Workgroup: Network Working Group
Internet-Draft:
draft-ietf-bess-mvpn-evpn-sr-p2mp-08
Published: 6 November 2023
Intended Status: Standards Track
Expires: 9 May 2024
Authors: R. Parekh, Cisco Systems, Inc.
C. Filsfils, Cisco Systems, Inc.
A. Venkateswaran, Cisco Systems, Inc.
H. Bidgoli, Nokia
D. Voyer, Bell Canada
Z. Zhang, Juniper Networks

Multicast and Ethernet VPN with Segment Routing P2MP and Ingress Replication

Abstract

A Point-to-Multipoint (P2MP) Tree in a Segment Routing domain carries traffic from a Root to a set of Leaves. This document describes extensions to BGP encodings and procedures for P2MP trees and Ingress Replication used in BGP/MPLS IP VPNs and Ethernet VPNs in a Segment Routing domain.

Requirements Language

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 RFC 2119 [RFC2119].

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Multicast in MPLS/BGP IP VPNs [RFC6513] and BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs [RFC6514] specify procedures that allow a Service Provider to provide Multicast VPN (MVPN) service to its customers. Multicast traffic from a customer is tunneled across the service provider network over Provider Tunnels (P-Tunnels). P-Tunnels can be instantiated via different technologies. A service provider network that uses Segment Routing can use a Point-to-Multipoint (SR P2MP) tree [I-D.ietf-pim-sr-p2mp-policy] or P2MP Ingress Replication to instantiate P-Tunnels for MVPN. SR P2MP P-Tunnels can be realized both for SR-MPLS [RFC8660] and SRv6 [RFC8986][RFC8754].

In a Segment Routing network, a P2MP tree allows efficient delivery of traffic from a Root to set of Leaf nodes. A SR P2MP tree is defined by a SR P2MP Policy and instantiated via a PCE. A P2MP Policy consists of a Root, a Set of Leaf Nodes and a set of candidate paths with optional set of constraints and/or optimization objectives to be satisfied by the P2MP tree. A unique Identifier, called Tree-SID, is associated with a P2MP tree. This Tree-SID can be an MPLS label or an IPv6 address.

This document describes extensions to BGP Auto-Discovery procedures specified in RFC 6514 for SR P2MP P-Tunnels. Use of PIM for Auto-Discovery is outside scope of this document. Support for customer BIDIR-PIM is outside the scope of this document.

For BGP MPLS Ethernet VPN specified in [RFC7432] and extensions to this document, P-Tunnels are advertised for handling multi-destination traffic. These P-Tunnels can be realized by SR-MPLS or SRv6 P2MP trees. SRv6 P2MP trees can also be used to support Multicast in Network Virtualization over Layer 3 [RFC8293].

The reader is expected to be familiar with concepts and terminology of RFC 6513, RFC 6514 and SR P2MP drafts.

2. SR P2MP P-Tunnels

For MVPN or EVPN, Provider Edge(PE) routers steer customer traffic into a P-Tunnel that can be instantiated by a SR-MPLS or SRv6 P2MP. A SR P2MP tree is defined by a SR P2MP policy [I-D.ietf-pim-sr-p2mp-policy].

Given a SR P2MP policy, a PCE computes and instantiates the SR P2MP tree on the nodes that are part of the tree by stitching Replication segments [I-D.ietf-spring-sr-replication-segment] at Root, Leaf and intermediate replication nodes. Tree-SID is an unique identifier for the tree. A Replication segment of a SR P2MP tree can be initiated by various methods (BGP, PCEP, others) which are outside the scope of this document.

A PCE provides conceptual APIs, listed below, to define and modify SR P2MP policies SR P2MP Policy Section 4.1.1. These APIs are invoked by a PCC, which is the root of P2MP tree, using various methods (BGP, PCEP, etc.) which are outside the scope of this document.

CreatePolicy: CreateSRP2MPPolicy<Root, Tree-ID>

DeletePolicy: DeleteSRP2MPPolicy<Root, Tree-ID>

UpdateLeafSet: SRP2MPPolicyLeafSetModify<Root, Tree-ID, {Leaf Set}>

The Root of a P2MP tree imposes the Tree-SID to steer the customer payload into the P2MP tree. Provider (P) routers replicate customer payload, using Replication segments, towards the Leaf nodes of the P2MP tree. Leaf nodes of the P2MP tree deliver the customer payload after disposing the Tree-SID.

An Ingress PE can deliver payload to egress PEs of the service using Ingress Replication. This payload is encapsulated in SR-MPLS or SRv6 and replicated to each egress PE.

