Selective Multicast in EVPN
draft-zhang-l2vpn-evpn-selective-mcast-01

Abstract

[I-D.ietf-l2vpn-evpn] describes a solution of carrying multicast traffic using replication ingress and inclusive trees (P2MP LSPs). It does not provide the usage of selective trees for carrying multicast traffic. This document provides protocols and procedures by which a selective tree can be used to carry traffic belonging only to a specified set of IP multicast streams from one or more EVPN instances.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

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 http://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 January 4, 2015.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.
Introduction

[I-D.ietf-l2vpn-evpn] describes a solution of carrying multicast traffic using replication ingress and inclusive trees (P2MP LSPs). It does not provide the usage of selective trees for carrying multicast traffic. When replication ingress or inclusive tree is used to carry multicast traffic, it is likely that a PE may receive...
multicast traffic for a multicast stream even if it doesn’t have any receivers that are interested in receiving traffic for that stream. If a particular stream has a large amount of traffic, it may result in highly non-optimal bandwidth utilization in the provider network. To improve bandwidth utilization for IP multicast streams, this document provides protocols and procedures by which a selective tree can be used to carry traffic belonging only to a specified set of IP multicast streams from one or more EVPN instances.

Optimizing multicast distribution needs the binding of particular IP multicast stream to an selective tree and the explicit tracking of particular IP multicast stream. The granularity of binding to a selective tree is <RD, PE, S, G> where S is an IP multicast source address and G is an IP multicast group address. Wildcard sources and wildcard groups are supported.

This document assumes that the exchanging of BGP messages is within the AS.

Section 3 describes the needed extensions of the current defined BGP EVPN NLRI. Section 4 describes the procedures of binding and explicit tracking.

2. Terminology

This document uses terminology described in [I-D.ietf-l2vpn-evpn].

3. BGP EVPN NLRI Extensions

The format of BGP EVPN NLRI is defined in [I-D.ietf-l2vpn-evpn]. This section describes the encoding of the BGP EVPN NLRI extensions required by this document.

This document defines the following two new route types for BGP EVPN NLRI:

+ 5 - Selective Multicast Auto-Discovery (A-D) Route
+ 6 - Leaf Auto-Discovery (A-D) Route

The detailed encoding are described in subsequent sections.

3.1. Selective Multicast A-D Route

A Selective Multicast A-D Route type specific EVPN NLRI consists of the following:
The RD and the Ethernet Tag ID are encoded as described in [I-D.ietf-l2vpn-evpn].

The Multicast Source field contains the C-S address i.e the address of the multicast source. If the Multicast Source field contains an IPv4 address, then the value of the Multicast Source Length field is 32. If the Multicast Source field contains an IPv6 address, then the value of the Multicast Source Length field is 128. The value of the Multicast Source Length field may be set to 0 to indicate a wildcard.

The Multicast Group field contains the C-G address i.e. the address of the multicast group. 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 Length field may be set to 0 to indicate a wildcard.

Usage of Selective Multicast A-D Route is described in section 4.

3.2. Leaf A-D Route

A Leaf A-D Route type specific EVPN NLRI consists of the following:

Leaf A-D routes may be originated as a result of processing a received Selective Multicast A-D route. A Leaf A-D route is originated in these situations only if the received route has a PMSI
Tunnel attribute whose "Leaf Information Required" bit is set to 1, and the Route Key of the Leaf A-D route is set to the NLRI of the received route.

Usage of Leaf A-D routes is described in Section 4.

4. Optimizing Multicast Distribution via Selective Trees

This section describes the binding (C-S, C-G) to a Selective Tree using Selective Multicast A-D Route and the explicit tracking of the IP multicast stream using Leaf A-D Route.

4.1. Originating Selective Multicast A-D Routes

The ingress PE informs all the PEs those are on the path to receivers of the (C-S, C-G) of the binding of the Selective tree to the (C-S, C-G), using Selective Multicast A-D routes. The format of the NLRI of this route is described in section 3. The following describes procedures for originating Selective Multicast A-D Route by a PE.

4.1.1. Construction of the Selective Multicast A-D Route

The settings of the RD, Ethernet Tag ID, Originating Router’s IP address, Next Hop field of the MP_REACH_NLRI, RT attributes of the Selective Multicast A-D Route are the same as the Inclusive Multicast Ethernet Tag Route described in [I-D.ietf-l2vpn-evpn].

The Multicast Source field MUST contain the source address associated with the C-multicast stream, and the Multicast Source Length field is set appropriately to reflect this. If the source address is a wildcard the source address is set to 0.

The Multicast Group field MUST contain the group address associated with the C-multicast stream, and the Multicast Group Length field is set appropriately to reflect this. If the group address is a wildcard the group address is set to 0.

4.1.2. P-Tunnel Identification

The selective multicast A-D route MUST carry a "PMSI Tunnel Attribute" as specified in [RFC6514]. Depending on the technology used for the selective tree for the EVPN instance on the PE, the PMSI Tunnel attribute of the selective multicast A-D Route is constructed differently.
4.1.2.1. P2MP LSPs

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

If in order to establish the P-Multicast tree the PE needs to know the leaves of the tree, then the PE obtains this information from the leaf A-D routes received from other PEs (as other PEs originate leaf A-D routes in response to receiving the selective multicast A-D route) by setting the Leaf Information Required flag in the PMSI Tunnel attribute to 1. This enables explicit tracking for the multicast stream(s) advertised by the selective multicast A-D route.

