Network Working Group
Internet Draft
Expiration Date: December 2006

R. Aggarwal
Juniper Networks

E. Rosen
Cisco Systems, Inc.

T. Morin
France Telecom

Y. Rekhter
Juniper Networks

C. Kodeboniya

June 2006

BGP Encodings for Multicast in MPLS/BGP IP VPNs

draft-raggarwa-l3vpn-2547bis-mcast-bgp-02.txt

Status of this Memo

Abstract

This document describes the BGP encodings for signaling the information elements required by Multicast in MPLS/BGP IP VPNs, as specified in [MVPN].

Table of Contents

1. Specification of requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Terminology

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

+ intra-AS auto-discovery route (auto-discovery route);

+ inter-AS auto-discovery route;

+ S-PMSI auto-discovery route;

+ intra-AS segment leaf auto-discovery route (leaf auto-discovery route);

+ Source Active auto-discovery route (SA auto-discovery 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;

+ Prune Source from Shared Tree 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 [PIM-SM], except that instead of a single TIB a PE maintains multiple MVPN-TIBs, one per each MVPN.

3. Introduction

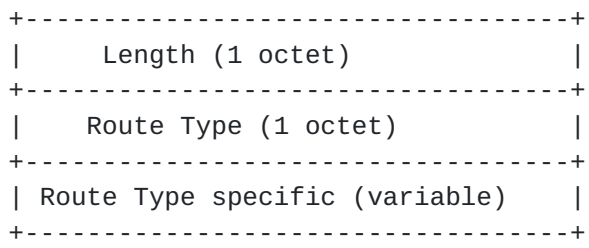
This document describes the BGP encodings for exchanging the information elements required by Multicast in MPLS/BGP IP VPNs, as specified in [MVPN]. This document assumes a thorough familiarity with procedures, concepts and terms described in [MVPN].

This document defines a new NLRI, MCAST-VPN NLRI. The MCAST-VPN NLRI is used for MVPN auto-discovery, advertising MVPN - I-PMSI tunnel binding, advertising <C-S, C-G> - S-PMSI tunnel binding, VPN customer multicast routing information exchange among PEs, choosing a single forwarder PE, and for procedures in support of co-locating a C-RP on a PE.

This document also specifies new BGP attribute, P-Multicast Service Interface Tunnel (PMSI Tunnel) attribute.

4. MCAST-VPN NLRI

This document defines a new BGP NLRI, called the MCAST-VPN NLRI. This NLRI is carried in BGP using BGP Multiprotocol Extensions [RFC2858]. Following is the format of the MCAST-VPN NLRI:



The Length field indicates the length in octets of MCAST-VPN NLRI, excluding the Length field itself.

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

This document defines the following Route Types for auto-discovery routes:

+ 1 - Intra-AS I-PMSI auto-discovery route (or just auto-discovery route);
+ 2 - Inter-AS I-PMSI auto-discovery route (or just inter-AS auto-discovery route);
+ 3 - S-PMSI auto-discovery route;
+ 4 - Intra-AS segment leaf auto-discovery route (or just leaf auto-discovery route).
+ 5 - Source Active auto-discovery route (or just SA auto-discovery route).

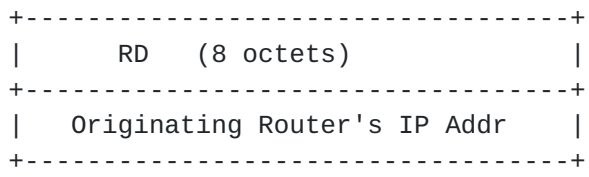
This document defines the following Route Types for C-multicast routes:

+ 6 - Shared Tree Join route;
+ 7 - Source Tree Join route;
+ 8 - Prune Source from Shared Tree route.

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 auto-discovery route

An intra-AS I-PMSI auto-discovery route type specific MCAST-VPN NLRI consists of the following:

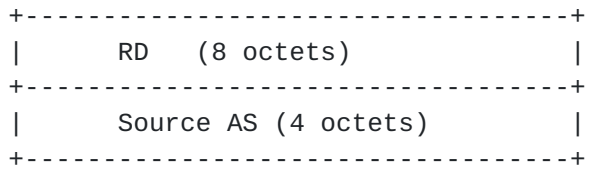


The RD is encoded as described in [RFC4364].

Usage of intra-AS I-PMSI auto-discovery routes is described in Section "MVPN Auto-Discovery/Binding - Intra-AS Operations".

4.2. Inter-AS I-PMSI auto-discovery route

An inter-AS I-PMSI auto-discovery route type specific MCAST-VPN NLRI consists of the following:



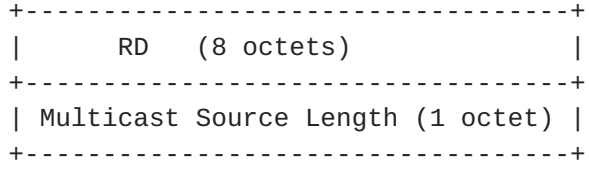
The RD is encoded as described in [RFC4364].

The Source AS contains an Autonomous System number. Two octets AS numbers are encoded in the low order two octets of the Source AS field.

Usage of inter-AS I-PMSI auto-discovery routes is described in Section "MVPN Auto-Discovery/Binding - Inter-AS Operations".

4.3. S-PMSI auto-discovery route

An S-PMSI auto-discovery route type specific MCAST-VPN NLRI consists of the following:



```
|  Multicast Source (Variable)      |
+-----------------------------------+
|  Multicast Group Length (1 octet) |
+-----------------------------------+
|  Multicast Group   (Variable)     |
+-----------------------------------+
|   Originating Router's IP Addr    |
+-----------------------------------+
```

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 Group Address 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.

Usage of S-PMSI auto-discovery routes is described in Section "Switching to S-PMSI".

4.4. Leaf auto-discovery route

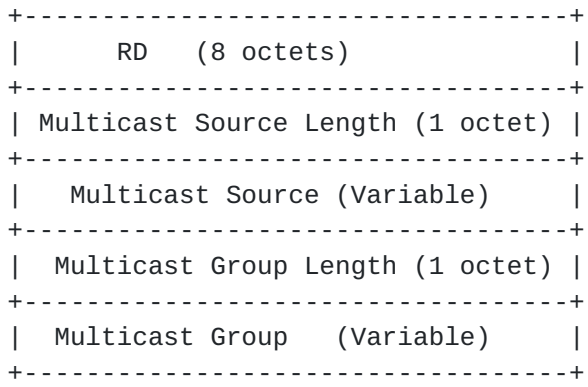
A leaf auto-discovery route type specific MCAST-VPN NLRI consists of the following:

```
+-----------------------------------+
|      Route Key (variable)         |
+-----------------------------------+
|   Originating Router's IP Addr    |
+-----------------------------------+
```

Usage of Leaf auto-discovery routes is described in Section "MVPN Auto-Discovery/Binding - Inter-AS Operations".

