BESS WorkGroup
Internet-Draft
Intended status: Standards Track
Expires: September 10, 2020

A. Sajassi
P. Brissette
Cisco Systems
R. Schell
Verizon
J. Drake
Juniper
J. Rabadan
Nokia

March 9, 2020

EVPN Virtual Ethernet Segment
draft-ietf-bess-evpn-virtual-eth-segment-06

Abstract

EVPN and PBB-EVPN introduce a family of solutions for multipoint Ethernet services over MPLS/IP network with many advanced features among which their multi-homing capabilities. These solutions introduce Single-Active and All-Active for an Ethernet Segment (ES), itself defined as a set of physical links between the multi-homed device/network and a set of PE devices that they are connected to. This document extends the Ethernet Segment concept so that an ES can be associated to a set of EVCs (e.g., VLANs) or other objects such as MPLS Label Switch Paths (LSPs) or Pseudowires (PWs), referred to as Virtual Ethernet Segments (vES). This draft describes the requirements and the extensions needed to support vES in EVPN and PBB-EVPN.

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] and RFC 8174 [RFC8174].

Table of Contents

1. Introduction 3
   1.1. Virtual Ethernet Segments in Access Ethernet Networks . . 3
   1.2. Virtual Ethernet Segments in Access MPLS Networks 5
2. Terminology 6
3. Requirements 7
   3.1. Single-Homed and Multi-Homed vES 7
   3.2. Scalability 8
   3.3. Local Switching 8
   3.4. EVC Service Types 9
   3.5. Designated Forwarder (DF) Election 9
   3.6. OAM 10
   3.7. Failure and Recovery 10
   3.8. Fast Convergence 10
4. Solution Overview 11
   4.1. EVPN DF Election for vES 12
5. Failure Handling and Recovery 13
   5.1. EVC Failure Handling for Single-Active vES in EVPN . . . 15
   5.2. EVC Failure Handling for Single-Active vES in PBB-EVPN . 15
   5.3. Port Failure Handling for Single-Active vESes in EVPN . . 16
   5.4. Port Failure Handling for Single-Active vESes in PBB-EVPN 17
   5.5. Fast Convergence in (PBB-)EVPN 18
6. Acknowledgements 20
7. Security Considerations 20
8. IANA Considerations 20

9. Intellectual Property Considerations 20
10. References 20
10.1. Normative References 20
10.2. Informative References 21
Authors' Addresses 21

1. Introduction

[RFC7432] and [RFC7623] introduce a family of solutions for multipoint Ethernet services over MPLS/IP network with many advanced features among which their multi-homing capabilities. These solutions introduce Single-Active and All-Active for an Ethernet Segment (ES), itself defined as a set of links between the multi-homed device/network and a set of PE devices that they are connected to.

This document extends the Ethernet Segment concept so that an ES can be associated to a set of EVCs (e.g., VLANs) or other objects such as MPLS Label Switch Paths (LSPs) or Pseudowires (PWs), referred to as Virtual Ethernet Segments (vES). This draft describes the requirements and the extensions needed to support vES in EVPN and PBB-EVPN.

1.1. Virtual Ethernet Segments in Access Ethernet Networks

Some Service Providers (SPs) want to extend the concept of the physical links in an ES to Ethernet Virtual Circuits (EVCs) where many of such EVCs (e.g., VLANs) can be aggregated on a single physical External Network-to-Network Interface (ENNI). An ES that consists of a set of EVCs instead of physical links is referred to as a virtual ES (vES). Figure 1 depicts two PE devices (PE1 and PE2) each with an ENNI where a number of vESes are aggregated on - each of which through its associated EVC.

```
                  Carrier
                  Ethernet
    +-----+       Network
    | CE11|EVC1   +---------+
    +-----+   \  |         |       +---+
    Cust. A    \-0=========0--ENNI1|   |
    +-----+      |         |  ENNI1|   |   +-------+   +---+
    | CE12|EVC2--0=========0--ENNI1|PE1|---|       |   |   |
    +-----+      |         |  ENNI1|   |   |       |---|PE3|-
                 |       ==0--ENNI1|   |   |IP/MPLS|   |   | \  +---+
    +-----+      |      /  |       +---+   |Network|   +---+  \-|   |
    | CE22|EVC3--0==== /   |               |       |            |CE4|
    +-----+      |    X    |               |       |   +---+    |   |
    Cust. B      |   / \   |       +---+   |       |   |   |  /-|   |
    +-----+     -0===   ===0--ENNI2|   |   |       |---|PE4|-/  +---+
    | CE3 |EVC4/ |         |  ENNI2|PE2|---|       |   |   |
    |     |EVC5--0=========0--ENNI2|   |   +-------+   +---+
    +-----+      |         |       +---+
    Cust. C      +---------+   /\
           /\                  ||
           ||                  ENNI
           EVCs             Interface
    <--------802.1Q---------->  <---- EVPN Network -----> <-802.1Q->
```