3. PMSI Tunnel Attribute for SR P2MP

BGP PMSI Tunnel Attribute (PTA) is defined in RFC 6514 to identify the P-Tunnel that is used to instantiate a Provider Multicast Service Interface (PMSI). The PTA is carried in Intra-AS I-PMSI, Inter-AS I-PMSI, Selective PMSI, and Leaf Auto-Discovery routes.

A P2MP tree PTA is constructed as specified below.

* Tunnel Type: The IANA assigned codepoint 0x0C for "SR-MPLS P2MP Tree" or codepoint TBDfor "SRv6 P2MP Tree", from the "P-Multicast Service Interface Tunnel (PMSI Tunnel) Tunnel Types" registry.

* Flags: See Section 4 for use of "Leaf Info Required bit".

* MPLS Label: See Section 3.1

* Tunnel Identifier: The SR P2MP P-Tunnel is identified by <Tree-ID, Root> where,

  - Tree-ID is a 32-bit unsigned value that identifies a unique P2MP tree at a Root.

  - Root is an IP address identifying the Root of a P2MP tree. This can be either an IPv4 or IPv6 address and can be inferred from the PTA length.

When a P-Tunnel is non-segmented, the PTA is created by PE router at the Root of a SR P2MP tree. For segmented P-Tunnels, each segment can be instantiated by a different technology. If a segment is instantiated using P2MP tree, the router at the root of a P2MP tree creates the PTA.

3.1. MPLS Label

[RFC6514] allows a PE to aggregate two or more MVPNs onto one P-Tunnel by advertising the same P-Tunnel in PTA of Auto-Discovery routes of different MVPNs. This section specifies how the "MPLS Label" field of PTA is filled to provide a context bound to a specific MVPN. For EVPN considerations, see SR P2MP Trees for EVPN section.

3.1.1. SR-MPLS

When a SR P2MP P-Tunnel is shared across two or more MVPNs in a SR MPLS domain [RFC8660], the "MPLS Label" field of a PTA advertised in an Auto-Discovery route MUST contain an upstream-assigned MPLS label that the advertising PE has bound to the MVPN, or a label assigned from a global context such as "Domain- wide Common Block" (DCB) as specified in [I-D.ietf-bess-mvpn-evpn-aggregation-label].

When a customer payload is steered into a shared SR P2MP P-Tunnel, this MPLS label MUST be imposed before the MPLS label representing the Tree-SID.

3.1.2. SRv6

When a SR P2MP P-Tunnel is shared across two or more MVPNs in a SRv6 domain [RFC8986], the "MPLS Label" field of a PTA advertised in an Auto-Discovery route MUST contain an upstream-assigned SRv6 Multicast Service SID Section 5.2.1 that the advertising PE has bound to the MVPN, or a SRv6 Multicast Service SID assigned from a global context; this follows same concept of "Domain- wide Common Block" (DCB) label as specified in [I-D.ietf-bess-mvpn-evpn-aggregation-label]. The high order 20 bits of this field carry the whole or a portion of the Function part of the SRv6 Multicast Service SID when Transposition Scheme of encoding as defined in [RFC9252] is used. When using the Transposition Scheme, the Transposition Length of SRv6 SID Structure Sub-Sub-TLV of SRv6 Prefix-SID attribute (see below) MUST be less than or equal to 20 and less than or equal to the Function Length. When Transposition shceme is not used, the label field MUST be set to zero and Transposition Length MUST be zero.

The advertising ingress PE attaches a BGP Prefix-SID attribute [RFC8669] to Intra-AS I-PMSI, Inter-AS I-PMSI or S-PMSI A-D routes with SRv6 L3 Service TLV [RFC9252] to signal SRv6 Multicast Service SID. The SRv6 SID Information Sub-TLV carries the SRv6 Multicast Service SID in SRv6 SID Value field. The SRv6 Endpoint Behavior of the SRv6 SID Information Sub-TLV encodes one of End.DTMC4, End.DTMC6 or End.DTMC46 codepoint values. The SRv6 SID Structure Sub-Sub-TLV encodes the structure of SRv6 Multicast Service SID. If Transposition scheme is used, the offset and length of SRv6 Multicast Endpoint function of SRv6 Multicast Service SID is set in Transposition Length and Transposition Offset fields of this sub-sub TLV. Otherwise, the Transposition Length and Offset fields MUST be set to zero.

