

Internet Working Group
Internet Draft
Category: Standards Track

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

Expires: August 26, 2016

February 26, 2018

EVPN Virtual Ethernet Segment
draft-sajassi-bess-evpn-virtual-eth-segment-03

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/lid-abstracts.html>

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

Copyright and License Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

EVPN and PBB-EVPN introduce a family of solutions for multipoint Ethernet services over MPLS/IP network with many advanced capabilities among which their multi-homing capabilities. These solutions define two types of multi-homing for an Ethernet Segment (ES): 1) Single-Active and 2) All-Active, where an Ethernet Segment is defined as a set of links between the multi-homed device/network and the set of PE devices that they are connected to.

Some Service Providers want to extend the concept of the physical links in an ES to Ethernet Virtual Circuits (EVCs) where many of such EVCs 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). This draft describes the requirements and the extensions needed to support vES in EVPN and 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	4
1.1 Virtual Ethernet Segments in Access Ethernet Networks . . .	4
1.2 Virtual Ethernet Segments in Access MPLS Networks	5
2. Terminology	7
3. Requirements	8
3.1. Single-Homed & Multi-Homed Virtual Ethernet Segments . . .	8
3.2. Scalability	8
3.3. Local Switching	9
3.4. EVC Service Types	9
3.5. Designated Forwarder (DF) Election	10
3.6. OAM	10
3.7. Failure & Recovery	10
3.8. Fast Convergence	11
4. Solution Overview	11
4.1. EVPN DF Election for vES	12
5. Failure Handling & Recovery	14

5.1. 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 vES's in EVPN . . .	16
5.4. Port Failure Handling for Single-Active vES's in PBB-EVPN .	17
5.5. Fast Convergence in PBB-EVPN	18
6. BGP Encoding	20
6.1. I-SID Extended Community	20
7. Acknowledgements	20
8. Security Considerations	21
9. IANA Considerations	21
10. Intellectual Property Considerations	21
11. Normative References	21
12. Informative References	21
13. Authors' Addresses	21

1. Introduction

[EVPN] and [PBB-EVPN] introduce a family of solutions for multipoint Ethernet services over MPLS/IP network with many advanced capabilities among which their multi-homing capabilities. These solutions define two types of multi-homing for an Ethernet Segment (ES): 1) Single-Active and 2) All-Active, where an Ethernet Segment is defined as a set of links between the multi-homed device/network and the 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 or other objects such as MPLS Label Switch Paths (LSP) or Pseudowires (PW).

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 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 below depicts two PE devices (PE1 and PE2) each with an ENNI where a number of vES's are aggregated on - each of which through its associated EVC.

