IGMP and MLD Proxy for EVPN
draft-ietf-bess-evpn-igmp-mld-proxy-06

Abstract

   Ethernet Virtual Private Network (EVPN) solution is becoming
   pervasive in data center (DC) applications for Network Virtualization
   Overlay (NVO) and DC interconnect (DCI) services, and in service
   provider (SP) applications for next generation virtual private LAN
   services.

   This draft describes how to support efficiently endpoints running
   IGMP for the above services over an EVPN network by incorporating
   IGMP proxy procedures on EVPN PEs.

Status of This Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF). Note that other groups may also distribute
   working documents as Internet-Drafts. The list of current Internet-
   Drafts is at https://datatracker.ietf.org/drafts/current/.

   Internet-Drafts are draft documents valid for a maximum of six months
   and may be updated, replaced, or obsoleted by other documents at any
   time. It is inappropriate to use Internet-Drafts as reference
   material or to cite them other than as "work in progress."

   This Internet-Draft will expire on July 26, 2021.

Copyright Notice

Table of Contents

1. Introduction

   Ethernet Virtual Private Network (EVPN) solution [RFC7432] is
   becoming pervasive in data center (DC) applications for Network
   Virtualization Overlay (NVO) and DC interconnect (DCI) services, and
   in service provider (SP) applications for next generation virtual
   private LAN services.

   In 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 standard
   way of inter-POD communication for both intra-DC and inter-DC. A
   subnet can span across multiple PODs and DCs. EVPN provides robust
   multi-tenant solution with extensive multi-homing capabilities to
   stretch a subnet (VLAN) across multiple PODs and DCs. There can be
   many hosts/VMs (several hundreds) attached to a subnet that is
   stretched across several PODs and DCs.

   These hosts/VMs express their interests in multicast groups on a
   given subnet/VLAN by sending IGMP membership reports (Joins) for
   their interested multicast group(s). Furthermore, an IGMP 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 draft accomplishes has three objectives:

   1. Reduce flooding of IGMP messages: just like the ARP/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 messages (both Queries and 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: to forward multicast traffic over EVPN
    network such that it only gets forwarded to the PEs that have
    interest in the multicast group(s), multicast traffic will not be
    forwarded to the PEs that have no receivers attached to them for
    that multicast group. This draft 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 P2MP tunnels can be found in [bum-
    procedure-updates]

The first two objectives are achieved by using IGMP/MLD proxy on the
PE and the third objective is achieved by setting up a multicast
tunnel (e.g., ingress replication) only among the PEs that have
interest in that 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

o POD: Point of Delivery

o ToR: Top of Rack

o NV: Network Virtualization

o NVO: Network Virtualization Overlay

o EVPN: Ethernet Virtual Private Network

o IGMP: Internet Group Management Protocol

o MLD: Multicast Listener Discovery

o EVI: An EVPN instance spanning the Provider Edge (PE) devices
   participating in that EVPN

o MAC-VRF: A Virtual Routing and Forwarding table for Media Access
   Control (MAC) addresses on a PE

o IR: Ingress Replication

   o Ethernet Segment (ES): When a customer site (device or network) is
      connected to one or more PEs via a set of Ethernet links, then
      that set of links is referred to as an 'Ethernet Segment'.

   o Ethernet Segment Identifier (ESI): A unique non-zero identifier
      that identifies an Ethernet Segment is called an 'Ethernet Segment
      Identifier'.

   o PE: Provider Edge.

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

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

   o 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.

   o 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.

   This document also assumes familiarity with the terminology of
   [RFC7432]. Though most of the place this document uses term IGMP
   membership request (Joins), the text applies equally for MLD
   membership request too. Similarly, text for IGMPv2 applies to MLDv1
   and text for IGMPv3 applies to MLDv2. IGMP / MLD version encoding in
   BGP update is stated in Section 9

4. **IGMP/MLD Proxy**

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

   1. Proxy for IGMP Reports.

   2. Proxy for IGMP Queries.

4.1. Proxy Reporting

   When IGMP protocol is used between hosts/VMs 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 Reports in the first hop PE, and
   translating and exchanging the relevant information among EVPN BGP
   speakers. The information is again translated back to 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 NLRI, the EVPN Selective
   Multicast Ethernet Tag route, along with its procedures to help
   exchange and register IGMP multicast groups Section 9.

