L2VPN Workgroup
INTERNET-DRAFT
Intended Status: Standards Track

Ali Sajassi
Samer Salam
Sami Boutros
Cisco

Wim Henderickx
Alcatel-Lucent

Jim Uttaro
AT&T

Aldrin Isaac
Bloomberg

Expires: April 21, 2014 October 21, 2013

E-TREE Support in EVPN & PBB-EVPN
draft-sajassi-l2vpn-evpn-etree-02

Abstract

The Metro Ethernet Forum (MEF) has defined a rooted-multipoint Ethernet service known as Ethernet Tree (E-Tree). [ETREE-FMWK] proposes a solution framework for supporting this service in MPLS networks. This document discusses how those functional requirements can be easily met with EVPN.

The list of Internet-Draft Shadow Directories can be accessed at
http://www.ietf.org/shadow.html

Copyright and License Notice

Table of Contents

1 Introduction

The Metro Ethernet Forum (MEF) has defined a rooted-multipoint Ethernet service known as Ethernet Tree (E-Tree). In an E-Tree service, endpoints are labeled as either Root or Leaf sites. Root sites can communicate with all other sites. Leaf sites can communicate with Root sites but not with other Leaf sites.

[ETREE-FMWK] proposes the solution framework for supporting E-Tree service in MPLS networks. The document identifies the functional components of the overall solution to emulate E-Tree services in addition to Ethernet LAN (E-LAN) services on an existing MPLS network.

[EVPN] is a solution for multipoint L2VPN services, with advanced multi-homing capabilities, using BGP for distributing customer/client MAC address reach-ability information over the MPLS/IP network. [PBB-EVPN] combines the functionality of EVPN with [802.1ah] Provider Backbone Bridging for MAC address scalability.

This document discusses how the functional requirements for E-Tree service can be easily met with EVPN and PBB-EVPN.

1.1 Terminology

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 [KEYWORDS].

2 E-Tree Scenarios and EVPN / PBB-EVPN Support

In this section, we will categorize support for E-Tree into three different scenarios, depending on the nature of the site association (Root/Leaf) per PE or per Ethernet Segment:

- Leaf OR Root site(s) per PE

- Leaf AND Root site(s) per PE

- Leaf AND Root site(s) per Ethernet Segment

2.1 Scenario 1: Leaf OR Root site(s) per PE

In this scenario, a PE may have Root sites OR Leaf sites for a given VPN instance, but not both concurrently. The PE may have both Root and Leaf sites albeit for different VPNs. Every Ethernet Segment

connected to the PE is uniquely identified as either a Root or a Leaf site.

```
                     +---------+            +---------+
                     |   PE1   |            |   PE2   |
    +---+            |  +---+  |  +-------+ |  +---+  |            +---+
    |CE1+-----ES1----+--+   |  |  | MPLS  | |  |   +--+----ES2-----+CE2|
    +---+    (Root)  |  | E |  |  |  /IP  | |  | E |  |   (Leaf)   +---+
                     |  | V |  |  |       | |  | V |  |
                     |  | I |  |  |       | |  | I |  |            +---+
                     |  |   |  |  |       | |  |   +--+----ES3-----+CE3|
                     |  +---+  |  +-------+ |  +---+  |   (Leaf)   +---+
                     +---------+            +---------+
```

Figure 1: Scenario 1

2.2 Scenario 2: Leaf AND Root site(s) per PE

In this scenario, a PE may have a set of one or more Root sites AND a set of one or more Leaf sites for a given VPN instance. Every Ethernet Segment connected to the PE is uniquely identified as either a Root or a Leaf site.

```
                     +---------+            +---------+
                     |   PE1   |            |   PE2   |
    +---+            |  +---+  |  +-------+ |  +---+  |            +---+
    |CE1+-----ES1----+--+   |  |  |       | |  |   +--+----ES2-----+CE2|
    +---+    (Leaf)  |  | E |  |  | MPLS  | |  | E |  |   (Leaf)   +---+
                     |  | V |  |  |  /IP  | |  | V |  |
                     |  | I |  |  |       | |  | I |  |            +---+
                     |  |   |  |  |       | |  |   +--+----ES3-----+CE3|
                     |  +---+  |  +-------+ |  +---+  |   (Root)   +---+
                     +---------+            +---------+
```