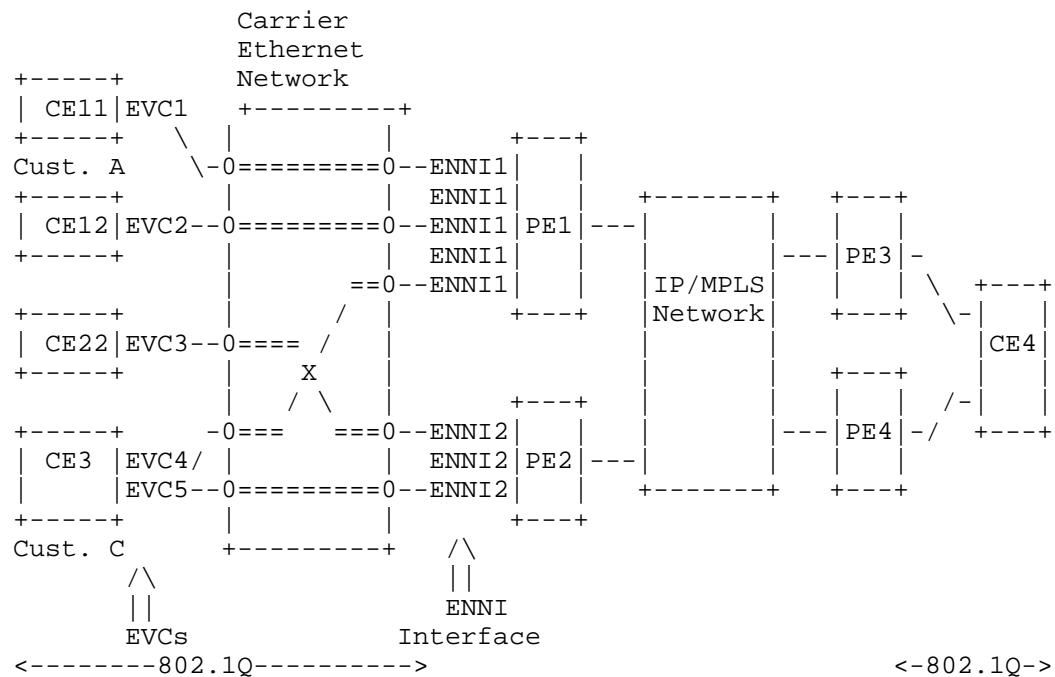


Figure 1: DHD/DHN (both SA/AA) and SH on same ENNI

E-NNIs 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 above figure). E-NNIs can aggregate traffic from hundreds to thousands of vES's; 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 customer's Service Level Agreements (SLA), SPs build redundancy via multiple E-PEs / ENNIs (as shown in figure above) where a given vES can be multi-homed to two or more PE devices (on two or more ENNIs) via their associated EVCs. Just like physical ES's in [EVPN] and [PBB-EVPN] solutions, these vES's 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 vES's, several hundreds of Single-Active vES's and it may also be associated with tens or hundreds of All-Active vES's.

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 (PW) or to MPLS Label Switched Paths (LSPs) per [EVPN-VPWS] in Access MPLS networks. Figure 2 illustrates this concept.

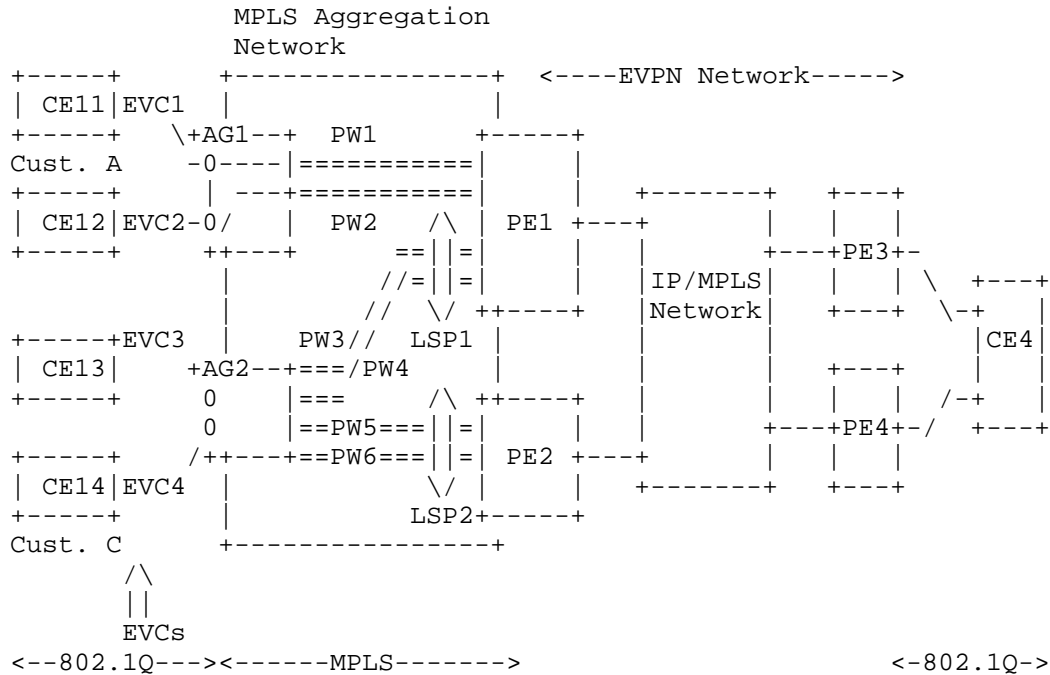


Figure 2: DHN and SH on Access MPLS networks

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.

An ES 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 ES 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 EVPN network, a common virtual ES can be defined for LSP1 and LSP2. This ES 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 ES 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.

An ES that consists of a set of LSPs or individual PWs is also referred as virtual ES (vES) in this document."

This draft describes requirements and the extensions needed to support vES in [EVPN] and [PBB-EVPN]. Section 3 lists the set of requirements for Virtual ES's. Section 4 describes the solution for [PBB-EVPN] to meet these requirements. Section 5 describes the failure handling and recovery for Virtual ES's in [PBB-EVPN]. Section 6 covers scalability and fast convergence required for Virtual ES's in [PBB-EVPN].

