            Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)

Abstract

   This document discusses how Ethernet Provider Backbone Bridging (PBB)
   can be combined with Ethernet VPN (EVPN) in order to reduce the
   number of BGP MAC Advertisement routes by aggregating Customer/Client
   MAC (C-MAC) addresses via Provider Backbone MAC (B-MAC) address,
   provide client MAC address mobility using C-MAC aggregation, 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.

Table of Contents

1. Introduction

   [RFC7432] introduces a solution for multipoint Layer 2 Virtual
   Private Network (L2VPN) services, with advanced multihoming
   capabilities, using BGP for distributing customer/client MAC address
   reachability information over the core MPLS/IP network. [PBB]
   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
   [RFC7080].

   In this document, we discuss how PBB can be combined with EVPN in
   order to: reduce the number of BGP MAC Advertisement routes by
   aggregating Customer/Client MAC (C-MAC) addresses via Provider
   Backbone MAC (B-MAC) address, provide client MAC address mobility
   using C-MAC aggregation, 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. Terminology

     ARP: Address Resolution Protocol
     BEB: Backbone Edge Bridge
     B-MAC: Backbone MAC
     B-VID: Backbone VLAN ID
     CE: Customer Edge
     C-MAC: Customer/Client MAC
     ES: Ethernet Segment
     ESI: Ethernet Segment Identifier
     EVI: EVPN Instance
     EVPN: Ethernet VPN
     I-SID: Service Instance Identifier (24 bits and global within a PBB
             network see [RFC7080])
     LSP: Label Switched Path
     MP2MP: Multipoint to Multipoint
     MP2P: Multipoint to Point
     NA: Neighbor Advertisement
     ND: Neighbor Discovery
     P2MP: Point to Multipoint
     P2P: Point to Point
     PBB: Provider Backbone Bridge
     PE: Provider Edge
     RT: Route Target
     VPLS: Virtual Private LAN Service

     Single-Active Redundancy Mode: When only a single PE, among a group
     of PEs attached to an Ethernet segment, is allowed to forward traffic
     to/from that Ethernet segment, then the Ethernet segment is defined
     to be operating in Single-Active redundancy mode.

     All-Active Redundancy Mode: When all PEs attached to an Ethernet
     segment are allowed to forward traffic to/from that Ethernet segment,
     then the Ethernet segment is defined to be operating in All-Active
     redundancy mode.

     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 BCP 14 [RFC2119].

3. Requirements

     The requirements for PBB-EVPN include all the requirements for EVPN
     that were described in [RFC7209], in addition to the following:

3.1. MAC Advertisement Route Scalability

   In typical operation, an EVPN PE sends a BGP MAC Advertisement route
   per C-MAC address. In certain applications, this poses scalability
   challenges, as is the case in data center interconnect (DCI)
   scenarios 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.

3.2. C-MAC Mobility Independent of B-MAC Advertisements

   Certain applications, such as virtual machine mobility, require
   support for fast C-MAC address mobility. For these applications,
   when using EVPN, the 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 ES to another. This means MAC address prefixes cannot be used in
   data center applications.

   In order to support C-MAC address mobility, while retaining the
   scalability benefits of MAC summarization, PBB technology is used.
   It defines a B-MAC address space that is independent of the C-MAC
   address space, and aggregates C-MAC addresses via a single B-MAC
   address.

3.3. C-MAC Address Learning and Confinement

   In EVPN, all the PE nodes participating in the same EVPN instance are
   exposed to all the C-MAC addresses learned by any one of these PE
   nodes because a C-MAC learned by one of the PE nodes is advertised in
   BGP to other PE nodes in that EVPN instance. This is the case even
   if some of the PE nodes for that EVPN 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 information base (RIB) table. 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 to only those PE nodes
   that are actively involved in forwarding traffic to, or from, these
   addresses.

3.4. 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. Both EVPN and PBB-
   EVPN are capable of providing this granularity of policy control. In
   the case where the policy needs to be at the granularity of per C-MAC
   address, then the C-MAC address should be learned in the control
   plane (in BGP) per [RFC7432].