4.1.1. IGMP/MLD Membership Report Advertisement in BGP

   When a PE wants to advertise an IGMP membership report (Join) using
   the BGP EVPN route, it follows the following rules (BGP encoding
   stated in Section 9):

   1. When the first hop PE receives several IGMP membership reports
       (Joins), belonging to the same IGMP version, from different
       attached hosts/VMs for the same (*,G) or (S,G), it only SHOULD
       send a single BGP message corresponding to the very first IGMP
       Join (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 Join is
       for (*,G), then 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 needs to be set to indicate
       that no source IP address to be excluded (include all
       sources"*"). If the IGMP Join is for (S,G), then besides setting
       multicast group address along with the version flag v3, the
       source IP address and the include/exclude 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 zero).

   2. When the first hop PE receives an IGMPv3 Join for (S,G) on a
       given BD, it SHOULD advertise the corresponding EVPN Selective
       Multicast Ethernet Tag (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 Join first for
       (*,G) and then later it receives an IGMP version-Y Join for the
       same (*,G), then it MUST re-advertise the same EVPN SMET route
       with 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
flag field is not part of BGP route key processing.

4. When the first hop PE receives an IGMP version-X Join first for
(*,G) and then later it receives an IGMPv3 Join for the same
multicast group address but for a specific source address S, then
the PE MUST advertise a new EVPN SMET route with v3 flag set (and
v2 reset). The include/exclude flag also need to be set
accordingly. Since source IP address is used as part of BGP
route key processing it is considered as a new BGP route
advertisement.

5. 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 source IP address
in the receive EVPN route. If there is, then an error SHOULD be
logged . If the v3 flag is set (in addition to v2), then the
include/exclude flag MUST indicate "exclude". If not, then an
error SHOULD be logged. The PE MUST generate an IGMP membership
report (Join) for that (*,G) and each IGMP version in the version
flag.

6. 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.

7. Upon receiving EVPN SMET route(s) and before generating the
corresponding IGMP Join(s), the PE checks to see whether it has
any CE multicast router for that BD on any of its ES's . 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 Join(s) are sent to those ACs. If it doesn't have
any of the router's AC, then no IGMP Join(s) needs to be
generated. This is because sending IGMP Joins 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 join 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 Join from host it will not forward
multicast data to it. In other words, an IGMPv2 Join 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 (Leave) or IGMPv3 "Leave" equivalent message, it
follows the following rules:

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-advertises EVPN SMET
    Multicast 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 withdraws the EVPN route for that (*,G).

2. 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) attached to the multicast router for that
    multicast group. It should be noted that the received EVPN route
    SHOULD at least have 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. Error MUST be
    considered as BGP error and the PE MUST apply the "treat-as-
    withdraw" procedure of [RFC7606].

3. 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 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 L2VPN service to
    act as distributed multicast router with Anycast IP address for
    all attached hosts/VMs in that subnet.

2. To enable suppression of IGMP membership reports and queries over
    MPLS/IP core.

5. Operation

Consider the EVPN network of Figure-1, where there is an EVPN
instance configured across the PEs shown in this figure (namely PE1,
PE2, and PE3). Let's consider that this EVPN instance consists of a
single bridge domain (single subnet) with all the hosts, sources, and
the multicast router connected to this subnet. PE1 only has hosts
connected to it. PE2 has a mix of hosts and a multicast source. PE3
has a mix of hosts, a multicast source, and a multicast router.
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:

