Segment-Based EVPN (S-EVPN)
draft-li-l2vpn-segment-evpn-01

Abstract

This document proposes an enhanced EVPN mechanism, segment-based EVPN (S-EVPN). It satisfies the requirements of PBB-EVPN but does not require PBB implementation on PE. The solution uses a global label for each Ethernet Segment (ES) in an EVPN. It inserts the source ES label into packets at ingress PE and learns C-MAC and source ES label binding at egress PE. The solution makes the implementation easier and closer to E-VPN compared to PBB-EVPN but has the PBB-EVPN benefits.

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 August 18, 2014.
1. Introduction

E-VPN [I-D.ietf-l2vpn-evpn] introduces a solution for multipoint L2VPN services. It has multi-homing capability and uses BGP for distributing customer/client MAC address reachability information.
over the core MPLS/IP network. PBB-EVPN [I-D.ietf-l2vpn-pbb-evpn] integrates PBB and E-VPN to achieve the following objects:

1. reduce the number of MAC advertisement routes in BGP;
2. provide client MAC address mobility;
3. confine the scope of C-MAC learning to only active flows;
4. offer per site policies and avoid C-MAC address flushing on topology changes.

This document discusses the challenges faced by PBB-EVPN in the implementation and operation. It proposes an enhanced E-VPN mechanism, i.e., segment-based EVPN (S-EVPN), that provides the same benefits as PBB-EVPN but does not require implementing PBB function on PE. S-EVPN mechanism allocates a global label for each Ethernet Segment in E-VPN, inserts the source ES label into the packet at ingress PE, and learns C-MAC and source ES label binding at egress PE. As a result, it is not necessary to determine the source of C-MAC according to the B-MAC encapsulation which is required in PBB-EVPN. S-EVPN has simpler operation and management of EVPN and better encapsulation efficiency of packets compared to PBB-EVPN. In addition, it is easy to enhance the E-VPN to support S-EVPN and S-EVPN can unify the unicast traffic forwarding no matter C-MACs are learned by control plane or data plane.

2. Terminology

- BEB: Backbone Edge Bridge
- B-MAC: Backbone MAC Address
- CE: Customer Edge
- C-MAC: Customer/Client MAC Address
- ES: Ethernet Segment
- ESI: Ethernet Segment Identifier
- E-VPN: Ethernet VPN
- EVI: Ethernet VPN Instance
- LACP: Link Aggregation Control Protocol
- P2P: Point to Point
3. Challenges of PBB-EVPN Implementation

PBB-EVPN has advantages in the following aspects as [I-D.ietf-l2vpn-pbb-evpn]:

-- MAC Advertisement Route Scalability
-- C-MAC Mobility with MAC Sub-netting
-- C-MAC Address Learning and Confinement
-- Seamless Interworking with TRILL and 802.1aq Access Networks
-- Per Site Policy Support
-- Avoiding C-MAC Address Flushing

However, there are some challenges to implement PBB-EVPN.

1. Creation and Management B-MAC

For PBB-EVPN, the choice of B-MAC address(es) for the PE nodes must be examined carefully as it has implications on the proper operation of multi-homing. These addresses are usually locally administered by the Service Provider which involves a lot of operation and management such as design, configuration and checking. Automating B-MAC Address Assignment can be applied, but for some scenarios the method cannot work and manual provision is inevitable. A more general automated solution can be proposed to reduce manual intervention.

2. Encapsulation Efficiency of PBB-EVPN

When PBB encapsulation (shown in the figure 1) is adopted in PBB-EVPN, the B-DA, I-Tag, etc. fields in the encapsulation are useless in PBB-EVPN which reduce the effective payload.
In the PBB encapsulation for PBB-EVPN, the source B-MAC is necessary since the egress PE need to learn the correspondence between C-MACs and B-MACs. The destination B-MAC is not necessary since the destination (egress PE) is reachable through the tunnel setup in advance instead of searching routes according to the destination B-MAC.

The I-SID is also not necessary any more. PBB divides the Ethernet network into two layers: I-Component and B-Component. In the egress PE, B-Component need identify I-Component through I-SID. For PBB-VPLS, MAC learning is through the data plane which is always to use broadcast or multicast for unknown unicast traffic. In order to indentify different forwarding instance, I-SID must be adopted. For PBB-EVPN, the forwarding instance is constructed through the control plane. That is, the forwarding instance is constructed through the RT matching of EVIs and identified by the label advertised. So I-SID information in PBB encapsulation for PBB-EVPN is not useful any more.

