Internet Working Group   Ali Sajassi
Internet Draft   Samer Salam
Category: Standards Track   Sami Boutros
                                                           Cisco

Florin Balus   Nabil Bitar
Wim Henderickx   Verizon
Alcatel-Lucent

                                                    Aldrin Isaac
Clarence Filsfils   Bloomberg
Dennis Cai
Cisco   Lizhong Jin
                                                            ZTE

Expires: August 25, 2013   February 25, 2013

PBB-EVPN
draft-ietf-l2vpn-pbb-evpn-04

Status of this Memo

Abstract

This document discusses how Ethernet Provider Backbone Bridging
[802.1ah] can be combined with E-VPN in order to reduce the number of
BGP MAC advertisement routes by aggregating Customer/Client MAC (C-
MAC) addresses via Provider Backbone MAC address (B-MAC), provide
client MAC address mobility using C-MAC aggregation and B-MAC sub-
netting, confine the scope of C-MAC learning to only active flows,
offer per site policies and avoid C-MAC address flushing on topology
changes. The combined solution is referred to as PBB-EVPN.

Conventions

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.

Table of Contents

1. Introduction

[E-VPN] introduces a solution for multipoint L2VPN services, with advanced multi-homing capabilities, using BGP for distributing customer/client MAC address reach-ability information over the core MPLS/IP network. [802.1ah] defines an architecture for Ethernet Provider Backbone Bridging (PBB), where MAC tunneling is employed to improve service instance and MAC address scalability in Ethernet as well as VPLS networks [PBB-VPLS].

In this document, we discuss how PBB can be combined with E-VPN in order to: reduce the number of BGP MAC advertisement routes by aggregating Customer/Client MAC (C-MAC) addresses via Provider Backbone MAC address (B-MAC), provide client MAC address mobility using C-MAC aggregation and B-MAC sub-netting, confine the scope of C-MAC learning to only active flows, offer per site policies and avoid C-MAC address flushing on topology changes. The combined solution is referred to as PBB-EVPN.

2. Contributors

In addition to the authors listed above, the following individuals also contributed to this document.

Keyur Patel, Cisco
Sam Aldrin, Huawei
Himanshu Shah, Ciena

3. Terminology

BEB: Backbone Edge Bridge
B-MAC: Backbone MAC Address
CE: Customer Edge
C-MAC: Customer/Client MAC Address
DHD: Dual-homed Device
DHN: Dual-homed Network
LACP: Link Aggregation Control Protocol
LSM: Label Switched Multicast
MDT: Multicast Delivery Tree
MP2MP: Multipoint to Multipoint
P2MP: Point to Multipoint
P2P: Point to Point
PE: Provider Edge
PoA: Point of Attachment
PW: Pseudowire
E-VPN: Ethernet VPN

4. Requirements

The requirements for PBB-EVPN include all the requirements for E-VPN
that were described in [EVPN-REQ], in addition to the following:

4.1. MAC Advertisement Route Scalability

In typical operation, an [E-VPN] PE sends a BGP MAC Advertisement
Route per customer/client MAC (C-MAC) address. In certain
applications, this poses scalability challenges, as is the case in
virtualized data center environments where the number of virtual
machines (VMs), and hence the number of C-MAC addresses, can be in
the millions. In such scenarios, it is required to reduce the number
of BGP MAC Advertisement routes by relying on a 'MAC summarization'
scheme, as is provided by PBB. Note that the MAC summarization
capability already built into E-VPN is not sufficient in those
environments, as will be discussed next.

4.2. C-MAC Mobility with MAC Summarization

Certain applications, such as virtual machine mobility, require
support for fast C-MAC address mobility. For these applications, it
is not possible to use MAC address summarization in E-VPN, i.e.
advertise reach-ability to a MAC address prefix. Rather, the exact
virtual machine MAC address needs to be transmitted in BGP MAC
Advertisement route. Otherwise, traffic would be forwarded to the
wrong segment when a virtual machine moves from one Ethernet segment
to another. This hinders the scalability benefits of summarization.

It is required to support C-MAC address mobility, while retaining the
scalability benefits of MAC summarization. This can be achieved by
leveraging PBB technology, which defines a Backbone MAC (B-MAC)
address space that is independent of the C-MAC address space, and
aggregate C-MAC addresses via a B-MAC address and then apply
summarization to B-MAC addresses.

