Internet Engineering Task Force (IETF)
Request for Comments: 9251
Category: Standards Track
ISSN: 2070-1721

A. Sajassi
S. Thoria
M. Mishra
Cisco Systems
K. Patel
Arrcus
J. Drake
W. Lin
Juniper Networks

June 2022

Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Proxies for Ethernet VPN (EVPN)

Abstract

This document describes how to support endpoints running the Internet Group Management Protocol (IGMP) or Multicast Listener Discovery (MLD) efficiently for the multicast services over an Ethernet VPN (EVPN) network by incorporating IGMP/MLD Proxy procedures on EVPN Provider Edges (PEs).

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9251.

Table of Contents

1. Introduction

In data center (DC) applications, a point of delivery (POD) can consist of a collection of servers supported by several top-of-rack (ToR) and spine switches. This collection of servers and switches are self-contained and may have their own control protocol for intra-POD communication and orchestration. However, EVPN is used as a standard way of inter-POD communication for both intra-DC and inter-DC. A subnet can span across multiple PODs and DCs. EVPN provides a robust multi-tenant solution with extensive multihoming capabilities to stretch a subnet (VLAN) across multiple PODs and DCs. There can be many hosts (several hundreds) attached to a subnet that is stretched across several PODs and DCs.

These hosts express their interests in multicast groups on a given subnet/VLAN by sending IGMP/MLD Membership Reports for their interested multicast group(s). Furthermore, an IGMP/MLD router periodically sends Membership Queries to find out if there are hosts on that subnet that are still interested in receiving multicast traffic for that group. The IGMP/MLD Proxy solution described in

this document accomplishes three objectives:

1. Reduce flooding of IGMP/MLD messages: Just like the ARP/Neighbor Discovery (ND) suppression mechanism in EVPN to reduce the flooding of ARP messages over EVPN, it is also desired to have a mechanism to reduce the flooding of IGMP/MLD messages (both Queries and Membership Reports) in EVPN.

2. Distributed anycast multicast proxy: It is desirable for the EVPN network to act as a distributed anycast multicast router with respect to IGMP/MLD Proxy function for all the hosts attached to that subnet.

3. Selective multicast: This describes forwarding multicast traffic over the EVPN network such that it only gets forwarded to the PEs that have interests in the multicast group(s). This document shows how this objective may be achieved when ingress replication is used to distribute the multicast traffic among the PEs. Procedures for supporting selective multicast using Point-to-Multipoint (P2MP) tunnels can be found in [EVPN-BUM].

The first two objectives are achieved by using the IGMP/MLD Proxy on the PE. The third objective is achieved by setting up a multicast tunnel among only the PEs that have interest in the multicast group(s) based on the trigger from IGMP/MLD Proxy processes. The proposed solutions for each of these objectives are discussed in the following sections.

2. Specification of Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminology

AC: Attachment Circuit

All-Active Redundancy Mode: When all PEs attached to an Ethernet segment are allowed to forward known unicast traffic to/from that Ethernet segment for a given VLAN, then the Ethernet segment is defined to be operating in All-Active redundancy mode.

BD: Broadcast Domain. As per [RFC7432], an EVPN instance (EVI) consists of a single BD or multiple BDs. In case of a VLAN bundle and a VLAN-aware bundle service model, an EVI contains multiple BDs. Also, in this document, BD and subnet are equivalent terms.

DC: Data Center

ES: Ethernet segment. This is when a customer site (device or network) is connected to one or more PEs via a set of Ethernet links.

ESI: Ethernet Segment Identifier. This is a unique non-zero identifier that identifies an Ethernet segment.

Ethernet Tag: It identifies a particular broadcast domain, e.g., a VLAN. An EVPN instance consists of one or more broadcast domains.

EVI: EVPN Instance. This spans the Provider Edge (PE) devices participating in that EVPN.

EVPN: Ethernet Virtual Private Network

IGMP: Internet Group Management Protocol

IR: Ingress Replication

MLD: Multicast Listener Discovery

OIF: Outgoing Interface for multicast. It can be a physical interface, virtual interface, or tunnel.

PE: Provider Edge

POD: Point of Delivery

S-PMSI: Selective P-Multicast Service Interface. This is a conceptual interface for a PE to send customer multicast traffic to some of the PEs in the same VPN.

Single-Active Redundancy Mode: When only a single PE, among all the PEs attached to an Ethernet segment, is allowed to forward traffic to/from that Ethernet segment for a given VLAN, then the Ethernet segment is defined to be operating in Single-Active redundancy mode.

SMET: Selective Multicast Ethernet Tag

ToR: Top of Rack

This document also assumes familiarity with the terminology of [RFC7432], [RFC3376], and [RFC2236]. When this document uses the term "IGMP Membership Report", the text equally applies to the MLD Membership Report. Similarly, text for IGMPv2 applies to MLDv1, and text for IGMPv3 applies to MLDv2. IGMP/MLD version encoding in the BGP update is stated in Section 9.

It is important to note that when there is text considering whether a PE indicates support for IGMP proxying, the corresponding behavior has a natural analog for indicating support for MLD proxying, and the analogous requirements apply as well.

4. IGMP/MLD Proxy

The IGMP Proxy mechanism is used to reduce the flooding of IGMP messages over an EVPN network, similar to the ARP proxy used in reducing the flooding of ARP messages over EVPN. It also provides a triggering mechanism for the PEs to set up their underlay multicast tunnels. The IGMP Proxy mechanism consists of two components:

1. Proxy for IGMP Membership Reports

2. Proxy for IGMP Membership Queries

The goal of IGMP and MLD proxying is to make the EVPN behave seamlessly for the tenant systems with respect to multicast operations while using a more efficient delivery system for signaling and delivery across the VPN. Accordingly, group state must be tracked synchronously among the PEs serving the VPN, with join and leave events propagated to the peer PEs and each PE tracking the state of each of its peer PEs with respect to whether there are locally attached group members (and in some cases, senders), what version(s) of IGMP/MLD are in use for those locally attached group members, etc. In order to perform this translation, each PE acts as an IGMP router for the locally attached domain, maintains the requisite state on locally attached nodes, sends periodic Membership Queries, etc. The role of EVPN Selective Multicast Ethernet Tag (SMET) route propagation is to ensure that each PE’s local state is

propagated to the other PEs so that they share a consistent view of the overall IGMP Membership Request and Leave Group state. It is important to note that the need to keep such local state can be triggered by either local IGMP traffic or BGP EVPN signaling. In most cases, a local IGMP event will need to be signaled over EVPN, though state initiated by received EVPN traffic will not always need to be relayed to the locally attached domain.

4.1. Proxy Reporting

When IGMP is used between hosts and their first hop EVPN router (EVPN PE), proxy reporting is used by the EVPN PE to summarize (when possible) reports received from downstream hosts and propagate them in BGP to other PEs that are interested in the information. This is done by terminating the IGMP Membership Reports in the first hop PE and translating and exchanging the relevant information among EVPN BGP speakers. The information is again translated back to an IGMP message at the recipient EVPN speaker. Thus, it helps create an IGMP overlay subnet using BGP. In order to facilitate such an overlay, this document also defines a new EVPN route type Network Layer Reachability Information (NLRI) and the EVPN SMET route, along with its procedures to help exchange and register IGMP multicast groups; see Section 9.