o only hosts/VMs

o mix of hosts/VMs and multicast source

o mix of hosts/VMs, multicast source, and multicast router

```
                        +-------------+
                        |             |
                        |             |
               +----+   |             |   +----+
    H1:(*,G1)v2 ---|    |  |             |   |    |---- H6(*,G1)v2
    H2:(*,G1)v2 ---| PE1|  |   IP/MPLS   | | PE2|---- H7(S2,G2)v3
    H3:(*,G1)v3 ---|    |  |   Network   | |    |---- S2
    H4:(S2,G2)v3 --|    |  |             | |    |
               +----+   |  |             |   +----+
                        |             |
               +----+   |             |
    H5:(S1,G1)v3 --|    |  |             |
            S1 ---| PE3|  |             |
            R1 ---|    |  |             |
               +----+   |             |
                        |             |
                        +-------------+
```
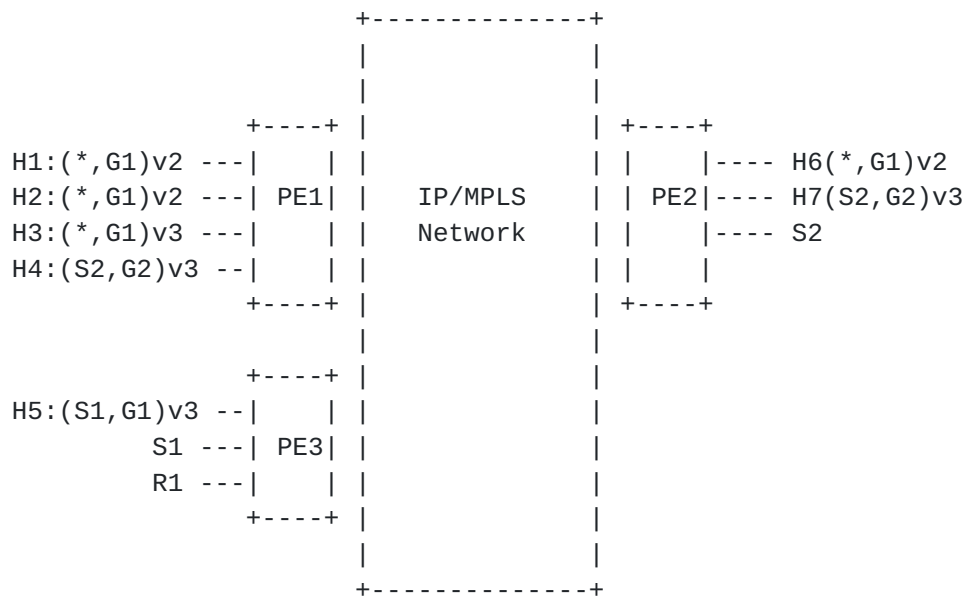

        Figure 1: EVPN network

5.1. PE with only attached hosts/VMs for a given subnet

   When PE1 receives an IGMPv2 Join Report from H1, it does not forward
   this join to any of its other ports (for this subnet) because all
   these local ports are associated with the hosts/VMs. PE1 sends an
   EVPN Multicast Group route corresponding to this join for (*,G1) and
   setting v2 flag. This EVPN route is received by PE2 and PE3 that are
   the members of the same BD (i.e., same EVI in case of VLAN-based
   service or EVI,VLAN in case of VLAN-aware bundle service). PE3
   reconstructs the IGMPv2 Join 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
   Join 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 Join 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 route because there is no change in
   information. However, when it receives the IGMPv3 Join 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 IMGMPv3 Join from H4 for (S2,G2), it
   advertises a new EVPN SMET route corresponding to it.

5.2. PE with a mix of attached hosts/VMs and multicast source

The main difference in this case is that when PE2 receives the IGMPv3
Join from H7 for (S2,G2), it does advertise it in BGP to support
source move 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 Join described in previous section.

5.3. PE with a mix of attached hosts/VMs, a multicast source and a router

The main difference in this case relative to the previous two
sections is that IGMP v2/v3 Join messages received locally needs to
be sent to the port associated with router R1. Furthermore, the
Joins received via BGP (SMET) need to be passed to the R1 port but
filtered for all other ports.

6. All-Active Multi-Homing

Because the 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 multi-homed PEs,
either DF or non-DF; i.e., different IGMP Join messages can arrive at
different PEs in the redundancy group and furthermore their
corresponding Leave messages can arrive at PEs that are different
from the ones that received the Join messages. Therefore, all PEs
attached to a given ES must coordinate IGMP Join and Leave Group
(x,G) state, where x may be either '*' or a particular source S, for
each BD on that ES. This allows the DF for that [ES,BD] to correctly
advertise or withdraw a Selective Multicast Ethernet Tag (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 Join Synchronization

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

The IGMP Join Synch route MUST carry the ES-Import 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 Join Synch route it
installs that route and if it doesn't already have IGMP Join (x,G)
state for that [ES,BD], it MUST instantiate that IGMP Join (x,G)
state - i.e., IGMP Join (x,G) state is the union of the local IGMP
Join (x,G) state and the installed IGMP Join 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 Join (x, G)
state for that [ES,BD], it MUST withdraw its BGP IGMP Join Synch
route for that [ES,BD].