In addition B-VID in PBB encapsulation is almost never used. In a summary, in the PBB encapsulation for PBB-EVPN, only source B-MAC is indispensable. The encapsulation efficiency can be optimized.

3. Combination of PBB and E-VPN

The issues are dealt with by PBB-EVPN through the combination of two distinct technologies: PBB (layer 2 technology) and MPLS technology. In order to reduce the number of BGP MAC advertisement routes in E-VPN, PBB-EVPN can aggregate Customer/Client MAC (C-MAC) addresses via Provider Backbone MAC address (B-MAC). In fact, C-MAC addresses can be aggregated via MPLS label. Thus the issue solved by PBB-VPN can be solved in the method that is based on only MPLS technology. That is, the method is similar as E-VPN which is only based on MPLS technology. In other word, we can enhance E-VPN according to the similar way to gain PBB-EVPN benefits but not implement PBB on PE, which is a cleaner and simpler solution than PBB-EVPN.
4. Architecture of S-EVPN

To implement C-MAC summarization scheme, Segment-based EVPN (S-EVPN) introduces a global label for each Ethernet Segment in an EVPN regardless single homed or multi-homed CE. BGP needs to advertise the global label and Ethernet Segment binding to all PEs. In data plane, the ingress PE inserts the source Ethernet Segment label into packets; the egress PE learns the C-MAC and source Ethernet Segment label binding upon receiving packets. S-EVPN purely relies on BGP IP/MPLS technology.

The encapsulation of S-EVPN is shown in figure 2 for the case that MPLS tunnel is used to transport EVPN traffic. The outmost label is the label for MPLS tunnel. The second label is the label which is allocated for Ethernet A-D route per EVI as E-VPN [I-D.ietf-l2vpn-evpn] and can identify a given <ESI, Ethernet Tag ID> tuple per EVI or per <ESI, EVI> (where the Ethernet Tag ID is set to 0). The third label is a global label which identifies an Ethernet Segment uniquely. The global label allocated for a specific Ethernet Segment will be described in section 4.2.

<table>
<thead>
<tr>
<th>Tunnel Label</th>
<th>EVI Label</th>
<th>Source ES Global Label</th>
<th>Payload</th>
</tr>
</thead>
</table>

Figure 2: S-EVPN Encapsulation

4.1. C-MAC Learning

In S-EVPN, C-MACs can be learned in the data plane to determine which source Ethernet Segment they are from and which EVI they belongs to. The forwarding entry to these learned C-MACs can be installed according to the source ES and EVI information.

In S-EVPN, the ingress PE needs to send unknown traffic with source C-MACs to all remote PEs according to the encapsulation as shown in figure 2. When a specific egress PE receives the packet:

1. it can learn the C-MAC and possible VLAN Tag in the payload;

2. it can learns the EVI which the C-MAC belongs to according to the EVI label which is allocated by the egress PE;

3. it can learns the Source Ethernet Segment which the C-MAC belongs to according to the advertised the global label and Ethernet Segment binding in BGP.
Then the egress PE needs to install the forwarding entry to the learned C-MAC. The forwarding entry to the C-MAC need two types of information: the reachability information to the ingress PE which the C-MAC belongs to; the identification for the Ethernet Segment of the EVI on the ingress PE through which the packet can send to the C-MAC.

1. Tunnel to the ingress PE: the egress PE determines PE which the Source Ethernet Segment belongs to according to the advertised the global label and Ethernet Segment binding in BGP. Then egress PE can determine the tunnel to the ingress PE.

2. Label for the Ethernet Segment of the EVI on the ingress PE: The ingress PE needs to allocate label for the <ESI, EVI, Ethernet Tag ID> tuple per EVI or per <ESI, EVI> and advertise the corresponding Ethernet A-D Route per EVI to remote PEs. The egress PE can determines the Source Ethernet Segment, the EVI and the possible VLAN which the learned the C-MAC belongs to. Then it can determine the label binded to the <ESI, EVI, Ethernet Tag ID> tuple per EVI or per <ESI, EVI> which is advertised though the Ethernet A-D Route per EVI by the ingress PE.

Besides the two types of forwarding information, when the egress PE sends a specific packet to the learned C-MAC, it needs to determine the Ethernet Segment from which the packet come and encapsulate the global label for the Ethernet Segment firstly in the packet.

According to above procedures in S-EVPN, the egress PE can learn C-MACs and install forwarding entries to these C-MACs.