4.3. C-MAC Address Learning and Confinement

In E-VPN, all the PE nodes participating in the same E-VPN instance
are exposed to all the C-MAC addresses learnt by any one of these PE
nodes because a C-MAC learned by one of the PE nodes is advertise in
BGP to other PE nodes in that E-VPN instance. This is the case even
if some of the PE nodes for that E-VPN instance are not involved in
forwarding traffic to, or from, these C-MAC addresses. Even if an
implementation does not install hardware forwarding entries for C-MAC
addresses that are not part of active traffic flows on that PE, the
device memory is still consumed by keeping record of the C-MAC
addresses in the routing table (RIB). In network applications with
millions of C-MAC addresses, this introduces a non-trivial waste of
PE resources. As such, it is required to confine the scope of

   visibility of C-MAC addresses only to those PE nodes that are
   actively involved in forwarding traffic to, or from, these addresses.

4.4. Per Site Policy Support

   In many applications, it is required to be able to enforce
   connectivity policy rules at the granularity of a site (or segment).
   This includes the ability to control which PE nodes in the network
   can forward traffic to, or from, a given site. PBB-EVPN is capable of
   providing this granularity of policy control. In the case where per
   C-MAC address granularity is required, the EVI can always continue to
   operate in E-VPN mode.

4.5. Avoiding C-MAC Address Flushing

   It is required to avoid C-MAC address flushing upon link, port or
   node failure for multi-homed devices and networks. This is in order
   to speed up re-convergence upon failure.

5. Solution Overview

   The solution involves incorporating IEEE 802.1ah Backbone Edge Bridge
   (BEB) functionality on the E-VPN PE nodes similar to PBB-VPLS, where
   BEB functionality is incorporated in the VPLS PE nodes. The PE
   devices would then receive 802.1Q Ethernet frames from their
   attachment circuits, encapsulate them in the PBB header and forward
   the frames over the IP/MPLS core. On the egress E-VPN PE, the PBB
   header is removed following the MPLS disposition, and the original
   802.1Q Ethernet frame is delivered to the customer equipment.

```
                      BEB    +--------------+  BEB
                      ||     |              |  ||
                      \/     |              |  \/
         +----+ AC1 +----+ |              | +----+   +----+
         | CE1|-----|    | |              | |    |---| CE2|
         +----+\    | PE1| |  IP/MPLS     | | PE3|   +----+
              \   +----+ |  Network     | +----+
               \         |              |
          AC2\ +----+ |              |
              \|    | |              |
               | PE2| |              |
               +----+ |              |
                 /\   +--------------+
                 ||
                BEB
         <-802.1Q-> <------PBB over MPLS------> <-802.1Q->
```
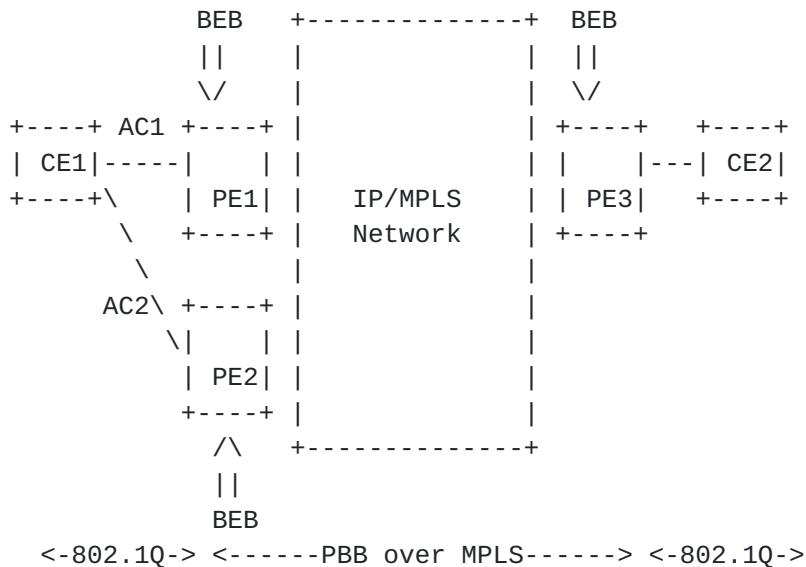
   Figure 1: PBB-EVPN Network

The PE nodes perform the following functions:- Learn customer/client
MAC addresses (C-MACs) over the attachment circuits in the data-
plane, per normal bridge operation.

- Learn remote C-MAC to B-MAC bindings in the data-plane from traffic
ingress from the core per [802.1ah] bridging operation.