3.5. No C-MAC Address Flushing for All-Active Multihoming

   Just as in [RFC7432], it is required to avoid C-MAC address flushing
   upon link, port, or node failure for All-Active multihomed segments.

4. Solution Overview

   The solution involves incorporating IEEE Backbone Edge Bridge (BEB)
   functionality on the EVPN 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 EVPN 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->
```
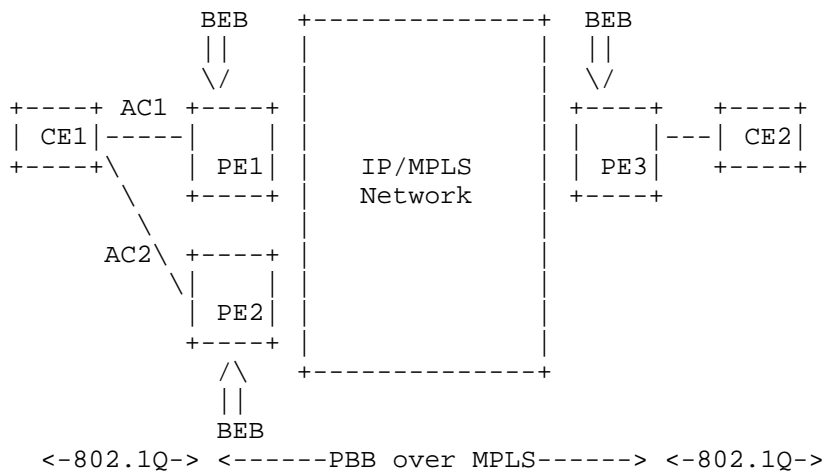
                      Figure 1: PBB-EVPN Network

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 for traffic
  received from the core per the bridging operation described in
  [PBB].

- Advertise local B-MAC address reachability information in BGP to
  all other PE nodes in the same set of service instances. Note
  that every PE has a set of B-MAC addresses that uniquely
  identifies the device. B-MAC address assignment is described in
  details in Section 6.2.2.

- 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).

5. BGP Encoding

   PBB-EVPN leverages the same BGP routes and attributes defined in
   [RFC7432], adapted as described below.

5.1. Ethernet Auto-Discovery Route

   This route and all of its associated modes are not needed in PBB-EVPN
   because PBB encapsulation provides the required level of indirection
   for C-MAC addresses -- i.e., an ES can be represented by a B-MAC
   address for the purpose of data-plane learning/forwarding.

   The receiving PE knows that it need not wait for the receipt of the
   Ethernet A-D (auto-discovery) 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.

5.2. MAC/IP Advertisement Route

   The EVPN MAC/IP 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 learned 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 to either 0 (for
single-homed segments or multihomed segments with per-I-SID load-
balancing) or to MAX-ESI (for multihomed segments with per-flow
load-balancing). All other values are not permitted.

- All other fields are set as defined in [RFC7432].

This route is tagged with the RT corresponding to its EVI. This EVI
is analogous to a B-VID.

5.3. Inclusive Multicast Ethernet Tag Route

This route is used for multicast pruning per I-SID. It is used for
auto-discovery of PEs participating in a given I-SID so that a
multicast tunnel (MP2P, P2P, P2MP, or MP2MP LSP) can be set up for
that I-SID. [RFC7080] uses multicast pruning per I-SID based on
[MMRP], which is a soft-state protocol. The advantages of multicast
pruning using this BGP route over [MMRP] are that a) it scales very
well for a large number of PEs and b) it works with any type of LSP
(MP2P, P2P, P2MP, or MP2MP); whereas, [MMRP] only works over P2P
pseudowires. The Inclusive Multicast Ethernet Tag route is encoded
as follows:

- The Ethernet Tag field is set with the appropriate I-SID value.

- All other fields are set as defined in [RFC7432].

This route is tagged with an RT. This RT SHOULD be set to a value