4.1.1. IGMP/MLD Membership Report Advertisement in BGP

When a PE wants to advertise an IGMP Membership Report using the BGP EVPN route, it follows the proceeding rules (BGP encoding is stated in Section 9). The first four rules are applicable to the originator PE, and the last three rules are applicable to remote PE processing SMET routes:

Processing at the BGP route originator:

1. When the first hop PE receives IGMP Membership Reports belonging to the same IGMP version from different attached hosts for the same (*,G) or (S,G), it SHOULD send a single BGP message corresponding to the very first IGMP Membership Request (BGP update as soon as possible) for that (*,G) or (S,G). This is because BGP is a stateful protocol, and no further transmission of the same report is needed. If the IGMP Membership Request is for (*,G), then the Multicast Group Address MUST be sent along with the corresponding version flag (v2 or v3) set. In case of IGMPv3, the exclude flag MUST also be set to indicate that no source IP address must be excluded (include all sources "*"). If the IGMP Membership Report is for (S,G), then besides setting the Multicast Group Address along with the v3 flag, the source IP address and the Include/Exclude (IE) flag MUST be set. It should be noted that, when advertising the EVPN route for (S,G), the only valid version flag is v3 (v2 flags MUST be set to 0).

2. When the first hop PE receives an IGMPv3 Membership Report for (S,G) on a given BD, it MUST advertise the corresponding EVPN SMET route, regardless of whether the source (S) is attached to itself or not, in order to facilitate the source move in the future.

3. When the first hop PE receives an IGMP version-X Membership Report first for (*,G) and then later receives an IGMP version-Y Membership Report for the same (*,G), then it MUST re-advertise the same EVPN SMET route with the flag for version-Y set in addition to any previously set version flag(s). In other words, the first hop PE MUST NOT withdraw the EVPN route before sending the new route because the Flags field is not part of BGP route key processing.

4. When the first hop PE receives an IGMP version-X Membership Report first for (*,G) and then later receives an IGMPv3 Membership Report for the same Multicast Group Address but for a specific source address S, then the PE MUST advertise a new EVPN SMET route with the v3 flag set (and v2 reset). The IE flag also needs to be set accordingly. Since the source IP address is used as part of BGP route key processing, it is considered to be a new BGP route advertisement. When different versions of IGMP Membership Report are received, the final state MUST be as per Section 5.1 of [RFC3376]. At the end of the route processing, local and remote group record state MUST be as per Section 5.1 of [RFC3376].

Processing at the BGP route receiver:

1. When a PE receives an EVPN SMET route with more than one version flag set, it will generate the corresponding IGMP Report for (*,G) for each version specified in the Flags field. With multiple version flags set, there must not be a source IP address in the received EVPN route. If there is, then an error SHOULD be logged. If the v3 flag is set (in addition to v2), then the IE flag MUST indicate "exclude". If not, then an error SHOULD be logged. The PE MUST generate an IGMP Membership Report for that (*,G) and each IGMP version in the version flag.

2. When a PE receives a list of EVPN SMET NLRIs in its BGP update message, each with a different source IP address and the same Multicast Group Address, and the version flag is set to v3, then the PE generates an IGMPv3 Membership Report with a record corresponding to the list of source IP addresses and the group address, along with the proper indication of inclusion/exclusion.

3. Upon receiving an EVPN SMET route(s) and before generating the corresponding IGMP Membership Request(s), the PE checks to see whether it has a Customer Edge (CE) multicast router for that BD on any of its ESs . The PE provides such a check by listening for PIM Hello messages on that AC, i.e., (ES,BD). If the PE does have the router’s ACs, then the generated IGMP Membership Request(s) is sent to those ACs. If it doesn’t have any of the router’s ACs, then no IGMP Membership Request(s) needs to be generated. This is because sending IGMP Membership Requests to other hosts can result in unintentionally preventing a host from joining a specific multicast group using IGMPv2, i.e., if the PE does not receive a Membership Report from the host, it will not forward multicast data to it. Per [RFC4541] , when an IGMPv2 host receives a Membership Report for a group address that it intends to join, the host will suppress its own Membership Report for the same group, and if the PE does not receive an IGMP Membership Report from the host, it will not forward multicast data to it. In other words, an IGMPv2 Membership Report MUST NOT be sent on an AC that does not lead to a CE multicast router. This message suppression is a requirement for IGMPv2 hosts. This is not a problem for hosts running IGMPv3, because there is no suppression of IGMP Membership Reports.

4.1.2. IGMP/MLD Leave Group Advertisement in BGP

When a PE wants to withdraw an EVPN SMET route corresponding to an IGMPv2 Leave Group or IGMPv3 "Leave" equivalent message, it follows the rules below. The first rule defines the procedure at the originator PE, and the last two rules talk about procedures at the remote PE:

Processing at the BGP route originator:

1. When a PE receives an IGMPv2 Leave Group or its "Leave"

equivalent message for IGMPv3 from its attached host, it checks to see if this host is the last host that is interested in this multicast group by sending a query for the multicast group. If the host was indeed the last one (i.e., no responses are received for the query), then the PE MUST re-advertise the EVPN SMET route with the corresponding version flag reset. If this is the last version flag to be reset, then instead of re-advertising the EVPN route with all version flags reset, the PE MUST withdraw the EVPN route for that (*,G).

Processing at the BGP route receiver:

1. When a PE receives an EVPN SMET route for a given (*,G), it compares the received version flags from the route with its per-PE stored version flags. If the PE finds that a version flag associated with the (*,G) for the remote PE is reset, then the PE MUST generate IGMP Leave for that (*,G) toward its local interface (if any), which is attached to the multicast router for that multicast group. It should be noted that the received EVPN route MUST have at least one version flag set. If all version flags are reset, it is an error because the PE should have received an EVPN route withdraw for the last version flag. An error MUST be considered as a BGP error, and the PE MUST apply the "treat-as-withdraw" procedure per [RFC7606].

2. When a PE receives an EVPN SMET route withdraw, it removes the remote PE from its OIF list for that multicast group, and if there are no more OIF entries for that multicast group (either locally or remotely), then the PE MUST stop responding to Membership Queries from the locally attached router (if any). If there is a source for that multicast group, the PE stops sending multicast traffic for that source.

4.2. Proxy Querier

As mentioned in the previous sections, each PE MUST have proxy querier functionality for the following reasons:

1. to enable the collection of EVPN PEs providing Layer 2 Virtual Private Network (L2VPN) service to act as a distributed multicast router with an anycast IP address for all attached hosts in that subnet

2. to enable suppression of IGMP Membership Reports and Membership Queries over MPLS/IP core

5. Operation

Consider the EVPN network in Figure 1, where there is an EVPN instance configured across the PEs (namely PE1, PE2, and PE3). Let’s consider that this EVPN instance consists of a single bridge domain (single subnet) with all the hosts and sources and the multicast router connected to this subnet. PE1 only has hosts (host denoted by Hx) connected to it. PE2 has a mix of hosts and a multicast source. PE3 has a mix of hosts, a multicast source (source denoted by Sx), and a multicast router (router denoted by Rx). Furthermore, let’s consider that for (S1,G1), R1 is used as the multicast router. The following subsections describe the IGMP Proxy operation in different PEs with regard to whether the locally attached devices for that subnet are:

* only hosts,

* a mix of hosts and a multicast source, or

* a mix of hosts, a multicast source, and a multicast router.