- Advertise local B-MAC address reach-ability information in BGP to
all other PE nodes in the same set of service instances. Note that
every PE has a set of local B-MAC addresses that uniquely identify
the device. More on the PE addressing in section 5.

- Build a forwarding table from remote BGP advertisements received
associating remote B-MAC addresses with remote PE IP addresses and
the associated MPLS label(s).

6. BGP Encoding

PBB-EVPN leverages the same BGP Routes and Attributes defined in [E-
VPN], adapted as follows:

6.1. BGP MAC Advertisement Route

The E-VPN MAC Advertisement Route is used to distribute B-MAC
addresses of the PE nodes instead of the C-MAC addresses of end-
stations/hosts. This is because the C-MAC addresses are learnt in the
data-plane for traffic arriving from the core. The MAC Advertisement
Route is encoded as follows:

- The MAC address field contains the B-MAC address.
- The Ethernet Tag field is set to 0.
- The Ethernet Segment Identifier field must be set either to 0 (for
single-homed Segments) or to MAX-ESI (for multi-homed Segments). All
other values are not permitted.

The route is tagged with the RT corresponding to the EVI associated
with the B-MAC address.

All other fields are set as defined in [E-VPN].

6.2. Ethernet Auto-Discovery Route

This route and all of its associated modes are not needed in PBB-
EVPN.

The receiving PE knows that it need not wait for the receipt of the
Ethernet A-D route for route resolution by means of the reserved ESI
encoded in the MAC Advertisement route: the ESI values of 0 and MAX-

ESI indicate that the receiving PE can resolve the path without an
Ethernet A-D route.

6.3. Per VPN Route Targets

PBB-EVPN uses the same set of route targets defined in [E-VPN]. The
future revision of this document will describe new RT types.

6.4. MAC Mobility Extended Community

This extended community is defined in [EVPN]. When used in PBB-EVPN,
it indicates that the C-MAC forwarding tables for the I-SIDs
associated with the RT tagging the MAC Advertisement route must be
flushed.

Note that all other BGP messages and/or attributes are used as
defined in [E-VPN].

7. Operation

This section discusses the operation of PBB-EVPN, specifically in
areas where it differs from [E-VPN].

7.1. MAC Address Distribution over Core

In PBB-EVPN, host MAC addresses (i.e. C-MAC addresses) need not be
distributed in BGP. Rather, every PE independently learns the C-MAC
addresses in the data-plane via normal bridging operation. Every PE
has a set of one or more unicast B-MAC addresses associated with it,
and those are the addresses distributed over the core in MAC
Advertisement routes.

7.2. Device Multi-homing

7.2.1 Flow-based Load-balancing

This section describes the procedures for supporting device multi-
homing in an all-active redundancy model with flow-based load-
balancing.

7.2.1.1 PE B-MAC Address Assignment

In [802.1ah] every BEB is uniquely identified by one or more B-MAC
addresses. These addresses are usually locally administered by the
Service Provider. 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. In particular, for the scenario

where a CE is multi-homed to a number of PE nodes with all-active
redundancy and flow-based load-balancing, a given C-MAC address would
be reachable via multiple PE nodes concurrently. Given that any given
remote PE will bind the C-MAC address to a single B-MAC address, then
the various PE nodes connected to the same CE must share the same B-
MAC address. Otherwise, the MAC address table of the remote PE nodes
will keep oscillating between the B-MAC addresses of the various PE
devices. For example, consider the network of Figure 1, and assume
that PE1 has B-MAC BM1 and PE2 has B-MAC BM2. Also, assume that both
links from CE1 to the PE nodes are part of an all-active multi-
chassis Ethernet link aggregation group. If BM1 is not equal to BM2,
the consequence is that the MAC address table on PE3 will keep
oscillating such that the C-MAC address CM of CE1 would flip-flop
between BM1 or BM2, depending on the load-balancing decision on CE1
for traffic destined to the core.