Figure 2: Scenario 2

2.3 Scenario 3: Leaf AND Root site(s) per Ethernet Segment

In this scenario, a PE may have a set of one or more Root sites AND a set of one or more Leaf sites for a given VPN instance. An Ethernet Segment connected to the PE may be identified as both a Root and a Leaf site concurrently.

```
                     +---------+            +---------+
                     |   PE1   |            |   PE2   |
    +---+            |  +---+  |  +------+  |  +---+  |            +---+
    |CE1+-----ES1----+--+   |  |  |      |  |  |   +--+----ES2-----+CE2|
    +---+ (Leaf/Root)|  | E |  |  | MPLS |  |  | E |  | (Leaf/Root)+---+
                     |  | V |  |  |  /IP |  |  | V |  |
                     |  | I |  |  |      |  |  | I |  |            +---+
                     |  |   |  |  |      |  |  |   +--+----ES3-----+CE3|
                     |  +---+  |  +------+  |  +---+  |   (Leaf)   +---+
                     +---------+            +---------+
```

Figure 3: Scenario 3

3 Operation for EVPN

[EVPN] defines the notion of an Ethernet Segment which can be readily used to identify a Root and/or Leaf site in E-TREE services. In other words, [EVPN] has inherent capability to support E-TREE services without defining any new BGP routes. It only requires a minor modification to the existing procedures and a modification to a BGP attribute as shown in this section.

The following procedures are used consistently for all the scenarios highlighted in the previous section.

3.1 Known Unicast Traffic

For known unicast traffic, the PE must advertise a Root/Leaf indication along with each MAC Advertisement route, to indicate whether the associated MAC address was learnt from a Root or a Leaf. This enables remote PEs to perform ingress filtering for known unicast traffic: On the ingress PE, the MAC destination address lookup yields, in addition to the forwarding adjacency, a flag which indicates whether the target MAC is associated with a Root or a Leaf site. The ingress PE cross-checks this flag with the status of the originating site, and if both are a Leaf, then the packet is not forwarded.

The PE places all Leaf Ethernet Segments of a given bridge domain in a single split-horizon group in order to prevent intra-PE forwarding among Leaf segments. This split-horizon function applies to BUM traffic as well.

To support the above ingress filtering functionality, a new Root/Leaf indication flag will added to the Tunnel Encapsulation Type Extended Community [RFC5512]. This extended community will be advertised with each EVPN MAC Advertisement route.

3.2 BUM Traffic

For BUM traffic, it is not possible to perform filtering on the ingress PE, as is the case with known unicast, because of the multi-destination nature of the traffic. As such, the solution relies on egress filtering. In order to apply the proper egress filtering, which varies based on whether a packet is sent from a Root or a Leaf, the MPLS-encapsulated frames MUST be tagged with an indication of whether they originated from a Root or a Leaf Ethernet Segment. This can be achieved in EVPN through the use of the ESI MPLS label, since this label identifies the Ethernet Segment of origin of a given frame. The egress PE determines whether or not to forward a particular frame to an Ethernet Segment depending on the split-horizon rule defined in [EVPN]:

- If the ESI Label indicates that the source Ethernet Segment is a Root, then the frame can be forwarded on a segment granted that it passes the split-horizon check.

- If the ESI Label indicates that the source Ethernet Segment is a Leaf, then the frame can be forwarded only on a Root segment, granted that it passes the split-horizon check.

When advertising the ESI MPLS label for a given Ethernet Segment, a PE must indicate whether the corresponding ESI is a Root or a Leaf site. This can be done by encoding the Root or Leaf indication in the Flags field of the ESI MPLS label Extended Community attribute ([EVPN] Section 8) to indicate Root/Leaf status.

In the case where a multi-homed Ethernet Segment has both Root and Leaf sites attached, two ESI MPLS labels are allocated and advertised: one ESI MPLS label denotes Root and the other denotes Leaf. The ingress PE imposes the right ESI MPLS label depending on whether the Ethernet frame originated from the Root or Leaf site on that Ethernet Segment. The mechanism by which the PE identifies whether a given frame originated from a Root or Leaf site on the segment is based on the Ethernet Tag associated with the frame. Other mechanisms of identification, beyond the Ethernet Tag, are outside the scope of this document. It should be noted that support for both Root and Leaf sites on a single Ethernet Segment requires that the PE performs the Ethernet Segment split-horizon check on a per Ethernet Tag basis. In the case where a multi-homed Ethernet Segment has either Root or Leaf sites attached, then a single ESI MPL label is allocated and advertised.

