Workgroup: BESS WorkGroup
Internet-Draft:
draft-ietf-bess-evpn-virtual-eth-segment-14
Published: 23 September 2023
Intended Status: Standards Track
Expires: 26 March 2024
Authors: A. Sajassi (Cisco Systems)
P. Brissette (Cisco Systems)
R. Schell (Verizon)
J. Drake (Juniper)
J. Rabadan (Nokia)

# EVPN Virtual Ethernet Segment

## Abstract

Etheret VPN (EVPN) and Provider Backbone EVPN (PBB-EVPN) introduce a family of solutions for Ethernet services over MPLS/IP network with many advanced features including multi-homing capabilities. These solutions introduce Single-Active and All-Active redundancy modes 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 Ethernet Virtual Circuits (EVCs e.g., VLANs) or other objects such as MPLS Label Switch Paths (LSPs) or Pseudowires (PWs). Such an ES is 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 [RFC2119] and [RFC8174].

## Status of This Memo

This Internet-Draft will expire on 26 March 2024.

## Copyright Notice



## Table of Contents

## 1. Introduction

Ethernet VPN (EVPN, [RFC7432]) and Provider Backbone EVPN (PBB-EVPN, [RFC7623]) introduce a family of solutions for Ethernet services over MPLS/IP network with many advanced features including multi-homing capabilities. These solutions introduce Single-Active and All-Active redundancy modes 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.

An Ethernet Segment, as defined in [RFC7432], represents a set of Ethernet links connecting customer site to one or more PE. This document extends the Ethernet Segment concept so that an ES can be associated to a set of Ethernet Virtual Circuits (EVCs e.g., VLANs) or other objects such as MPLS Label Switch Paths (LSPs) or Pseudowires (PWs). Such an ES is 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 that aggregates several EVCs. Some of the EVCs on a given ENNI can be associated with vESes. For example, the multi-homed vES in Figure-1 consists of EVC4 on ENNI1 and EVC5 on ENNI2.