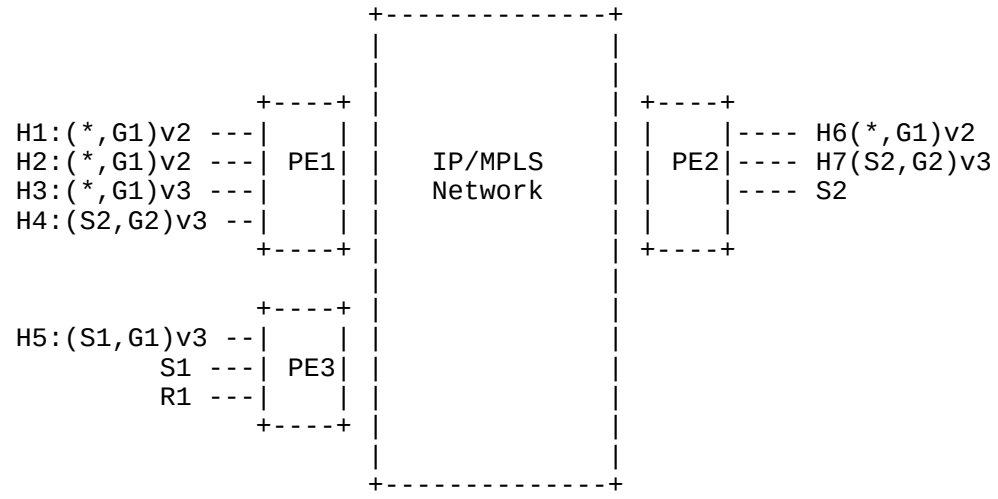


Figure 1: EVPN Network

5.1. PE with Only Attached Hosts for a Given Subnet

When PE1 receives an IGMPv2 Membership Report from H1, it does not forward this Membership Report to any of its other ports (for this subnet) because all these local ports are associated with the hosts. PE1 sends an EVPN SMET route corresponding to this Membership Report for (*,G1) and sets the v2 flag. This EVPN route is received by PE2 and PE3, which are the members of the same BD (i.e., same EVI in case of a VLAN-based service or EVI and VLAN in case of a VLAN-aware bundle service). PE3 reconstructs the IGMPv2 Membership Report from this EVPN BGP route and only sends it to the port(s) with multicast routers attached to it (for that subnet). In this example, PE3 sends the reconstructed IGMPv2 Membership Report for (*,G1) only to R1. Furthermore, even though PE2 receives the EVPN BGP route, it does not send it to any of its ports for that subnet (viz., ports associated with H6 and H7).

When PE1 receives the second IGMPv2 Membership Report from H2 for the same multicast group (*,G1), it only adds that port to its OIF list, but it doesn’t send any EVPN BGP routes because there is no change in information. However, when it receives the IGMPv3 Membership Report from H3 for the same (*,G1), besides adding the corresponding port to its OIF list, it re-advertises the previously sent EVPN SMET route with the v3 and exclude flag set.

Finally, when PE1 receives the IGMPv3 Membership Report from H4 for (S2,G2), it advertises a new EVPN SMET route corresponding to it.

5.2. PE with a Mix of Attached Hosts and a Multicast Source

The main difference in this case is that when PE2 receives the IGMPv3 Membership Report from H7 for (S2,G2), it advertises it in BGP to support the source moving, even though PE2 knows that S2 is attached to its local AC. PE2 adds the port associated with H7 to its OIF list for (S2,G2). The processing for IGMPv2 received from H6 is the same as the IGMPv2 Membership Report described in the previous section.

5.3. PE with a Mix of Attached Hosts, a Multicast Source, and a Router

The main difference in this case relative to the previous two sections is that IGMPv2/v3 Membership Report messages received locally need to be sent to the port associated with router R1. Furthermore, the Membership Reports received via BGP (SMET) need to be passed to the R1 port but filtered for all other ports.

6. All-Active Multihoming

Because the Link Aggregation Group (LAG) flow hashing algorithm used by the CE is unknown at the PE, in an All-Active redundancy mode, it must be assumed that the CE can send a given IGMP message to any one of the multihomed PEs, either Designated Forwarder (DF) or non-DF, i.e., different IGMP Membership Request messages can arrive at different PEs in the redundancy group. Furthermore, their corresponding Leave messages can arrive at PEs that are different from the ones that received the Membership Report. Therefore, all PEs attached to a given Ethernet segment (ES) must coordinate the IGMP Membership Request and Leave Group (x,G) state, where x may be either "*" or a particular source S for each BD on that ES. Each PE has a local copy of that state, and the EVPN signaling serves to synchronize that state across PEs. This allows the DF for that (ES,BD) to correctly advertise or withdraw a SMET route for that (x,G) group in that BD when needed. All-Active multihoming PEs for a given ES MUST support IGMP synchronization procedures described in this section if they need to perform IGMP Proxy for hosts connected to that ES.

6.1. Local IGMP/MLD Membership Report Synchronization

When a PE, either DF or non-DF, receives an IGMP Membership Report for (x,G) on a given multihomed ES operating in All-Active redundancy mode, it determines the BD to which the IGMP Membership Report belongs. If the PE doesn’t already have the local IGMP Membership Request (x,G) state for that BD on that ES, it MUST instantiate that local IGMP Membership Request (x,G) state and MUST advertise a BGP IGMP Membership Report Synch route for that (ES,BD). The local IGMP Membership Request (x,G) state refers to the IGMP Membership Request (x,G) state that is created as a result of processing an IGMP Membership Report for (x,G).

The IGMP Membership Report Synch route MUST carry the ES-Import Route Target (RT) for the ES on which the IGMP Membership Report was received. Thus, it MUST only be imported by the PEs attached to that ES and not any other PEs.

When a PE, either DF or non-DF, receives an IGMP Membership Report Synch route, it installs that route, and if it doesn’t already have the IGMP Membership Request (x,G) state for that (ES,BD), it MUST instantiate that IGMP Membership Request (x,G) state, i.e., the IGMP Membership Request (x,G) state is the union of the local IGMP Membership Report (x,G) state and the installed IGMP Membership Report Synch route. If the DF did not already advertise (originate) a SMET route for that (x,G) group in that BD, it MUST do so now.

When a PE, either DF or non-DF, deletes its local IGMP Membership Request (x,G) state for that (ES,BD), it MUST withdraw its BGP IGMP Membership Report Synch route for that (ES,BD).

When a PE, either DF or non-DF, receives the withdrawal of an IGMP Membership Report Synch route from another PE, it MUST remove that route. When a PE has no local IGMP Membership Request (x,G) state and it has no installed IGMP Membership Report Synch routes, it MUST remove that IGMP Membership Request (x,G) state for that (ES,BD). If the DF no longer has the IGMP Membership Request (x,G) state for that BD on any ES for which it is the DF, it MUST withdraw its SMET route for that (x,G) group in that BD.

In other words, a PE advertises a SMET route for that (x,G) group in that BD when it has the IGMP Membership Request (x,G) state on at least one ES for which it is the DF, and it withdraws that SMET route when it does not have an IGMP Membership Request (x,G) state in that

BD on any ES for which it is the DF.

6.2. Local IGMP/MLD Leave Group Synchronization

When a PE, either DF or non-DF, receives an IGMP Leave Group message for (x,G) from the attached CE on a given multihomed ES operating in All-Active redundancy mode, it determines the BD to which the IGMPv2 Leave Group belongs. Regardless of whether it has the IGMP Membership Request (x,G) state for that (ES,BD), it initiates the (x,G) leave group synchronization procedure, which consists of the following steps:

1. It computes the Maximum Response Time, which is the duration of the (x,G) leave group synchronization procedure. This is the product of two locally configured values, Last Member Query Count and Last Member Query Interval (described in Section 3 of [RFC2236]), plus a delta corresponding to the time it takes for a BGP advertisement to propagate between the PEs attached to the multihomed ES (delta is a consistently configured value on all PEs attached to the multihomed ES).

2. It starts the Maximum Response Time timer. Note that the receipt of subsequent IGMP Leave Group messages or BGP Leave Synch routes for (x,G) do not change the value of a currently running Maximum Response Time timer and are ignored by the PE.

3. It initiates the Last Member Query procedure described in Section 3 of [RFC2236]; viz., it sends a number of Group-Specific Query (x,G) messages (Last Member Query Count) at a fixed interval (Last Member Query Interval) to the attached CE.

4. It advertises an IGMP Leave Synch route for that (ES,BD). This route notifies the other multihomed PEs attached to the given multihomed ES that it has initiated an (x,G) leave group synchronization procedure, i.e., it carries the ES-Import RT for the ES on which the IGMP Leave Group was received. It also contains the Maximum Response Time.