4.5. Source Active auto-discovery route (SA auto-discovery route)

A Source Active (SA) auto-discovery route type specific MCAST-VPN NLRI consists of the following:



The RD is encoded as described in [RFC4364].

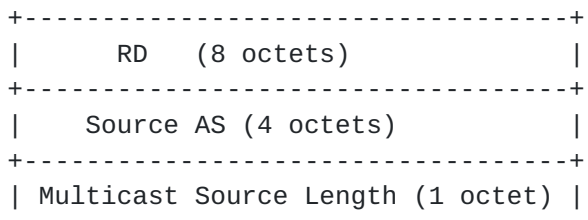
The Multicast Source field contains the C-S address. If the Multicast Source field contains an IPv4 address, then the value of the Multicast Source Length field is 32. If the Multicast Source field contains an IPv6 address, then the value of the Multicast Source Length field is 128.

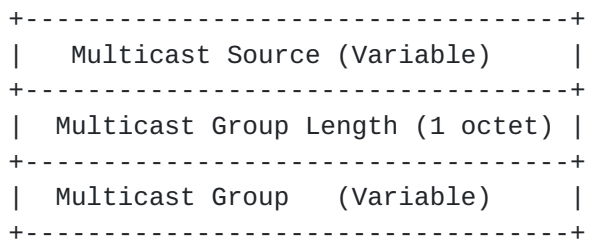
The Group Address field contains the C-G address. If the Multicast Group field contains an IPv4 address, then the value of the Multicast Group Length field is 32. If the Multicast Group field contains an IPv6 address, then the value of the Multicast Group Length field is 128.

Usage of Source Active auto-discovery routes is described in Sections "Choosing a single forwarder PE when switching from RPT to SPT", and "Co-locating C-RPs on a PE".

4.6. C-Multicast Route

A Shared Tree Join route, a Source Tree Join route, and a Prune Source from Shared Tree route type specific MCAST-VPN NLRI have the following encoding:





The RD is encoded as described in [RFC4364].

The Source AS contains an Autonomous System number. Two octets AS numbers are encoded in the low order two octets of the Source AS field.

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 Source field contains the C-S address.

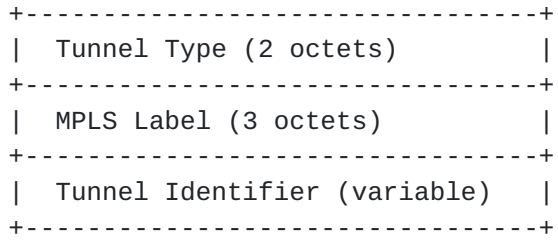
If the Multicast Group field contains an IPv4 address, then the value of the Multicast Group Length field is 32. If the Multicast Group field contains an IPv6 address, then the value of the Multicast Group Length field is 128.

Usage of C-multicast routes is described in Section "VPN C-Multicast Routing Information Exchange among PEs".

The Group Address field contains the C-G address.

5. P-Multicast Service Interface Tunnel (PMSI Tunnel) attribute

This document defines and uses a new BGP attribute, called 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 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:

+ 1 - Leaf Information Required
+ 2 - RSVP-TE P2MP LSP
+ 3 - LDP P2MP LSP
+ 4 - PIM-SSM Tree
+ 5 - PIM-SM Tree
+ 6 - PIM-Bidir Tree
+ 7 - Ingress Replication
+ 8 - LDP 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 MPLS Label is indicated by setting the MPLS Label field to zero.

When the type is set to Leaf Information Required, the Tunnel Identifier field is empty (zero length), and the MPLS Label field is zero.

When the type is set to RSVP-TE P2MP LSP, the Tunnel Identifier contains the RSVP-TE P2MP LSP's SESSION Object.

When the type is set to LDP P2MP LSP, the Tunnel Identifier is <P-Root Node Address, Variable length opaque identifier>.

When the type is set to PIM-SM Tree, the Tunnel Identifier MUST include <Sender Address, P-Multicast Group>, and MAY include P-RP Node Address. 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 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.

When the type is set to PIM-Bidir 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 type is set to Ingress Replication the Tunnel Identifier carries the unicast tunnel endpoint.

When the type is set to LDP MP2MP LSP, the Tunnel Identifier is <P-

Root Node Address, Variable length opaque identifier>.

If LDP MP2MP LSPs are used as PMSI tunnels, the router that transmitted a given packet into the tunnel cannot be identified. As a result, LDP MP2MP LSPs do not support aggregation, and therefore can only be used as unaggregated tunnels. Support of aggregation with LDP MP2MP LSPs is a matter for further study. In addition, if an LDP MP2MP LSP is used within a given AS as an intra-AS segment of an inter-AS tunnel, a single ASBR within that AS must be chosen to be the one which transmits packets to and from the upstream segment of the inter-AS tunnel. In the absence of the procedures for doing this, LDP MP2MP LSPs can not be used for intra-AS segments of inter-AS tunnels. Procedures for choosing a single ASBR are a matter for further study. Finally, use of LDP MP2MP LSPs makes choosing a single forwarder PE (see section "Choosing a single forwarder PE") mandatory.

The PMSI Tunnel attribute is only used in conjunction with intra-AS and inter-AS I-PMSI auto-discovery routes, with S-PMSI auto-discovery routes, and with leaf auto-discovery routes.

6. Source AS Extended Community

This document defines a new extended community called Source AS.

The Source AS is an AS specific extended community.

The Source AS extended community is of an extended type, and is transitive across AS boundaries.

To support MVPN a PE that originates a (unicast) route to VPN-IPv4 addresses MUST include in the BGP Update message that carries this route the Source AS extended community, except if it is known a priori that none of these addresses will act as multicast sources and/or RP, in which case the (unicast) route need not carry the Source AS extended community. The Global Administrator field of this community MUST be set to the autonomous system number of the PE. The Local Administrator field of this community SHOULD be set to 0.

Usage of the Source AS extended community is described in Sections "PIM as the C-Multicast protocol", "mLDP as the C-Multicast protocol", and "Receiving Source Active auto-discovery route".

7. Route Import Extended Community

This document defines a new extended community called Route Import.

The Route Import is an IPv4 address specific extended community.

The Route Import is of an extended type, and is transitive across AS boundaries.