Figure 1: DHD/DHN (both SA/AA) and SH on same ENNI

ENNIs are commonly used to reach off-network / out-of-franchise customer sites via independent Ethernet access networks or third-party Ethernet Access Providers (EAP) (see Figure 1). ENNIs can aggregate traffic from hundreds to thousands of vESes, where each vES is represented by its associated EVC on that ENNI. As a result, ENNIs and their associated EVCs are a key element of SP off-networks that are carefully designed and closely monitored.

In order to meet customers' Service Level Agreements (SLA), SPs build redundancy via multiple EVPN PEs and across multiple ENNIs (as shown in Figure 1) where a given vES can be multi-homed to two or more EVPN PE devices (on two or more ENNIs) via their associated EVCs. Just like physical ES's in [RFC7432] and [RFC7623] solutions, these vESes can be single-homed or multi-homed ES's and when multi-homed, then can operate in either Single-Active or All-Active redundancy modes. In a typical SP off-network scenario, an ENNI can be associated with several thousands of single-homed vESes, several hundreds of Single-Active vESes and it may also be associated with tens or hundreds of All-Active vESes.

1.2. Virtual Ethernet Segments in Access MPLS Networks

Other Service Providers (SPs) want to extend the concept of the physical links in an ES to individual Pseudowires (PWs) or to MPLS Label Switched Paths (LSPs) in Access MPLS networks - i.e., a vES consisting of a set of PWs or a set of LSPs. Figure 2 illustrates this concept.

```
                 MPLS Aggregation
                 Network
   +-----+      +-----------------+
   | CE11|EVC1  |                 |
   +-----+   \+AG1--+  PW1      +-----+
   Cust. A    -0----|===========|     |
   +-----+     | ---+===========|     |   +-------+   +---+
   | CE12|EVC2-0/   |  PW2   /\ | PE1 +---+       |   |   |
   +-----+     ++---+      ==||=|     |   |       +---+PE3+-
                |         //=||=|     |   |IP/MPLS|   |   | \  +---+
                |        //  \/ ++----+   |Network|   +---+  \-+   |
   +-----+EVC3  |    PW3//  LSP1  |       |       |            |CE4|
   | CE13|    +AG2--+===/PW4      |       |       |   +---+    |   |
   +-----+     0    |===     /\ ++----+   |       |   |   |  /-+   |
               0    |==PW5===||=|     |   |       +---+PE4+-/  +---+
   +-----+    /++---+==PW6===||=| PE2 +---+       |   |   |
   | CE14|EVC4  |            \/ |     |   +-------+   +---+
   +-----+      |           LSP2+-----+
   Cust. C      +-----------------+
          /\
          ||
          EVCs
   <--802.1Q---><------MPLS------> <---- EVPN Network ---> <-802.1Q->
```

Figure 2: DHN and SH on Access MPLS networks

In some cases, Service Providers use Access MPLS Networks that belong to separate administrative entities or third parties as a way to get access to the their own IP/MPLS network infrastructure. This is the case illustrated in Figure 2.

In such scenarios, a virtual ES (vES) is defined as a set of individual PWs if they cannot be aggregated into a common LSP. If the aggregation of PWs is possible, the vES can be associated to an LSP in a given PE. In the example of Figure 2, EVC3 is connected to a VPWS instance in AG2 that is connected to PE1 and PE2 via PW3 and

PW5 respectively. EVC4 is connected to a separate VPWS instance on AG2 that gets connected to an EVI on PE1 and PE2 via PW4 and PW6, respectively. Since the PWs for the two VPWS instances can be aggregated into the same LSPs going to the MPLS network, a common virtual ES can be defined for LSP1 and LSP2. This vES will be shared by two separate EVIs in the EVPN network.

In some cases, this aggregation of PWs into common LSPs may not be possible. For instance, if PW3 were terminated into a third PE, e.g. PE3, instead of PE1, the vES would need to be defined on a per individual PW on each PE, i.e. PW3 and PW5 would belong to ES-1, whereas PW4 and PW6 would be associated to ES-2.

For MPLS/IP access networks where a vES represents a set of PWs or LSPs, this document extends Single-Active multi-homing procedures of [RFC7432] and [RFC7623] to vES. The vES extension to All-Active multi- homing is outside of the scope of this document for MPLS/IP access networks.