corresponding to its EVI (which is analogous to a B-VID). The RT for
this route MAY also be auto-derived from the corresponding Ethernet
Tag (I-SID) based on the procedure specified in Section 5.1.2.1 of
[OVERLAY].

5.4. Ethernet Segment Route

This route is used for auto-discovery of PEs belonging to a given
redundancy group (e.g., attached to a given ES) per [RFC7432].

5.5. ESI Label Extended Community

This extended community is not used in PBB-EVPN. In [RFC7432], this
extended community is used along with the Ethernet A-D route to
advertise an MPLS label for the purpose of split-horizon filtering.
Since in PBB-EVPN, the split-horizon filtering is performed natively
using B-MAC source address, there is no need for this extended
community.

5.6. ES-Import Route Target

   This RT is used as defined in [RFC7432].

5.7. MAC Mobility Extended Community

   This extended community is defined in [RFC7432] and it is used with a
   MAC route (B-MAC route in case of PBB-EVPN). The B-MAC route is
   tagged with the RT corresponding to its EVI (which is analogous to a
   B-VID). When this extended community is used along with a B-MAC
   route in PBB-EVPN, it indicates that all C-MAC addresses associated
   with that B-MAC address across all corresponding I-SIDs must be
   flushed.

   When a PE first advertises a B-MAC, it MAY advertise it with this
   extended community where the sticky/static flag is set to 1 and the
   sequence number is set to zero. In such cases where the PE wants to
   signal the stickiness of a B-MAC, then when a flush indication is
   needed, the PE advertises the B-MAC along with the MAC Mobility
   extended community where the sticky/static flag remains set and the
   sequence number is incremented.

5.8. Default Gateway Extended Community

   This extended community is not used in PBB-EVPN.

6. Operation

   This section discusses the operation of PBB-EVPN, specifically in
   areas where it differs from [RFC7432].

6.1. 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.

6.2. Device Multihoming

6.2.1. Flow-Based Load-Balancing

   This section describes the procedures for supporting device
   multihoming in an All-Active redundancy mode (i.e., flow-based load-
   balancing).

6.2.1.1. PE B-MAC Address Assignment

   In [PBB], 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 multihoming. In particular, for the scenario
   where a CE is multihomed to a number of PE nodes with All-Active
   redundancy mode, 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 address BM1 and PE2 has B-MAC address BM2. Also, assume
   that both links from CE1 to the PE nodes are part of the same
   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 M1 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
   multihomed 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 +---------+
```
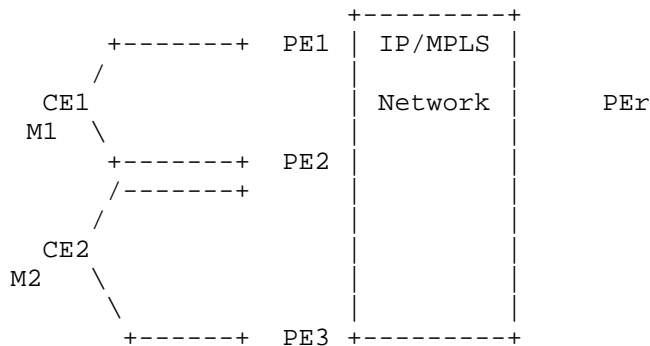
                 Figure 2: B-MAC Address Assignment