The ingress PE MUST encapsulate customer payload, steered into a shared SR P2MP P-Tunnel, in an outer IPv6 header with SRH in which the SRv6 Multicast Service SID MUST be the last segment in the segment list (note the SRv6 Multicast Service SID may be the only segment in SRH) . If Transposition scheme is used, ingress PE MUST merge Function in MPLS Label field of PTA with SRv6 SID in SID Information TLV using the Transposition Offset and Length fields from SID structure sub-sub TLV to create SRv6 Multicast Service SID

The Egress PEs of a shared SR P2MP P-Tunnel use the SRv6 Multicast Service SID in SRH to determine the MVPN in which the customer payload is to be delivered. An Egress PE, in role of Leaf or Bud Node of Replication Segment associated with shared SR-P2MP P-Tunnel tree, uses "look at next SID in SRH" [I-D.ietf-spring-sr-replication-segment] behavior to process the SRv6 Multicast Service SID.

4. MVPN Auto-Discovery and Binding Procedures for P2MP Trees

RFC 6514 defines procedures for discovering PEs participating in a given MVPN and binding customer multicast flows to specific P-Tunnels. This section specifies modifications to these procedures for SR P2MP tree P-Tunnels. In this section, the term "SR P2MP" refers to both SR-MPLS and SRv6.

4.1. Intra-AS I-PMSI

Intra-AS I-PMSI A-D routes are exchanged to discover PEs participating in a MVPN within an AS, or across different ASes when non-segmented P-Tunnels are used for inter-AS MVPNs.

4.1.1. Originating Intra-AS I-PMSI routes

RFC 6514 Section 9.1.1 describes procedures for originating Intra-AS I-PMSI A-D routes. For SR P2MP P-Tunnels, these procedures remain unchanged except as described in the following paragraphs.

When a PE originates an Intra-AS I-PMSI A-D route with a PTA having SR P2MP P-Tunnel Type, it MUST create a P2MP policy by invoking CreatePolicy API of the PCE. When the PCE instantiates the P2MP tree on the PE, the Tree-SID MUST be imposed for customer flow(s) steered into the P2MP tree. The Leaf nodes of P2MP tree are discovered using procedures described in Section 4.1.2.

For a PE in "Receiver Sites set", condition (c) is modified to include P2MP tree; such a PE MUST originate an Intra-AS I-PMSI A-D route when some PEs of the MVPN have VRFs that use SR P2MP tree but MUST NOT create a SR P2MP policy as described above.

When a PE withdraws an Intra-AS I-PMSI A-D route, advertised with a PTA having SR P2MP P-Tunnel Type, the Tree-SID imposition state at the PE MUST be removed.

A PE MAY aggregate two or more Intra-AS I-PMSIs from different MVPNs onto the same SR P2MP P-Tunnel. When a PE withdraws the last Intra-AS I-PMSI A-D route, advertised with a PTA identifying a SR P2MP P-Tunnel , it SHOULD remove the SR P2MP policy by invoking DeletePolicy API of the PCE.

4.1.2. Receiving Intra-AS I-PMSI A-D routes

Procedure for receiving Intra-AS I-PMSI A-D routes, as described in RFC 6514 Section 9.1.2, remain unchanged for SR P2MP P-Tunnels except as described in the following paragraphs.

When a PE that advertises a SR P2MP P-Tunnel in the PTA of its Intra-AS I-PMSI A-D route, imports an Intra-AS I-PMSI A-D route from some PE, it MUST add that PE as a Leaf node of the P2MP tree. The Originating IP Address of the Intra-AS i-PMSI A-D route is used as the Leaf Address when invoking UpdateLeafSet API of the PCE. This procedure MUST also be followed for all Intra-AS I-PMSI routes that are already imported when the PE advertises a SR P2MP P-Tunnel in PTA of its Intra-AS I-PMSI A-D route.

A PE that imports and processes an Intra-AS I-PMSI A-D route from another PE with PTA having SR P2MP P-Tunnel MUST program the Tree-SID of the P2MP tree identified in the PTA of the route for disposition. Note that an Intra-AS I-PMSI A-D route from another PE can be imported before the P2MP tree identified in the PTA of the route is instantiated by the PCE at the importing PE. In such case, the PE MUST correctly program Tree-SID for disposition. A PE in "Sender Sites set" MAY avoid programming the Tree-SID for disposition.

When an Intra-AS I-PMSI A-D route, advertised with a PTA having SR P2MP P-Tunnel Type is withdrawn, a PE MUST remove the disposition state of the Tree-SID associated with P2MP tree.

A PE MAY aggregate two or more Intra-AS I-PMSIs from different MVPNs onto the same SR P2MP P-Tunnel. When a remote PE withdraws an Intra-AS I-PMSI A-D route from a MVPN, and if this is the last MVPN sharing a SR P2MP P-Tunnel, a PE must remove the originating PE as a Leaf from the P2MP tree, by invoking UpdateLeafSet API.