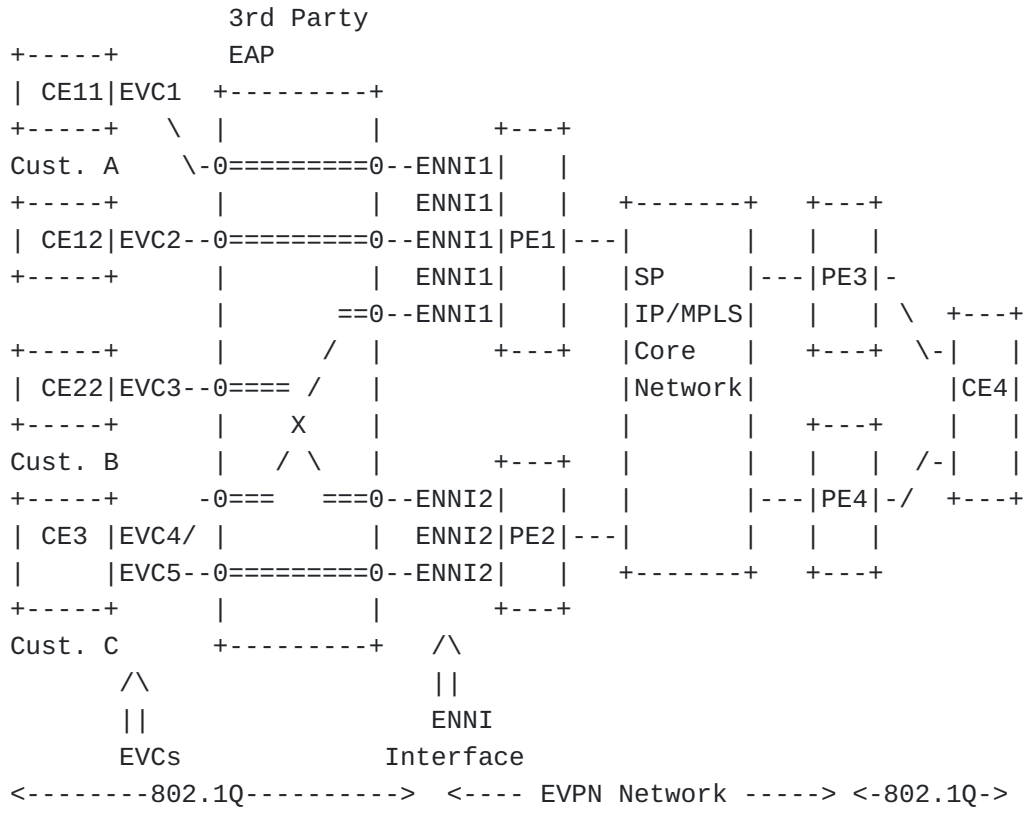


Figure 1: Dual-homed Device/Network (both SA/AA) and SH on same ENNI

ENNIs are commonly used to reach remote customer sites via independent Ethernet access networks or third- party Ethernet Access Providers (EAP). ENNIs can aggregate traffic from many vESes (e.g., hundreds to thousands), 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 external boundaries that are carefully designed and closely monitored. As a reminder, the ENNI is the demarcation between the SP (IP/MPLS Core Network) and the third-party Ethernet Access Provider.

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 ESs in [RFC7432] and [RFC7623] solutions, these vESes can be single-homed or multi-homed ESs and when multi-homed, then can operate in either Single-Active or All-Active redundancy modes. In a typical SP external-boundary scenario (e.g., with an EAP), 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. Specific numbers (hundreds, thousands, etc.) being used through this document

are used to describe the relation of various elements between them at time of writing.