   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
[RFC7432]. 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 PBB bridge learning
procedure, 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.

6.2.1.2. Automating B-MAC Address Assignment

   The PE B-MAC address used for single-homed or Single-Active sites can
   be automatically derived from the hardware (using for example the
   backplane's address and/or PE's reserved MAC pool). However, the
   B-MAC address used for All-Active sites must be coordinated among the
   redundancy group 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 Link Aggregation Control Protocol (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 All-
   Active CE use the same value for the B-MAC address.

   Note that with this automatic provisioning of the B-MAC address
   associated with All-Active CEs, it is not possible to support the
   uncommon scenario where a CE has multiple link bundles within the
   same LACP session 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 a 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.

6.2.1.3. Split Horizon and Designated Forwarder Election

   [RFC7432] relies on split-horizon filtering for a multi-homed ES,
   where the ES 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, ES 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
   [PBB] 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 filtering beyond what is already defined in
   [PBB].

   The Designated Forwarder (DF) election procedures are defined in
   [RFC7432].

6.2.2. Load-Balancing based on I-SID

   This section describes the procedures for supporting device
   multihoming in a Single-Active redundancy mode with per-I-SID load-
   balancing.

6.2.2.1. PE B-MAC Address Assignment

   In the case where per-I-SID load-balancing is desired among the PE
   nodes in a given redundancy group, multiple unicast B-MAC addresses
   are allocated per multihomed ES: Each PE connected to the multihomed
   segment is assigned a unique B-MAC. Every PE then advertises its
   B-MAC address using the BGP MAC Advertisement route. In this mode of
   operation, two B-MAC address-assignment models are possible:

   - The PE may use a shared B-MAC address for all its single-homed
      segments and/or all its multi-homed Single-Active segments (e.g.,
      segments operating in per-I-SID load-balancing mode).

   - The PE may use a dedicated B-MAC address for each ES operating
      with per-I-SID load-balancing mode.

   A PE implementation MAY choose to support either the shared B-MAC
   address model or the dedicated B-MAC address model without causing
   any interoperability issues. The advantage of the dedicated B-MAC
   over the shared B-MAC address for multi-homed Single-Active segments,
   is that when C-MAC flushing is needed, fewer C-MAC addresses are
   impacted. Furthermore, it gives the disposition PE the ability to

avoid C-MAC destination address lookup even though source C-MAC learning is still required in the data plane. Its disadvantage is that there are additional B-MAC advertisements in BGP.

A remote PE initially floods traffic to a destination C-MAC address, located in a given multihomed ES, to all the PE nodes configured with that I-SID. 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.

6.2.2.2. 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.

6.2.2.3. Handling Failure Scenarios

When a PE connected to a multihomed ES loses connectivity to the segment, due to link or port failure, it needs to notify the remote PEs to trigger C-MAC address flushing. This can be achieved in one of two ways, depending on the B-MAC assignment model:

- If the PE uses a shared B-MAC address for multiple Ethernet segments, then the C-MAC flushing is signaled by means of having the failed PE re-advertise the MAC Advertisement route for the associated B-MAC, tagged with the MAC Mobility extended community attribute. The value of the Counter field in that attribute must be incremented prior to advertisement. 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.

- If the PE uses a dedicated B-MAC address for each ES operating under per-I-SID load-balancing mode, the failed PE simply 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.

When a PE connected to a multihomed ES fails (i.e., node failure) or when the PE becomes completely isolated from the EVPN network, the remote PEs will start purging the MAC Advertisement routes that were advertised by the failed PE. This is done either as an outcome of the remote PEs detecting that the BGP session to the failed PE has gone down, or by having a Route Reflector withdrawing all the routes that were advertised by the failed PE. The remote PEs, in this case,

   will perform C-MAC address flushing as an outcome of the MAC
   Advertisement route withdrawals.

   For all failure scenarios (link/port failure, node failure, and PE
   node isolation), when the fault condition clears, the recovered PE
   re-advertises the associated ES route to other members of its
   redundancy group. This triggers the backup PE(s) in the redundancy
   group to block the I-SIDs for which the recovered PE is a DF. When a
   backup PE blocks the I-SIDs, it triggers a C-MAC address flush
   notification to the remote PEs by re-advertising the MAC
   Advertisement route for the associated B-MAC, with the MAC Mobility
   extended community attribute. The value of the Counter field in that
   attribute must be incremented prior to advertisement. 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/re-advertised MAC route.

6.3. Network Multihoming

   When an Ethernet network is multihomed to a set of PE nodes running
   PBB-EVPN, Single-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 just like a Single-Active multihomed device
   scenario (but unlike an All-Active multihomed device scenario where
   DF filtering is limited to multi-destination frames in the core-to-
   access direction). Similar to a Single-Active multihomed device
   scenario, with load-balancing based on I-SID, a unique B-MAC address
   is assigned to each of the PE nodes connected to the multihomed
   network (segment).