5. When the Maximum Response Time timer expires, the PE that has advertised the IGMP Leave Synch route withdraws it.

6.2.1. Remote Leave Group Synchronization

When a PE, either DF or non-DF, receives an IGMP Leave Synch route, it installs that route and it starts a timer for (x,G) on the specified (ES,BD), whose value is set to the Maximum Response Time in the received IGMP Leave Synch route. Note that the receipt of subsequent IGMPv2 Leave Group messages or BGP Leave Synch routes for (x,G) do not change the value of a currently running Maximum Response Time timer and are ignored by the PE.

6.2.2. Common Leave Group Synchronization

If a PE attached to the multihomed ES receives an IGMP Membership Report for (x,G) before the Maximum Response Time timer expires, it advertises a BGP IGMP Membership Report Synch route for that (ES,BD). If it doesn’t already have the local IGMP Membership Request (x,G) state for that (ES,BD), it instantiates that local IGMP Membership Request (x,G) state. If the DF is not currently advertising (originating) a SMET route for that (x,G) group in that BD, it does so now.

If a PE attached to the multihomed ES receives an IGMP Membership Report Synch route for (x,G) before the Maximum Response Time timer expires, it installs that route, and if it doesn’t already have the IGMP Membership Request (x,G) state for that BD on that ES, it

instantiates that IGMP Membership Request (x,G) state. If the DF has not already advertised (originated) a SMET route for that (x,G) group in that BD, it does so now.

When the Maximum Response Time timer expires, a PE that has advertised an IGMP Leave Synch route withdraws it. Any PE attached to the multihomed ES, which started the Maximum Response Time and has no local IGMP Membership Request (x,G) state and no installed IGMP Membership Report Synch routes, removes the IGMP Membership Request (x,G) state for that (ES,BD). If the DF no longer has the IGMP Membership Request (x,G) state for that BD on any ES for which it is the DF, it withdraws its SMET route for that (x,G) group in that BD.

6.3. Mass Withdraw of the Multicast Membership Report Synch Route in Case of Failure

A PE that has received an IGMP Membership Request would have synced the IGMP Membership Report by the procedure defined in Section 6.1. If a PE with the local Membership Report state goes down or the PE to CE link goes down, it would lead to a mass withdraw of multicast routes. Remote PEs (PEs where these routes were remote IGMP Membership Reports) SHOULD NOT remove the state immediately; instead, General Query SHOULD be generated to refresh the states. There are several ways to detect failure at a peer, e.g., using IGP next-hop tracking or ES route withdraw.

7. Single-Active Multihoming

Note that to facilitate state synchronization after failover, the PEs attached to a multihomed ES operating in Single-Active redundancy mode SHOULD also coordinate the IGMP Membership Report (x,G) state. In this case, all IGMP Membership Report messages are received by the DF and distributed to the non-DF PEs using the procedures described above.

8. Selective Multicast Procedures for IR Tunnels

If an ingress PE uses ingress replication, then for a given (x,G) group in a given BD:

1. It sends (x,G) traffic to the set of PEs not supporting IGMP or MLD Proxies. This set consists of any PE that has advertised an Inclusive Multicast Ethernet Tag (IMET) route for the BD without a Multicast Flags Extended Community or with a Multicast Flags Extended Community in which neither the IGMP Proxy support nor the MLD Proxy support flags are set.

2. It sends (x,G) traffic to the set of PEs supporting IGMP or MLD Proxies and has listeners for that (x,G) group in that BD. This set consists of any PE that has advertised an IMET route for the BD with a Multicast Flags Extended Community in which the IGMP Proxy support and/or the MLD Proxy support flags are set and that has advertised a SMET route for that (x,G) group in that BD.

9. BGP Encoding

This document defines three new BGP EVPN routes to carry IGMP Membership Reports. The route types are known as:

6 - Selective Multicast Ethernet Tag Route

7 - Multicast Membership Report Synch Route

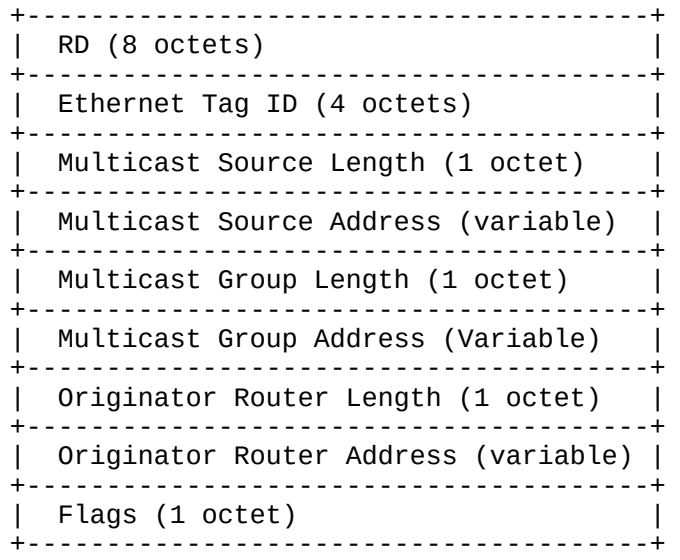
8 - Multicast Leave Synch Route

The detailed encoding and procedures for these route types are

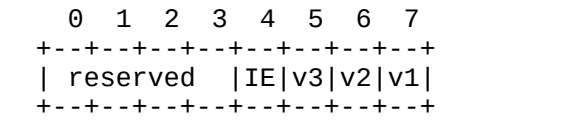
described in subsequent sections.

9.1. Selective Multicast Ethernet Tag Route

A SMET route-type-specific EVPN NLRI consists of the following:



For the purpose of BGP route key processing, all the fields are considered to be part of the prefix in the NLRI, except for the 1-octet Flags field. The Flags fields are defined as follows:



* The least significant bit (bit 7) indicates support for IGMP version 1. Since IGMPv1 is being deprecated, the sender MUST set it to 0 for IGMP and the receiver MUST ignore it.

* The second least significant bit (bit 6) indicates support for IGMP version 2.

* The third least significant bit (bit 5) indicates support for IGMP version 3.

* The fourth least significant bit (bit 4) indicates whether the (S,G) information carried within the route type is of an Include Group type (bit value 0) or an Exclude Group type (bit value 1). The Exclude Group type bit MUST be ignored if bit 5 is not set.

* This EVPN route type is used to carry tenant IGMP multicast group information. The Flags field assists in distributing the IGMP Membership Report of a given host for a given multicast route. The version bits help associate the IGMP version of receivers participating within the EVPN domain.

* The IE bit helps in creating filters for a given multicast route.

* If the route is used for IPv6 (MLD), then bit 7 indicates support for MLD version 1. The second least significant bit (bit 6) indicates support for MLD version 2. Since there is no MLD version 3, in case of IPv6 routes, the third least significant bit MUST be 0. In case of IPv6 routes, the fourth least significant bit MUST be ignored if bit 6 is not set.

* Reserved bits MUST be set to 0 by the sender, and the receiver

MUST ignore the Reserved bits.

9.1.1. Constructing the Selective Multicast Ethernet Tag Route

This section describes the procedures used to construct the SMET route.

The Route Distinguisher (RD) SHOULD be a Type 1 RD [RFC4364]. The value field comprises an IP address of the PE (typically, the loopback address), followed by a number unique to the PE.

The Ethernet Tag ID MUST be set, as per the procedure defined in [RFC7432].

The Multicast Source Length MUST be set to the length of the Multicast Source Address in bits. If the Multicast Source Address field contains an IPv4 address, then the value of the Multicast Source Length field is 32. If the Multicast Source Address field contains an IPv6 address, then the value of the Multicast Source Length field is 128. In case of a (*,G) Membership Report, the Multicast Source Length is set to 0.