Considering that there could be multiple sites (e.g. CEs) that are
multi-homed to the same set of PE nodes, then it is required for all
the PE devices in a Redundancy Group to have a unique B-MAC address
per site. This way, it is possible to achieve fast convergence in the
case where a link or port failure impacts the attachment circuit
connecting a single site to a given PE.

```
                           +---------+
            +-------+  PE1 | IP/MPLS |
           /            |  |         |
        CE1             |  | Network |      PEr
      M1   \            |  |         |
            +-------+  PE2 |         |
           /-------+      |  |         |
          /           |  |         |
        CE2           |  |         |
      M2    \         |  |         |
             \        |  |         |
             +------+  PE3 +---------+
```
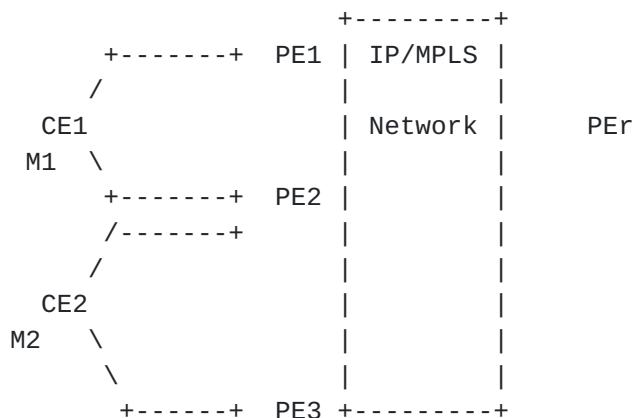
Figure 2: B-MAC Address Assignment

In the example network shown in Figure 2 above, two sites
corresponding to CE1 and CE2 are dual-homed to PE1/PE2 and PE2/PE3,
respectively. Assume that BM1 is the B-MAC used for the site
corresponding to CE1. Similarly, BM2 is the B-MAC used for the site
corresponding to CE2. On PE1, a single B-MAC address (BM1) is
required for the site corresponding to CE1. On PE2, two B-MAC
addresses (BM1 and BM2) are required, one per site. Whereas on PE3, a
single B-MAC address (BM2) is required for the site corresponding to
CE2. All three PE nodes would advertise their respective B-MAC
addresses in BGP using the MAC Advertisement routes defined in [E-

VPN]. The remote PE, PEr, would learn via BGP that BM1 is reachable
via PE1 and PE2, whereas BM2 is reachable via both PE2 and PE3.
Furthermore, PEr establishes via the normal bridge learning that C-
MAC M1 is reachable via BM1, and C-MAC M2 is reachable via BM2. As a
result, PEr can load-balance traffic destined to M1 between PE1 and
PE2, as well as traffic destined to M2 between both PE2 and PE3. In
the case of a failure that causes, for example, CE1 to be isolated
from PE1, the latter can withdraw the route it has advertised for
BM1. This way, PEr would update its path list for BM1, and will send
all traffic destined to M1 over to PE2 only.

For single-homed sites, it is possible to assign a unique B-MAC
address per site, or have all the single-homed sites connected to a
given PE share a single B-MAC address. The advantage of the first
model over the second model is the ability to avoid C-MAC destination
address lookup on the disposition PE (even though source C-MAC
learning is still required in the data-plane). Also, by assigning the
B-MAC addresses from a contiguous range, it is possible to advertise
a single B-MAC subnet for all single-homed sites, thereby rendering
the number of MAC advertisement routes required at par with the
second model.

In summary, every PE may use a unicast B-MAC address shared by all
single-homed CEs or a unicast B-MAC address per single-homed CE and,
in addition, a unicast B-MAC address per dual-homed CE. In the latter
case, the B-MAC address MUST be the same for all PE nodes in a
Redundancy Group connected to the same CE.

7.2.1.2. Automating B-MAC Address Assignment

The PE B-MAC address used for single-homed sites can be automatically
derived from the hardware (using for e.g. the backplane's address).
However, the B-MAC address used for multi-homed sites must be
coordinated among the RG members. To automate the assignment of this
latter address, the PE can derive this B-MAC address from the MAC
Address portion of the CE's LACP System Identifier by flipping the
'Locally Administered' bit of the CE's address. This guarantees the
uniqueness of the B-MAC address within the network, and ensures that
all PE nodes connected to the same multi-homed CE use the same value
for the B-MAC address.

Note that with this automatic provisioning of the B-MAC address
associated with multi-homed CEs, it is not possible to support the
uncommon scenario where a CE has multiple bundles towards the PE
nodes, and the service involves hair-pinning traffic from one bundle
to another. This is because the split-horizon filtering relies on B-
MAC addresses rather than Site-ID Labels (as will be described in the
next section). The operator must explicitly configure the B-MAC

address for this fairly uncommon service scenario.

Whenever a B-MAC address is provisioned on the PE, either manually or automatically (as an outcome of CE auto-discovery), the PE MUST transmit an MAC Advertisement Route for the B-MAC address with a downstream assigned MPLS label that uniquely identifies that address on the advertising PE. The route is tagged with the RTs of the associated EVIs as described above.