To support MVPN in addition to the import/export Route Target(s) used by the unicast routing, each VRF on a PE MUST have an import Route Target that is unique to this VRF, except if it is known a priori that none of the (local) MVPN sites associated with the VRF contain multicast source(s) and/or RP, in which case the VRF need not have this import Route Target. This Route Target MUST be IP address specific, and is constructed as follows:

+ The Global Administrator field of the Route Target MUST be set to an IP address of the PE. This address MUST be a routable IP address. This address MAY be common for all the VRFs on the PE (e.,g., this address may be PE's loopback address).

+ The Local Administrator field of the Route Target associated with a given VRF contains a 2 octets long 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 outside the scope of this document).

A PE that originates a (unicast) route to VPN-IPv4 addresses MUST include in the BGP Updates message that carries this route the Route Import extended community that has the value of this Route Target, except if it is known a priori that none of these addresses will act as multicast sources and/or RP, in which case the (unicast) route need not carry the Route Import extended community.

If a PE uses Route Target Constrain [RT-CONSTRAIN], the PE SHOULD advertise all such import Route Targets using Route Target Constrains (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 Constrain routes to the AS of the advertising PE these routes SHOULD carry the NO_EXPORT Community ([RFC1997]).

Usage of Route Import extended community is described in Sections "PIM as the C-Multicast protocol", "mLDP as the C-Multicast protocol", and "Receiving Source Active auto-discovery route".

8. 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.

MVPN auto-discovery/binding consists of two components: intra-AS and inter-AS. The former provides MVPN auto-discovery/binding within a single AS. The latter provides MVPN auto-discovery/binding across multiple ASes.

Note that the inter-AS component applies only to the ASes that use either option (b) or (c) for unicast inter-AS operations, as specified in Section "Multi-AS Backbones" of [BGP-VPN]. ASes that use option (a) need only the intra-AS component.

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

This section describes exchanges of auto-discovery routes originated/received by PEs within the same AS.

8.1.1. Originating (intra-AS) auto-discovery routes

To participate in the MVPN auto-discovery/binding a PE router that has a given VRF of a given MVPN MUST originate an auto-discovery 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 Route Import extended community of the VPN-IPv4 routes advertised by the PE. Note that the <RD, Originating Router's IP address> tuple uniquely identifies a given multicast VRF.

Depending on the type of a P-Multicast tree used to instantiate the provider tunnel for the MVPN on the PE, the PMSI Tunnel attribute of the intra-AS auto-discovery route is constructed as follows.

+ If a P-Multicast tree is used to instantiate the provider tunnel for the MVPN on the PE, and either (a) this tree exists at the time of discovery, or (b) the PE doesn't need to know the leaves of the tree beforehand in order to advertise the P-Multicast tree identifier, then the advertising PE SHOULD advertise the type and the identity of the P-Multicast tree in the PMSI Tunnel attribute of the route.

+ If a P-Multicast tree is used to instantiate the provider tunnel for the MVPN on the PE, and in order to advertise the P-Multicast tree identifier the advertising PE needs to know the leaves of the tree beforehand, the PE first discovers the leaves by advertising an auto-discovery route without the PMSI Tunnel attribute. Once the PE obtains the information about the leaves (this information is obtained from the auto-discovery routes received by the PE), the PE then advertises the binding of the tree to the MVPN using the same route as the one used for the auto-discovery, with the addition of carrying in the route the PMSI Tunnel attribute that contains the type and the identity of the P-Multicast tree. In other words, in the first phase the PE advertises an auto-discovery route, but with no binding, and receives auto-discovery routes from other PEs. Once this is done, in the second phase the PE advertises binding based on the information acquired in the first phase. If at some later point a new PE advertises participation in the same MVPN, the initial binding PMSI Tunnel binding information SHOULD NOT change (though the leaves of the corresponding P-Multicast tree may change).

+ When the MVPN is aggregated with other MVPNs onto the same P-Multicast tree, advertised in the PMSI Tunnel attribute, the PMSI Tunnel attribute MUST carry a MPLS upstream assigned label [MPLS-UPSTREAM] that is associated with the MVPN.

+ If the PE that originates the advertisement uses ingress replication to instantiate the provider 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.

+ Discovery of PE capabilities in terms of what tunnels 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.

The Next Hop field of the MP_REACH_NLRI attribute of the route SHOULD 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 auto-discovery routes is controlled by the same Route Targets as the ones used for the distribution of VPN-IPv4 unicast routes. That is, by default the auto-discovery route MUST carry the export Route Target used by the

unicast routing. If any other PE has one of these Route Targets configured for a VRF, 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 auto-discovery routes being distinct from the ones used for the VPN-IPv4 unicast routes (see also section "Non-congruent Unicast and Multicast Connectivity").

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]).

8.1.2. Receiving (intra-AS) auto-discovery routes

When a PE receives a BGP Update message that carries an auto-discovery route such that (a) the route was originated by some other PE within the same AS as the local PE, (b) 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, (c) the BGP route selection determines that this is the best route with respect to the NLRI carried by the route, and (d) the route carries the PMSI Tunnel attribute, the PE performs the following.

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 LDP P2MP LSP, or PIM-SSM tree, or PIM-SM tree, or PIM-Bidir tree, the PE SHOULD join 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, the receiving PE has to establish the appropriate state to properly handle the traffic received over that LSP. 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 MAY be established after the local PE receives the route.

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 auto-discovery route.

If the PMSI Tunnel attribute has the Tunnel Type set to LDP P2MP LSP, or PIM-SSM tree, or PIM-SM tree, or PIM-Bidir 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 Target that matches one of the Route Targets of the received auto-discovery route.

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

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

+ An ASBR MUST be 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 auto-discovery 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).

+ The ASBR MUST be configured with an import Route Target that 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 inter-AS auto-discovery routes and S-PMSI auto-discovery routes advertised by this ASBR (if the ASBR uses different Next Hops, then the ASBR MUST be 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 ASBR supports Route Target Constrain [RT-CONSTRAIN], the ASBR SHOULD advertise this import Route Target within its own AS using Route Target Constrains. Note that this Route Target controls acceptance of leaf auto-discovery routes and C-multicast routes by the ASBR, and is used to constrain distribution of both leaf auto-discovery routes and C-multicast routes (see Section "VPN C-Multicast Routing Information Exchange among PEs"). To constrain distribution of the Route Target Constrain 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.

+ If the ASBR originates an inter-AS auto-discovery route for a particular MVPN present on some of the PEs within its own AS, the ASBR MUST be configured with an RD for that MVPN. To allow aggregation of inter-AS auto-discovery routes it is RECOMMENDED that all the ASBRs within an AS that are configured to originate an inter-AS auto-discovery 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).

