route(s) is reachable through some other PE, then for each such route the PE MUST originate a Source Tree Join C-multicast route. If there is one or more such matching routes, and the best path to C-S carried in the matching route(s) is reachable through a CE connected to the PE, then for each such route the PE MUST originate a PIM Join (C-S,C-G) towards the CE.

When, as a result of receiving a new Source Active A-D route, a PE updates its VRF with the route, the PE MUST check if the newly received route matches any (C-*,C-G) entries. If there is a matching entry, and the best path to C-S carried in the (A-D) route is reachable through some other PE, the PE MUST originate a Source Tree Join C-multicast route for the (C-S,C-G) carried by the route. If there is a matching entry, and the best path to C-S carried in the (A-D) route is reachable through a CE connected to the PE, the PE MUST originate a PIM Join (C-S,C-G) towards the CE.

Construction and distribution of the Source Tree Join C-multicast route follows the procedures specified in Section 11.1.1.1, except that the Multicast Source Length, Multicast Source, Multicast Group

Length, and Multicast Group fields in the MCAST-VPN NLRI of the Source Tree Join C-multicast route are copied from the corresponding field in the Source Active A-D route.

A PE MUST withdraw a Source Tree Join C-multicast route for (C-S,C-G) if, as a result of having received PIM messages from one of its CEs, the PE creates a Prune (C-S,C-G,rpt) upstream state in one of its MVPN-TIBs but has no (C-S,C-G) Joined state in that MVPN-TIB and had previously advertised the said route. (This is even if the VRF associated with the MVPN-TIB still has a (C-S,C-G) Source Active A-D route.)

A PE MUST withdraw a Source Tree Join C-multicast route for (C-S,C-G) if the Source Active A-D route that triggered the advertisement of the C-multicast route is withdrawn.

When a PE deletes the (C-*,C-G) state (e.g., due to receiving PIM Prune (C-*,C-G) from its CEs), the PE MUST withdraw all the Source Tree Join C-multicast routes for C-G that have been advertised by the PE, except for the routes for which the PE still maintains the corresponding (C-S,C-G) state.

Even though PIM is used as a C-multicast protocol, procedures described in Section 11.1.1.2 do not apply here, as only the Source Tree Join C-multicast routes are exchanged among PEs.

14.3. Receiving C-Multicast Routes by a PE

In this model, the only valid type of a C-multicast route that a PE could receive is a Source Tree Join C-multicast route. Processing of such a route follows the procedures specified in Section 11.3.1.1.

15. Carrier's Carrier

A way to support the Carrier's Carrier model is provided by using mLDP as the CE-PE multicast routing and label distribution protocol, as specified in this document.

To improve scalability, it is RECOMMENDED that for the Carrier's Carrier scenario within an AS, all the S-PMSIs of a given MVPN be aggregated into a single P-multicast tree (by using upstream-assigned labels).

16. Scalability Considerations

A PE should use Route Target Constraint [RT-CONSTRAIN] to advertise the Route Targets that the PE uses for the VRF Route Imports Extended Community (note that doing this requires just a single Route Target

Constraint advertisement by the PE). This allows each C-multicast route to reach only the relevant PE, rather than all the PEs participating the an MVPN.

To keep the intra-AS membership/binding information within the AS of the advertising router the BGP Update message originated by the advertising router SHOULD carry the NO_EXPORT Community [RFC1997].

An Inter-AS I-PMSI A-D route originated by an ASBR aggregates Intra-AS I-PMSI A-D routes originated within the ASBR's own AS. Thus, while the Intra-AS I-PMSI A-D routes originated within an AS are at the granularity of <PE, MVPN> within that AS, outside of that AS the (aggregated) Inter-AS I-PMSI A-D routes are at the granularity of <AS, MVPN>. An Inter-AS I-PMSI A-D route for a given <AS, MVPN> indicates the presence of one or more sites of the MVPN connected to the PEs of the AS.

For a given inter-AS tunnel, each of its intra-AS segments could be constructed by its own mechanism. Moreover, by using upstream-assigned labels within a given AS, multiple intra-AS segments of different inter-AS tunnels of either the same or different MVPNs may share the same P-multicast tree.

Since (aggregated) Inter-AS I-PMSI A-D routes may have a granularity of <AS, MVPN>, an MVPN that is present in N ASes would have total of N inter-AS tunnels. Thus, for a given MVPN, the number of inter-AS tunnels is independent of the number of PEs that have this MVPN.

Within each Autonomous System, BGP route reflectors can be partitioned among MVPNs present in that Autonomous System so that each partition carries routes for only a subset of the MVPNs supported by the service provider. Thus, no single route reflector is required to maintain routes for all MVPNs. Moreover, route reflectors used for MVPN do not have to be used for VPN-IP routes (although they may be used for VPN-IP routes as well).