7.2.1.3 Split Horizon and Designated Forwarder Election

[E-VPN] relies on access split horizon, where the Ethernet Segment Label is used for egress filtering on the attachment circuit in order to prevent forwarding loops. In PBB-EVPN, the B-MAC source address can be used for the same purpose, as it uniquely identifies the originating site of a given frame. As such, Segment Labels are not used in PBB-EVPN, and the egress split-horizon filtering is done based on the B-MAC source address. It is worth noting here that [802.1ah] defines this B-MAC address based filtering function as part of the I-Component options, hence no new functions are required to support split-horizon beyond what is already defined in [802.1ah]. Given that the Segment label is not used in PBB-EVPN, the PE sets the Label field in the Ethernet Segment Route to 0.

The Designated Forwarder election procedures are defined in [I-D-Segment-Route].

7.2.2 I-SID Based Load-balancing

This section describes the procedures for supporting device multi-homing in an all-active redundancy model with per-ISID load-balancing.

7.2.2.1 PE B-MAC Address Assignment

In the case where per-ISID load-balancing is desired among the PE nodes in a given redundancy group, multiple unicast B-MAC addresses are allocated per multi-homed Ethernet Segment: Each PE connected to the multi-homed segment is assigned a unique B-MAC. Every PE then advertises its B-MAC address using the BGP MAC advertisement route.

A remote PE initially floods traffic to a destination C-MAC address, located in a given multi-homed Ethernet Segment, to all the PE nodes connected to that segment. Then, when reply traffic arrives at the remote PE, it learns (in the data-path) the B-MAC address and associated next-hop PE to use for said C-MAC address. When a PE connected to a multi-homed Ethernet Segment loses connectivity to the segment, due to link or port failure, it withdraws the B-MAC route

previously advertised for that segment. This causes the remote PE
nodes to flush all C-MAC addresses associated with the B-MAC in
question. This is done across all I-SIDs that are mapped to the EVI
of the withdrawn MAC route.

7.2.2.2 **Split Horizon and Designated Forwarder Election The procedures
are similar to the flow-based load-balancing case, with the only**
difference being that the DF filtering must be applied to unicast as
well as multicast traffic, and in both core-to-segment as well as
segment-to-core directions.

7.3. **Network Multi-homing**

When an Ethernet network is multi-homed to a set of PE nodes running
PBB-EVPN, an all-active redundancy model can be supported with per
service instance (i.e. I-SID) load-balancing. In this model, DF
election is performed to ensure that a single PE node in the
redundancy group is responsible for forwarding traffic associated
with a given I-SID. This guarantees that no forwarding loops are
created. Filtering based on DF state applies to both unicast and
multicast traffic, and in both access-to-core as well as core-to-
access directions (unlike the multi-homed device scenario where DF
filtering is limited to multi-destination frames in the core-to-
access direction). Similar to the multi-homed device scenario, with
I-SID based load-balancing, a unique B-MAC address is assigned to
each of the PE nodes connected to the multi-homed network (Segment).

7.4. **Frame Forwarding**

The frame forwarding functions are divided in between the Bridge
Module, which hosts the [802.1ah] Backbone Edge Bridge (BEB)
functionality, and the MPLS Forwarder which handles the MPLS
imposition/disposition. The details of frame forwarding for unicast
and multi-destination frames are discussed next.

7.4.1. **Unicast**

Known unicast traffic received from the AC will be PBB-encapsulated
by the PE using the B-MAC source address corresponding to the
originating site. The unicast B-MAC destination address is determined
based on a lookup of the C-MAC destination address (the binding of
the two is done via transparent learning of reverse traffic). The
resulting frame is then encapsulated with an LSP tunnel label and the
MPLS label which uniquely identifies the B-MAC destination address on
the egress PE. If per flow load-balancing over ECMPs in the MPLS core
is required, then a flow label is added as the end of stack label.

For unknown unicast traffic, the PE forwards these frames over MPLS

core. When these frames are to be forwarded, then the same set of
options used for forwarding multicast/broadcast frames (as described
in next section) are used.

7.4.2. Multicast/Broadcast

Multi-destination frames received from the AC will be PBB-
encapsulated by the PE using the B-MAC source address corresponding
to the originating site. The multicast B-MAC destination address is
selected based on the value of the I-SID as defined in [802.1ah]. The
resulting frame is then forwarded over the MPLS core using one out of
the following two options:

Option 1: the MPLS Forwarder can perform ingress replication over a
set of MP2P tunnel LSPs. The frame is encapsulated with a tunnel LSP
label and the E-VPN ingress replication label advertised in the
Inclusive Multicast Route.