This draft describes requirements and the extensions needed to support a vES in [RFC7432] and [RFC7623]. Section 3 lists the set of requirements for a vES. Section 4 describes extensions for a vES that are applicable to EVPN solutions including [RFC7432] and [RFC7209]. Furthermore, these extensions meet the requirements described in Section 3. Section 4 gives solution overview and Section 5 describes failure handling, recovery, scalability, and fast convergence of [RFC7432] and [RFC7623] for vESes.

2. Terminology

AC: Attachment Circuit

BEB: Backbone Edge Bridge

B-MAC: Backbone MAC Address

CE: Customer Edge

CFM: Connectivity Fault Management (802.1ag)

C-MAC: Customer/Client MAC Address

DHD: Dual-homed Device

DHN: Dual-homed Network

ENNI: External Network-Network Interface

ES: Ethernet Segment

ESI: Ethernet Segment Identifier

EVC: Ethernet Virtual Circuit

EVPN: Ethernet VPN

I-SID: Service Instance Identifier (24 bits and global within a PBB network see [RFC7080])

LACP: Link Aggregation Control Protocol

PBB: Provider Backbone Bridge

PBB-EVPN: Provider Backbone Bridge EVPN

PE: Provider Edge

SH: Single-Homed

Single-Active Redundancy Mode (SA): 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 (AA): 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.

3. Requirements

This section describes the requirements specific to virtual Ethernet Segment (vES) for (PBB-)EVPN solutions. These requirements are in addition to the ones described in [RFC8214], [RFC7432], and [RFC7623].

3.1. Single-Homed and Multi-Homed vES

A PE needs to support the following types of vESes:

(R1a) A PE MUST handle single-homed vESes on a single physical port (e.g., single ENNI)

(R1b) A PE MUST handle a mix of Single-Homed vESes and Single-Active multi-homed vESes simultaneously on a single physical port (e.g.,

single ENNI). Single-Active multi-homed vESes will be simply referred to as Single-Active vESes through the rest of this document.

(R1c) A PE MAY handle All-Active multi-homed vESes on a single physical port. All-Active multi-homed vESes will be simply referred to as All-Active vESes through the rest of this document.

(R1d) A PE MAY handle a mixed of All-Active vESes along with other types of vESes on a single physical port.

(R1e) A Multi-Homed vES (Single-Active or All-Active) can be spread across two or more ENNIs, on any two or more PEs.

3.2. Scalability

A single physical port (e.g., ENNI) can be associated with many vESes. The following requirements give a quantitative measure for each vES type.

(R2a) A PE SHOULD handle very large number of Single-Homed vESes on a single physical port (e.g., thousands of vESes on a single ENNI).

(R2b) A PE SHOULD handle large number of Single-Active vESes on a single physical port (e.g., hundreds of vESes on a single ENNI).

(R2c) A PE MAY handle large number of All-Active vESes on a single physical port (e.g., hundreds of vESes on a single ENNI).

(R2d) A PE SHOULD handle the above scale for a mix of Single-homed vESes and Single-Active vESes simultaneously on a single physical port (e.g., single ENNI).

(R2e) A PE MAY handle the above sale for a mixed of All-Active vESes along with other types of vESes on a single physical port.

3.3. Local Switching

Many vESes of different types can be aggregated on a single physical port on a PE device and some of these vES can belong to the same service instance (or customer). This translates into the need for supporting local switching among the vESes of the same service instance on the same physical port (e.g., ENNI) of the PE.

(R3a) A PE MUST support local switching among different vESes belonging to the same service instance (or customer) on a single physical port. For example, in Figure 1, PE1 MUST support local switching between CE11 and CE12 (both belonging to customer A) that are mapped to two Single-homed vESes on ENNI1. In case of Single-

Active vESes, the local switching is performed among active EVCs belonging to the same service instance on the same ENNI.

3.4. EVC Service Types

A physical port (e.g., ENNI) of a PE can aggregate many EVCs each of which is associated with a vES. Furthermore, an EVC may carry one or more VLANs. Typically, an EVC carries a single VLAN and thus it is associated with a single broadcast domain. However, there is no restriction on an EVC to carry more than one VLAN.

(R4a) An EVC can be associated with a single broadcast domain - e.g., VLAN-based service or VLAN bundle service.

(R4b) An EVC MAY be associated with several broadcast domains - e.g., VLAN-aware bundle service.

In the same way, a PE can aggregate many LSPs and PWs. In the case of individual PWs per vES, typically a PW is associated with a single broadcast domain, but there is no restriction on the PW to carry more than one VLAN if the PW is of type Raw mode.