4.2. Using S-PMSIs for binding customer flows to P2MP Segments

RFC 6514 specifies procedures for binding (C-S,C-G) customer flows to P-Tunnels using S-PMSI A-D routes. Wildcards in Multicast VPN Auto-Discovery Routes [RFC6625] specifies additional procedures to binding aggregate customer flows to P-Tunnels using "wildcard" S-PMSI A-D routes. This section describes modification to these procedures for SR P2MP P-Tunnels.

4.2.1. Originating S-PMSI A-D routes

RFC 6514 Section 12.1 describes procedures for originating S-PMSI A-D routes. For SR P2MP P-Tunnels, these procedures remain unchanged except as described in the following paragraphs.

When a PE originates S-PMSI A-D route with a PTA having SR P2MP P-Tunnel Type, it MUST set the "Leaf Info Required bit" in the PTA. The PE MUST create a SR P2MP policy by invoking CreatePolicy API of the PCE. When the PCE instantiates the P2MP tree on the PE, the Tree-SID MUST be imposed for customer flows steered into the SR P2MP P-Tunnel.

The Leaf nodes of P2MP tree are discovered by Leaf A-D routes using procedures described in Section 4.4.2. When a PE originates S-PMSI A-D route with a PTA having SR P2MP P-Tunnel Type, it is possible the PE might have imported Leaf A-D routes whose route keys match the S-PMSI A-D route. The PE MUST re-apply procedures of Section 4.4.2 to these Leaf A-D routes.

When a PE withdraws a S-PMSI A-D route, advertised with PTA having P2MP tree P-Tunnel type, the Tree-SID imposition state MUST be removed.

A PE MAY aggregate two or more S-PMSIs onto the same SR P2MP P-Tunnel. When a PE withdraws the last S-PMSI A-D route, advertised with a PTA identifying a specific SR P2MP P-Tunnel , it SHOULD remove the SR P2MP policy by invoking DeletePolicy API of the PCE.

4.2.2. Receiving S-PMSI A-D routes

RFC 6514 Section 12.3 describes procedures for receiving S-PMSI A-D routes. For SR P2MP P-Tunnels, these procedures remain unchanged except as described in the following paragraphs.

The procedure to join SR P2MP P-Tunnel of S-PMSI A-D route by using a Leaf A-D route is described in Section 4.4.1. If P2MP tree identified in PTA of S-PMSI A-D route is already instantiated by PCE, the PE MUST program Tree-SID for disposition. If the P2MP tree is instantiated later, the Tree-SID MUST be programmed for disposition at that time.

When a S-PMSI A-D route, whose SR P2MP P-Tunnel has been joined by a PE, is withdrawn, or when conditions (see RFC 6514 Section 12.3) required to join that P-Tunnel are no longer satisfied, the PE MUST leave the P-Tunnel. The PE MUST withdraw the Leaf A-D route it had originated and remove the Tree-SID disposition state.

4.3. Inter-AS P-tunnels using P2MP Segments

A segmented inter-AS P-Tunnel consists of one or more intra-AS segments, one in each AS, connected by inter-AS segments between ASBRs of different ASes https://tools.ietf.org/html/rfc6514#section-9.2. These segments are constructed by PEs/ASBRs originating or re-advertising Inter-AS I-PMSI A-D routes. This section describes procedures for instantiating intra-AS segments using SR P2MP trees.

4.3.1. Advertising Inter-AS I-PMSI routes into iBGP

RFC 6514 Section 9.2.3.2 specifies procedures for advertising an Inter-AS I-PMSI A-D route to construct an intra-AS segment. The PTA of the route identifies the type and identifier of the P-Tunnel instantiating the intra-AS segment. The procedure for creating SR P2MP P-Tunnel for intra-AS segment are same as specified in Section 4.2.1 except that instead of S-PMSI A-D routes, the procedures apply to Inter-AS I-PMSI A-D routes.

4.3.2. Receiving Inter-AS I-PMSI A-D routes in iBGP

RFC 6514 Section 9.2.3.2 specifies procedures for processing an Inter-AS I-PMSI A-D route received via iBGP. If the PTA of the Inter-AS I-PMSI A-D route has SR P2MP P-Tunnel Type, the procedures are same as specified in Section 4.2.2 except that instead of S-PMSI A-D routes, the procedures apply to Inter-AS I-PMSI A-D routes. If the receiving router is an ASBR, the Tree-SID is stitched to the inter-AS segments to ASBRs in other ASes.

4.4. Leaf A-D routes for P2MP Segment Leaf Discovery

This section describes procedures for originating and processing Leaf A-D routes used for Leaf discovery of SR P2MP trees.