Furthermore, a PE advertises two special ESI MPLS labels: one for Root and another for Leaf. These are used by remote PEs for traffic originating from single-homed segments and for multi-homed segments

that are not connected to the advertising PE. Note that these special labels are advertised on a per PE basis (i.e. each PE advertises only two such special labels).

In addition to egress filtering (which is a MUST requirement), an EVPN PE implementation MAY provide topology constraint among the PEs belonging to the same EVI associated with an E-TREE service. The purpose of this topology constraint is to avoid having PEs with only host Leaf sites importing and processing BGP MAC routes from each other, thereby unnecessarily exhausting their RIB tables. However, as soon as a Root site is added to a Leaf PE, then that PE needs to process MAC routes from all other Leaf PEs and add them to its forwarding table. To support such topology constrain in EVPN, two BGP Route-Targets (RTs) are used for every EVPN Instance (EVI): one RT is associated with the Root sites and the other is associated with the Leaf sites. On a per EVI basis, every PE exports the single RT associated with its type of site(s). Furthermore, a PE with Root site(s) imports both Root and Leaf RTs, whereas a PE with Leaf site(s) only imports the Root RT. If for a given EVI, the PEs can eventually have both Leaf and Root sites attached, even though they may start as Root-only or Leaf-only PEs, then it is recommended to use a single RT per EVI and avoid additional configuration and operational overhead. If the number of EVIs is very large (e.g., more than 32K or 64K), then RT type 0 as defined in [RFC4360] SHOULD be used; otherwise, RT type 2 is sufficient.

3.3 E-TREE Traffic Flows for EVPN

Per [ETREE-FMWK], a generic E-Tree service supports all of the following traffic flows:

- Ethernet Unicast from Root to Roots & Leaf
- Ethernet Unicast from Leaf to Root
- Ethernet Broadcast/Multicast from Root to Roots & Leafs
- Ethernet Broadcast/Multicast from Leaf to Roots

A particular E-Tree service may need to support all of the above types of flows or only a select subset, depending on the target application. In the case where unicast flows need not be supported, the L2VPN PEs can avoid performing any MAC learning function.

In the subsections that follow, we will describe the operation of EVPN to support E-Tree service with and without MAC learning.

3.3.1 E-Tree with MAC Learning

The PEs implementing an E-Tree service must perform MAC learning when

unicast traffic flows must be supported from Root to Leaf or from Leaf to Root sites. In this case, the PE with Root sites performs MAC learning in the data-path over the Ethernet Segments, and advertises reachability in EVPN MAC Advertisement routes. These routes will be imported by PEs that have Leaf sites as well as by PEs that have Root sites, in a given EVI. Similarly, the PEs with Leaf sites perform MAC learning in the data-path over their Ethernet Segments, and advertise reachability in EVPN MAC Advertisement routes which are imported only by PEs with at least one Root site in the EVI. A PE with only Leaf sites will not import these routes. PEs with Root and/or Leaf sites may use the Ethernet A-D routes for aliasing (in the case of multi-homed segments) and for mass MAC withdrawal.

To support multicast/broadcast from Root to Leaf sites, either a P2MP tree rooted at the PE(s) with the Root site(s) or ingress replication can be used. The multicast tunnels are set up through the exchange of the EVPN Inclusive Multicast route, as defined in [EVPN].

To support multicast/broadcast from Leaf to Root sites, ingress replication should be sufficient for most scenarios where there is a single Root or few Roots. If the number of Roots is large, a P2MP tree rooted at the PEs with Leaf sites may be used.

3.3.2 E-Tree without MAC Learning

The PEs implementing an E-Tree service need not perform MAC learning when the traffic flows between Root and Leaf sites are multicast or broadcast. In this case, the PEs do not exchange EVPN MAC Advertisement routes. Instead, the Ethernet A-D routes are used to exchange the EVPN labels.

The fields of the Ethernet A-D route are populated per the procedures defined in [EVPN], and the route import rules are as described in previous sections.