2. Terminology

AC: Attachment Circuit
BEB: Backbone Edge Bridge
B-MAC: Backbone MAC Address
CE: Customer Edge
CFM: Connectivity Fault Management
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
LACP: Link Aggregation Control Protocol
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 [EVPN-REQ], [EVPN], and [PBB-EVPN].

3.1. Single-Homed & Multi-Homed Virtual Ethernet Segments

A PE needs to support the following types of vES's:

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

(R1b) A PE MUST handle a mix of Single-Homed vES's and Single-Active multi-homed vES's simultaneously on a single physical port (e.g., single ENNI). Single-Active multi-homed vES's will be simply referred to as Single-Active vES's through the rest of this document.

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

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

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

3.2. Scalability

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

(R2a) A PE MUST handle thousands or tens of thousands of Single-homed vES's on a single physical port (e.g., single ENNI)

(R2b) A PE MUST handle hundreds of Single-Active vES's on a single physical port (e.g., single ENNI)

(R2c) A PE MAY handle tens or hundreds of All-Active Multi-Homed

vES's on a single physical port (e.g., single ENNI)

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

(R4e) A PE MAY handle the above scale for a mixed of All-Active Multi-Homed vES's along with other types of vES's on a single physical port

3.3. Local Switching

Many vES's 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 vES's 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 vES's belonging to the same service instance (or customer) on a single physical port. For example, in the above figure (1), PE1 MUST support local switching between CE11 and CE12 (both belonging to customer A) that are mapped to two Single-homed vES's on ENNI1.

In case of Single-Active vES's, 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 VLANs.

(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 aggregated 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 defined as vc-type VLAN.

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

(R4b) 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 [EVPN] describes the default procedure for DF election in EVPN which is also used in [PBB-EVPN]. This default DF election procedure is performed at the granularity of <ESI, EVI>. In case of a vES, the same EVPN default procedure for DF election also applies; however, at the granularity of <vESI, EVI>; where vESI is the virtual Ethernet Segment Identifier. As in [EVPN], this default procedure for DF election at the granularity of <vESI, EVI> is also referred to as "service carving"; where, EVI is represented by an I-SID in PBB-EVPN and by a EVI service-id/vpn-id in EVPN. With service carving, it is possible to evenly distribute the DFs for different vES's among different PEs, thus distributing the traffic among different PEs. The following list the requirements apply to DF election of vES's for EVPN.

(R5a) A vES with m EVCs can be distributed among n ENNIs belonging to p PEs in any arbitrary order; where $n \geq p \geq m$. 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 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 & Recovery

(R7a) Failure and failure recovery of an EVC for a Single-homed vES SHALL NOT impact any other EVCs for its own 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 Multi-Homed vES, failure and failure

recovery of an EVC for that vES SHALL NOT impact any other EVCs for its own 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 & failure recovery of an EVC for a Single-Active vES SHALL only impact 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 & 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 vES's) are aggregated via a single physical port (e.g., ENNI), then the failure of that physical port impacts large number of vES's 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 [EVPN] 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 [EVPN] and [PBB-EVPN] are leveraged as is with one simple modification and that is the ESI assignment is performed for a group of EVCs instead of a group of links. In other words, the ESI is associated with a virtual ES (vES) and that's why it will be referred to as vESI.

For EVPN solution, everything basically remains the same except for the handling of physical port failure where many vES's can be impacted. Section 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 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 [PBB-EVPN] with the following refinements:

- One shared BMAC address is used per PE for the single-homed vES's. In other words, a single BMAC is shared for all single-homed vES's on that PE.
- One shared BMAC address should be used per PE per physical port (e.g., ENNI) for the Single-Active vES's. In other words, a single BMAC is shared for all Single-Active vES's that shared the same ENNI.
- One shared BMAC address can be used for all Single-Active vES's on that PE.
- One BMAC address is used per EVC per physical port per PE for each All-Active multi-homed vES. In other words, a single BMAC address is used per vES for All-Active multi-homing scenarios.
- A single BMAC address may also be used per vES per PE for Single-Active multi-homing scenarios.