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 defined in this document.

Moreover, in addition to the above the ASBR performs the following procedures.

8.2.1. Originating Inter-AS MVPN Auto-Discovery 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 auto-discovery route and advertises it in 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 Autonomous System number of the ASBR.

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

+ By default the route MUST carry the export Route Target used by the unicast routing of that VPN. The default could be modified via configuration by having a set of Route Targets used for the inter-AS auto-discovery routes being distinct from the ones used by the unicast routing of that VPN (see also section "Non-congruent Unicast and Multicast Connectivity").

An inter-AS auto-discovery route for a given <AS, MVPN> indicates presence of the MVPN sites connected to one or more PEs of the AS.

An inter-AS auto-discovery route originated by an ASBR aggregates (intra-AS) auto-discovery routes originated within the ASBR's own AS. Thus while the auto-discovery routes originated within an AS are at the granularity of <PE, MVPN> within that AS, outside of that AS the (aggregated) inter-AS auto-discovery routes could be at the granularity of <AS, MVPN>.

8.2.2. Propagating Inter-AS MVPN Auto-Discovery routes

An inter-AS auto-discovery route for a given MVPN, originated by an ASBR within a given AS, is propagated via BGP to other ASes. The precise rules for distributing and processing the inter-AS auto-discovery routes are given in subsequent sections.

Suppose that an ASBR A installs an inter-AS auto-discovery route for MVPN V that originated at a particular AS, AS1. The BGP next hop of that route becomes A's "upstream neighbor" on a multicast distribution tree for V that is rooted at AS1. When the inter-AS auto-discovery 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 auto-discovery 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 which 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 auto-discovery 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 instantiating a particular 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 will be present in the inter-AS auto-discovery routes. But if the P-Multicast tree instantiating 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 auto-discovery" routes. The precise conditions under which leaf auto-discover routes need to be used are described in subsequent sections.

Since (aggregated) inter-AS auto-discovery routes could have granularity of <AS, MVPN>, an MVPN that is present in N ASes could 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.

The following sections specify procedures for propagation of (aggregated) inter-AS auto-discovery routes across ASes.

8.2.2.1. Inter-AS Auto-Discovery Route received via EBGP

When an ASBR receives from one of its EBGP neighbors a BGP Update message that carries an inter-AS auto-discovery 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 to the destination carried in the NLRI of the route, the ASBR re-advertises this auto-discovery route to other PEs and ASBRs within its own AS.

When re-advertising an inter-AS auto-discovery route the ASBR MUST set the Next Hop field of the MP_REACH_NLRI attribute to a routable IP address of the ASBR.

Depending on the type of a P-Multicast tree used to instantiate the intra-AS segment of the inter-AS tunnel, the PMSI Tunnel attribute of the re-advertised inter-AS auto-discovery route is constructed as follows:

+ If the ASBR uses ingress replication to instantiate the intra-AS segment of the inter-AS tunnel, the re-advertised route SHOULD 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 to instantiate the intra-AS segment of the inter-AS tunnel, and in order to advertise the P-Multicast tree identifier the ASBR does not need to know the leaves of the tree beforehand, then the advertising ASBR SHOULD advertise the type and the identity of the P-Multicast tree in the PMSI Tunnel attribute of the route. This, in effect, creates a binding between the inter-AS auto-discovery route and the P-Multicast tree.

+ If the ASBR uses a P-Multicast tree to instantiate the intra-AS segment of the inter-AS tunnel, and in order to advertise the P-Multicast tree identifier the advertising ASBR needs to know the leaves of the tree beforehand, the inter-AS auto-discovery route re-advertised by the ASBR MUST include the PMSI Tunnel attribute with the Tunnel Type set to Leaf Information Required. As a result the ASBR first discovers the leaves using the procedures specified in "Leaf Auto-Discovery route received via IBGP". The ASBR then advertises the binding of the tree to the inter-AS auto-discovery route using the original inter-AS auto-discovery route, with the addition of carrying in the route the PMSI Tunnel attribute that contains the type and the identity of the tree. In other words, in the first phase the ASBR advertises inter-AS auto-discovery routes, but with no binding. Once this is done, in the second phase the ASBR advertises binding based on the information acquired in the first phase.

+ When multiple inter-AS auto-discovery routes are bound onto the same P-Multicast tree advertised in the PMSI Tunnel attribute, the PMSI Tunnel attribute MUST carry a MPLS upstream assigned label [MPLS-UPSTREAM] that is associated with the inter-AS auto-discovery route.

In addition the ASBR MUST send to the EBGP neighbor, from whom it receives the inter-AS auto-discovery route, a BGP Update message that carries a leaf auto-discovery 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 auto-discovery 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 auto-discovery 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 to demultiplex the MVPN traffic received over a unicast tunnel by the advertising router.

+ The ASBR constructs an IP-based Route Target community by placing the IP address carried in the next hop of the received inter-AS auto-discovery route in the Global Administrator field of the community, with the Local Administrator field of this community set to 0, and sets the Extended Community attribute of the leaf auto-discovery route to that community.

+ The Next Hop field of the MP_REACH_NLRI attribute of the route SHOULD 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]).

8.2.2.2. Leaf Auto-Discovery Route received via EBGP

When an ASBR receives via EBGP a leaf auto-discovery route, the ASBR accepts the route only if the Route Target carried in the Extended Community attribute of the route matches one of the import Route Target configured on the ASBR.

If the ASBR accepts the leaf auto-discovery route, the ASBR finds an inter-AS auto-discovery route whose MCAST-VPN NLRI has the same value as the Route Key field of the the leaf auto-discovery route.

The MPLS label carried in the PMSI Tunnel attribute of the leaf auto-discovery route is used to stitch a one hop ASBR-ASBR LSP to the tail of the intra-AS tunnel segment associated with the found inter-AS auto-discovery route.

8.2.2.3. Inter-AS Auto-Discovery 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 auto-discovery 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 auto-discovery 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 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 auto-discovery route carries the PMSI Tunnel attribute with the Tunnel Type set to LDP P2MP LSP, or PIM-SSM tree, or PIM-SM tree, or PIM-Bidir tree, the PE/ASBR SHOULD join the P-Multicast tree whose identity is carried in the Tunnel Identifier.

If the received inter-AS auto-discovery 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/ASBRas a leaf. This LSP MAY have been established before the local PE/ASBR receives the route, or MAY be established after the local PE receives the route.