4.4.1. Originating Leaf A-D routes

The procedures for originating Leaf A-D route in response to receiving a S-PMSI or Inter-AS I-PMSI A-D route with PTA having SR P2MP P-Tunnel Type are same as specified in RFC 6514 Section 9.2.3.4.1.

4.4.2. Receiving Leaf A-D routes

Procedures for processing a received Leaf A-D route are specified in RFC 6514 Section 9.2.3.5. These procedures remain unchanged for discovering Leaf nodes of P2MP trees except for considerations described in following paragraphs. These procedures apply to Leaf A-D routes received in response to both S-PMSI and Inter-AS I-PMSI A-D routes, shortened to "A-D routes" in this section

A Root PE/ASBR MAY use the same SR P2MP P-Tunnel in PTA of two or more A-D routes. For such aggregated P2MP trees, the PE/ASBR may receive multiple Leaf A-D routes from a Leaf PE. The P2MP tree for which a Leaf A-D is received can be identified by examining the P2MP tunnel Identifier in the PTA of A-D route that matches "Route Key" field of the Leaf A-D route. When the PE receives the first Leaf A-D route from a Leaf PE, identified by the Originating Router's IP address field, it MUST add that PE as Leaf of the P2MP tree by invoking the UpdateLeafSet API of the PCE.

When a Leaf PE withdraws the last Leaf A-D route for a given SR P2MP P-Tunnel, the Root PE MUST remove the Leaf PE from the P2MP tree by invoking UpdateLeafSet API of PCE. Note that Root PE MAY remove the P2MP tree, via the DeletePolicyAPI, before the last Leaf A-D is withdrawn. In this case, the Root PE MAY decide to not invoke the UpdateLeafSet API.

5. MVPN with Ingress Replication over Segment Routing

A PE can provide MVPN service using Ingress Replication over Segment Routing. Customer payload is encapsulated in SR-MPLS or IPv6 (SRv6) at Ingress PE. The encapsulated payload is replicated and a unicast copy is sent to each egress PE.

Ingress Replication Tunnels in Multicast VPN [RFC7988] specifies procedures that can be used to provide MVPN service with Ingress Replication in a Segment Routing domain. A PE advertises Intra-AS, Inter-AS, Selective PMSI BGP Auto-Discovery routes with PTA for Ingress Replication. Egress PEs join asLeaf Nodes using Intrra-AS I-PMSI or Leaf Auto-Discovery routes.

RFC 7988 procedures allow an ingress PE to deliver MVPN traffic to egress PEs using best-effort unicast connectivity. For MVPN service with an underlay SLA from ingress PE to an egress PE, the egress PE colors the Leaf Auto-Discovery route with a Color Extended Community as specified in [I-D.ietf-idr-segment-routing-te-policy]. The ingress PE replicates MVPN customer payload to that egress PE by steering traffic into a SR-TE policy (Color, egress PE) according to section 8 of [RFC9256].

5.1. SR-MPLS

Procedures of RFC 7988 are sufficient to create a SR-MPLS Ingress Replication for MVPN service.

If an egress PE colors the Leaf A-D route with Color Extended Community, the ingress PE encapsulates the payload packet into segment list of (Color, egress PE) SR-TE policy along with IR label received from the egress PE. Suppose the egress PE, say PE2, sends Leaf A-D route with extended color community C1 and IR label L10. Assume the segment list of SR-TE policy (C1, PE2) at ingress PE1 is <L1, L2, L3>, PE1 will encapsulate MVPN customer payload into MPLS label stack <L1, L2, L3, L10> with L10 as BoS label.

5.2. SRv6

Procedures of RFC 7988, along with modifications described in this Section, are sufficient to create a SRv6 Ingress Replication for MVPN service.

The PTA carried in Intra-AS, Inter-AS, Selective PMSI and Leaf Auto-Discovery routes is constructed as specified in RFC 7988 with modifications as below:

*Tunnel Type: "Ingress Replication" as per RFC 6514.

* MPLS Label: The high order 20 bits of this field carry the whole or a portion of the Function part of the SRv6 Multicast Service SID when ingress replication is used with the Transposition Scheme of encoding as defined in [RFC9252]. When using the Transposition Scheme, the Transposition Length of SRv6 SID Structure Sub-Sub-TLV of SRv6 Prefix-SID attribute (see below) MUST be less than or equal to 20 and less than or equal to the Function Length. When Transposition shceme is not used, the label field MUST be set to zero and Transposition Length MUST be zero.

Section 6 and 7 of RFC 7988 describe considerations and procedures for allocating MPLS labels for IR P-Tunnel. For SRv6 Ingress Replication, these sections apply to SRv6 Multicast Service SID.