6.4. Frame Forwarding

   The frame-forwarding functions are divided in between the Bridge
   Module, which hosts the [PBB] 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.

6.4.1. Unicast

   Known unicast traffic received from the Attachment Circuit (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 an EVPN label associated with the B-MAC
   destination address. If per flow load-balancing over ECMPs in the
   MPLS core is required, then a flow label is added below the label
   associated with the B-MAC address in the label stack.

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

6.4.2. 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 [PBB]. The
   resulting frame is then forwarded over the MPLS core using one of the
   following two options:

   Option 1: the MPLS Forwarder can perform ingress replication over a
      set of MP2P or P2P tunnel LSPs. The frame is encapsulated with a
      tunnel LSP label and the EVPN ingress replication label advertised
      in the Inclusive Multicast Ethernet Tag [RFC7432].

   Option 2: the MPLS Forwarder can use P2MP tunnel LSP per the
      procedures defined in [RFC7432]. This includes either the use of
      Inclusive or Aggregate Inclusive trees. Furthermore, the MPLS
      Forwarder can use MP2MP tunnel LSP if Inclusive trees are used.

   Note that the same procedures for advertising and handling the
   Inclusive Multicast Ethernet Tag defined in [RFC7432] apply here.

6.5. MPLS Encapsulation of PBB Frames

   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    |
                      +-----------------+
```
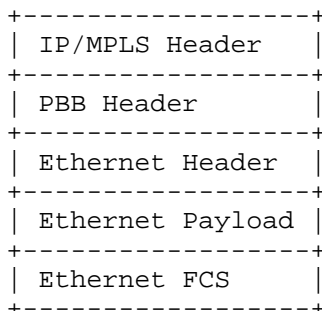
                 Figure 3: PBB over MPLS Encapsulation

7. Minimizing ARP/ND Broadcast

   The PE nodes MAY implement an ARP/ND-proxy function in order to
   minimize the volume of ARP/ND traffic that is broadcasted over the
   MPLS network. In case of ARP proxy, 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 target
   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 P2MP LSPs. In case of ND proxy, this is achieved
   similar to the above but with ND/NA messages per [RFC4389].

8. 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
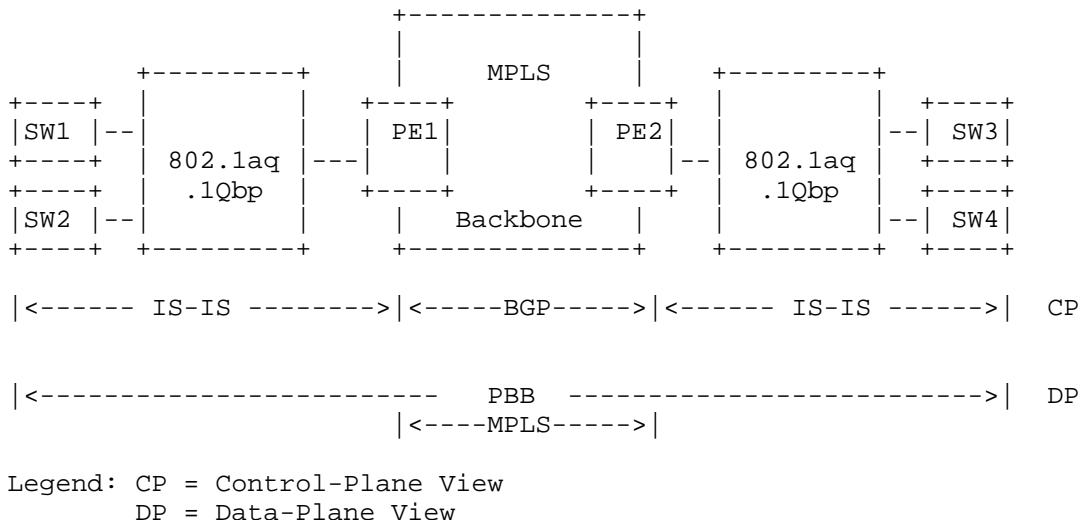
         Figure 4: Interconnecting 802.1aq / 802.1Qbp Networks with PBB-EVPN

8.1. B-MAC Address Assignment

   The B-MAC addresses need to be globally unique across all networks
   including PBB-EVPN and IEEE 802.1aq / 802.1Qbp networks. The B-MAC
   addresses used for single-homed and Single-Active Ethernet segments
   should be unique because they are typically auto-derived from the
   PE's pools of reserved MAC addresses that are unique. The B-MAC
   addresses used for All-Active Ethernet segments should also be unique
   given that each network operator typically has its own assigned
   Organizationally Unique Identifier (OUI) and thus can assign its own
   unique MAC addresses.

8.2. IEEE 802.1aq / 802.1Qbp B-MAC Address Advertisement

   B-MAC addresses associated with 802.1aq / 802.1Qbp switches are
   advertised using the EVPN MAC/IP route advertisement already defined
   in [RFC7432].

8.3. 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.

9. 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.

9.1. MAC Advertisement Route Scalability

   In PBB-EVPN, the number of MAC Advertisement routes is a function of