If the received inter-AS auto-discovery route carries the PMSI Tunnel attribute with the Tunnel Type set to LDP P2MP LSP, or RSVP-TE P2MP LSP, or PIM-SSM, or PIM-SM tree, or PIM-Bidir 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 auto-discovery route and rooted at the AS that originated the inter-AS auto-discovery 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 received router is an ASBR, this intra-AS segment may further be stitched to 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.

If the Tunnel Type in the PMSI Tunnel attribute of the received inter-AS auto-discovery route is set to either Leaf Information Required or Ingress Replication, then the PE/ASBR originates a new leaf auto-discovery 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 auto-discovery 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).

+ If the received inter-AS auto-discovery route carries the PMSI Tunnel attribute with the Tunnel Type set to Ingress Replication, then the leaf auto-discovery 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 community by placing the IP address carried in the next hop of the received inter-AS auto-discovery route in the Global Administrator field of the community, with the Local Administrator field of this community set to 0, and sets the Extended Community attribute of the leaf auto-discovery route to that community.

+ The Next Hop field of the MP_REACH_NLRI attribute of the route SHOULD 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 BGP community ([RFC1997]).

+ Once the leaf auto-discovery route is constructed, the PE/ASBR advertises this route into IBGP.

8.2.2.4. Leaf Auto-Discovery route received via IBGP

When an ASBR receives via IBGP a leaf auto-discovery route, the ASBR accepts the route only if the Route Target carried in the Extended Community attribute of the route matches one of the import Route Target configured on the ASBR.

If the ASBR accepts the leaf auto-discovery route, the ASBR finds an inter-AS auto-discovery route whose MCAST-VPN NLRI has the same value as the Route Key field of the the leaf auto-discovery 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 auto-discovery route. The IP address of a leaf is the IP address carried in the Originating Router's IP address field of the received leaf auto-discovery 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. Non-congruent Unicast and Multicast Connectivity

If the multicast connectivity of a MVPN is congruent to its unicast connectivity, the VRF of that MVPN, as referred to by this document, means the VRF of that VPN used for unicast routing.

If the multicast connectivity of a MVPN is non-congruent to its unicast connectivity, the VRF of that MVPN, as referred to by this document, means the VRF that is distinct from the VRF of that VPN used for unicast routing. On a given PE such a VRF may have its own import and export Route Targets, different from the ones used by the VRF used for unicast routing. These Route Targets are used for the auto-discovery routes. The export Route Targets are added to the Route Targets used for unicast routing when originating VPN-IPv4 routes. The export Route Targets control the set of sites that could receive multicast traffic originated by the sources within the local site. The import Route Targets associated with a given VRF are used to determine which of the received VPN-IPv4 routes should be accepted into the VRF. The import Route Targets control the set of sites that contain sources of multicast traffic that could be received within the local site.

If an MVPN site is single-homed to a PE, then on this PE the VRF associated with the site should use the same RD as the one used by the VRF used for unicast routing of that VPN. If an MVPN site is multi-homed to several PEs, then to support non-congruent unicast and multicast connectivity on each of these PEs the VRF of the MVPN should use its own distinct RD (although on a given PE the RD used by the VRF of the MVPN should be the same as the one used by the VRF used for unicast routing of that VPN).

If for a given MVPN all of its sites connected to a given PE are known a priori to have no multicast sources, then this PE is NOT REQUIRED to originate an auto-discovery route for that MVPN at all,

unless either (a) some other PEs that have VRFs in that MVPN use RSVP TE P2MP LSPs, in which case the PE originates an auto-discovery route, but with no PMSI Tunnel attribute, or (b) the PE uses ingress replication for incoming multicast traffic, in which case the PE originates an auto-discovery route with the PMSI Tunnel attribute indicating ingress replication.

10. VPN C-Multicast Routing Information Exchange among PEs

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

10.1. Originating C-Multicast Routes by a PE

10.1.1. Constructing MCAST-VPN NLRI

Procedures for constructing MCAST-VPN NLRI depend on the multicast routing protocol between CE and PE (C-multicast protocol).

10.1.1.1. PIM as the C-Multicast protocol

The following specifies 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 <C-S, C-G> or <C-*, C-G> or <C-S, C-G, RPT-bit> or <C-*, C-*> state learnt by a PE via the C-multicast protocol.

Whenever a PE creates a new <C-S,C-G> state in one of its MVPN-TIBs, if C-S is reachable through some other PE the local PE originates a C-multicast route. 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. The MCAST-VPN NLRI is carried as a Source Tree Join route type. The semantics of the route is 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 PE deletes a previously created <C-S, C-G> state that had resulted in originating a C-multicast route, the PE withdraws the route (the withdrawn 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.

Whenever a PE creates a new <C-*, C-G> state in one of its MVPN-TIBs, if the C-RP for C-G is reachable through some other PE the local PE originates a C-multicast route. 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. The MCAST-VPN NLRI of the route is carried as the Shared Tree Join route type. The semantics of the route is 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 PE deletes a previously created <C-*, C-G> state that had resulted in originating a C-multicast route, the PE withdraws the route (the withdrawn route has the <C-*, C-G> Prune semantics). The MCAST-VPN NLRI of the withdrawn route is carried in the MP_UNREACH_NLRI attribute.

Whenever a PE creates a new <C-S, C-G, RPT-bit> state in one of its MVPN-TIBs, except for the case when this state is created when the PE that has an existing <C-*, C-G> state switches from the C-RP based tree to the C-S based tree for C-G, if the C-RP is reachable through some other PE, the local PE originates a C-multicast route. The Multicast Source field in the MCAST-VPN NLRI of the route is set to C-S, and the Multicast Group field is set to C-G. The MCAST-VPN NLRI of the route is carried as the Prune Source from Shared Tree route type. The semantics of the route is to indicate that the receivers for <C-S,C-G> no longer want to receive <C-S, C-G> traffic via the C-RP based tree (the route has the <C-S,C-G, RPT> Prune semantics). If later on the PE deletes a previously created <C-S, C-G, RPT-bit> state, the PE withdraws the route (the semantics of the withdrawn route is to indicate that <C-S, C-G> traffic should be carried via the C-RP-based tree). The MCAST-VPN NLRI of the withdrawn route is carried in the MP_UNREACH_NLRI attribute.

Whenever a PE creates a <C-*, C-*> state in one of its MVPN-TIBs, if the C-RP is reachable through some other PE the local PE originates a C-multicast route. 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 NLRI MUST be set to a wildcard i.e. 0. The MCAST-VPN NLRI of the route is carried as the Shared Tree Join route type. Whenever a PE deletes a previously created <C-*, C-*> state that had resulted in originating a C-multicast route, the PE withdraws the route (the withdrawn route has the (C-*, C-*) Prune semantics). The MCAST-VPN NLRI of the withdrawn route is carried in the MP_UNREACH_NLRI attribute.