Option 2: the MPLS Forwarder can use P2MP tunnel LSP per the
procedures defined in [E-VPN]. This includes either the use of
Inclusive or Aggregate Inclusive trees.

Note that the same procedures for advertising and handling the
Inclusive Multicast Route defined in [E-VPN] apply here.

8. Minimizing ARP Broadcast

The PE nodes implement an ARP-proxy function in order to minimize the
volume of ARP traffic that is broadcasted over the MPLS network. This
is achieved by having each PE node snoop on ARP request and response
messages received over the access interfaces or the MPLS core. The PE
builds a cache of IP / MAC address bindings from these snooped
messages. The PE then uses this cache to respond to ARP requests
ingress on access ports and targeting hosts that are in remote sites.
If the PE finds a match for the IP address in its ARP cache, it
responds back to the requesting host and drops the request.
Otherwise, if it does not find a match, then the request is flooded
over the MPLS network using either ingress replication or LSM.

9. Seamless Interworking with IEEE 802.1aq/802.1Qbp

```
                        +--------------+
                        |              |
            +---------+  |    MPLS      |    +---------+
 +----+     |          |  +----+        +----+  |          |  +----+
 |SW1 |--|  |          | | PE1|        | PE2| |          |--| SW3|
 +----+  | 802.1aq |---|    |         |    |--| 802.1aq |  +----+
 +----+  |  .1Qbp  |  +----+        +----+  |  .1Qbp  |  +----+
 |SW2 |--|          |  |    Backbone  |  |          |--| SW4|
 +----+  +---------+  +--------------+   +---------+  +----+


 |<------- IS-IS -------->|<-----BGP----->|<------ IS-IS ------>|  CP


 |<-----------------------  PBB  -------------------------->|  DP
                         |<----MPLS----->|
```
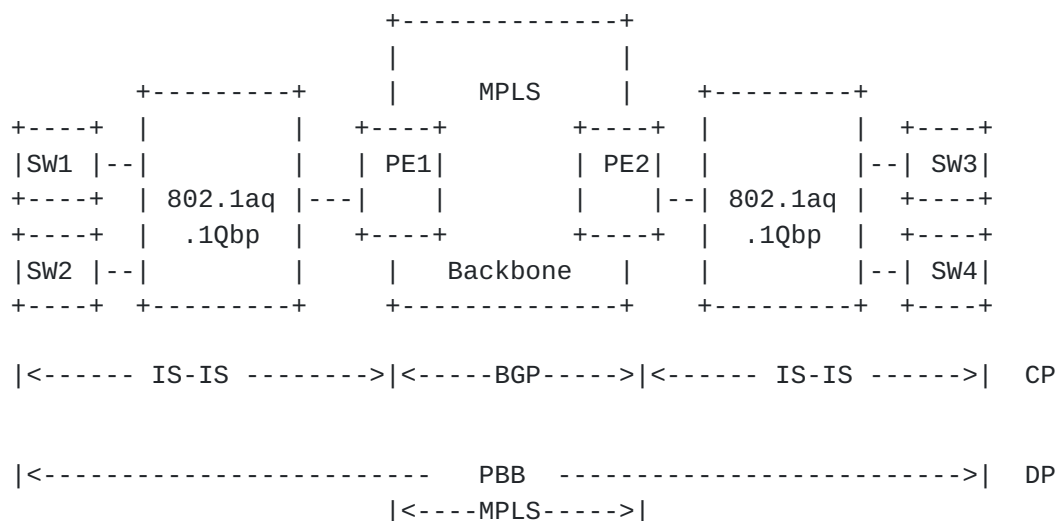
      Legend: CP = Control Plane View
              DP = Data Plane View

      Figure 7: Interconnecting 802.1aq/802.1Qbp Networks with PBB-EVPN

9.1 **B-MAC Address Assignment**

   For the same reasons cited in the TRILL section, the B-MAC addresses
   need to be globally unique across all the IEEE 802.1aq / 802.1Qbp
   networks. The same hierarchical address assignment scheme depicted
   above is proposed for B-MAC addresses as well.

9.2 **IEEE 802.1aq / 802.1Qbp B-MAC Advertisement Route**

   B-MAC addresses associated with 802.1aq / 802.1Qbp switches are
   advertised using the BGP MAC Advertisement route already defined in
   [E-VPN].