4.2. ES Global Label Assignment

In S-EVPN, C-MAC summarization is done per an Ethernet Segment. The global ES label is introduced to identify the Ethernet Segment. The advantages of using global label are:

1. identify the ES globally;

2. leverage existing MPLS label stack implementation;

3. the label can be allocated dynamically to automate provision.
In order to allocate a global label for an Ethernet Segment, there should be a centralized control point. Route Reflector (RR) of BGP may serve as this role and we call this type of RR as RR+. The S-EVPN network is shown in the figure 3. All PEs of S-EVPN connects with RR+. The procedure is as follows:

1. Auto-Discovery of Ethernet Segment

RR+ can learn Ethernet Segment through the Ethernet A-D route per Ethernet Segment defined by [I-D.ietf-l2vpn-evpn]. Note that, in S-EVPN, every ES MUST has a unique identifier including the single-homed CEs. That is, ESI 0 cannot denote for a single-homed CE in S-EVPN. The ESI for the single-homed CE MUST be unique network wide and can be created automatically. The ESI is encoded as a ten octets integer. One way to generate ESI value for a single-homed CE is to use the MAC address of the Ethernet Segment with the low order 4 octets filled by value 0. The ESI value generation for multi-homed CE is specified in EVPN and can be reused in S-EVPN. Through Ethernet A-D route per Ethernet Segment, RR+ can learn all Ethernet Segments on all PEs.

2. ES Global Label Allocation

[I-D.li-mpls-global-label-framework] specifies the framework for the global label allocation. In S-EVPN, RR+ allocates global labels for the Ethernet Segments discovered and advertises <Ethernet Segment, label> pair to all PEs. The PEs that are members of E-VPN keep track of the global label/Ethernet Segment mappings.
4.3. Ethernet A-D Route Per EVI

The procedures defined for Ethernet A-D router per EVI in [I-D.ietf-l2vpn-evpn] will be reused by S-EVPN. In S-EVPN, both single home CE and multi-home CE have a unique ES identification. So for both single-homed CEs and multi-homed CEs, PEs needs to allocate MPLS label for the \(<ESI, EVI, Ethernet Tag ID>\) tuple per EVI or per \(<ESI, EVI>\) and advertise corresponding Ethernet A-D routes per EVI. The MPLS label is used to identify a specific Ethernet Segment in an EVI.

4.4. Ethernet A-D Route Per ES

In S-EVPN, support of Ethernet A-D Route per Ethernet Segment is still MANDATORY. PEs can learn Ethernet Segments through this type of route as E-VPN. In S-EVPN, RR+ which all PEs connect to can also learn Ethernet Segments. When constructing the Ethernet A-D Route per Ethernet Segment, there are following differences from E-VPN:

-- The ESI for the single-homed CE in this route MUST be unique network wide instead of 0.

-- The "ESI Label Extended Community" MUST be included in the route and the "Active-Standby" bit in the flags MUST be set accordingly. But the MPLS label in the extended community can be set as 0 (Invalid MPLS label value) since the ES global label is introduced in S-EVPN which can substitute ESI label.

5. Improvement on EVPN

When S-EVPN process is introduced, the E-VPN process defined by [I-D.ietf-l2vpn-evpn] can also be improved. The improvement includes split horizon, unifying unicast and multicast forwarding.

5.1. Split Horizon

ES global label is introduced to identify the Ethernet Segment globally. Thus S-EVPN can fulfill requirements proposed by PBB-EVPN. Besides this, the ES global label can also be used for split horizon in EVPN. In order to achieve split horizon function, E-VPN adopts ESI label to encapsulate it in every BUM packet originating from a non-DF PE to identify the Ethernet Segment of origin. ES global label can use for the same purpose since it can identify the Ethernet Segment. Every BUM packet originating from a non-DF PE is encapsulated as the encapsulation which is shown in the figure 2. Since the original ESI label in E-VPN can be substituted by the ES global label, the ESI label in the ESI Label Extended Community can be an invalid label value. For the reason of compatibility, the ESI
Label Extended Community can carry a valid ESI label. Both ESI label and ES global label SHOULD be used for split horizon no matter which label is encapsulated in the packet.