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

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

6.2. Local IGMP/MLD Leave Group Synchronization

When a PE, either DF or non-DF, receives, on a given multihomed ES
operating in All-Active redundancy mode, an IGMP Leave Group message
for (x,G) from the attached CE, it determines the BD to which the
IGMPv2 Leave Group belongs. Regardless of whether it has IGMP Join
(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
    (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 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 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 Join Synch route for that [ES,BD]. If it
doesn't already have local IGMP Join (x, G) state for that [ES, BD],
it instantiates local IGMP Join (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 Join Synch
route for (x,G) before the Maximum Response Time timer expires, it
installs that route and if it doesn't already have IGMP Join (x,G)
state for that BD on that ES, it instantiates that IGMP Join (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 Timer expires a PE that has advertised an
IGMP Leave Synch route, withdraws it. Any PE attached to the
multihomed ES, that started the Maximum Response Time and has no
local IGMP Join (x,G) state and no installed IGMP Join Synch routes,
it removes IGMP Join (x,G) state for that [ES,BD]. If the DF no
longer has IGMP Join (x,G) state for that BD on any ES for which it
is DF, it withdraws its SMET route for that (x,G) group in that BD.

6.3. Mass Withdraw of Multicast join Sync route in case of failure

A PE which has received an IGMP Join, would have synced the IGMP Join
by the procedure defined in section 6.1. If a PE with local join
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 Joins) SHOULD not remove the state
immediately; instead General Query SHOULD be generated to refresh the
states. There are several ways to Some of the way to detect failure
at a peer, e.g. using IGP next hop tracking or ES route withdraw.

7. Single-Active Multi-Homing

Note that to facilitate state synchronization after failover, the PEs
attached to a mutihomed ES operating in Single-Active redundancy mode
SHOULD also coordinate IGMP Join (x,G) state. In this case all IGMP
Join 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
    Proxy. This set consists of any PE that has advertised an
    Inclusive Multicast Tag route for the BD without the "IGMP Proxy
    Support" flag.

2. It sends (x,G) traffic to the set of PEs supporting IGMP Proxy
    and having listeners for that (x,G) group in that BD. This set
    consists of any PE that has advertised an Inclusive Multicast Tag
    route for the BD with the "IGMP Proxy Support" flag and that has
    advertised a SMET route for that (x,G) group in that BD.

If an ingress PE's Selective P-Tunnel for a given BD uses P2MP and
all of the PEs in the BD support that tunnel type and IGMP proxy,
then for a given (x,G) group in a given BD it sends (x,G) traffic
using the Selective P-Tunnel for that (x,G) group in that BD. This
tunnel includes those PEs that have advertised a SMET route for that

(x,G) group on that BD (for Selective P-tunnel) but it may include
other PEs as well (for Aggregate Selective P-tunnel).

9. BGP Encoding

This document defines three new BGP EVPN routes to carry IGMP
membership reports. The route type is known as:

+ 6 - Selective Multicast Ethernet Tag Route

+ 7 - Multicast Join Synch Route

+ 8 - Multicast Leave Synch Route

The detailed encoding and procedures for this route type are
described in subsequent sections.

9.1. Selective Multicast Ethernet Tag Route

A Selective Multicast Ethernet Tag route type specific EVPN NLRI
consists of the following:

```
+---------------------------------------+
|  RD (8 octets)                        |
+---------------------------------------+
|  Ethernet Tag ID (4 octets)           |
+---------------------------------------+
|  Multicast Source Length (1 octet)    |
+---------------------------------------+
|  Multicast Source Address (variable)  |
+---------------------------------------+
|  Multicast Group Length (1 octet)     |
+---------------------------------------+
|  Multicast Group Address (Variable)   |
+---------------------------------------+
|  Originator Router Length (1 octet)   |
+---------------------------------------+
|  Originator Router Address (variable) |
+---------------------------------------+
|  Flags (1 octet)                      |
+---------------------------------------+
```

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 one-
octet flag field. The Flags fields are defined as follows:

```
            0  1  2  3  4  5  6  7
           +--+--+--+--+--+--+--+--+
           | reserved  |IE|v3|v2|v1|
           +--+--+--+--+--+--+--+--+
```

o The least significant bit, bit 7 indicates support for IGMP
   version 1. Since IGMP V1 is being deprecated , sender MUST set it
   as 0 for IGMP and receiver MUST ignore it.