   The encapsulation for the transport of PBB frames over MPLS is
   similar to that of classical Ethernet, albeit with the additional PBB
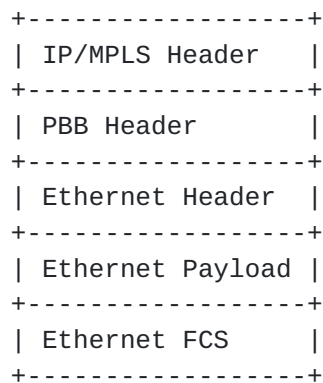   header, as shown in the figure below:

```
+------------------+
| IP/MPLS Header   |
+------------------+
| PBB Header       |
+------------------+
| Ethernet Header  |
+------------------+
| Ethernet Payload |
+------------------+
| Ethernet FCS     |
+------------------+
```

Figure 8: PBB over MPLS Encapsulation

9.3 Operation:

When a PE receives a PBB-encapsulated Ethernet frame from the access
side, it performs a lookup on the B-MAC destination address to
identify the next hop. If the lookup yields that the next hop is a
remote PE, the local PE would then encapsulate the PBB frame in MPLS.
The label stack comprises of the VPN label (advertised by the remote
PE), followed by an LSP/IGP label. From that point onwards, regular
MPLS forwarding is applied.

On the disposition PE, assuming penultimate-hop-popping is employed,
the PE receives the MPLS-encapsulated PBB frame with a single label:
the VPN label. The value of the label indicates to the disposition PE
that this is a PBB frame, so the label is popped, the TTL field (in
the 802.1Qbp F-Tag) is reinitialized and normal PBB processing is
employed from this point onwards.

10. Solution Advantages

In this section, we discuss the advantages of the PBB-EVPN solution
in the context of the requirements set forth in section 3 above.

10.1. MAC Advertisement Route Scalability

In PBB-EVPN the number of MAC Advertisement Routes is a function of
the number of segments (sites), rather than the number of
hosts/servers. This is because the B-MAC addresses of the PEs, rather
than C-MAC addresses (of hosts/servers) are being advertised in BGP.
And, as discussed above, there's a one-to-one mapping between multi-
homed segments and B-MAC addresses, whereas there's a one-to-one or
many-to-one mapping between single-homed segments and B-MAC addresses
for a given PE. As a result, the volume of MAC Advertisement Routes
in PBB-EVPN is multiple orders of magnitude less than E-VPN.

10.2. C-MAC Mobility with MAC Sub-netting

   In PBB-EVPN, if a PE allocates its B-MAC addresses from a contiguous
   range, then it can advertise a MAC prefix rather than individual 48-
   bit addresses. It should be noted that B-MAC addresses can easily be
   assigned from a contiguous range because PE nodes are within the
   provider administrative domain; however, CE devices and hosts are
   typically not within the provider administrative domain. The
   advantage of such MAC address sub-netting can be maintained even as
   C-MAC addresses move from one Ethernet segment to another. This is
   because the C-MAC address to B-MAC address association is learnt in
   the data-plane and C-MAC addresses are not advertised in BGP. To
   illustrate how this compares to E-VPN, consider the following
   example:

   If a PE running E-VPN advertises reachability for a MAC subnet that
   spans N addresses via a particular segment, and then 50% of the MAC
   addresses in that subnet move to other segments (e.g. due to virtual
   machine mobility), then in the worst case, N/2 additional MAC
   Advertisement routes need to be sent for the MAC addresses that have
   moved. This defeats the purpose of the sub-netting. With PBB-EVPN, on
   the other hand, the sub-netting applies to the B-MAC addresses which
   are statically associated with PE nodes and are not subject to
   mobility. As C-MAC addresses move from one segment to another, the
   binding of C-MAC to B-MAC addresses is updated via data-plane
   learning.

10.3. C-MAC Address Learning and Confinement

   In PBB-EVPN, C-MAC address reachability information is built via
   data-plane learning. As such, PE nodes not participating in active
   conversations involving a particular C-MAC address will purge that
   address from their forwarding tables. Furthermore, since C-MAC
   addresses are not distributed in BGP, PE nodes will not maintain any
   record of them in control-plane routing table.