The Multicast Source Address is the source IP address from the IGMP Membership Report. In case of a (*,G) Membership Report, this field is not used.

The Multicast Group Length MUST be set to the length of the Multicast Group Address in bits. If the Multicast Group Address field contains an IPv4 address, then the value of the Multicast Group Length field is 32. If the Multicast Group Address field contains an IPv6 address, then the value of the Multicast Group Length field is 128.

The Multicast Group Address is the group address from the IGMP or MLD Membership Report.

The Originator Router Length is the length of the Originator Router Address in bits.

The Originator Router Address is the IP address of the router originating this route. The SMET Originator Router IP address MUST match that of the IMET (or S-PMSI Authentic Data (AD)) route originated for the same EVI by the same downstream PE.

The Flags field indicates the version of IGMP from which the Membership Report was received. It also indicates whether the multicast group had the Include/Exclude bit set.

Reserved bits MUST be set to 0. They can be defined by other documents in the future.

IGMP is used to receive group membership information from hosts by Top-of-the-Rack (ToR) switches. Upon receiving the host’s expression of interest in a particular group membership, this information is then forwarded using the SMET route. The NLRI also keeps track of the receiver’s IGMP version and any source filtering for a given group membership. All EVPN SMET routes are announced per EVI Route Target extended communities (EVI-RT ECs).

9.1.2. Reconstructing IGMP/MLD Membership Reports from the Selective Multicast Route

This section describes the procedures used to reconstruct IGMP/MLD Membership Reports from the SMET route.

* If the Multicast Group Length is 32, the route is translated to the IGMP Membership Request. If the Multicast Group Length is

128, the route is translated to an MLD Membership Request.

* The Multicast Group Address field is translated to the IGMP/MLD group address.

* If the Multicast Source Length is set to 0, it is translated to any source (*). If the Multicast Source Length is non-zero, the Multicast Source Address field is translated to the IGMP/MLD source address.

* If flag bit 7 is set, it translates the Membership Report to be IGMPv1 or MLDv1.

* If flag bit 6 is set, it translates the Membership Report to be IGMPv2 or MLDv2.

* Flag bit 5 is only valid for the IGMP Membership Report; if it is set, it translates to the IGMPv3 report.

* If the IE flag is set, it translates to the IGMP/MLD Exclude mode Membership Report. If the IE flag is not set (0), it translates to the Include mode Membership Report.

9.1.3. Default Selective Multicast Route

If there is a multicast router connected behind the EVPN domain, the PE MAY originate a default SMET (*,*) to get all multicast traffic in the domain.

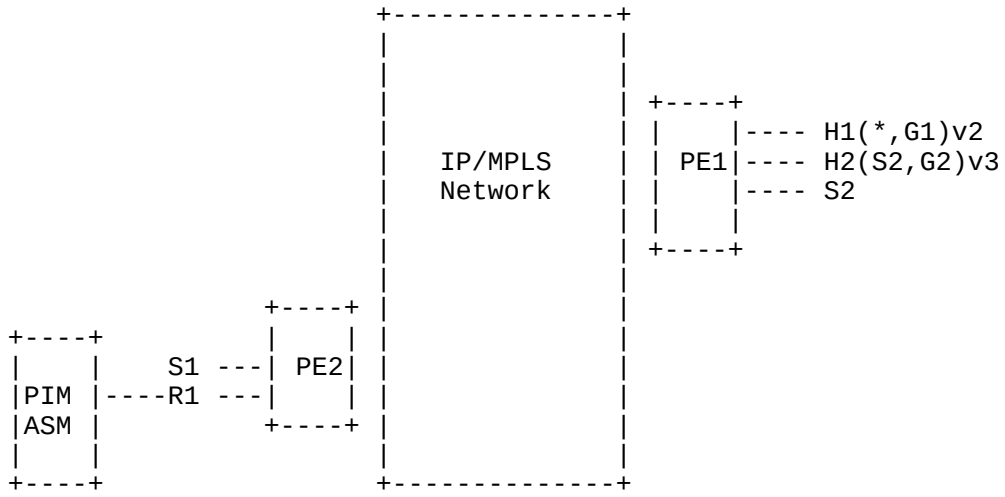


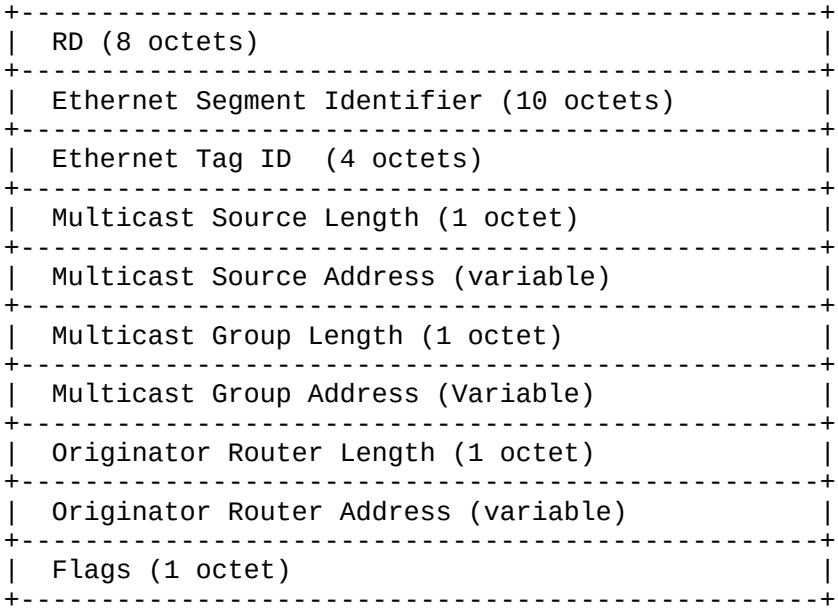
Figure 2: Multicast Router behind the EVPN Domain

Consider the EVPN network in Figure 2, where there is an EVPN instance configured across the PEs. Let’s consider that PE2 is connected to multicast router R1 and there is a network running PIM ASM behind R1. If there are receivers behind the PIM ASM network, the PIM Join would be forwarded to the PIM Rendezvous Point (RP). If receivers behind the PIM ASM network are interested in a multicast flow originated by multicast source S2 (behind PE1), it is necessary for PE2 to receive multicast traffic. In this case, PE2 MUST originate a (*,*) SMET route to receive all of the multicast traffic in the EVPN domain. To generate wildcard (*,*) routes, the procedure from [RFC6625] MUST be used.

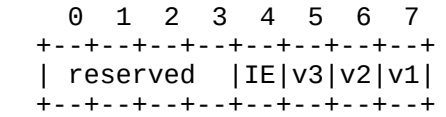
9.2. Multicast Membership Report Synch Route

This EVPN route type is used to coordinate the IGMP Membership Report (x,G) state for a given BD between the PEs attached to a given ES operating in All-Active (or Single-Active) redundancy mode, and it

consists of the following:



For the purpose of BGP route key processing, all the fields are considered to be part of the prefix in the NLRI, except for the 1-octet Flags field, whose fields are defined as follows:



* The least significant bit (bit 7) indicates support for IGMP version 1.

* The second least significant bit (bit 6) indicates support for IGMP version 2.

* The third least significant bit (bit 5) indicates support for IGMP version 3.

* The fourth least significant bit (bit 4) indicates whether the (S, G) information carried within the route type is of an Include Group type (bit value 0) or an Exclude Group type (bit value 1). The Exclude Group type bit MUST be ignored if bit 5 is not set.

* Reserved bits MUST be set to 0.

The Flags field assists in distributing the IGMP Membership Report of a given host for a given multicast route. The version bits help associate the IGMP version of receivers participating within the EVPN domain. The Include/Exclude bit helps in creating filters for a given multicast route.