o The second least significant bit, bit 6 indicates support for IGMP
   version 2.

o The third least significant bit, bit 5 indicates support for IGMP
   version 3.

o 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.

o This EVPN route type is used to carry tenant IGMP multicast group
   information. The flag field assists in distributing IGMP
   membership interest of a given host/VM for a given multicast
   route. The version bits help associate IGMP version of receivers
   participating within the EVPN domain.

o The include/exclude bit helps in creating filters for a given
   multicast route.

o If 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 route 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.

o Reserve bit SHOULD be set to 0 by sender. And receiver SHOULD
   ignore the reserve bit.

9.1.1. Constructing the Selective Multicast Ethernet Tag route

   This section describes the procedures used to construct the Selective
   Multicast Ethernet Tag (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 follows:

   o EVI is VLAN-Based or VLAN Bundle service - set to 0

   o EVI is VLAN-Aware Bundle service without translation - set to the
      customer VID for that BD

   o EVI is VLAN-Aware Bundle service with translation - set to the
      normalized Ethernet Tag ID - e.g., normalized VID

   The Multicast Source Length MUST be set to 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) Join, 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), this field is not used.

   The Multicast Group Length MUST be set to length of 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 router originating
   this route. The SMET Originator Router IP address MUST match that of
   the IMET (or SPMSI AD) route originated for the same EVI by the same
   downstream PE.

The Flags field indicates the version of IGMP protocol from which the membership report was received. It also indicates whether the multicast group had the INCLUDE or EXCLUDE bit set.

Reserve bit MUST be set to 0. They can be defined in future by other document.

IGMP is used to receive group membership information from hosts/VMs by TORs. Upon receiving the hosts/VMs expression of interest of a particular group membership, this information is then forwarded using SMET route. The NLRI also keeps track of receiver's IGMP protocol version and any source filtering for a given group membership. All EVPN SMET routes are announced with per- EVI Route Target extended communities.

9.1.2. Default Selective Multicast Route

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

```
                              +--------------+
                              |              |
                              |              |
                              |              | +----+
                              |              | |    |---- H1(*,G1)v2
                              |   IP/MPLS    | | PE1|---- H2(S2,G2)v3
                              |   Network    | |    |---- S2
                              |              | |    |
                              |              | +----+
                              |              |
                    +----+    |              |
        +----+      |    | |  |              |
        |    |  S1 ---| PE2| |              |
        |PIM |----R1 ---|    | |              |
        |ASM |        +----+ |              |
        |    |               |              |
        +----+               +--------------+


         Figure 2: Multicast Router behind EVPN domain
```

Consider the EVPN network of Figure-2, where there is an EVPN instance configured across the PEs. Lets consider 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 RP (Rendezvous Point). If receivers
behind 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.

9.2. Multicast Join Synch Route

This EVPN route type is used to coordinate IGMP Join (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
following:

```
+----------------------------------------------------+
|  RD (8 octets)                                     |
+----------------------------------------------------+
| Ethernet Segment Identifier (10 octets)            |
+----------------------------------------------------+
|  Ethernet Tag ID  (4 octets)                       |
+----------------------------------------------------+
|  Multicast Source Length (1 octet)                 |
+----------------------------------------------------+
|  Multicast Source Address (variable)               |
+----------------------------------------------------+
|  Multicast Group Length (1 octet)                  |
+----------------------------------------------------+
|  Multicast Group Address (Variable)                |
+----------------------------------------------------+
|  Originator Router Length (1 octet)                |
+----------------------------------------------------+
|  Originator Router Address (variable)              |
+----------------------------------------------------+
|  Flags (1 octet)                                   |
+----------------------------------------------------+
```

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 one-
octet Flags field, whose fields are defined as follows:

```
                      0  1  2  3  4  5  6  7
                     +--+--+--+--+--+--+--+--+
                     | reserved  |IE|v3|v2|v1|
                     +--+--+--+--+--+--+--+--+
```

o The least significant bit, bit 7 indicates support for IGMP
   version 1.

o The second least significant bit, bit 6 indicates support for IGMP
   version 2.

o The third least significant bit, bit 5 indicates support for IGMP
   version 3.

o The fourth least significant bit, bit 4 indicates whether the (S,
   G) information carried within the route-type is of 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.

o Reserve bit MUST be set to 0. They can be defined in future by
   other document.