To join a SRv6 Ingres Replication P-Tunnel advertised in PTA of Inra-AS, Inter-AS, or Selective S-PMSI A-D routes, an egress PE constructs a Leaf A-D or Intra-AS I-PMSI route as described in RFC 7988 with modified PTA above. The egress PE attaches a BGP Prefix-SID attribute [RFC8669] in Leaf A-D or Intra-AS I-PMSI route with SRv6 L3 Service TLV [RFC9252] to signal SRv6 Multicast Service SID . The SRv6 SID Information Sub-TLV carries the SRv6 Multicast Service SID in SRv6 SID Value field. The SRv6 Endpoint Behavior of the SRv6 SID Information Sub-TLV encodes one of End.DTMC4, End.DTMC6 or End.DTMC46 codepoint value. The SRv6 SID Structure Sub-Sub-TLV encodes the structure of SRv6 Multicast Service SID. If Transposition scheme is used, the offset and length of SRv6 Multicast Endpoint function of SRv6 Multicast Service SID is set in Transposition Length and Transposition Offset fields of this sub-sub TLV. Otherwise, the Transposition Length and Offset fields MUST be set to zero. The BGP Prefix SID attribute with SRv6 L3 Service TLV in Intra-AS I-PMSI or Leaf A-D route indicates to ingress PE that egress PE supports SRv6.

The SRv6 Multicast Service SID SHOULD be routable within the AS of the egress PE. As per RFC 7988, the Ingress PE uses the Tunnel Identifier of PTA to determine the unicast tunnel to use in order to send data to the egress PE. This document requires the ingress PE to use the SRv6 Multicast Service SID to determine the unicast tunnel to be used. For best-effort MVPN service or SLA based MVPN service using IGP Flexible Algorithm, the ingress PE MUST encapsulate payload in an outer IPv6 header with the SRv6 Multicast Service SID provided by the egress PE as the destination address. If Transposition scheme is used, ingress PE MUST merge Function in MPLS Label field of PTA with SRv6 SID in SID Information TLV using the Transposition Offset and Length fields from SID structure sub-sub TLV to create SRv6 Multicast Service SID

If an egress PE colors a Leaf A-D route with Color Extended Community, the ingress PE SHOULD encapsulate the payload packet into outer IPv6 header with segment list of (Color, egress PE) SR-TE policy along with SRv6 Multicast Service SID received with Leaf A-D route from the egress PE using SRH. Suppose the egress PE, say PE2, sends Leaf A-D route with extended color community C1 and SRv6 Multicast Service SID S10. Assume the segment list of SR-TE policy (C1, PE2) at ingress PE1 is <S1, S2, S3>, PE1 will encapsulate MVPN customer payload into IPv6 header with SRH (PE1, S1) (S10, S3, S2; SL=3) (payload). If SRv6 SID compression is used, the ingress PE SHOULD use CSID containers for the policy segments.

5.2.1. SRv6 Multicast Endpoint Behaviors

The following behaviors can be associated with SRv6 Multicast Service SID.

5.2.1.1. End.DTMC4: Decapsulation and Specific IPv4 Multicast Table Lookup

The "Endpoint with decapsulation and specific IPv4 Multicast table lookup" behavior ("End.DTMC4" for short) is similar to End.DT4 behavior of RFC 8986 except the lookup is in IPv4 multicast table.

5.2.1.2. End.DTMC6: Decapsulation and Specific IPv6 Multicast Table Lookup

The "Endpoint with decapsulation and specific IPv6 Multicast table lookup" behavior ("End.DTMC6" for short) is similar to End.DT6 behavior of RFC 8986 except the lookup is in IPv6 multicast table.

5.2.1.3. End.DTMC46: Decapsulation and Specific IP Multicast Table Lookup

The "Endpoint with decapsulation and specific IP Multicast table lookup" behavior ("End.DTMC46" for short) is similar to End.DT4 and End.DT6 behaviors of RFC 8986 except the lookup is in IP multicast table.

6. Dampening of MVPN routes

When P2MP trees are used as P-Tunnels for S-PMSI A-D routes, change in group membership of receivers connected to PEs has direct impact on the Leaf node set of a P2MP tree. If group membership changes frequently for a large number of groups with a lot of receivers across sites connected to different PEs, it can have an impact on the interaction between PEs and the PCE.

Since Leaf A-D routes are used to discover Leaf PE of a P2MP tree, it is RECOMMENDED that PEs SHOULD damp Leaf A-D routes as described in Section 6.1 of RFC 7899 [RFC7899]. PEs MAY also implement procedures for damping other Auto-Discovery and BGP C-multicast routes as described in [RFC7899].