(R4c) A PW can be associated with a single broadcast domain - e.g., VLAN-based service or VLAN bundle service.

(R4d) An PW MAY be associated with several broadcast domains - e.g., VLAN-aware bundle service.

3.5. Designated Forwarder (DF) Election

Section 8.5 of [RFC7432] describes the default procedure for DF election in EVPN which is also used in [RFC7623] and [RFC8214]. This default DF election procedure is performed at the granularity of (ESI, Ethernet Tag). In case of a vES, the same EVPN default procedure for DF election also applies; however, at the granularity of (vESI, Ethernet Tag); where vESI is the virtual Ethernet Segment Identifier and the Ethernet Tag field is represented by and I-SID in PBB-EVPN and by a VLAN ID (VID) in EVPN. As in [RFC7432], this defult procedure for DF election at the granularity of (vESI, Ethernet Tag) is also referred to as "service carving". With service carving, it is desireable to evenly partition the DFs for different vES's among different PEs, thus evenly distributing the traffic among different PEs. The following list the requirements apply to DF election of vES's for (PBB-)EVPN.

(R5a) A vES with m EVCs can be distributed among n ENNIs belonging to p PEs in any arbitrary order; where n >= p >= m. For example, if there is an vES with 2 EVCs and there are 5 ENNIs on 5 PEs (PE1

through PE5), then vES can be dual-homed to PE2 and PE4 and the DF election must be performed between PE2 and PE4.

(R5b) Each vES MUST be identified by its own virtual ESI (vESI).

3.6. OAM

In order to detect the failure of an individual EVC and perform DF election for its associated vES as the result of this failure, each EVC should be monitored independently.

(R6a) Each EVC SHOULD be monitored for its health independently.

(R6b) A single EVC failure (among many aggregated on a single physical port/ENNI) MUST trigger DF election for its associated vES.

3.7. Failure and Recovery

(R7a) Failure and failure recovery of an EVC for a Single-homed vES SHALL NOT impact any other EVCs within its service instance or any other service instances. In other words, for PBB-EVPN, it SHALL NOT trigger any MAC flushing both within its own I-SID as well as other I-SIDs.

(R7b) In case of All-Active vES, failure and failure recovery of an EVC for that vES SHALL NOT impact any other EVCs within its service instance or any other service instances. In other words, for PBB-EVPN, it SHALL NOT trigger any MAC flushing both within its own I-SID as well as other I-SIDs.

(R7c) Failure and failure recovery of an EVC for a Single-Active vES SHALL impact only its own service instance. In other words, for PBB-EVPN, MAC flushing SHALL be limited to the associated I-SID only and SHALL NOT impact any other I-SIDs.

(R7d) Failure and failure recovery of an EVC for a Single-Active vES MAY only impact C-MACs associated with MHD/MHNs for that service instance. In other words, MAC flushing SHOULD be limited to single service instance (I-SID in the case of PBB-EVPN) and only CMACs for Single-Active MHD/MHNs.

3.8. Fast Convergence

Since large number of EVCs (and their associated vESes) are aggregated via a single physical port (e.g., ENNI), then the failure of that physical port impacts large number of vESes and triggers large number of ES route withdrawals. Formulating, sending, receiving, and processing such large number of BGP messages can

introduce delay in DF election and convergence time. As such, it is highly desirable to have a mass-withdraw mechanism similar to the one in the [RFC7432] for withdrawing large number of Ethernet A-D routes.

(R8a) There SHOULD be a mechanism equivalent to EVPN mass-withdraw such that upon an ENNI failure, only a single BGP message is needed to indicate to the remote PEs to trigger DF election for all impacted vES associated with that ENNI.

4. Solution Overview

The solutions described in [RFC7432] and [RFC7623] are leveraged as-is with the modification that the ESI assignment is performed for an EVC or a group of EVCs or LSPs/PWs instead of a link or a group of physical links. In other words, the ESI is associated with a virtual ES (vES), hereby referred to as vESI.

For the EVPN solution, everything basically remains the same except for the handling of physical port failure where many vESes can be impacted. Sections 5.1 and 5.3 below describe the handling of physical port/link failure for EVPN. In a typical multi-homed operation, MAC addresses are learned behind a vES and are advertised with the ESI corresponding to the vES (i.e., vESI). EVPN aliasing and mass- withdraw operations are performed with respect to vES. In other words, the Ethernet A-D routes for these operations are advertised with vESI instead of ESI.

For PBB-EVPN solution, the main change is with respect to the BMAC address assignment which is performed similar to what is described in section 7.2.1.1 of [RFC7623] with the following refinements:

- One shared BMAC address SHOULD be used per PE for the single-homed vESes. In other words, a single BMAC is shared for all single-homed vESes on that PE.