The Flags field assists in distributing IGMP membership interest of a
given host/VM for a given multicast route. The version bits help
associate IGMP version of receivers participating within the EVPN
domain. The include/exclude bit helps in creating filters for a
given multicast route.

If 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 IPv6 route third least significant bit MUST be 0. In
case of IPv6 route, the fourth least significant bit MUST be ignored
if bit 6 is not set.

9.2.1. Constructing the Multicast Join Synch Route

This section describes the procedures used to construct the IGMP Join
Synch route. Support for this route type is optional. If a PE does
not support this route, then it MUST NOT indicate that it supports
'IGMP proxy' in the Multicast Flag extended community for the EVIs
corresponding to its multi-homed Ethernet Segments (ESs).

An IGMP Join Synch route MUST carry exactly one ES-Import Route
Target extended community, the one that corresponds to the ES on
which the IGMP Join was received. It MUST also carry exactly one

EVI-RT EC, the one that corresponds to the EVI on which the IGMP Join was received. See Section 9.5 for details on how to encode and construct the EVI-RT EC.

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

The Ethernet Tag ID MUST be set as follows:

o EVI is VLAN-Based or VLAN Bundle service - set to 0

o EVI is VLAN-Aware Bundle service without translation - set to the customer VID for the BD

o EVI is VLAN-Aware Bundle service with translation - set to the normalized Ethernet Tag ID - e.g., normalized VID

The Multicast Source length MUST be set to length of 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) Join, 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) Join, this field does not exist.

The Multicast Group length MUST be set to length of 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 Router Originating the prefix.

The Flags field indicates the version of IGMP protocol from which the
membership report was received. It also indicates whether the
multicast group had INCLUDE or EXCLUDE bit set.

Reserve bit MUST be set to 0. They can be defined in future by other
document.

9.3. Multicast Leave Synch Route

This EVPN route type is used to coordinate IGMP Leave Group (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
following:

```
        +----------------------------------------------------+
        |  RD (8 octets)                                     |
        +----------------------------------------------------+
        | Ethernet Segment Identifier (10 octets)           |
        +----------------------------------------------------+
        |  Ethernet Tag ID  (4 octets)                      |
        +----------------------------------------------------+
        |  Multicast Source Length (1 octet)                |
        +----------------------------------------------------+
        |  Multicast Source Address (variable)              |
        +----------------------------------------------------+
        |  Multicast Group Length (1 octet)                 |
        +----------------------------------------------------+
        |  Multicast Group Address (Variable)               |
        +----------------------------------------------------+
        |  Originator Router Length (1 octet)               |
        +----------------------------------------------------+
        |  Originator Router Address (variable)             |
        +----------------------------------------------------+
        |  Reserved (4 octet)                               |
        +----------------------------------------------------+
        |  Maximum Response Time (1 octet)                  |
        +----------------------------------------------------+
        |  Flags (1 octet)                                  |
        +----------------------------------------------------+
```

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 the one-octet Flags field, whose
fields are defined as follows:

```
                    0  1  2  3  4  5  6  7
                  +--+--+--+--+--+--+--+--+
                  | reserved   |IE|v3|v2|v1|
                  +--+--+--+--+--+--+--+--+
```


   o The least significant bit, bit 7 indicates support for IGMP
      version 1.

   o The second least significant bit, bit 6 indicates support for IGMP
      version 2.

   o The third least significant bit, bit 5 indicates support for IGMP
      version 3.

   o The fourth least significant bit, bit 4 indicates whether the (S,
      G) information carried within the route-type is of 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.

   o Reserve bit MUST be set to 0. They can be defined in future by
      other document.

   The Flags field assists in distributing IGMP membership interest of a
   given host/VM for a given multicast route. The version bits help
   associate IGMP version of receivers participating within the EVPN
   domain. The include/exclude bit helps in creating filters for a
   given multicast route.

   If 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 IPv6 route third least significant bit MUST be 0. In
   case of IPv6 route, the fourth least significant bit MUST be ignored
   if bit 6 is not set.

   Reserve bit in flag MUST be set to 0. They can be defined in future
   by other document.