7. SR P2MP Trees for EVPN

BGP MPLS Ethernet VPN specified in RFC 7432 specifies Inclusive Multicast Ethernet Tag route to support Broadcast, Unknown Unicast and Multicast (BUM) traffic. This IMET route is the equivalent of MVPN Intra-AS I-PMSI route and is advertised with a PMSI Tunnel Attribute (PTA) as specified in RFC 6514 to advertise the inclusive P-Tunnels.

[I-D.ietf-bess-evpn-bum-procedure-updates] updates BUM procedures to support selective P-Tunnels and P-Tunnel segmentation in EVPN. That document specifies new route types that are advertised with PTA, including Selective PMSI (S-PMSI) Auto-Discovery route.

These inclusive/selective P-Tunnels can be realized by SR P2MP trees. As with other types of P2MP P-Tunnels, the ESI label used for split horizon MUST be either upstream assigned by PE advertising the IMET or S-PMSI route, or assigned from a global context such as "Domain- wide Common Block" (DCB) as specified in [I-D.ietf-bess-mvpn-evpn-aggregation-label].

[I-D.ietf-bess-evpn-irb-mcast] specifies procedures to support Inter-Subnet Multicast. [I-D.ietf-bess-evpn-mvpn-seamless-interop] specifies how MVPN SAFI routes can be used to support Inter-Subnet Multicast. The P-Tunnels advertised in PTA of either EVPN and MVPN routes as specified in these documents respectively can be realized by SR P2MP trees.

SRv6 P2MP trees can serve as an underlay multicast as described in RFC 8293 Section 3.4. A NVE encapsulates a tenant packet in an SRv6 header and deliver it over SRv6 P2MP trees to other NVEs.

The same procedures specified for MVPN are used to collect the leaf information of corresponding SR P2MP tree (either based on IMET route or Leaf A-D routes in response to x-PMSI routes), to pass the tree information to the PCE controller, and to get back tree forwarding state used for customer multicast traffic forwarding.

8. IANA Considerations

IANA has assigned the value 0x0C for "SR-MPLS P2MP Tree" in the "P-Multicast Service Interface Tunnel (PMSI Tunnel) Tunnel Types" registry https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#pmsi-tunnel-types [RFC 7385] in the "Border Gateway Protocol (BGP) Parameters" registry.

IANA is requested to assign codepoint for "SRv6 P2MP Tree" in the "P-Multicast Service Interface Tunnel (PMSI Tunnel) Tunnel Types" registry https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#pmsi-tunnel-types [RFC 7385] in the "Border Gateway Protocol (BGP) Parameters" registry. A proposed value is 0x0D.

This document requests IANA to allocate the following codepoints in "SRv6 Endpoint Behaviors" sub-registry of "Segment Routing Parameters" top-level registry.

Value	Hex	Endpoint behavior	Reference
76	0x004C	End.DTMC4	[This.ID]
77	0x004D	End.DTMC6	[This.ID]
78	0x004E	End.DTMC46	[This.ID]

Table 1: IETF - SRv6 Endpoint Behaviors

9. Security Considerations

The procedures in this document do not introduce any additional security considerations beyond those mentioned in [RFC6513] and [RFC6514]. For general security considerations applicable to P2MP trees, please refer to [I-D.ietf-pim-sr-p2mp-policy] .

10. Acknowledgements

The authors would like to acknowledge Luc Andre Burdett reviewing the document..

11. Contributors

Zafar Ali Cisco Systems, Inc. US

Email: zali@cisco.com

Ehsan Hemmati Cisco Systems, Inc. US

Email: ehemmati@cisco.com

Jayant Kotalwar Nokia Mountain View US

Email: jayant.kotalwar@nokia.com

Tanmoy Kundu Nokia Mountain View US

Email: tanmoy.kundu@nokia.com

Clayton Hassen Bell CanadaVancouver CA

Email: clayton.hassen@bell.ca

12. References

12.1. Normative References

[I-D.ietf-pim-sr-p2mp-policy] Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., and Z. J. Zhang, "Segment Routing Point-to-Multipoint Policy", Work in Progress, Internet-Draft, draft-ietf-pim-sr-p2mp-policy-07, 11 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-pim-sr-p2mp-policy-07>.

[I-D.ietf-spring-sr-replication-segment] Voyer, D., Filsfils, C., Parekh, R., Bidgoli, H., and Z. J. Zhang, "SR Replication segment for Multi-point Service Delivery", Work in Progress, Internet-Draft, draft-ietf-spring-sr-replication-segment-19, 28 August 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-sr-replication-segment-19>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC6513] Rosen, E., Ed. and R. Aggarwal, Ed., "Multicast in MPLS/BGP IP VPNs", RFC 6513, DOI 10.17487/RFC6513, February 2012, <https://www.rfc-editor.org/info/rfc6513>.

[RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012, <https://www.rfc-editor.org/info/rfc6514>.

[RFC7988] Rosen, E., Ed., Subramanian, K., and Z. Zhang, "Ingress Replication Tunnels in Multicast VPN", RFC 7988, DOI 10.17487/RFC7988, October 2016, <https://www.rfc-editor.org/info/rfc7988>.

[RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <https://www.rfc-editor.org/info/rfc8660>.

[RFC8669] Previdi, S., Filsfils, C., Lindem, A., Ed., Sreekantiah, A., and H. Gredler, "Segment Routing Prefix Segment Identifier Extensions for BGP", RFC 8669, DOI 10.17487/RFC8669, December 2019, <https://www.rfc-editor.org/info/rfc8669>.

[RFC8754]
Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

[RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <https://www.rfc-editor.org/info/rfc8986>.

[RFC9252] Dawra, G., Ed., Talaulikar, K., Ed., Raszuk, R., Decraene, B., Zhuang, S., and J. Rabadan, "BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)", RFC 9252, DOI 10.17487/RFC9252, July 2022, <https://www.rfc-editor.org/info/rfc9252>.

12.2. Informative References

[I-D.ietf-bess-evpn-bum-procedure-updates]
Zhang, Z. J., Lin, W., Rabadan, J., Patel, K., and A. Sajassi, "Updates on EVPN BUM Procedures", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-bum-procedure-updates-14, 18 November 2021, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-evpn-bum-procedure-updates-14>.

[I-D.ietf-bess-evpn-irb-mcast] Lin, W., Zhang, Z. J., Drake, J., Rosen, E. C., Rabadan, J., and A. Sajassi, "EVPN Optimized Inter-Subnet Multicast (OISM) Forwarding", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-irb-mcast-09, 21 February 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-evpn-irb-mcast-09>.

[I-D.ietf-bess-evpn-mvpn-seamless-interop]
Sajassi, A., Thiruvenkatasamy, K., Thoria, S., Gupta, A., and L. Jalil, "Seamless Multicast Interoperability between EVPN and MVPN PEs", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-mvpn-seamless-interop-06, 23 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-evpn-mvpn-seamless-interop-06>.

[I-D.ietf-bess-mvpn-evpn-aggregation-label] Zhang, Z. J., Rosen, E. C., Lin, W., Li, Z., and I. Wijnands, "MVPN/EVPN Tunnel Aggregation with Common Labels", Work in Progress, Internet-Draft, draft-ietf-bess-mvpn-evpn-aggregation-label-14, 4 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-mvpn-evpn-aggregation-label-14>.

[I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-segment-routing-te-policy-26, 23 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-segment-routing-te-policy-26>.

[RFC6625] Rosen, E., Ed., Rekhter, Y., Ed., Hendrickx, W., and R. Qiu, "Wildcards in Multicast VPN Auto-Discovery Routes", RFC 6625, DOI 10.17487/RFC6625, May 2012, <https://www.rfc-editor.org/info/rfc6625>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

[RFC7899] Morin, T., Ed., Litkowski, S., Patel, K., Zhang, Z., Kebler, R., and J. Haas, "Multicast VPN State Damping", RFC 7899, DOI 10.17487/RFC7899, June 2016, <https://www.rfc-editor.org/info/rfc7899>.

[RFC8293] Ghanwani, A., Dunbar, L., McBride, M., Bannai, V., and R. Krishnan, "A Framework for Multicast in Network Virtualization over Layer 3", RFC 8293, DOI 10.17487/RFC8293, January 2018, <https://www.rfc-editor.org/info/rfc8293>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

Authors' Addresses

Rishabh Parekh
Cisco Systems, Inc.
170 W. Tasman Drive
San Jose, CA 95134
United States of America

Email: riparekh@cisco.com

Clarence Filsfils
Cisco Systems, Inc.
Brussels
Belgium

Email: cfilsfil@cisco.com

Arvind Venkateswaran
Cisco Systems, Inc.
170 W. Tasman Drive
San Jose, CA 95134
United States of America

Email: arvvenka@cisco.com

Hooman Bidgoli
Nokia
Ottawa
Canada

Email: hooman.bidgoli@nokia.com

Daniel Voyer
Bell Canada
Montreal
Canada

Email: daniel.voyer@bell.ca

Zhaohui Zhang
Juniper Networks

Email: zzhang@juniper.net