- One shared BMAC address SHOULD be used per PE per physical port (e.g., ENNI) for the Single-Active vESes. In other words, a single BMAC is shared for all Single-Active vESes that share the same ENNI.

- One shared BMAC address MAY be used for all Single-Active vESes on that PE.

- One BMAC address SHOULD be used per set of EVCs representing an All-Active vES. In other words, a single BMAC address is used per vES for All-Active scenarios.

o A single BMAC address MAY also be used per vES per PE for Single-Active scenarios.

```
                 BEB   +---------------+  BEB
                  ||   |               |   ||
                  \/   |               |   \/
    +----+ EVC1 +----+ |               | +----+   +----+
    | CE1|------|    | |               | |    |---| CE2|
    +----+\     | PE1| |   IP/MPLS     | | PE3|   +----+
           \    +----+ |   Network     | +----+
            \          |               |
         EVC2\  +----+ |               |
              \ |    | |               |
               \| PE2| |               |
                +----+ |               |
                  /\   +---------------+
                  ||
                  BEB
    <--802.1Q--><---------- PBB-EVPN --------><--802.1Q->
```

Figure 3: PBB-EVPN Network

4.1. EVPN DF Election for vES

The procedure for service carving for virtual Ethernet Segments is the same as the one outlined in section 8.5 of [RFC7432] except for the fact that ES is replaced with vES. For the sake of clarity and completeness, this procedure is repeated below:

1. When a PE discovers the vESI or is configured with the vESI associated with its attached vES, it advertises an Ethernet Segment route with the associated ES-Import extended community attribute.

2. The PE then starts a timer (default value = 3 seconds) to allow the reception of Ethernet Segment routes from other PE nodes connected to the same vES. This timer value MUST be same across all PEs connected to the same vES.

3. When the timer expires, each PE builds an ordered list of the IP addresses of all the PE nodes connected to the vES (including itself), in increasing numeric value. Each IP address in this list is extracted from the "Originator Router's IP address" field of the advertised Ethernet Segment route. Every PE is then given

an ordinal indicating its position in the ordered list, starting with 0 as the ordinal for the PE with the numerically lowest IP address. The ordinals are used to determine which PE node will be the DF for a given EVPN instance on the vES using the following rule: Assuming a redundancy group of N PE nodes, the PE with ordinal i is the DF for an EVPN instance with an associated Ethernet Tag value of V when (V mod N) = i. It should be noted that using "Originator Router's IP address" field in the Ethernet Segment route to get the PE IP address needed for the ordered list, allows for a CE to be multi-homed across different ASes if such need ever arises.

4. The PE that is elected as a DF for a given EVPN instance will unblock traffic for that EVPN instance. Note that the DF PE unblocks all traffic in both ingress and egress directions for Single-Active vES and unblocks multi-destination in egress direction for All-Active Multi-homed vES. All non-DF PEs block all traffic in both ingress and egress directions for Single-Active vES and block multi-destination traffic in the egress direction for All-Active vES.

In the case of an EVC failure, the affected PE withdraws its Virtual Ethernet Segment route if there are no more EVCs associated to the vES in the PE. This will re-trigger the DF Election procedure on all the PEs in the Redundancy Group. For PE node failure, or upon PE commissioning or decommissioning, the PEs re-trigger the DF Election Procedure across all affected vESes. In case of a Single-Active, when a service moves from one PE in the Redundancy Group to another PE as a result of DF re-election, the PE, which ends up being the elected DF for the service, SHOULD trigger a MAC address flush notification towards the associated vES. This can be done, for e.g. using IEEE 802.1ak MVRP 'new' declaration.

For LSP-baesd and PW-based vES, the non-DF PE SHOULD signal PW-status 'standby' to the Aggregation PE (e.g., AG PE in Figure 2), and a new DF PE MAY send an LDP MAC withdraw message as a MAC address flush notification. It should be noted that the PW-status is signaled for the scenarios where there is a one-to-one mapping between EVI/BD and the PW.

5. Failure Handling and Recovery

There are a number of failure scenarios to consider such as:

A: CE uplink port failure

B: Ethernet Access Network failure

C: PE access-facing port or link failure

D: PE node failure

E: PE isolation from IP/MPLS network

[RFC7432], [RFC7623], and [RFC8214] solutions provide protection against such failures as described in the corresponding references. In the presence of virtual Ethernet Segments (vESes) in these solutions, besides the above failure scenarios, EVC failure is an additional scenario to consider. Handling vES failure scenarios implies that individual EVCs or PWs need to be monitored and upon detection of failure or restoration of services, appropriate DF election and failure recovery mechanisms are executed.