If a PE originates Selective Multicast A-D routes with the Leaf Information Required flag in the PMSI Tunnel attribute set to 1, then the PE MUST be (auto)configured with an import Route Target, which controls acceptance of leaf A-D routes by the PE. This Route Target is IP address specific. The Global Administrator field of this Route Target MUST be set to the IP address carried in the Next Hop of all the selective multicast A-D routes advertised by this PE (if the PE uses different Next Hops, then the PE MUST be (auto)configured with multiple import RTs, one per each such Next Hop). The Local Administrator field of this Route Target MUST be set to 0.

A PE MAY aggregate two or more selective trees originated by the PE onto the same P-Multicast tree. If the PE already advertises selective multicast A-D routes for these selective trees, then aggregation requires the PE to re-advertise these routes. The re-advertised routes MUST be the same as the original ones, except for the PMSI tunnel attribute. If the PE has not previously advertised selective multicast A-D routes for these selective trees, then the aggregation requires the PE to advertise (new) selective multicast A-D routes for these selective trees. The PMSI Tunnel attribute in the newly advertised/re-advertised routes MUST carry the identity of the P-Multicast tree that aggregates the selective trees. If at least some of the selective trees aggregated onto the same P-Multicast tree belong to different EVPN instances, then these routes MUST carry an MPLS upstream assigned label [RFC5331]. If all these aggregated selective trees belong to the same EVPN, then the routes MAY carry an MPLS upstream assigned label [RFC5331]. The labels can be assigned per EVPN instance or per selective multicast A-D route.
4.1.2.2. MP2MP LSPs

Construction of MP2MP LSPs P-Tunnel identification is the same as specified in section 4.1.2.1 except the upstream assigned label demultiplexing the EVPN multicast traffic received over the selective tree. The allocation of this label will be described in future version.

4.2. Receiving Selective Multicast A-D Route by PE

When a PE receives a BGP Update message that carries a Selective Multicast A-D route such that (a) at least one of the Route Targets of the route matches one of the import Route Targets configured for a particular EVI on the local PE, (b) the BGP route selection determines that this is the best route with respect to the NLRI carried by the route, and (c) the multicast state snooped on the PE-CE interfaces match the Selective Multicast A-D route. (The procedures of determining the multicast snooped state will be described in a future version.), the PE performs the following:

+ If the Tunnel Type in the PMSI Tunnel attribute is set to P2MP LSP, the PE SHOULD join the P-Multicast tree whose identity is carried in the PMSI Tunnel attribute. (The procedures of building P2MP LSP are described in [RFC6388].). If the PMSI Tunnel attribute does not carry a label, then all packets that are received on the P-Multicast tree, as identified by the PMSI Tunnel attribute, are forwarded using the EVI that has at least one of its import Route Targets that matches one of the Route Targets of the received Selective Multicast A-D route. If the PMSI Tunnel attribute carries an MPLS label, then the egress PE MUST treat this as an upstream-assigned label, and all packets that are received on the P-Multicast tree, as identified by the PMSI Tunnel attribute, with that upstream label are forwarded using the EVI that has at least one of its import Route Target that matches one of the Route Targets of the received Selective Multicast A-D route.

+ The process in case of Tunnel Type set to mLDP MP2MP LSP will be described in future version.

If the received Selective Multicast A-D route has a PMSI Tunnel attribute with the Leaf Information Required flag set to 1, then the PE originates a Leaf A-D route constructed as follows (The format of the NLRI of the Leaf A-D route is described in section 3.).

+ The route carries a single EVPN NLRI with the Route Key field set to the EVPN NLRI of the Selective Multicast A-D route received.
from that neighbor and the Originating Router’s IP address set to the IP address of the PE (this MUST be a routable IP address).

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

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

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

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

In addition to the procedures specified above, the PE MUST set up its forwarding path to receive traffic, for each multicast stream in the matching snooped state, from the tunnel advertised by the Selective Multicast A-D route.

When a new snooped state is created by a PE then the PE MUST first determine if there is a Selective Multicast route that matches the snooped state. If such a Selective Multicast A-D route is found, then the PE MUST follow the procedures described in this section, for that particular Selective Multicast A-D route. If later on the snooped state ages out and is deleted from the PE, the PE SHOULD withdraw the leaf A-D route that it had originated in response to the Selective Multicast A-D route, and SHOULD prune itself off the associated selective tree.

The snooped state is said to "match" the Selective Multicast A-D route if any of the following is true:

+ The Selective Multicast A-D route carries (C-S, C-G) and the snooped state is for (C-S, C-G) or for (C-*, C-G), OR

+ The Selective Multicast A-D route carries (C-*, C-G) and (a) the snooped state is for (C-*, C-G) OR (b) the snooped state is for at least one multicast join with the multicast group address equal to C-G and there doesn’t exist another Selective Multicast A-D route that carries (CS, C-G) where C-S is the source address of the snooped state.
+ The Selective Multicast A-D route carries (C-S, C-*) and (a) the snooped state is for at least one multicast join with the multicast source address equal to C-S, and (b) there doesn’t exist another Selective Multicast A-D route that carries (C-S, C-G) where C-G is the group address of the snooped state.

+ The Selective Multicast A-D route carries (C-*, C-*) and there is no other Selective Multicast A-D route that matches the snooped state as per the above conditions.

5. IANA Considerations

This document requires IANA to assign two new route type values for EVPN NLRI.

6. Security Considerations

There are no additional security aspects beyond those of EVPN.

7. Reference

[I-D.ietf-l2vpn-evpn]


Authors’ Addresses