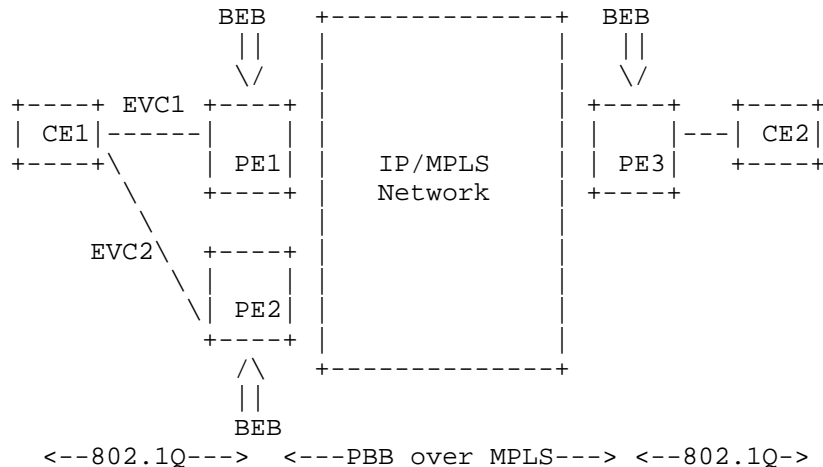


Figure 2: PBB-EVPN Network

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 [EVPN] 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 ESI or is configured with the ESI 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 EVI ID value of V when $(V \bmod N) = i$.

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 multi-homed vES.

In the case of an EVC failure, the affected PE withdraws its Ethernet Segment route. This will re-trigger the service carving procedures on all the PEs in the RG. For PE node failure, or upon PE commissioning or decommissioning, the PEs re-trigger the service carving across all affected vES's. In case of a Single-Active multi-homing, when a service moves from one PE in the RG to another PE as a result of re-carving, 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 and PW based vES, the non-DF PE SHOULD signal PW-status 'standby' signaling to the AG PE, and the new DF MAY send an LDP MAC withdraw message as a MAC address flush notification.

5. Failure Handling & 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

[EVPN] and [PBB-EVPN] solutions provide protection against such failures as described in the corresponding references. In the presence of virtual Ethernet Segments (vES's) in these solutions, besides the above failure scenarios, there is one more scenario to consider and that is EVC failure. This implies that individual EVCs need to be monitored and upon their failure detection, appropriate DF election procedures and failure recovery mechanism need to be 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 addition extensions are needed to failure handling and recovery procedures of [PBB-EVPN] in order to meet the above requirements. These extensions are describe in the next section.

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

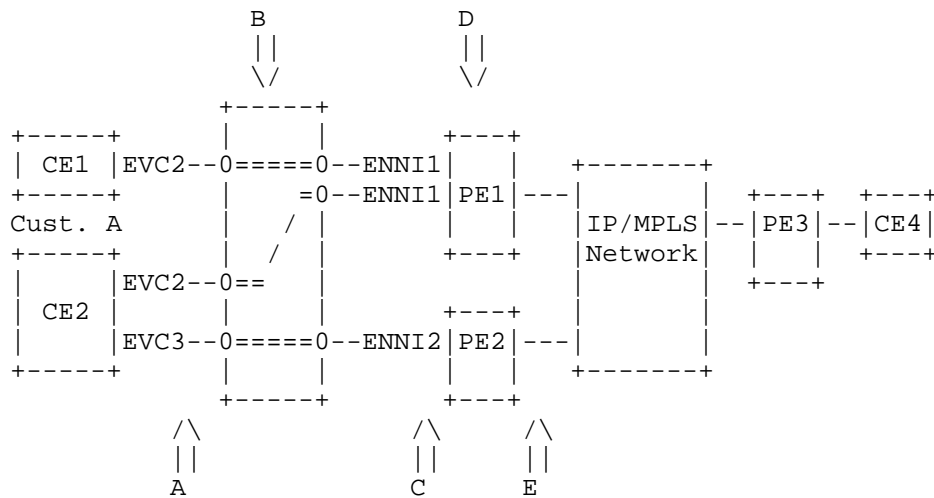


Figure 3: Failure Scenarios A,B,C,D and E

5.1. Failure Handling for Single-Active vES in EVPN

When a PE connected to a Single-Active multi-homed 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 mass-withdraw message using Ethernet A-D per-ES route. To be precise, there is no MAC flush per-se if there is only one backup PE for a given ES - i.e., only an update of the forwarding entries per backup-path procedure in [RFC 7432].