   the number of Ethernet segments (e.g., 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. As discussed above, there's a one-to-one mapping between
   All-Active multihomed segments and their associated B-MAC addresses;
   there can be either a one-to-one or many-to-one mapping between
   Single-Active multihomed segments and their associated B-MAC
   addresses; and finally there is a many-to-one mapping between single-
   home sites and their associated B-MAC addresses on a given PE. This
   means a single B-MAC is associated with one or more segments where
   each segment can be associated with many C-MAC addresses. As a
   result, the volume of MAC Advertisement routes in PBB-EVPN may be
   multiple orders of magnitude less than EVPN.

9.2. C-MAC Mobility Independent of B-MAC Advertisements

   As described above, in PBB-EVPN, a single B-MAC address can aggregate
   many C-MAC addresses. Given that B-MAC addresses are associated with
   segments attached to a PE or to the PE itself, their locations are
   fixed and thus not impacted what so ever by C-MAC mobility.
   Therefore, C-MAC mobility does not affect B-MAC addresses (e.g., any
   re-advertisements of them). This is because the association of C-MAC
   address to B-MAC address is learned in the data-plane and C-MAC
   addresses are not advertised in BGP. Aggregation via B-MAC addresses
   in PBB-EVPN performs much better than EVPN.

To illustrate how this compares to EVPN, consider the following
example:

> If a PE running EVPN advertises reachability for N MAC addresses
> via a particular segment, and then 50% of the MAC addresses in
> that segment move to other segments (e.g., due to virtual machine
> mobility), then N/2 additional MAC Advertisement routes need to be
> sent for the MAC addresses that have moved. With PBB-EVPN, on the
> other hand, the B-MAC addresses that are statically associated
> with PE nodes 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 in PBB-EVPN.

9.3. 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 the control-plane routing table.

9.4. Seamless Interworking with 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
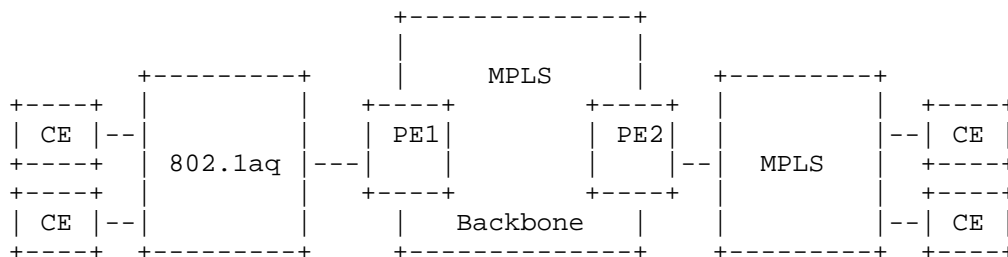   network. The figure below shows such an example network.

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

                  Figure 5: Interoperability with 802.1aq

   If the MPLS backbone network employs EVPN, 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.

9.5. Per-Site Policy Support

In PBB-EVPN, the per-site policy can be supported via B-MAC addresses
via assigning a unique B-MAC address for every site/segment
(typically multihomed but can also be single-homed). Given that the
B-MAC addresses are sent in BGP MAC/IP route advertisement, it is
possible to define per-site (i.e., B-MAC) forwarding policies
including policies for E-TREE service.

9.6. No C-MAC Address Flushing for All-Active Multihoming

Just as in [RFC7432], with PBB-EVPN, it is possible to avoid C-MAC
address flushing upon topology change affecting an All-Active
multihomed segment. To illustrate this, consider the example network
of Figure 1. Both PE1 and PE2 advertise 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 learned 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.

10. Security Considerations

All the security considerations in [RFC7432] apply directly to this
document because this document leverages the control plane described