4 Operation for PBB-EVPN

In PBB-EVPN, the PE must advertise a Root/Leaf indication along with each MAC Advertisement route, to indicate whether the associated B-MAC address corresponds to a Root or a Leaf site. Similar to the EVPN case, this flag will be added to the Tunnel Encapsulation Type Extended Community [RFC5512], and advertised with each MAC Advertisement route.

In the case where a multi-homed Ethernet Segment has both Root and Leaf sites attached, two B-MAC addresses are allocated and advertised: one B-MAC address denotes Root and the other denotes Leaf. The ingress PE uses the right B-MAC source address depending on

whether the Ethernet frame originated from the Root or Leaf site on that Ethernet Segment. The mechanism by which the PE identifies whether a given frame originated from a Root or Leaf site on the segment is based on the Ethernet Tag associated with the frame. Other mechanisms of identification, beyond the Ethernet Tag, are outside the scope of this document. It should be noted that support for both Root and Leaf sites on a single Ethernet Segment requires that the PE performs the Ethernet Segment split-horizon check on a per Ethernet Tag basis. In the case where a multi-homed Ethernet Segment has either Root or Leaf sites attached, then a single B-MAC address is allocated and advertised per segment.

Furthermore, a PE advertises two global B-MAC addresses: one for Root and another for Leaf, and tags them as such in the MAC Advertisement routes. These B-MAC addresses are used as source addresses for traffic originating from single-homed segments.

4.1 Known Unicast Traffic

For known unicast traffic, the PEs perform ingress filtering: On the ingress PE, the C-MAC destination address lookup yields, in addition to the target B-MAC address and forwarding adjacency, a flag which indicates whether the target B-MAC is associated with a Root or a Leaf site. The ingress PE cross-checks this flag with the status of the originating site, and if both are a Leaf, then the packet is not forwarded.

The PE places all Leaf Ethernet Segments of a given bridge domain in a single split-horizon group in order to prevent intra-PE forwarding among Leaf segments. This split-horizon function applies to BUM traffic as well.

4.2 BUM Traffic

For BUM traffic, the PEs must perform egress filtering. When a PE receives a MAC advertisement route, it updates its Ethernet Segment egress filtering function (based on the B-MAC source address), as follows:

- If the MAC Advertisement route indicates that the advertised B-MAC is a Leaf, and the local Ethernet Segment is a Leaf as well, then the source B-MAC address is added to the B-MAC filtering list.

- Otherwise, the B-MAC filtering list is not updated.

When the egress PE receives the packet, it examines the B-MAC source address to check whether it should filter or forward the frame. Note that this uses the same filtering logic as baseline [PBB-EVPN] and

does not require any additional flags in the data-plane.

5 Acknowledgement

We would like to thank Sami Boutros and Dennis Cai for their comments.

6 Security Considerations

Same security considerations as [EVPN].

7 IANA Considerations

Allocation of Extended Community Type and Sub-Type for EVPN.

8 References

8.1 Normative References

[KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4360] S. Sangli et al, ""BGP Extended Communities Attribute", February, 2006.

[RFC5512] Mohapatra, P. and E. Rosen, "The BGP Encapsulation Subsequent Address Family Identifier (SAFI) and the BGP Tunnel Encapsulation Attribute", RFC 5512, April 2009.

8.2 Informative References

[ETREE-FMWK] Key et al., "A Framework for E-Tree Service over MPLS Network", draft-ietf-l2vpn-etree-frwk-03, work in progress, September 2013.

[EVPN] Sajassi et al., "BGP MPLS Based Ethernet VPN", draft-ietf-l2vpn-evpn-04.txt, work in progress, July, 2013.

[PBB-EVPN] Sajassi et al., "PBB-EVPN", draft-ietf-l2vpn-pbb-evpn-05.txt, work in progress, October, 2013.

Authors' Addresses

Ali Sajassi
Cisco
Email: sajassi@cisco.com

Samer Salam
Cisco
Email: ssalam@cisco.com

Wim Henderickx
Alcatel-Lucent
Email: wim.henderickx@alcatel-lucent.com

Jim Uttaro
AT&T
Email: ju1738@att.com

Aldrin
Bloomberg Issac
Email: aisaac71@bloomberg.net

Sami Boutros
Cisco
Email: sboutros@cisco.com