[I-D.ietf-l2vpn-evpn-req] specifies the multicast optimization requirements to use MP2MP LSPs in EVPN. The ES global label can also solve the possible issue for split horizon when MP2MP LSP is used to transport BUM traffic. In E-VPN, when P2MF LSPs is used the upstream label assignment mechanism is introduced for split horizon. When PE received the packet, it decapsulates the top MPLS label and forwards the packet using the context label space determined by the top label. If the next label is the ESI label allocated by the ingress PE for a specific Ethernet Segment, the received PE will not forward the packet on the corresponding ES. In the MP2MP LSP scenarios, there are multiple roots and the upstream label allocated for Ethernet Segment maybe the same. So the received PE cannot determine a correct context label space according the top label for the MP2MP LSP. That is, the upstream label assignment mechanism for split horizon introduced in the P2MP LSP scenario can not be reused in the MP2MP LSP. But if the ES global label is used, in the MP2MP LSP scenario the received PE can also determine not to forward the packet on the specific ES which is identified by the ES global label. In one word, no matter ingress replication, P2MP LSP, or MP2MP, S-EVPN provides a unified solution for split horizon based on the ES global label. It can reduce the complexity of the split horizon mechanism in E-VPN.

5.2. Unifying MPLS Forwarding

S-EVPN adopts MPLS forwarding for C-MAC learning. In the control plane, it is just to add one new route type for E-VPN. It is a smooth upgrading of E-VPN and can switch easily between C-MAC learning through control plane and C-MAC learning through data plane.

When C-MACs is learned through the control plane, the unicast forwarding uses the label for the MAC route which is shown as follows:

```
+--------------+-----------+---------------+
| Tunnel Label | MAC Label |    Payload    |
+--------------+-----------+---------------+
```

Figure 4: E-VPN Unicast Forwarding Encapsulation

When C-MACs is learned through the data plane, the unicast forwarding uses the EVI label and the Segment global label which is shown in figure 2. In fact even if the C-MAC is learned through the data plane, the data plane can also use following encapsulation. In this case, the label in MAC advertisement route should not be used. From
the comparison, we can see that when E-VPN and S-EVPN are introduced, the forwarding encapsulation can be unified no matter which way C-MACs are learned by.

+--------------+---------------+------------------------+---------------+
<table>
<thead>
<tr>
<th>Tunnel Label</th>
<th>EVI Label</th>
<th>Source ES Global Label</th>
<th>Payload</th>
</tr>
</thead>
</table>

Figure 5: Unicast Forwarding Encapsulation without MAC Label

6. BGP E-VPN NLRI Extensions

6.1. ES Global Label Request Extended Community

ES Global Label Request Extended Community may be advertised along Ethernet A-D route per Ethernet Segment. ES Global Label Request Extended Community can reuse ESI Label Extended Community defined in [I-D.ietf-l2vpn-evpn] which is shown in the following figure:

<table>
<thead>
<tr>
<th>Type=0x06</th>
<th>Sub-Type=0x01</th>
<th>Flags (One Octet)</th>
<th>Reserved=0</th>
</tr>
</thead>
</table>

Reserved = 0

<table>
<thead>
<tr>
<th>Reserved = 0</th>
<th>ESI Label</th>
</tr>
</thead>
</table>

There defines a new bit of the flag octet as the "Global Label Request" bit.

+++++++
|*|*|*|*|2|1|0|
+++++++

Bit0: "Active-Standby" bit
Bit1: "Root-Leaf" bit
Bit2: "Global Label Request" bit

The third low order bit of the flags octet is defined as the "Global Label Request". A value of 0 means there is no global label request for the Ethernet A-D route. A value of 1 means that global label request is associated with the Ethernet A-D route.

6.2. ES Global Label Mapping Route

A new route type is defined for E-VPN NLRI to allocate global label for Ethernet Segment:

+5 - ES Global Label Mapping Route
An ES Global Label Mapping route type specific E-VPN NLRI consists of the following:

+---------------------------------------+
|      RD   (8 octets)                  |
|-------------------------------------+
|Ethernet Segment Identifier (10 octets)|
|-------------------------------------+
|   Ethernet Tag ID (4 octets)         |
|-------------------------------------+
|  MPLS Global Label (3 octets)        |
|-------------------------------------+
|               .......                |
|-------------------------------------+
|  MPLS Global Label (3 octets)        |
|-------------------------------------+

7. Operations

7.1. ES Global Label Request

Global label request is only for the Ethernet A-D route per Ethernet Segment. The Ethernet A-D route per Ethernet Segment is constructed as defined by [I-D.ietf-l2vpn-evpn]. The Ethernet Segment Identifier MUST be a unique ten octet entity. Even if the CE is single-homed, the corresponding Ethernet Segment Identifier MUST NOT be the reserved value 0.