[ETH-OAM] is used for monitoring EVCs and upon failure detection of a given EVC, DF election procedure per section [4.1] is executed. For PBB-EVPN, some extensions are needed to handle the failure and recovery procedures of [RFC7623] in order to meet the above requirements. These extensions are described in the next section.

[MPLS-OAM] and [PW-OAM] are used for monitoring the status of LSPs and/or PWs associated to vES.

```
                         B            D
                         ||           ||
                         \/           \/
                       +-----+
          +-----+      |     |       +---+
          | CE1 |EVC2--0=====0--ENNI1|   |   +-------+
          +-----+      |    =0--ENNI1|PE1|---|       |  +---+  +---+
          Cust. A      |   / |       |   |   |IP/MPLS|--|PE3|--|CE4|
          +-----+      |  /  |       +---+   |Network|  |   |  +---+
          |     |EVC2--0==   |               |       |  +---+
          | CE2 |      |     |       +---+   |       |
          |     |EVC3--0=====0--ENNI2|PE2|---|       |
          +-----+      |     |       |   |   +-------+
                       +-----+       +---+
                    /\             /\     /\
                    ||             ||     ||
                    A              C      E
```

Figure 4: Failure Scenarios A,B,C,D and E

5.1. EVC Failure Handling for Single-Active vES in EVPN

In RFC7432, when a DF PE connected to a Single-Active multi-homed Ethernet Segment loses connectivity to the segment, due to link or port failure, it signals to the remote PEs to withdraw all MAC addresses associated with that Ethernet Segment. This is done by advertising a mass-withdraw message using Ethernet A-D per-ES route. It should be noted that for dual-homing use cases where there is only a single backup path, MAC withdraw can be avoided by the remote PEs as they can simply update their nexthop associated with the affected MAC entries to the backup path per procedure described in section 8.2 of [RFC7432].

In case of an EVC failure which impacts a single vES, the exact same EVPN procedure is used. In this case, the message using Ethernet A-D per-vES route carries the vESI representing the vES which in turn is associated with the failed EVC. The remote PEs upon receiving this message perform the same procedures outlined in section 8.2 of [RFC7432].

5.2. EVC Failure Handling for Single-Active vES in PBB-EVPN

In [RFC7432], when a PE connected to a Single-Active Ethernet Segment loses connectivity to the segment, due to link or port failure, it signals the remote PE to flush all CMAC addresses associated with that Ethernet Segment. This is done by advertising a BMAC route along with MAC Mobility Extended community.

In case of an EVC failure that impacts a single vES, if the above PBB-EVPN procedure is used, it results in excessive CMAC flushing because a single physical port can support large number of EVCs (and their associated vESes) and thus advertising a BMAC corresponding to the physical port with MAC mobility Extended community will result in flushing CMAC addresses not just for the impacted EVC but for all other EVCs on that port.

In order to reduce the scope of CMAC flushing to only the impacted service instances (the service instance(s) impacted by the EVC failure), the PBB-EVPN CMAC flushing needs to be adapted on a per service instance basis (i.e., per I-SID). [I-D.ietf-bess-pbb-evpn-isid-cmacflush] introduces BMAC/I-SID route where existing PBB-EVPN BMAC route is modified to carry an I-SID in the "Ethernet Tag ID" field instead of NULL value. This field indicates to the receiving PE, to flush all CMAC addresses associated with that I-SID for that BMAC. This CMAC flushing mechanism per I-SID SHOULD be used in case of EVC failure impacting a vES. Since typically an EVC maps to a single broadcast domain and thus a single service instance, the affected PE only needs to advertise a single

BMAC/I-SID route. However, if the failed EVC carries multiple VLANs each with its own broadcast domain, then the affected PE needs to advertise multiple BMAC/I-SID routes - one for each VLAN (broadcast domain) - i.e., one for each I-SID. Each BMAC/I-SID route basically instructs the remote PEs to perform flushing for CMACs corresponding to the advertised BMAC only for the advertised I-SID.

The CMAC flushing based on BMAC/I-SID route works fine when there are only a few VLANs (e.g., I-SIDs) per EVC. However if the number of I-SIDs associated with a failed EVC is large, then it is recommended to assign a BMAC per vES and upon EVC failure, the affected PE simply advertise BMAC withdraw message to other PEs.