10.4. Seamless Interworking with TRILL and 802.1aq Access Networks

   Consider the scenario where two access networks, one running MPLS and
   the other running 802.1aq, are interconnected via an MPLS backbone
   network. The figure below shows such an example network.

```
                               +--------------+
                               |              |
                 +---------+   |    MPLS      |   +---------+
         +----+  |         |   +----+         +----+  |         |  +----+
         | CE |--|         |   | PE1|         | PE2|  |         |--| CE |
         +----+  | 802.1aq |---|    |         |    |--|  MPLS   |  +----+
         +----+  |         |   +----+         +----+  |         |  +----+
         | CE |--|         |   |    Backbone  |       |         |--| CE |
         +----+  +---------+   +--------------+       +---------+  +----+
```

Figure 9: Interoperability with 802.1aq

If the MPLS backbone network employs E-VPN, then the 802.1aq data-
plane encapsulation must be terminated on PE1 or the edge device
connecting to PE1. Either way, all the PE nodes that are part of the
associated service instances will be exposed to all the C-MAC
addresses of all hosts/servers connected to the access networks.
However, if the MPLS backbone network employs PBB-EVPN, then the
802.1aq encapsulation can be extended over the MPLS backbone, thereby
maintaining C-MAC address transparency on PE1. If PBB-EVPN is also
extended over the MPLS access network on the right, then C-MAC
addresses would be transparent to PE2 as well.

Interoperability with TRILL access network will be described in
future revision of this draft.

10.5. Per Site Policy Support

In PBB-EVPN, a unique B-MAC address can be associated with every site
(single-homed or multi-homed). Given that the B-MAC addresses are
sent in BGP MAC Advertisement routes, it is possible to define per
site (i.e. B-MAC) forwarding policies including policies for E-TREE
service.

10.6. Avoiding C-MAC Address Flushing

With PBB-EVPN, it is possible to avoid C-MAC address flushing upon
topology change affecting a multi-homed device. To illustrate this,
consider the example network of Figure 1. Both PE1 and PE2 advertize
the same B-MAC address (BM1) to PE3. PE3 then learns the C-MAC
addresses of the servers/hosts behind CE1 via data-plane learning. If
AC1 fails, then PE3 does not need to flush any of the C-MAC addresses
learnt and associated with BM1. This is because PE1 will withdraw the
MAC Advertisement routes associated with BM1, thereby leading PE3 to
have a single adjacency (to PE2) for this B-MAC address. Therefore,
the topology change is communicated to PE3 and no C-MAC address
flushing is required.

[11]. Acknowledgements

TBD.

[12]. Security Considerations

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

[13]. IANA Considerations

This document requires IANA to assign a new SAFI value for L2VPN_MAC SAFI.

[14]. Intellectual Property Considerations

This document is being submitted for use in IETF standards discussions.

[15]. Normative References

[802.1ah] "Virtual Bridged Local Area Networks Amendment 7: Provider Backbone Bridges", IEEE Std. 802.1ah-2008, August 2008.

[16]. Informative References

[PBB-VPLS] Sajassi et al., "VPLS Interoperability with Provider Backbone Bridges", draft-ietf-l2vpn-vpls-pbb-interop-02.txt, work in progress, July, 2011.

[EVPN-REQ] Sajassi et al., "Requirements for Ethernet VPN (E-VPN)", draft-sajassi-raggarwa-l2vpn-evpn-req-01.txt, work in progress, July, 2011.

[E-VPN] Aggarwal et al., "BGP MPLS Based Ethernet VPN", draft-ietf-l2vpn-evpn-00.txt, work in progress, February, 2012.

[17]. Authors' Addresses

Ali Sajassi
Cisco
170 West Tasman Drive
San Jose, CA 95134, US
Email: sajassi@cisco.com

Samer Salam
Cisco
595 Burrard Street, Suite 2123
Vancouver, BC V7X 1J1, Canada
Email: ssalam@cisco.com

Sami Boutros
Cisco
170 West Tasman Drive
San Jose, CA 95134, US
Email: sboutros@cisco.com

Nabil Bitar
Verizon Communications
Email : nabil.n.bitar@verizon.com

Aldrin Isaac
Bloomberg
Email: aisaac71@bloomberg.net

Florin Balus
Alcatel-Lucent
701 E. Middlefield Road
Mountain View, CA, USA 94043
Email: florin.balus@alcatel-lucent.com

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

Clarence Filsfils
Cisco
Email: cfilsfil@cisco.com

Dennis Cai
Cisco
Email: dcai@cisco.com

Lizhong Jin
ZTE Corporation

   889, Bibo Road
   Shanghai, 201203, China
   Email: lizhong.jin@zte.com.cn