If the route is being prepared for IPv6 (MLD), then bit 7 indicates support for MLD version 1. The second least significant bit (bit 6) indicates support for MLD version 2. Since there is no MLD version 3, in case of the IPv6 route, the third least significant bit MUST be 0. In case of the IPv6 route, the fourth least significant bit MUST be ignored if bit 6 is not set.

9.2.1. Constructing the Multicast Membership Report Synch Route

This section describes the procedures used to construct the IGMP Membership Report Synch route. Support for these route types is

optional. If a PE does not support this route, then it MUST NOT indicate that it supports "IGMP Proxy" in the Multicast Flags Extended Community for the EVIs corresponding to its multihomed ESs.

An IGMP Membership Report Synch route MUST carry exactly one ES-Import Route Target extended community, i.e., the one that corresponds to the ES on which the IGMP Membership Report was received. It MUST also carry exactly one EVI-RT EC, i.e., the one that corresponds to the EVI on which the IGMP Membership Report was received. See Section 9.5 for details on how to encode and construct the EVI-RT EC.

The RD SHOULD be Type 1 [RFC4364]. The value field comprises an IP address of the PE (typically, the loopback address), followed by a number unique to the PE.

The Ethernet Segment Identifier (ESI) MUST be set to the 10-octet value defined for the ES.

The Ethernet Tag ID MUST be set, as per the procedure defined in [RFC7432].

The Multicast Source Length MUST be set to the length of the Multicast Source Address in bits. 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. In case of a (*,G) Membership Report, the Multicast Source Length is set to 0.

The Multicast Source is the source IP address of the IGMP Membership Report. In case of a (*,G) Membership Report, this field does not exist.

The Multicast Group Length MUST be set to the length of the Multicast Group Address in bits. 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.

The Multicast Group is the group address of the IGMP Membership Report.

The Originator Router Length is the length of the Originator Router Address in bits.

The Originator Router Address is the IP address of the router originating the prefix.

The Flags field indicates the version of IGMP from which the Membership Report was received. It also indicates whether the multicast group had the Include/Exclude bit set.

Reserved bits MUST be set to 0.

9.2.2. Reconstructing IGMP/MLD Membership Reports from a Multicast Membership Report Synch Route

This section describes the procedures used to reconstruct IGMP/MLD Membership Reports from the Multicast Membership Report Synch route.

* If the Multicast Group Length is 32, the route is translated to the IGMP Membership Request. If the Multicast Group Length is 128, the route is translated to an MLD Membership Request.

* The Multicast Group Address field is translated to the IGMP/MLD

group address.

* If the Multicast Source Length is set to 0, it is translated to any source (*). If the Multicast Source Length is non-zero, the Multicast Source Address field is translated to the IGMP/MLD source address.

* If flag bit 7 is set, it translates the Membership Report to be IGMPv1 or MLDv1.

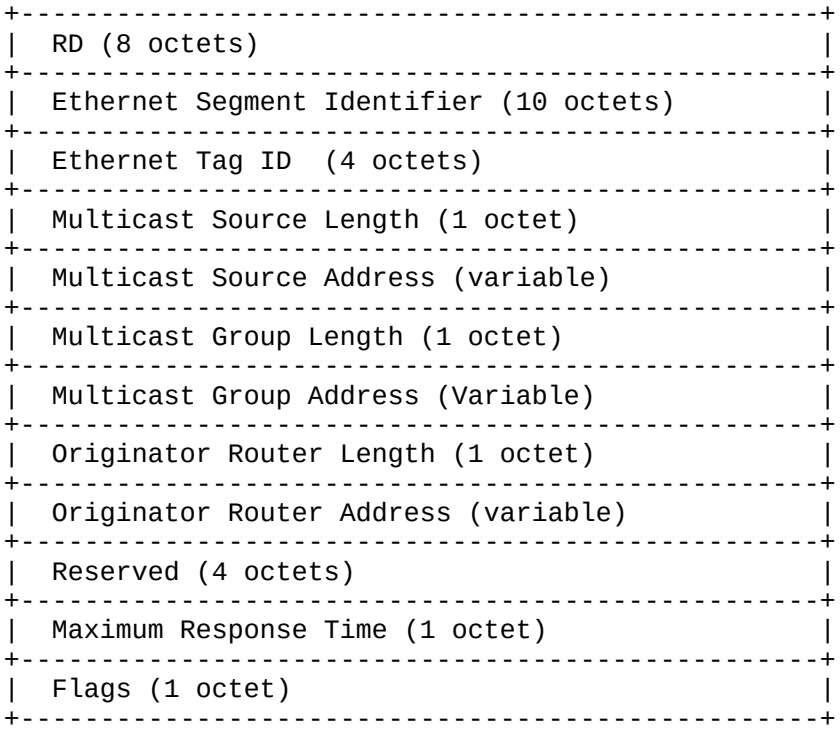
* If flag bit 6 is set, it translates the Membership Report to be IGMPv2 or MLDv2.

* Flag bit 5 is only valid for the IGMP Membership Report; if it is set, it translates to the IGMPv3 report.

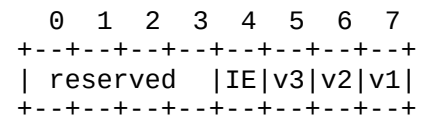
* If the IE flag is set, it translates to the IGMP/MLD Exclude mode Membership Report. If the IE flag is not set (0), it translates to the Include mode Membership Report.

9.3. Multicast Leave Synch Route

This EVPN route type is used to coordinate the IGMP Leave Group (x,G) state for a given BD between the PEs attached to a given ES operating in an All-Active (or Single-Active) redundancy mode, and it consists of the following:



For the purpose of BGP route key processing, all the fields are considered to be part of the prefix in the NLRI, except for the Reserved, Maximum Response Time, and 1-octet Flags fields, which are defined as follows:



* The least significant bit (bit 7) indicates support for IGMP version 1.

* The second least significant bit (bit 6) indicates support for IGMP version 2.

* The third least significant bit (bit 5) indicates support for IGMP version 3.

* The fourth least significant bit (bit 4) indicates whether the (S, G) information carried within the route type is of an Include Group type (bit value 0) or an Exclude Group type (bit value 1). The Exclude Group type bit MUST be ignored if bit 5 is not set.

* Reserved bits MUST be set to 0. They can be defined by other documents in the future.

The Flags field assists in distributing the IGMP Membership Report of a given host for a given multicast route. The version bits help associate the IGMP version of the receivers participating within the EVPN domain. The Include/Exclude bit helps in creating filters for a given multicast route.

If the route is being prepared for IPv6 (MLD), then bit 7 indicates support for MLD version 1. The second least significant bit (bit 6) indicates support for MLD version 2. Since there is no MLD version 3, in case of the IPv6 route, the third least significant bit MUST be 0. In case of the IPv6 route, the fourth least significant bit MUST be ignored if bit 6 is not set.

Reserved bits in the flag MUST be set to 0. They can be defined by other documents in the future.

9.3.1. Constructing the Multicast Leave Synch Route

This section describes the procedures used to construct the IGMP Leave Synch route. Support for these route types is optional. If a PE does not support this route, then it MUST NOT indicate that it supports "IGMP Proxy" in the Multicast Flags Extended Community for the EVIs corresponding to its multihomed Ethernet segments.

An IGMP Leave Synch route MUST carry exactly one ES-Import Route Target extended community, i.e., the one that corresponds to the ES on which the IGMP Leave was received. It MUST also carry exactly one EVI-RT EC, i.e., the one that corresponds to the EVI on which the IGMP Leave was received. See Section 9.5 for details on how to form the EVI-RT EC.

The RD SHOULD be Type 1 [RFC4364]. The value field comprises an IP address of the PE (typically, the loopback address), followed by a number unique to the PE.

The ESI MUST be set to the 10-octet value defined for the ES.

The Ethernet Tag ID MUST be set, as per the procedure defined in [RFC7432].