When request a global label for a specific Ethernet Segment, ES Global Label Request Extended Community MUST be used for the Ethernet A-D route. ES Global Label Request Extended Community of S-EVPN can reuse the ESI Label Extended Community. The "Global Label Request" bit of the flag octet MUST be set as 1 for Global Label Request. According to Section 5 "Improvement on E-VPN", if ES global label is introduced, the original ESI label MAY NOT be used. The "root-leaf" bit of the flag octet and the ESI Label value in the ESI Label Extended Community can always be 0 to simplify the process.

One or more Route Target (RT) MUST be carried with the Ethernet A-D route. These RTs are the set of RTs associated with all the EVIs to which the Ethernet Segment belongs. Since the Global label is allocated per Ethernet Segment, RTs carried by the Ethernet A-D route will be ignored by the RR+ when allocate global label for the Ethernet Segment specified in the Ethernet A-D routes. The global label per Ethernet Segment is advertised to all PEs. For multi-homed Ethernet Segment, if one EVI on one PE requests label allocation for the Ethernet Segment and the ES Global Label Mapping Route has been
advertised corresponding to the Ethernet Segment, other EVIs on other PEs SHOULD NOT send the global label request for the Ethernet Segment again, that is, the "Global Label Request" bit SHOULD set as 0 when advertise Ethernet A-D routes for the Ethernet Segment by these EVIs.

7.2. ES Global Label Allocation

When RR+ receives the Ethernet A-D route per Ethernet Segment and the "Global Label Request" bit of the ES Global Label Request Extended Community is set as 1, RR+ MUST allocate global label for the Ethernet Segment and advertise the ES Global Mapping route to all PEs.

The ES Global Label Mapping route is constructed as follows:

RD, Ethernet Segment Identifier and Ethernet Tag ID values can be directly derived from the corresponding Ethernet A-D route per Ethernet Segment.

The MPLS Global Label field carries one or more labels (that corresponds to the stack of labels [MPLS-ENCAPS]). Each label is encoded as 3 octets, where the high-order 20 bits contain the label value, and the low order bit contains "Bottom of Stack" (as defined in [MPLS-ENCAPS]).

One or more Route Target(RT) MUST be carried with the ES Global Label Mapping route. These RTs can be directly derived from the RTs associated with the corresponding Ethernet A-D route.

For multi-homed Ethernet Segment, there maybe multiple global label request for the same Ethernet Segment advertised to RR+ by different PEs. When RR+ receives them, if RTs for these routes are same, owing to the Ethernet Segment Identifier is the same, it SHOULD advertise only one corresponding ES Global Label Mapping Route to all PEs. That is, the subsequent global label request for the same Ethernet Segment SHOULD be ignored. If RTs carried with the Ethernet A-D routes for the Ethernet Segment are different, RR+ SHOULD advertise multiple ES Global Label Mapping Routes with the same global label value and different RTs.

8. Solution Advantages

S-EVN has following advantages:

1. Remove the requirement of automating B-MAC address assignment to simplify provision of PBB-EVPN.

2. Improve the encapsulation efficiency of PBB-EVPN.
3. Seamless MPLS thoughts to solve the issue dealt with by PBB-EVPN instead of combination of two distinct technologies.

4. Be able to unify the split horizon mechanisms for ingress replication, P2MP LSP, and MP2MP LSP in E-VPN.

5. Be able to unify unicast traffic forwarding of E-VPN to implement seamless switch between C-MACs learning through control plane and C-MACs learning through data plane.

9. IANA Considerations

This document requires IANA to assign a new route type value for E-VPN NLRI.

10. Security Considerations

There are no additional security aspects beyond those of VPLS/H-VPLS that need to be discussed here.

11. References

11.1. Normative References

[I-D.ietf-l2vpn-evpn-req]
Sajassi, A., Aggarwal, R., Bitar, N., and A. Isaac,
"Requirements for Ethernet VPN (EVPN)", draft-ietf-l2vpn-evpn-req-07 (work in progress), February 2014.

[I-D.ietf-l2vpn-evpn]


11.2. References

[I-D.ietf-l2vpn-pbb-evpn]

[I-D.li-mpls-global-label-framework]
Authors’ Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiing Rd.
Beijing 100095
China

Email: lizhenbin@huawei.com

Lucy Yong
Huawei Technologies
1700 Alma Dr. Suite 500
Plano, TX 75075
USA

Email: lucyyong@huawei.com

Junlin Zhang
Huawei Technologies
Huawei Bld., No.156 Beiing Rd.
Beijing 100095
China

Email: jackey.zhang@huawei.com