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

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

When a PE connected to a Single-Active multi-homed 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 vES's) 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 BGP flush message is sent along with a list of impacted I-SID(s) represented by the new EVPN I-SID Extended Community as defined in section 6. Since typically an EVC maps to a single broadcast domain and thus a single service instance, the list only contains a single I-SID. However, if the failed EVC carries multiple VLANs each with its own broadcast domain, then the list contains several I-SIDs - one for each broadcast domain. This new BGP flush message basically instructs the remote PE to perform flushing for CMACs corresponding to the advertised BMAC only across the advertised list of I-ISIDs (which is typically one).

The above BMAC route that is advertised with the MAC Mobility Extended Community, can either represent the MAC address of the physical port that the failed EVC is associated with, or it can represent the MAC address of the PE. In the latter case, this is the dedicated MAC address used for all Single-Active vES's on that PE. The former one performs better than the latter one in terms of reducing the scope of flushing as described below and thus it is the recommended approach.

Advertising the BMAC route that represent the physical port (e.g., ENNI) on which the failed EVC reside along with MAC Mobility and I-SID extended communities provide the most optimum mechanism for CMAC flushing upon EVC failure in PBB-EVPN for Single-Active vES because:

- 1) Only CMAC addresses for the impacted service instances are flushed.
- 2) Only a subset of CMAC addresses for the impacted service instances are flushed - only the ones that are learned over the BMAC associated with the failed EVC. In other words, only a small fraction of the CMACs for the impacted service instance(s) are flushed.

5.3. Port Failure Handling for Single-Active vES's 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 vES's. If the

number of EVCs corresponding to the Single-Active vES's 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) A PE when advertises an Ether-AD per ES route for a given vES, it colors it with the MAC address of the physical port which is associated with that vES. The receiving PEs take note of this color and create a list of vES's 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 is used with the MAC field set to the MAC address of the port and the 3-octet local discriminator field set to 0xFFFFFFFF. 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 they can fit into a single attribute, then a set of Ethernet A-D per ESroutes are advertised. The remote PEs upon receiving this message, realize that this is a special mass-withdraw message and they access the list of the vES's for the specified color. Next, they initiate mass-withdraw procedure for each of the vES's in the list.

5.4. Port Failure Handling for Single-Active vES's 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 vES's. If the number of EVCs corresponding to the Single-Active vES's for that physical port is in thousands, then thousands of service instances (I-SIDs) are impacted. Therefore, the BGP flush message need to be sent with a list of thousands of I-SIDs. The new I-SID Extended Community provides a way to encode upto 24 I-SIDs in each Extended Community if the impacted I-SIDs are sequential (the base I-SID value plus the next 23 I-SID values). So, the packing efficiency can range from 1 to 24 and there can be up to 400 such Extended Community sent along with a BGP flush message for a total of 400 to 9600 I-SIDs. If the number of I-SIDs is large enough to not fit in a single Attribute, then either a number of BGP flush messages (with different RDs) can be transmitted or a single BGP flush message without the I-SID list can be transmitted. If the BGP flush message is transmitted without the I-SID list, then it instructs the receiving PEs to flush CMACs associated with that BMAC across all I-SIDs. For simplicity, we opt for the latter option in this document. In other words, if the number of impacted I-SIDs exceed that of a single BGP flush message,

then the flush message is sent without the I-SID list.

As also described in [PBB-EVPN], there are two ways to signal flush message upon a physical port failure:

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.

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; where each EVC corresponds to a vES, then the port failure impacts all the associated EVCs and their corresponding vES's. Two actions must be taken as the result of such port failure:

- Flushing of all CMACs associated with the BMAC of the failed port for the impacted I-SIDs
- DF election for all impacted vES's associated with the failed port

Section 5.4 describes how to flush CMAC address in the most optimum way - e.g., to flush least number of CMAC addresses for the impacted I-SIDs. This section describes how to perform DF election in the most optimum way - e.g., to trigger DF election for all impacted vES's (which can be in thousands) among the participating PEs via a single BGP message as opposed to sending thousands of BGP messages - one per vES.

In order to devise such fast convergence mechanism that can be triggered via a single BGP message, all vES's associated with a given physical port (e.g., ENNI) are colored with the same color representing that physical port. The MAC address of the physical port is used for this coloring purposes and when the PE advertises an ES route for a vES associated with that physical port, it advertises it with an EVPN MAC Extended Community indicating the color of that port.

The receiving PEs take note of this color and for each such color,

they create a list of vES's associated with this color (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 vES's associated with that color (from the above list) and re-execute DF election procedures for all the impacted vES's.