The Multicast Source Length MUST be set to the length of the Multicast Source Address in bits. 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. In case of a (*,G) Membership Report, the Multicast Source Length is set to 0.

The Multicast Source is the source IP address of the IGMP Membership Report. In case of a (*,G) Membership Report, this field does not exist.

The Multicast Group Length MUST be set to the length of the Multicast Group Address in bits. 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.

The Multicast Group is the group address of the IGMP Membership Report.

The Originator Router Length is the length of the Originator Router Address in bits.

The Originator Router Address is the IP address of the router originating the prefix.

The Reserved field is not part of the route key. The originator MUST set the Reserved field to 0; the receiver SHOULD ignore it, and if it needs to be propagated, it MUST propagate it unchanged.

The Maximum Response Time is the value to be used while sending a query, as defined in [RFC2236].

The Flags field indicates the version of IGMP from which the Membership Report was received. It also indicates whether the multicast group had an Include/Exclude bit set.

9.3.2. Reconstructing IGMP/MLD Leave from a Multicast Leave Synch Route

This section describes the procedures used to reconstruct IGMP/MLD Leave from the Multicast Leave Synch route.

* If the Multicast Group Length is 32, the route is translated to IGMP Leave. If the Multicast Group Length is 128, the route is translated to MLD Leave.

* The Multicast Group Address field is translated to an IGMP/MLD group address.

* If the Multicast Source Length is set to 0, it is translated to any source (*). If the Multicast Source Length is non-zero, the Multicast Source Address field is translated to the IGMP/MLD source address.

* If flag bit 7 is set, it translates the Membership Report to be IGMPv1 or MLDv1.

* If flag bit 6 is set, it translates the Membership Report to be IGMPv2 or MLDv2.

* Flag bit 5 is only valid for the IGMP Membership Report; if it is set, it translates to the IGMPv3 report.

* If the IE flag is set, it translates to the IGMP/MLD Exclude mode Leave. If the IE flag is not set (0), it translates to the Include mode Leave.

9.4. Multicast Flags Extended Community

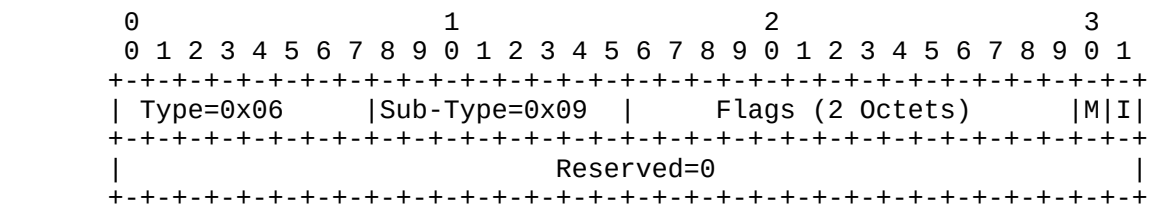
The Multicast Flags Extended Community is a new EVPN Extended Community. EVPN Extended Communities are transitive extended communities with a Type Value of 0x06. IANA has assigned 0x09 to Multicast Flags Extended Community in the "EVPN Extended Community Sub-Types" subregistry.

A PE that supports IGMP and/or the MLD Proxy on a given BD MUST attach this extended community to the IMET route it advertises for

that BD, and it MUST set the IGMP and/or MLD Proxy Support flags to 1. Note that a PE compliant with [RFC7432] will not advertise this extended community, so its absence indicates that the advertising PE does not support either IGMP or MLD Proxies.

The advertisement of this extended community enables a more efficient multicast tunnel setup from the source PE specially for ingress replication, i.e., if an egress PE supports the IGMP Proxy but doesn't have any interest in a given (x,G), it advertises its IGMP Proxy capability using this extended community, but it does not advertise any SMET route for that (x,G). When the source PE (ingress PE) receives such advertisements from the egress PE, it does not replicate the multicast traffic to that egress PE; however, it does replicate the multicast traffic to the egress PEs that don't advertise such capability, even if they don't have any interests in that (x,G).

A Multicast Flags Extended Community is encoded as an 8-octet value as follows:



The low-order (least significant) 2 bits are defined as the "IGMP Proxy Support" and "MLD Proxy Support" bits (see Section 12.3. The absence of this extended community also means that the PE does not support the IGMP Proxy, where:

* The Type is 0x06, as registered with IANA for EVPN Extended Communities.

* The Sub-Type is 0x09.

* Flags are 2-octet values.

  - Bit 15 (shown as I) defines IGMP Proxy Support. The value of 1 for bit 15 means that the PE supports the IGMP Proxy. The value of 0 for bit 15 means that the PE does not support the IGMP Proxy.

  - Bit 14 (shown as M) defines MLD Proxy Support. The value of 1 for bit 14 means that the PE supports the MLD Proxy. The value of 0 for bit 14 means that the PE does not support the MLD Proxy.

  - Bits 0 to 13 are reserved for the future. The sender MUST set it to 0, and the receiver MUST ignore it.

* Reserved bits are set to 0. The sender MUST set it to 0, and the receiver MUST ignore it.

If a router does not support this specification, it MUST NOT add the Multicast Flags Extended Community in the BGP route. When a router receives a BGP update, if both M and I flags are 0, the router MUST treat this update as malformed. The receiver of such an update MUST ignore the extended community.

9.5. EVI-RT Extended Community

In EVPN, every EVI is associated with one or more Route Targets. These RTs serve two functions:

1. Distribution control: RTs control the distribution of the routes. If a route carries the RT associated with a particular EVI, it will be distributed to all the PEs on which that EVI exists.

2. EVI identification: Once a route has been received by a particular PE, the RT is used to identify the EVI to which it applies.

An IGMP Membership Report Synch or IGMP Leave Synch route is associated with a particular combination of ES and EVI. These routes need to be distributed only to PEs that are attached to the associated ES. Therefore, these routes carry the ES-Import RT for that ES.

Since an IGMP Membership Report Synch or IGMP Leave Synch route does not need to be distributed to all the PEs on which the associated EVI exists, these routes cannot carry the RT associated with that EVI. Therefore, when such a route arrives at a particular PE, the route's RTs cannot be used to identify the EVI to which the route applies. Some other means of associating the route with an EVI must be used.

This document specifies four new ECs that can be used to identify the EVI with which a route is associated but do not have any effect on the distribution of the route. These new ECs are known as "Type 0 EVI-RT EC", "Type 1 EVI-RT EC", "Type 2 EVI-RT EC", and "Type 3 EVI-RT EC".

1. A Type 0 EVI-RT EC is an EVPN EC (type 6) of sub-type 0xA.

2. A Type 1 EVI-RT EC is an EVPN EC (type 6) of sub-type 0xB.

3. A Type 2 EVI-RT EC is an EVPN EC (type 6) of sub-type 0xC.

4. A Type 3 EVI-RT EC is an EVPN EC (type 6) of sub-type 0xD

Each IGMP Membership Report Synch or IGMP Leave Synch route MUST carry exactly one EVI-RT EC. The EVI-RT EC carried by a particular route is constructed as follows. Each such route is the result of having received an IGMP Membership Report or an IGMP Leave message from a particular BD. The route is said to be associated with that BD. For each BD, there is a corresponding RT that is used to ensure that routes "about" that BD are distributed to all PEs attached to that BD. So suppose a given IGMP Membership Report Synch or Leave Synch route is associated with a given BD, say BD1, and suppose that the corresponding RT for BD1 is RT1. Then:

* If RT1 is a Transitive Two-Octet AS-specific EC, then the EVI-RT EC carried by the route is a Type 0 EVI-RT EC. The value field of the Type 0 EVI-RT EC is identical to the value field of RT1.