The (local) PE uses its VRF to determine (a) the autonomous system number of the (remote) PE that originates the (unicast) route to C-S/C-RP, and (b) the import Route Target community associated with the VRF on the remote PE which was used to originate the route (this information is available from the Route Import extended community carried in the unicast VPN-IPv4 routing advertisements by the remote PE). Note that for a C-multicast route that the PE originates in response to creating a <C-S, C-G, RPT-bit> state, the remote PE is the PE that originates the route to C-RP, not to C-S. The Source AS

field in the C-multicast route is set to the found autonomous system. The Extended Community attribute of the C-multicast route is set to the found Route Target.

If there is more than one (remote) PE that originates the (unicast) route to C-S/C-RP, then the procedures for selecting an upstream PE to reach C-S/C-RP are as specified in [MVPN].

10.1.1.2. mLDP as the C-Multicast protocol

The following specifies construction of 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 yes, 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 as follows:

+ The PE constructs MCAST-VPN NLRI of the route by placing X into the Multicast Source field, and placing Y into the Multicast Group field.

+ The PE uses its VRF to determine (a) the autonomous system number of the (remote) PE that originates the (unicast) VPN-IPv4 route to X, and (b) the import Route Target community associated with the VRF on the remote PE which was used to originate the route (this information is available from the Route Import extended community carried in the unicast VPN-IPv4 routing advertisements by the remote PE). The Source AS field in the C-multicast route is set to the found autonomous system. The Extended Community attribute of the C-multicast route is set to the found Route Target.

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.

10.1.2. 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).

If the local and the remote PEs are in the same AS, then the RD of the advertised MCAST-VPN NLRI is set to the RD of the VPN-IPv4 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 remote PEs are in different ASes, then the local PE finds in its VRF an inter-AS auto-discovery route whose Source AS field carries the autonomous system number of the remote PE that originates the (unicast) route to the address carried in the Multicast Source field. The RD of the found inter-AS auto-discovery route is used as the RD of the advertised C-multicast route. The local PE constructs an IP-based Route Target community by placing the next hop of the found inter-AS auto-discovery route in the Global Administrator field of this community, with the Local Administrator field of this community set to 0, and adds this community to the Extended Community attribute of the C-multicast route.

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

If the next hop of the found inter-AS auto-discovery 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 auto-discovery route is within the same AS as the local PE, then the PE advertises the C-multicast route into IBGP.

10.1.3. Unicast Route Changes

Whenever unicast route used for determining PE connected to C-S/C-RP changes, the local PE updates and re-originates the previously originated C-multicast routes, as appropriate.

10.2. 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 import Route Targets configured on the ASBR, the ASBR finds an inter-AS auto-discovery 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 PE discards the received C-multicast route. Otherwise (if a matching route is found) the PE proceeds as follows.

When an ASBR receives a BGP Update message that carries a C-multicast route, 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 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 further down.

When an ASBR receives a BGP Update message that carries a withdraw of a previously advertised C-multicast route, the ASBR first checks if it already has at least one 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 withdrawn route. Otherwise, the ASBR re-advertise the withdraw of a previously advertised C-multicast route, as described below.

If the next hop for the found inter-AS auto-discovery route is an EBGP neighbor of the ASBR, then the ASBR re-advertises the C-multicast route to that neighbor. If the next hop for the found inter-AS auto-discovery 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 auto-discovery route in the first place, then the ASBR just re-advertises the C-multicast route into IBGP.

Unless it is the ASBR that originating the found inter-AS auto-discovery route in the first place, then before re-advertising the C-multicast route, the ASBR modifies the Extended Community attribute of the C-multicast route as follows. If the Route Target of the route that matches one of the import Route Targets configured on the ASBR is an IP-based Route Target with the Global Administrator field set to the IP address of ASBR, then the ASBR replaces this Route Target with a newly constructed IP-based Route Target that has the Global Administrator field set to the Next Hop of the found inter-AS auto-discovery route, and Local Administrator field of this community set to 0. The rest of the Extended Community attribute of the route SHOULD be passed unmodified.

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

10.3. Receiving C-Multicast Routes by a PE

When a PE receives a C-multicast route the PE checks if any of the Route Target communities carried in the Extended Community attribute of the route match any of the import Route Target communities associated with the VRFs 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-IPv4 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.

Procedures for a PE to process received C-multicast routes depend on the multicast routing protocol between CE and PE (C-multicast protocol).

10.3.1. PIM as the C-Multicast protocol

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

10.3.1.1. 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 there is no S-PMSI for <C-S, C-G> then the PE adds an I-PMSI to the outgoing interface list of the state if it is not already there. If there is an S-PMSI for <C-S, C-G> then the PE add S-PMSI to the outgoing interface list of the state if it is not already there.

10.3.1.2. 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. The PE adds I-PMSI to the outgoing interface list of the state if it is not already there.

10.3.1.3. Prune Source from Shared Tree C-Multicast route

If the received route has the route type set to Prune Source from Shared Tree, then the PE creates a new <C-S, C-G, RPT-bit> 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.

10.3.2. mLDP as the C-Multicast protocol

The following described 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 creates a new <X, Y> 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 there is no S-PMSI for <X, Y> then the PE creates and advertises an S-PMSI for <X, Y>, as described in section "Switching to S-PMSI". If there is an S-PMSI for <X, Y> then the PE add S-PMSI to the outgoing interface list of the state if it is not already there.

10.4. C-multicast Routes Aggregation

Note that C-multicast routes are "de facto" aggregated by BGP. The MCAST-VPN NLRIs advertised by multiple PEs, for a C-multicast route, for a particular MVPN, C-S and C-G, 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) of a given MVPN into a single C-multicast route.

Further a BGP receiver, that 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. As per the procedures described in the section "Receiving C-Multicast Routes by a PE", this forwarding state will be the same as the state that would have been created based an other route with same NLRI.

11. Switching to S-PMSI

[MVPN] describes a BGP based procedures for switching to S-PMSI. S-PMSI auto-discovery routes are used for this purpose.

Procedures for handling an S-PMSI auto-discovery route outside of the autonomous system of the PE that originates the route are the same as specified in "Propagating Inter-AS MVPN Auto-Discovery Information", except that instead of inter-AS auto-discovery routes the procedures apply to S-PMSI auto-discovery routes.

Procedures for receiving an S-PMSI auto-discovery route by a PE within the same autonomous system as the PE that originates the route are the same as specified in "Inter-AS Auto-Discovery Route received via IBGP", except that instead of inter-AS auto-discovery routes the procedures apply to S-PMSI auto-discovery routes.