As described in Section 11.4, C-multicast routes for a given (S,G) of a given MVPN originated by PEs that are clients of a given route reflector are aggregated by the route reflector. Therefore, even if, within a route reflector cluster, there are multiple C-multicast routes for a given (S,G) of a given MVPN, outside of the cluster, all these routes are aggregated into a single C-multicast route. Additional aggregation of C-multicast routes occurs at ASBRs, where an ASBR aggregates all the received C-multicast routes for a given (S,G) of a given MVPN into a single C-multicast route. Moreover, both route reflectors and ASBRs maintain C-multicast routes only in the control plane, but not in the data plane.

16.1. Dampening C-Multicast Routes

The rate of C-multicast routing changes advertised by a PE is not necessarily directly proportional to the rate of multicast routing changes within the MVPN sites connected to the PE, as after the first (C-S,C-G) Join originated within a site, all the subsequent Joins for same (C-S,C-G) originated within the sites of the same MVPN connected to the PE do not cause origination of new C-multicast routes by the PE.

Depending on how multicast VPN is engineered, dynamic addition and removal of P2MP RSVP-TE leaves through advertisement/withdrawal of Leaf A-D routes will happen. Dampening techniques can be used to limit corresponding processing.

To lessen the control plane overhead associated with the processing of C-multicast routes, this document proposes OPTIONAL route dampening procedures similar to what is described in [RFC2439]. The following OPTIONAL procedures can be enabled on a PE, ASBR, or BGP Route Reflector advertising or receiving C-multicast routes.

16.1.1. Dampening Withdrawals of C-Multicast Routes

A PE/ASBR/route reflector can OPTIONALLY delay the advertisement of withdrawals of C-multicast routes. An implementation SHOULD provide the ability to control the delay via a configurable timer, possibly with some backoff algorithm to adapt the delay to multicast routing activity.

Dampening of withdrawals of C-multicast routes does not impede the multicast Join latency observed by MVPN customers, and it also does not impede the multicast leave latency observed by a CE, as multicast forwarding from the VRF will stop as soon as C-multicast state is removed in the VRF.

The potential drawbacks of dampening of withdrawals of C-multicast routes are as follows:

+ Until the withdrawals are actually sent, multicast traffic for the C-multicast routes in question will be continued to be transmitted to the PE, which will just have to discard it. Note that the PE may receive useless (multicast) traffic anyway, irrespective of dampening of withdrawals of C-multicast routes due to the use of I-PMSIs.

+ Any state in the upstream PEs that would be removed as a result of processing the withdrawals will remain until the withdrawals are sent.

Discussion on whether the potential drawbacks mentioned above are of any practical significance is outside the scope of this document.

16.1.2. Dampening Source/Shared Tree Join C-Multicast Routes

A PE/ASBR/route reflector can OPTIONALLY delay the advertisement of Source/Shared Tree Join C-multicast routes. An implementation SHOULD provide the ability to control the delay via a configurable timer, possibly with some backoff algorithm to adapt the delay to multicast routing activity.

Dampening Source/Shared Tree Join C-multicast routes will not impede multicast Join latency observed by a given MVPN, except if the PE advertising the Source/Shared Tree Join C-multicast route for a particular C-S/C-RP is the first to do so for all the sites of the MVPN that share the same upstream PE with respect to the C-S/C-RP.

16.2. Dampening Withdrawals of Leaf A-D Routes

Similar to the procedures proposed above for withdrawal of C-multicast routes, dampening can be applied to the withdrawal of Leaf A-D routes.

17. Security Considerations

The mechanisms described in this document could reuse the existing BGP security mechanisms [RFC4271] [RFC4272]. The security model and threats specific to Provider Provisioned VPNs, including L3VPNs, are discussed in [RFC4111]. [MVPN] discusses additional threats specific to the use of multicast in L3VPNs. There is currently work in progress to improve the security of TCP authentication. When the document is finalized as an RFC, the method defined in [RFC5925] SHOULD be used where authentication of BGP control packets is needed.

A PE router MUST NOT accept, from CEs routes, with MCAST-VPN SAFI.

If BGP is used as a CE-PE routing protocol, then when a PE receives a route from a CE, if this route carries the VRF Route Import Extended Community, the PE MUST remove this Community from the route before turning it into a VPN-IP route. Routes that a PE advertises to a CE MUST NOT carry the VRF Route Import Extended Community.

It is important to protect the control plane resources within the PE to prevent any one VPN from hogging excessive resources. This is the subject of the remainder of the Security Considerations section.