9.3.1. Constructing the Multicast Leave Synch Route

   This section describes the procedures used to construct the IGMP
   Leave Synch route. Support for this route type is optional. If a PE
   does not support this route, then it MUST not indicate that it
   supports 'IGMP proxy' in Multicast Flag extended community for the
   EVIs corresponding to its multi-homed Ethernet Segments.

An IGMP Leave Synch route MUST carry exactly one ES-Import Route
Target extended community, the one that corresponds to the ES on
which the IGMP Leave was received. It MUST also carry exactly one
EVI-RT EC, 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 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 Segment Identifier (ESI) MUST be set to the 10-octet
value defined for the ES.

The Ethernet Tag ID MUST be set as follows:

o EVI is VLAN-Based or VLAN Bundle service - set to 0

o EVI is VLAN-Aware Bundle service without translation - set to the
   customer VID for the BD

o EVI is VLAN-Aware Bundle service with translation - set to the
   normalized Ethernet Tag ID - e.g., normalized VID

The Multicast Source length MUST be set to length of 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) Join, 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) Join, this field does not exist.

The Multicast Group length MUST be set to length of 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 Router Originating
the prefix.

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

Maximum Response Time is value to be used while sending query as
defined in [RFC2236]

The Flags field indicates the version of IGMP protocol from which the
membership report was received. It also indicates whether the
multicast group had INCLUDE or EXCLUDE bit set.

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 field value of 6. IANA will assign a Sub-
Type from the 'EVPN Extended Community Sub-Types' registry.

A PE that supports IGMP proxy on a given BD MUST attach this extended
community to the Inclusive Multicast Ethernet Tag (IMET) route it
advertises for that BD and it MUST set the IGMP Proxy Support flag to
1. Note that an [RFC7432] compliant PE will not advertise this
extended community so its absence indicates that the advertising PE
does not support IGMP Proxy.

The advertisement of this extended community enables more efficient
multicast tunnel setup from the source PE specially for ingress
replication - i.e., if an egress PE supports 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:

```
                          1                   2                   3
         0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         | Type=0x06   | Sub-Type=0x09|      Flags (2 Octets)    |M|I|
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         |                        Reserved=0                           |
         +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   The low-order (lease significant) two bits are defined as the "IGMP
   Proxy Support and MLD Proxy Support" bit. The absence of this
   extended community also means that the PE does not support IGMP
   proxy. where:

   o Type is 0x06 as registered with IANA for EVPN Extended
      Communities.

   o Sub-Type : 0x09

   o Flags are two Octets value.

      * Bit 15 (shown as I) defines IGMP Proxy Support. Value of 1 for
         bit 15 means that PE supports IGMP Proxy. Value of 0 for bit
         15 means that PE does not supports IGMP Proxy.

      * Bit 14 (shown as M) defines MLD Proxy Support. Value of 1 for
         bit 14 means that PE supports MLD Proxy. Value of 0 for bit 14
         means that PE does not support MLD proxy.

      * Bit 0 to 13 are reserved for future. Sender MUST set it 0 and
         receiver MUST ignore it.

   o Reserved bits are set to 0. Sender MUST set it to 0 and receiver
      MUST ignore it.

   If a router does not support this specification, it MUST not add
   Multicast Flags Extended Community in BGP route. A router receiving
   BGP update , if M and I both flag are zero (0), the router MUST treat
   this Update as malformed . Receiver of such update MUST ignore the
   extended community.

9.5. EVI-RT Extended Community

   In EVPN, every EVI is associated with one or more Route Targets
   (RTs). These Route Targets 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 Join 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 Join 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 Extended Communities (EC) that can
be used to identify the EVI with which a route is associated, but
which do not have any effect on the distribution of the route. These
new ECs are known as the "Type 0 EVI-RT EC", the "Type 1 EVI-RT EC",
the "Type 2 EVI-RT EC", and the "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 Join 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 Join or an IGMP Leave message from a particular BD.
The route is said to be associated 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 Join Synch or Leave Synch route is associated
with a given BD, say BD1, and suppose that the corresponding RT for
BD1 is RT1. Then:

o 0. 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.

o 1. 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.

o 2. If RT1 is a Transitive Four-Octet-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.

o 3. 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 Join Synch or Leave Synch route MUST carry exactly one EVI-RT
EC.

Suppose a PE receives a particular IGMP Join 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 of [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.