5.3. Port Failure Handling for Single-Active vESes in EVPN

When a large number of EVCs are aggregated via a single physical port on a PE; where each EVC corresponds to a vES, then the port failure impacts all the associated EVCs and their corresponding vESes. If the number of EVCs corresponding to the Single-Active vESes for that physical port is in thousands, then thousands of service instances are impacted. Therefore, the BGP flush message need to be inclusive of all these impacted service instances. In order to achieve this, the following extensions are added to the baseline EVPN mechanism:

1. When a PE advertises an Ethernet A-D per-ES route for a given vES, it colors it with the MAC address of the physical port which is associated with that vES using EVPN Router's MAC Extended Community per [EVPN-IRB]. The receiving PEs take note of this color and create a list of vESes for this color.

2. Upon a port failure (e.g., ENNI failure), the PE advertise a special mass-withdraw message with the MAC address of the failed port (i.e., the color of the port) encoded in the ESI field. For this encoding, type 3 ESI (RFC7432 section 5) is used with the MAC field set to the MAC address of the port and the 3-octet local discriminator field set to 0xFFFFFF. This mass-withdraw route is advertised with a list of Route Targets corresponding to the impacted service instances. If the number of Route Targets is more than can fit into a single attribute, then a set of Ethernet A-D per ES routes are advertised.

3. Upon a port failure (e.g., ENNI failure), the PE advertise a special mass-withdraw message with the MAC address of the failed port.

4. The remote PEs upon receiving this message, based on ESI Type 3 and 0xFFFFFF Local Discrimnator values, detect the special vES mass-withdraw message. The remote PEs then access the list of

the vES's for the specified color created in (1) and initialte locally mass-withdraw procedures for each of the vES's in the list.

In scenarios where a logical ENNI is used the above procedure equally applies. The logical ENNI is represented by a Type 3 ESI and the MAC address used in the ENNI's ESI is used as a color for vESes as described above.

5.4. Port Failure Handling for Single-Active vESes in PBB-EVPN

When a large number of EVCs are aggregated via a single physical port on a PE, where each EVC corresponds to a vES, then the port failure impacts all the associated EVCs and their corresponding vESes. If the number of EVCs corresponding to the Single-Active vESes for that physical port is in thousands, then thousands of service instances (I-SIDs) are impacted. In such failure scenarios, the following two MAC flushing mechanisms per [RFC7623] can be performed.

1. If the MAC address of the physical port is used for PBB encapsulation as BMAC SA, then upon the port failure, the PE MUST use the EVPN MAC route withdrawal message to signal the flush.

2. If the PE shared MAC address is used for PBB encapsulation as BMAC SA, then upon the port failure, the PE MUST re-advertise this MAC route with the MAC Mobility Extended Community to signal the flush.

The first method is recommended because it reduces the scope of flushing the most.

If there are large number of service instances (i.e., I-SIDs) associated with each EVC, and if there is a BMAC assigned per vES as recommended in the above section, then in order to handle port failure efficiently, each vES MAY be color with another MAC representing the physical port similar to the coloring mechanism for EVPN. In other words, each BMAC representing a vES is advertised with the EVPN Router's MAC Extended Community carrying the MAC address of the physical port.The difference between coloring mechanism for EVPN and PBB-EVPN is that for EVPN, the extended community is advertised with the Ethernet A-D per ES route; whereas, for PBB-EVPN, the extended community is advertised with the BMAC route. As noted above, the advertisement of the extended community along with BMAC route for coloring purpoes is optional and only recommended when there are many vESes per physical port and each vES is associated with very large number of service instances (i.e., large numbe of I-SIDs).

When coloring mechanism is used, the receiving PEs take note of the color being advertised along with the BMAC route and for each such color, they create a list of vESes associated with this color (i.e., associated with this MAC address). Now, when a port failure occurs, the impacted PE needs to notify the other PEs of this color so that these PEs can identify all the impacted vESes associated with this color (from the above list) and flush CMACs associated with the failed physical port. This is accomplished by withdrawing the MAC route associated with the failed port.

5.5. Fast Convergence in (PBB-)EVPN

As described above, when a large number of EVCs are aggregated via a physical port on a PE, and where each EVC corresponds to a vES, then the port failure impacts all the associated EVCs and their corresponding vESes. Two actions must be taken as the result of such port failure:

- o For EVPN initiate mass-withdraw procedure for all vESes associated with the failed port and for PBB-EVPN flush all CMACs associated with the failed port across all vESes and the impacted I-SIDs