The following describes procedures for originating S-PMSI auto-discovery routes by a PE.

11.1. Originating S-PMSI auto-discovery routes

An S-PMSI auto-discovery route instantiated for a given <C-S, C-G> multicast stream is constructed as follows.

The MCAST-VPN NLRI of the route is constructed as follows.

+ The RD in this NLRI is set to the RD of the 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 PE places in the Global Administrator field of the Route Import extended community of the VPN-IPv4 routes advertised by the PE. Note that the <RD, Originating Router's IP address> tuple uniquely identifies a given multicast VRF.

Depending on the type of a P-Multicast tree used to instantiate the S-PMSI, the PMSI Tunnel attribute of the S-PMSI auto-discovery route is constructed as follows:

+ If a P-Multicast tree is used to instantiate the S-PMSI, and in order to advertise the P-Multicast tree identifier the PE does not need to know the leaves of the tree within its own AS beforehand, then the advertising PE SHOULD advertise the type and the identity of the P-Multicast tree in the PMSI Tunnel attribute of the route. This, in effect, creates a binding between the S-PMSI auto-discovery route and the P-Multicast tree.

+ If a P-Multicast tree is used to instantiate the S-PMSI, and in order to advertise the P-Multicast tree identifier the advertising PE needs to know the leaves of the tree within its own AS beforehand, the S-PMSI auto-discovery route advertised by the PE MUST include the PMSI Tunnel attribute with the Tunnel Type set to Leaf Information Required. As a result the PE first discovers the leaves using the procedures specified in "Leaf Auto-Discovery route received via IBGP", except that instead of inter-AS auto-discovery routes the procedures applied to S-PMSI auto-discovery routes. The PE then advertises the binding of the P-Multicast tree to the S-PMSI auto-discovery route using the original S-PMSI auto-discovery route with the addition of carrying in the route the PMSI Tunnel attribute that contains the type and the identity of the tree. In other words, in the first phase the PE advertises an S-PMSI auto-discovery route, but with no binding. Once this is done, in the second phase the PE advertises binding based on the information acquired in the first phase.

+ When within a given AS multiple S-PMSI auto-discovery routes are bound onto the same P-Multicast tree advertised in the PMSI Tunnel attribute, the PMSI Tunnel attribute MUST carry an MPLS upstream assigned label [MPLS- UPSTREAM] that is associated with the individual S-PMSI auto-discovery route.

The Next Hop field of the MP_REACH_NLRI attribute of the route SHOULD

be set to the same IP address as the one carried in the Originating Router's IP Address field.

The route SHOULD carry the same set of Route Targets as the intra-AS auto-discovery route of the MVPN originated by the PE

12. 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. Use of RSVP-TE and/or BGP as the CE-PE multicast routing and label distribution protocol is for further study.

To improve scalability it is strongly 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).

13. Choosing a single forwarder PE when switching from RPT to SPT

In the scenario where an MVPN customer switches from an 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 consistant upstream PE, as described in [MVPN], for <C-*, C-G> state. Further this upstream PE, say PE1, is using an 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

Whenever a PE creates an <C-S,C-G> state as a result of receiving a Source Tree Join C-multicast route for <C-S, C-G> from some other PE, the PE that creates the state SHOULD originate a Source Active auto-discovery route. 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 route SHOULD carry the same set of Route Targets as the intra-AS auto-discovery route of the MVPN originated by the PE.

Using the normal BGP procedures the Source Active auto-discovery route is propagated to all the PEs of the MVPN.

Whenever the PE deletes the <C-S, C-G> state that was previousely created as a result of receiving a C-multicast route for <C-S, C-G> from some other PE, the PE that deletes the state also withdraws the Source Active auto-discovery route, if such a route was advertised when the state was created.

13.2. Receiving Source Active auto-discovery route

When a PE receives a new Source Active auto-discovery route that carries a given <C-S, C-G>, the PE finds a VRF whose import Route Targets match one or more of the Route Targets carried by the route. The PE then checks for an <*, C-G> or an <C-S, C-G> entry in that VRF with a non-empty outgoing interface list that contains one or more PE-CE interfaces (interfaces from the PE to its directly connected CEs). Presence of such an entry implies that some system in the sites (directly) connected to the PE is interested in the <C-S, C-G> carried in the Source Active auto-discovery route, in which case the PE originates a Source Tree Join C-multicast route for <C-S, C-G>.

If a PE receives from one of its directly connected CEs a PIM Join message for a new group C-G, the PE SHOULD originate a Source Tree Join C-multicast route for each Source Active auto-discovery route that contains C-G and is present in the VRF associated with the CE.

Construction and distribution of the Source Tree Join C-multicast route follows the procedures specified in "VPN C-Multicast Routing Information Exchange among PEs", except that procedures specified in "Constructing MCAST-VPN NLRI" are replaced with the following:

+ The Multicast Source Length, Multicast Source, Multicast Group Length, and Multicast Group fields are copied from the corresponding field in the Source Active auto-discovery route.

+ The (local) PE uses its VRF to find (a) the autonomous system number of the (remote) PE that originates the (unicast) route to C-S carried in the Multicast Source field of the Source Active auto-discovery route, and (b) the import Route Target community associated with the VRF on the remote PE which was used to originate the (unicast) route (this information is available from the Route Import extended community carried in the unicast VPN-IPv4 routing advertisements by the remote PE). The Source AS field in the C-multicast route is set to the found autonomous system. The Extended Community attribute of the C-multicast route is set to the found Route Target.

  If there is more than one (remote) PE that originates the (unicast) route to C-S, then the procedures for selecting an upstream PE to reach C-S are as specified in [MVPN].

If the incoming interface list (iif) for the <*, C-G> state in the MVPN-TIB on the PE contains one of the PE-CE interfaces (interfaces from the PE to its directly connected CEs), then the PE creates in the MVPN-TIB a <C-S, C-G, RPT-bit> state, if it does not already exist. C-S of this state is set to the address carried in the Multicast Source field of the Source Active auto-discovery route, and C-G of this state is set to the address carried in the Multicast Group field of the route. Creating this state results in pruning <C-S, C-G> traffic off the shared (RPT) tree.

14. Co-locating C-RPs on a PE

This section describes the procedures for co-locating a C-RP on a PE. The procedures are for the anycast RP based on C-(*, G) scheme described in [MVPN].

14.1. Source Within a Site - Source Active Advertisement

When a PE that acts as an anycast RP for a given MVPN first learns of a new (multicast) sender within that MVPN, e.g., via PIM register messages originated within that MVPN, the PE follows the normal PIM procedures. In addition, the PE constructs a Source Active auto-discovery 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. The Multicast Source Length field is set appropriately to reflect this.