```
                          1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       | Type=0x06   |  Sub-Type=n  |        RT associated with EVI   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |              RT associated with the EVI  (cont.)            |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where the value of 'n' is 0x0A, 0x0B, 0x0C, or 0x0D corresponding to
EVI-RT type 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 some such situations, 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 provisioned
with a policy that removes the RTs that are meaningful in AS1 and
replaces them with the corresponding (i.e., RTs corresponding to the

same EVIs) RTs that are meaningful in AS2. This is known as RT-
rewriting.

Note that if a given route's RTs are rewritten, and the route carries
an EVI-RT EC, the EVI-RT EC needs to be rewritten as well.

9.7. BGP Error Handling

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

If a received BGP update is malformed such that BGP route keys cannot
be extracted, then BGP update MUST be considered as invalid.
Receiving PE MUST apply the "Session reset" procedure of [RFC7606].

10. IGMP/MLD Immediate Leave

IGMP MAY be configured with immediate leave option. This allows the
device to remove the group entry from the multicast routing table
immediately upon receiving a IGMP leave message for (x,G). In case
of all active multi-homing while synchronizing the IGMP Leave state
to redundancy peers, Maximum Response Time MAY be filled in as Zero.
Implementations SHOULD have identical configuration across multi-
homed peers. In case IGMP Leave Synch route is received with Maximum
Response Time Zero, irrespective of local IGMP configuration it MAY
be processed as an immediate leave.

11. IGMP Version 1 Membership Request

This document does not provide any detail about IGMPv1 processing.
Multicast working group are in process of deprecating uses of IGMPv1.
Implementations MUST only use IGMPv2 and above for IPv4 and MLDv1 and
above for IPv6. IGMP V1 routes MUST be considered as invalid and the
PE MUST apply the "treat-as-withdraw" procedure as per [RFC7606]

12. Security Considerations

Same security considerations as [RFC7432] ,[RFC2236] ,[RFC3376] ,
[RFC2710], [RFC3810].

13. IANA Considerations

IANA has allocated the following codepoints from the EVPN Extended
Community sub-types registry.

```
           0x09    Multicast Flags Extended Community   [this document]
           0x0A    EVI-RT Type 0                        [this document]
           0x0B    EVI-RT Type 1                        [this document]
           0x0C    EVI-RT Type 2                        [this document]
```

IANA is requested to allocate a new codepoint from the EVPN Extended
Community sub-types registry for the following.

```
               0x0D    EVI-RT Type 3                       [this document]
```

IANA has allocated the following EVPN route types from the EVPN Route
Type registry.

```
               6 - Selective Multicast Ethernet Tag Route
               7 - Multicast Join Synch Route
               8 - Multicast Leave Synch Route
```

The Multicast Flags Extended Community contains a 16-bit Flags field.
The bits are numbered 0-15, from high-order to low-order.

```
          The registry should be initialized as follows:
       Bit          Name                          Reference
       ----         --------------                -------------
       0 - 13       Unassigned
       14           MLD Proxy Support             This document
       15           IGMP Proxy Support            This document
```

The registration policy should be "First Come First Served".

14. Acknowledgement

The authors would like to thank Stephane Litkowski, Jorge Rabadan,
Anoop Ghanwani, Jeffrey Haas, Krishna Muddenahally Ananthamurthy for
reviewing and providing valuable comment.

15. Contributors

Mankamana Mishra

Cisco systems

Email: mankamis@cisco.com

Derek Yeung

Arrcus

Email: derek@arrcus.com

16. References

16.1. 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>.

   [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>.

16.2. Informative References

   [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>.

Authors' Addresses

   Ali Sajassi
   Cisco Systems
   821 Alder Drive,
   MILPITAS, CALIFORNIA 95035
   UNITED STATES

   Email: sajassi@cisco.com

   Samir Thoria
   Cisco Systems
   821 Alder Drive,
   MILPITAS, CALIFORNIA 95035
   UNITED STATES

   Email: sthoria@cisco.com

   Keyur PAtel
   Arrcus
   UNITED STATES

   Email: keyur@arrcus.com

John Drake
Juniper Networks

Email: jdrake@juniper.net

Wen Lin
Juniper Networks

Email: wlin@juniper.net