In PBB-EVPN, there are two ways to convey this color to other PEs upon a port failure - one corresponding to each method for signaling flush message as described in section 5.4. If for PBB encapsulation, the MAC address of the physical port is used as BMAC SA, then upon the port failure, the PE sends MAC withdrawal message with the MAC address of the failed port as the color. However, if for PBB encapsulation, the shared MAC address of the PE (dedicated for all Single-Active vES's) is used as BMAC SA, then upon the port failure, the PE re-advertises the MAC route (that carries the shared BMAC) along with this new EVPN MAC Extended Community to indicate the color along with MAC Mobility Extended Community.

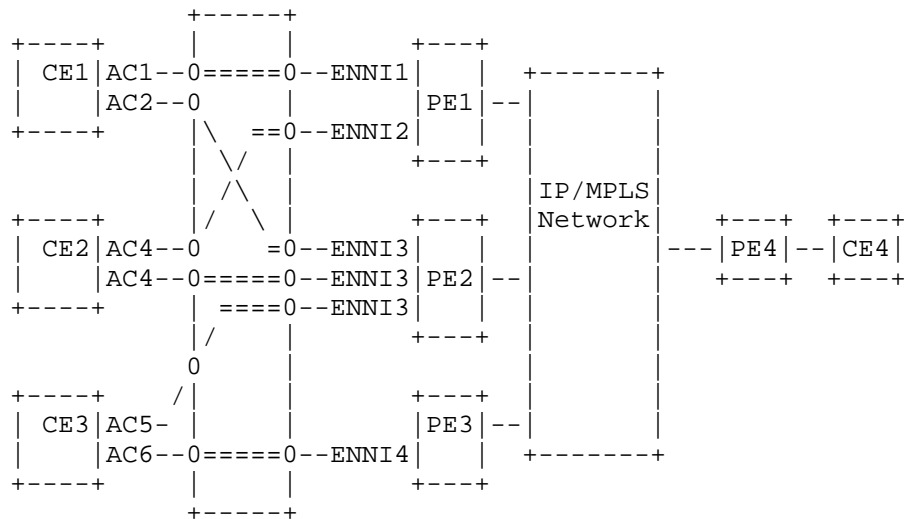


Figure 4: Fast Convergence Upon ENNI Failure

The following describes the procedure for coloring vES's and fast convergence using this color in more details:

- 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 sends the flush message in one of the two ways described above indicating this color.

4- On reception of the flush message, other PEs use this info to flush their impacted CMACs and to initiate DF election procedures across all their affected vES's.

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

6. BGP Encoding

This document defines one new BGP Extended Community for EVPN.

6.1. I-SID Extended Community

A new EVPN BGP Extended Community called I-SID is introduced. This new extended community is a transitive extended community with the Type field of 0x06 (EVPN) and the Sub-Type of 0x04.

The I-SID Extended Community is encoded as an 8-octet value as follows:

0	1	2	3
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+			
Type=0x06		Sub-Type=0x03	
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+		Base I-SID	
Cont.		Bit Map (24 bits)	
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+			

This extended community is used to indicate the list of I-SIDs associated with a given Ethernet Segment.

24-bit map represents the next 24 I-SID after the base I-SID. For example based I-SID of 10025 with 24-bit map of zero means, only a single I-SID of 10025. I-SID of 10025 with bit map of 0x000001 means there are two I-SIDs, 10025 and 10026.

7. Acknowledgements

TBD

8. Security Considerations This document does not introduce any additional security constraints.

9. IANA Considerations

TBD

10. Intellectual Property Considerations

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

11. Normative References

[PBB] Clauses 25 and 26 of "IEEE Standard for Local and metropolitan area networks - Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks", IEEE Std 802.1Q, 2013.

12. Informative References

[RFC7209] Sajassi, et al., "Requirements for Ethernet VPN (EVPN)", RFC7209, May 2014.

[EVPN] Sajassi, et al., "BGP MPLS Based Ethernet VPN", draft-ietf-l2vpn-evpn-07.txt, work in progress, May 7, 2014.

[PBB-EVPN] Sajassi, et al., "PBB-EVPN", draft-ietf-l2vpn-pbb-evpn-07.txt, work in progress, June 18, 2014.

13. Authors' Addresses

Ali Sajassi
Cisco Systems
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

Tapraj Singh
Juniper
Email: tsingh@juniper.net

Jorge Rabadan
ALU
Email: jorge.rabadan@alcatel-lucent.com