in [RFC7432] and their associated procedures -- although not the
complete set but rather a subset.

This document does not introduce any new security considerations
beyond that of [RFC7432] and [RFC4761] because advertisements and
processing of B-MAC addresses follow that of [RFC7432] and processing
of C-MAC addresses follow that of [RFC4761] -- i.e, B-MAC addresses
are learned in the control plane and C-MAC addresses are learned in
data plane.

11. IANA Considerations

There are no additional IANA considerations for PBB-EVPN beyond what
is already described in [RFC7432].

12. References

12.1. Normative References

   [PBB]   IEEE, "IEEE Standard for Local and metropolitan area
                networks - Media Access Control (MAC) Bridges and Virtual
                Bridged Local Area Networks", Clauses 25 and 26, IEEE Std
                802.1Q, DOI 10.1109/IEEESTD.2011.6009146.

   [RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
                Requirement Levels", BCP 14, RFC 2119,
                DOI 10.17487/RFC2119, March 1997,
                <http://www.rfc-editor.org/info/rfc2119>.

   [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, <http://www.rfc-editor.org/info/rfc7432>.

12.2. Informative References

   [MMRP]   IEEE, "IEEE Standard for Local and metropolitan area
                networks - Media Access Control (MAC) Bridges and Virtual
                Bridged Local Area Networks", Clause 10, IEEE Std 802.1Q,
                DOI 10.1109/IEEESTD.2011.6009146.

   [OVERLAY]   Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Isaac, A.,
                Uttaro, J., Henderickx, W., Shekhar, R., Salam, S., Patel,
                K., Rao, D., and S. Thoria, "A Network Virtualization
                Overlay Solution using EVPN",
                draft-ietf-bess-evpn-overlay-01, February 2015.

   [RFC4389]   Thaler, D., Talwar, M., and C. Patel, "Neighbor Discovery
                Proxies (ND Proxy)", RFC 4389, DOI 10.17487/RFC4389, April
                2006, <http://www.rfc-editor.org/info/rfc4389>.

   [RFC4761]   Kompella, K., Ed., and Y. Rekhter, Ed., "Virtual Private
                LAN Service (VPLS) Using BGP for Auto-Discovery and
                Signaling", RFC 4761, DOI 10.17487/RFC4761, January 2007,
                <http://www.rfc-editor.org/info/rfc4761>.

   [RFC7080]   Sajassi, A., Salam, S., Bitar, N., and F. Balus, "Virtual
                Private LAN Service (VPLS) Interoperability with Provider
                Backbone Bridges", RFC 7080, DOI 10.17487/RFC7080,
                December 2013, <http://www.rfc-editor.org/info/rfc7080>.

   [RFC7209] Sajassi, A., Aggarwal, R., Uttaro, J., Bitar, N.,
              Henderickx, W., and A. Isaac, "Requirements for Ethernet
              VPN (EVPN)", RFC 7209, DOI 10.17487/RFC7209, May 2014,
              <http://www.rfc-editor.org/info/rfc7209>.

Contributors

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

   Lizhong Jin, ZTE
   Sami Boutros, Cisco
   Dennis Cai, Cisco
   Keyur Patel, Cisco
   Sam Aldrin, Huawei
   Himanshu Shah, Ciena
   Jorge Rabadan, ALU

Authors' Addresses

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

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

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

    Aldrin Isaac
    Juniper
    Email: aisaac@juniper.net

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