- o DF election for all impacted vESes associated with the failed port

Section 5.3 already describes how perform mass-withdraw for all affected vESes using a single BGP advertisment. Section 5.4 describes how to only flush CMAC address associated with the failed physical port (e.g., optimum CMAC flushing). This section describes how to perform DF election in the most optimum way - e.g., to trigger DF election for all impacted vESes (which can be very large) among the participating PEs via a single BGP message as opposed to sending large number of BGP messages - one per vES. This section assumes that the MAC flushing mechanism described in section 5.4, bullet (1) is used.

```
                     +-----+
          +----+     |     |       +---+
          | CE1|AC1--0=====0--ENNI1|   |  +-------+
          |    |AC2--0     |       |PE1|--|       |
          +----+     |\  ==0--ENNI2|   |  |       |
                     | \/  |       +---+  |       |
                     | /\  |              |IP/MPLS|
          +----+     |/  \ |       +---+  |Network|   +---+  +---+
          | CE2|AC4--0    =0--ENNI3|   |  |       |---|PE4|--|CE4|
          |    |AC4--0=====0--ENNI3|PE2|--|       |   +---+  +---+
          +----+     | ====0--ENNI3|   |  |       |
                     |/    |       +---+  |       |
                     0     |              |       |
          +----+    /|     |       +---+  |       |
          | CE3|AC5- |     |       |PE3|--|       |
          |    |AC6--0=====0--ENNI4|   |  +-------+
          +----+     |     |       +---+
                     +-----+
```

Figure 5: Fast Convergence Upon ENNI Failure

The following describes the procedure for coloring vESes and fast convergence for DF election using this color:

1. When a vES is configured, the PE colors the vES with the MAC address of the corresponding physical port and advertises the Ethernet Segment route for this vES with this color.

2. All other PEs (in the redundancy group) take note of this color and add the vES to the list for this color.

3. Upon the occurrence of a port failure (e.g., an ENNI failure), the PE withdraw the previously advertised MAC address associated with the failed port. The PE should prioritize sending this MAC address withdraw message over vES route withdrawal messages of impacted vESes.

4. On reception of this MAC withdraw message, other PEs in the redundancy group use this info to initiate DF election procedures across all their affected vESes.

5. The PE with the physical port failure (ENNI failure), also sends vES route withdrawal for every impacted vESes. The other PEs upon receiving these messages, clear up their BGP tables. It should be noted the vES route withdrawal messages are not used for executing DF election procedures by the receiving PEs.

6. Acknowledgements

The authors would like to thanks Mei Zhang, Jose Liste, and Luc Andre Burdet for their reviews and feedbacks of this document.

7. Security Considerations

All the security considerations in [RFC7432] and [RFC7623] apply directly to this document because this document leverages the control and data plane procedures described in those documents.

This document does not introduce any new security considerations beyond that of [RFC7432] and [RFC7623] because advertisements and processing of Ethernet Segment route for vES in this document follows that of physical ES in those RFCs.

8. IANA Considerations

IANA has allocated sub-type value 7 in the "EVPN Extended Community Sub-Types" registry defined in "https://www.iana.org/assignments/bgp-extended-communities/bgp-extended-communities.xhtml#evpn" as follows:

SUB-TYPE	NAME	Reference
0x07	I-SID Ext Comm	[draft-ietf-bess-evpn-virtual-eth-segment]

It is requested from IANA to update the reference to this document.

9. Intellectual Property Considerations

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

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://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, <https://www.rfc-editor.org/info/rfc7432>.

[RFC7623] Sajassi, A., Ed., Salam, S., Bitar, N., Isaac, A., and W. Henderickx, "Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)", RFC 7623, DOI 10.17487/RFC7623, September 2015, <https://www.rfc-editor.org/info/rfc7623>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8214] Boutros, S., Sajassi, A., Salam, S., Drake, J., and J. Rabadan, "Virtual Private Wire Service Support in Ethernet VPN", RFC 8214, DOI 10.17487/RFC8214, August 2017, <https://www.rfc-editor.org/info/rfc8214>.

10.2. Informative References

[I-D.ietf-bess-pbb-evpn-isid-cmacflush] Rabadan, J., Sathappan, S., Nagaraj, K., Miyake, M., and T. Matsuda, "PBB-EVPN ISID-based CMAC-Flush", draft-ietf-bess-pbb-evpn-isid-cmacflush-00 (work in progress), October 2019.

[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, <https://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, <https://www.rfc-editor.org/info/rfc7209>.

Authors' Addresses

Ali Sajassi
Cisco Systems
MILPITAS, CALIFORNIA 95035
UNITED STATES

Email: sajassi@cisco.com

Patrice Brissette
Cisco Systems

Email: pbrisset@cisco.com

Rick Schell
Verizon

Email: richard.schell@verizon.com

John E Drake
Juniper

Email: jdrake@juniper.net

Jorge Rabadan
Nokia

Email: jorge.rabadan@nokia.com