* If RT1 is a Transitive IPv4-Address-specific EC, then the EVI-RT EC carried by the route is a Type 1 EVI-RT EC. The value field of the Type 1 EVI-RT EC is identical to the value field of RT1.

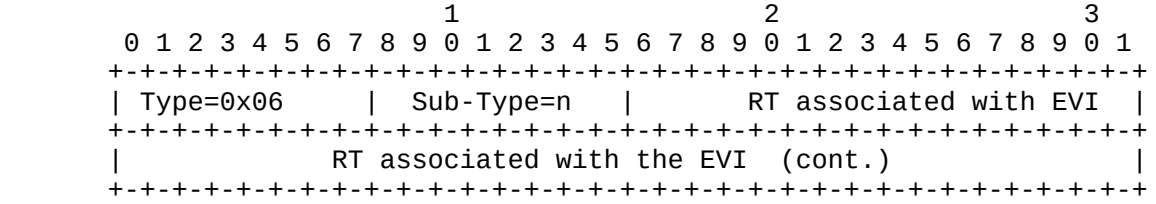
* If RT1 is a Transitive Four-Octet AS-specific EC, then the EVI-RT EC carried by the route is a Type 2 EVI-RT EC. The value field of the Type 2 EVI-RT EC is identical to the value field of RT1.

* If RT1 is a Transitive IPv6-Address-specific EC, then the EVI-RT EC carried by the route is a Type 3 EVI-RT EC. The value field of the Type 3 EVI-RT EC is identical to the value field of RT1.

An IGMP Membership Report Synch or Leave Synch route MUST carry exactly one EVI-RT EC.

Suppose a PE receives a particular IGMP Membership Report Synch or IGMP Leave Synch route, say R1, and suppose that R1 carries an ES-Import RT that is one of the PE's Import RTs. If R1 has no EVI-RT EC or has more than one EVI-RT EC, the PE MUST apply the "treat-as-withdraw" procedure per [RFC7606].

Note that an EVI-RT EC is not a Route Target extended community, is not visible to the RT Constrain mechanism [RFC4684], and is not intended to influence the propagation of routes by BGP.



The value of "n" is 0x0A, 0x0B, 0x0C, or 0x0D, corresponding to EVI-RT types 0, 1, 2, or 3, respectively.

9.6. Rewriting of RT ECs and EVI-RT ECs by ASBRs

There are certain situations in which an ES is attached to a set of PEs that are not all in the same AS, or not all operated by the same provider. In this situation, the RT that corresponds to a particular EVI may be different in each AS. If a route is propagated from AS1 to AS2, an ASBR at the AS1/AS2 border may be configured with a policy that replaces the EVI RTs for AS1 with the corresponding EVI RTs for AS2. This is known as RT-rewriting.

If an ASBR is configured to perform RT-rewriting of the EVI RTs in EVPN routes, it MUST be configured to perform RT-rewriting of the corresponding EVI-RT extended communities in IGMP Join Synch and IGMP Leave Synch Routes.

9.7. BGP Error Handling

If a received BGP update contains Flags not in accordance with the IGMP/MLD version-X expectation, the PE MUST apply the "treat-as-withdraw" procedure per [RFC7606].

If a received BGP update is malformed such that BGP route keys cannot be extracted, then the BGP update MUST be considered invalid. The receiving PE MUST apply the "session reset" procedure per [RFC7606].

10. IGMP Version 1 Membership Report

This document does not provide any detail about IGMPv1 processing. Implementations are expected to only use IGMPv2 and above for IPv4 and MLDv1 and above for IPv6. IGMPv1 routes are considered invalid, and the PE MUST apply the "treat-as-withdraw" procedure per [RFC7606].

11. Security Considerations

This document describes a means to efficiently operate IGMP and MLD on a subnet constructed across multiple PODs or DCs via an EVPN solution. The security considerations for the operation of the underlying EVPN and BGP substrates are described in [RFC7432], and specific multicast considerations are outlined in [RFC6513] and [RFC6514]. The EVPN and associated IGMP Proxy provides a single broadcast domain so the same security considerations of IGMPv2 [RFC2236], IGMPv3 [RFC3376], MLD [RFC2710], or MLDv2 [RFC3810] apply.

12. IANA Considerations

12.1. EVPN Extended Community Sub-Types Registration

IANA has allocated the following codepoints in the "EVPN Extended Community Sub-Types" subregistry under the "Border Gateway Protocol (BGP) Extended Communities" registry.

Sub-Type Value	Name	Reference
0x09	Multicast Flags Extended Community	RFC 9251
0x0A	EVI-RT Type 0	RFC 9251
0x0B	EVI-RT Type 1	RFC 9251
0x0C	EVI-RT Type 2	RFC 9251
0x0D	EVI-RT Type 3	RFC 9251

Table 1: EVPN Extended Community Sub-Types Subregistry Allocated Codepoints

12.2. EVPN Route Types Registration

IANA has allocated the following EVPN route types in the "EVPN Route Types" subregistry.

6 - Selective Multicast Ethernet Tag Route

7 - Multicast Membership Report Synch Route

8 - Multicast Leave Synch Route

12.3. Multicast Flags Extended Community Registry

IANA has created and now maintains a new subregistry called "Multicast Flags Extended Community" under the "Border Gateway Protocol (BGP) Extended Communities" registry. The registration procedure is First Come First Served [RFC8126]. For the 16-bit Flags field, the bits are numbered 0-15, from high order to low order. The registry was initialized as follows:

Bit	Name	Reference	Change Controller
0-13	Unassigned		
14	MLD Proxy Support	RFC 9251	IETF
15	IGMP Proxy Support	RFC 9251	IETF

Table 2: Multicast Flags Extended Community

13. References

13.1. Normative References

[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>.

[RFC2236] Fenner, W., "Internet Group Management Protocol, Version

2", RFC 2236, DOI 10.17487/RFC2236, November 1997, <https://www.rfc-editor.org/info/rfc2236>.

[RFC2710] Deering, S., Fenner, W., and B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", RFC 2710, DOI 10.17487/RFC2710, October 1999, <https://www.rfc-editor.org/info/rfc2710>.

[RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, DOI 10.17487/RFC3376, October 2002, <https://www.rfc-editor.org/info/rfc3376>.

[RFC3810] Vida, R., Ed. and L. Costa, Ed., "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, DOI 10.17487/RFC3810, June 2004, <https://www.rfc-editor.org/info/rfc3810>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, DOI 10.17487/RFC4684, November 2006, <https://www.rfc-editor.org/info/rfc4684>.

[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>.

[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>.

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <https://www.rfc-editor.org/info/rfc7606>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

13.2. Informative References

[EVPN-BUM] Zhang, Z., 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>.

[RFC4541] Christensen, M., Kimball, K., and F. Solensky, "Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches", RFC 4541, DOI 10.17487/RFC4541, May 2006, <https://www.rfc-editor.org/info/rfc4541>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

Acknowledgements

The authors would like to thank Stephane Litkowski, Jorge Rabadan, Anoop Ghanwani, Jeffrey Haas, Krishna Muddenahally Ananthamurthy, and Swadesh Agrawal for their reviews and valuable comments.

Contributors

Derek Yeung
Arrcus
Email: derek@arrcus.com

Authors' Addresses

Ali Sajassi
Cisco Systems
821 Alder Drive
Milpitas, CA 95035
United States of America
Email: sajassi@cisco.com

Samir Thoria
Cisco Systems
821 Alder Drive
Milpitas, CA 95035
United States of America
Email: sthoria@cisco.com

Mankamana Mishra
Cisco Systems
821 Alder Drive
Milpitas, CA 95035
United States of America
Email: mankamis@cisco.com

Keyur Patel
Arrcus
United States of America
Email: keyur@arrcus.com

John Drake
Juniper Networks
Email: jdrake@juniper.net

Wen Lin
Juniper Networks
Email: wlin@juniper.net