When C-multicast routing information is exchanged among PEs using BGP, an implementation SHOULD provide the ability to rate limit BGP messages used for this exchange. This SHOULD be provided on a per-PE, per-MVPN granularity.

An implementation SHOULD provide capabilities to impose an upper bound on the number of S-PMSI A-D routes, as well as on how frequently they may be originated. This SHOULD be provided on a per-PE, per-MVPN granularity.

In conjunction with the procedures specified in Section 14, an implementation SHOULD provide capabilities to impose an upper bound on the number of Source Active A-D routes, as well as on how frequently they may be originated. This SHOULD be provided on a per-PE, per-MVPN granularity.

In conjunction with the procedures specified in Section 13 limiting the amount of (C-S,C-G) state would limit the amount of Source Active A-D route, as in the context of this section, Source Active A-D routes are created in response to Source Tree Join C-multicast routes, and Source Tree Join C-multicast routes are created as a result of creation of (C-S,C-G) state on PEs. However, to provide an extra level of robustness in the context of these procedures, an implementation MAY provide capabilities to impose an upper bound on the number of Source Active A-D routes, as well as on how frequently they may be originated. This MAY be provided on a per-PE, per-MVPN granularity.

Section 16.1.1 describes optional procedures for dampening withdrawals of C-multicast routes. It is RECOMMENDED that an implementation support such procedures.

Section 16.1.1 describes optional procedures for dampening withdrawals of Leaf A-D routes. It is RECOMMENDED that an implementation support such procedures.

18. IANA Considerations

This document defines a new BGP Extended Community called "Source AS". This Community is of an extended type and is transitive. The Type value for this Community has been allocated from the two-octet AS-Specific Extended Community registry as 0x0009 and from the four-octet AS-Specific Extended Community registry as 0x0209.

This document defines a new BGP Extended Community called "VRF Route Import" (Type value 0x010b). This Community is IP address specific, of an extended type, and is transitive.

This document defines a new NLRI, called "MCAST-VPN", to be carried in BGP using multiprotocol extensions. It has been assigned SAFI 5. Also, SAFI 129 has been assigned to "Multicast for BGP/MPLS IP Virtual Private Networks (VPNs)".

This document defines a new BGP optional transitive attribute, called "PMSI_TUNNEL". IANA has assigned the codepoint 22 in the "BGP Path Attributes" registry to the PMSI_TUNNEL attribute.

This document defines a new BGP optional transitive attribute, called "PE Distinguisher Labels". IANA has assigned the codepoint 27 in the "BGP Path Attributes" registry to the PE Distinguisher Labels attribute.

19. Acknowledgements

We would like to thank Chaitanya Kodeboniya for helpful discussions. We would also like to thank members of the L3VPN IETF Working Group for insightful comments and review.

20. References

20.1. Normative References

[IANA-SAFI] IANA, "Subsequent Address Family Identifiers (SAFI) Parameters", http://www.iana.org.

[MVPN] Rosen, E., Ed. and R. Aggarwal, Ed., "Mulitcast in MPLS/BGP IP VPNs", RFC 6513, February 2012.

[RFC1997] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, August 1996.

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

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC 4601, August 2006.

[RFC4659] De Clercq, J., Ooms, D., Carugi, M., and F. Le Faucheur, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", RFC 4659, September 2006.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.

20.2. Informative References

[mLDP] Wijnands, IJ., Ed., Minei, I., Ed., Kompella, K., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", RFC 6388, November 2011.

[MSDP] Fenner, B., Ed., and D. Meyer, Ed., "Multicast Source Discovery Protocol (MSDP)", RFC 3618, October 2003.

[PIM-ANYCAST-RP] Farinacci, D. and Y. Cai, "Anycast-RP Using Protocol Independent Multicast (PIM)", RFC 4610, August 2006.

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, August 2008.

[RT-CONSTRAIN] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, November 2006.

[RFC2439] Villamizar, C., Chandra, R., and R. Govindan, "BGP Route Flap Damping", RFC 2439, November 1998.

[RFC4111] Fang, L., Ed., "Security Framework for Provider-Provisioned Virtual Private Networks (PPVPNs)", RFC 4111, July 2005.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, January 2006.

[RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", RFC 4607, August 2006.

[RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.

[RFC5015] Handley, M., Kouvelas, I., Speakman, T., and L. Vicisano, "Bidirectional Protocol Independent Multicast (BIDIR-PIM)", RFC 5015, October 2007.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.

Authors' Addresses

Rahul Aggarwal
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089
EMail: raggarwa_1@yahoo.com

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

Thomas Morin
France Telecom - Orange
2, avenue Pierre-Marzin
22307 Lannion Cedex
France
EMail: thomas.morin@orange.com

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