+ The Multicast Group field MUST be set to the group address of the multicast data packet carried in the PIM-Register message. The Multicast Group Length field is set appropriately to reflect this.

The route SHOULD carry the same set of Route Targets as the intra-AS auto-discovery route of the MVPN originated by the PE.

Using the normal BGP procedures the Source Active auto-discovery route is propagated to all the PEs of the MVPN.

When a PE that previousely advertised a Source Active auto-discovery route for a given sender learns that the sender is no longer active the PE withdraws the previously advertised Source Active route.

14.2. Receiver(s) Within a Site

When a PE receives a new Source Active auto-discovery route that carries a given <C-S, C-G>, the PE finds a VRF whose import Route Targets match one or more of the Route Targets carried by the route. The PE then checks for an <*, C-G> or an <C-S, C-G> entry in that VRF with a non-empty outgoing interface list that contains one or more PE-CE interfaces (interfaces from the PE to its directly connected CEs). Presence of such an entry implies that some system in the sites (directly) connected to the PE is interested in the <C-S, C-G> carried in the Source Active auto-discovery route, in which case the PE originates a Source Tree Join C-multicast route for <C-S, C-G>.

If a PE receives from one of its directly connected CEs a PIM Join message for a new group C-G, the PE SHOULD originate a Source Tree Join C-multicast route for each Source Active auto-discovery route that contains C-G and is present in the VRF associated with the CE.

Construction and distribution of the Source Tree Join C-multicast route follows the procedures specified in "VPN C-Multicast Routing Information Exchange among PEs", except that procedures specified in "Constructing MCAST-VPN NLRI" are replaced with the following:

+ The Multicast Source Length, Multicast Source, Multicast Group Length, and Multicast Group fields are copied from the corresponding field in the Source Active auto-discovery route.

+ The (local) PE uses its VRF to find (a) the autonomous system number of the (remote) PE that originates the (unicast) route to C-S carried in the Multicast Source field of the Source Active auto-discovery route, and (b) the import Route Target community associated with the VRF on the remote PE which was used to originate the (unicast) route (this information is available from the Route Import extended community carried in the unicast VPN-IPv4 routing advertisements by the remote PE). The Source AS field in the C-multicast route is set to the found autonomous system. The Extended Community attribute of the C-multicast route is set to the found Route Target.

  If there is more than one (remote) PE that originates the (unicast) route to C-S, then the procedures for selecting an upstream PE to reach C-S are as specified in [MVPN].

A PE which receives C-Join for <C-*, C-G> or <C-S, C-G> from one of its connected CEs that belong to a given MVPN does not send the information that it has receiver(s) for C-G until it receives an Source Active auto-discovery route from some other PE indicating that there are active sources for that C-G.

14.3. Receiving C-multicast routes

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 "Source Tree Join C-Multicast route".

15. Scalability Considerations

A PE should use Route Target Constrain [RT-CONSTRAIN] to advertise the Route Targets that the PE uses for the 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 auto-discovery route originated by an ASBR aggregates auto-discovery routes originated within the ASBR's own AS. Thus while the auto-discovery routes originated within an AS are at the granularity of <PE, MVPN> within that AS, outside of that AS the (aggregated) inter-AS auto-discovery routes are at the granularity of <AS, MVPN>. An inter-AS auto-discovery route for a given <AS, MVPN> indicates the presence of or 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 auto-discovery routes have 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-IPv4 routes (although they may be used for VPN-IPv4 routes as well).

As described in Section "C-multicast routes aggregation", 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. Dampening of C-multicast routes

The rate of C-multicast routing changes advertised by a PE is not 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 auto-discovery routes, will happen. Dampening techniques can be used to limit corresponding processing.

To lessen the control plane overhead associated with 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. Dampening of C-multicast prunes

A PE/ASBR/Route Reflector can OPTIONALLY delay the advertisement of C-multicast prune routes (routes resulting from the removal of multicast state). 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 C-multicast prune routes does not impede the multicast join latency observed by MVPN customers, and 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 only potential drawback of dampening C-multicast prune routes is that the PE that performs the dampening may receive useless (multicast) traffic for some period of time. Note that the PE may receive useless (multicast) traffic anyway, irrespective of dampening C-multicast prune routes due to the use of I-PMSIs.

16.2. Dampening of C-multicast joins

A PE/ASBR/Route Reflector can OPTIONALLY delay the advertisement of C-multicast join routes (routes resulting from the creation of multicast state). 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 C-multicast join routes will not impede multicast join latency observed by a given MVPN, except if the PE advertising the C-multicast join route is the first for all the sites of the MVPN to do so.

16.3. Dampening of leaf auto-discovery routes

As described in section "Switching to S-PMSI, leaf auto-discovery routes are used for S-PMSI tunnel for which the root need to know the leaves of the tree.

Similarly to the procedures proposed above for C-multicast routes, dampening can be applied aggressively to the withdrawal of such auto-discovery routes.

17. IANA Consideration

This document defines a new BGP Extended Community called Source AS. This community is 2-octet AS specific, of an extended type, and is transitive.

This document defines a new BGP Extended Community called Route Import. This community is IPv4 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 is assigned its own SAFI.

This document defines a new BGP optional transitive attribute, called PMSI Tunnel.

18. Security Considerations

The mechanisms described in this document could re-use the existing BGP security mechanisms.

19. Acknowledgement

Some of the text in Section "Co-locating C-RPs on a PE" has been taken almost verbatim from RFC3618.

20. References

20.1. Normative References

[MVPN] E. Rosen, R. Aggarwal [Editors], "Multicast in MPLS/BGP IP VPNs", draft-ietf-l3vpn-2547bis-mcast-01.txt,

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

[RFC2858] Bates, T., Rekhter, Y., Chandra, R., and D. Katz, "Multiprotocol Extensions for BGP-4", RFC 2858, June 2000.

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

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

20.2. Informative References

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

[PIM-SM] B. Fenner et. al., "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", draft-ietf-pim-sm-v2-new-11.txt

[RT-CONSTRAIN] P. Marques et. al., ",Constrained VPN Route Distribution" draft-ietf-l3vpn-rt-constrain-02

[mLDP] I. Minei et. al., "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", draft-ietf-mpls-ldp-p2mp-00

21. Author Information

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

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

Thomas Morin
France Telecom R & D
2, avenue Pierre-Marzin
22307 Lannion Cedex
France
Email: thomas.morin@francetelecom.com

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

Chaitanya Kodeboniya

22. Intellectual Property Statement

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

23. Copyright Notice