### 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.

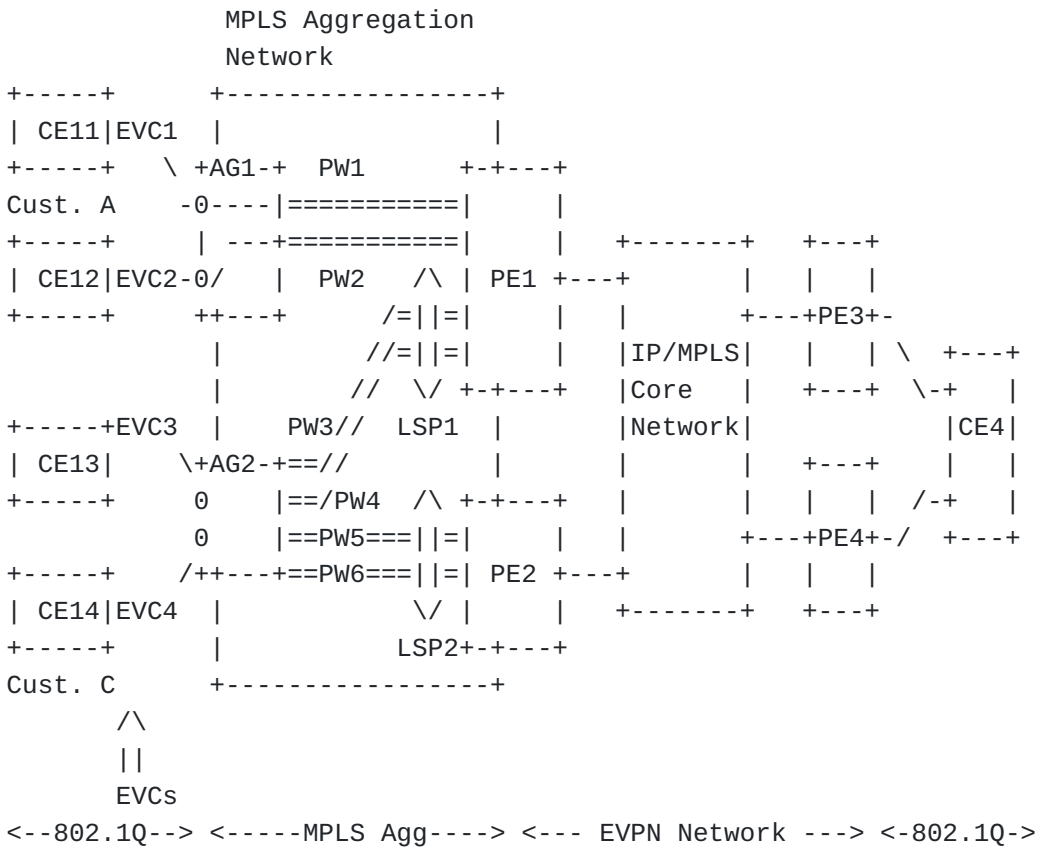


Figure 2: Dual-Homed and Single-homed Network on MPLS Aggregation networks

In some cases, Service Providers use MPLS Aggregation Networks that belong to separate administrative entities or third parties to get access to their own IP/MPLS Core 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. If the aggregation of PWs is possible, the vES can be associated to a group of PWs that

share the same unidirectional LSP pair (by LSP pair we mean the ingress and egress LSPs between the same endpoints).

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 another VPWS instance on AG2 that is connected to PE1 and PE2 via PW4 and PW6, respectively. Since the PWs for the two VPWS instances can be aggregated into the same LSP pair going to and coming from the MPLS network, a common virtual ES (vES) can be defined for the four mentioned PWs. In Figure 2, LSP1 and LSP2 represent the two LSP pairs between PE1 and AG2, and between PE2 and AG2, respectively. The vES consists of these two LSP pairs (LSP1 and LSP2) and each LSP pair has two PWs. This vES will be shared by two separate EVPN instances (e.g., EVI-1 and EVI-2) in the EVPN network. PW3 and PW4 are associated with EVI-1 and EVI-2 respectively on PE1, and PW5 and PW6 are associated with EVI-1 and EVI-2 respectively on PE2.

In some cases, the aggregation of PWs that share the same LSP pair 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.

For MPLS/IP access networks where a vES represents a set of LSP pairs or a set of PWs, 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 defines the concept of a vES and additional 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

**B-MAC:** Backbone MAC Address

**CE:** Customer Edge Device

**C-MAC:** Customer/Client MAC Address

**DF:** Designated Forwarder

**ENNI:**
External Network-Network Interface

**ES:** Ethernet Segment

**ESI:** Ethernet Segment Identifier

**Ethernet A-D:** Ethernet Auto-Discovery Route

**EVC:** Ethernet Virtual Circuit, [MEF63]

**EVI:** EVPN Instance

**EVPN:** Ethernet VPN

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

**PBB:** Provider Backbone Bridge

**PBB-EVPN:** Provider Backbone Bridge EVPN

**PE:** Provider Edge Device

**VPWS:** Virtual Pseudowire Service

**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 mix 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. Local Switching

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

(R3a) A PE that supports vES function, 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.

### 3.3. 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 preventing an EVC from carrying 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 preventing the PW from carrying 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.

### 3.4. 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]. [RFC8584] describes the additional procedures for DF election in EVPN. These DF election procedures 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 default procedure for DF election at the granularity of (vESI, Ethernet Tag) is also referred to as "service carving". With service carving, it is desirable to evenly partition the DFs for different vESes among different PEs, thus evenly distributing the traffic among different PEs. The following list the requirements apply to DF election of vESes for (PBB-)EVPN.

(R5a) A PE that supports vES function, MUST support a vES with m EVCs among n ENNIs belonging to p PEs in any arbitrary order; where n >= p >= m >=2. For example, if there is a 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.

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

### 3.5. OAM

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.

### 3.6. 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 MUST only impact C-MACs associated with multi-homed device/network for that service instance. In other words, MAC flushing MUST be limited to single service instance (I-SID in the case of PBB-EVPN) and only C-MACs for Single-Active multi-homed device/network.

### 3.7. Fast Convergence

Since many EVCs (and their associated vESes) are aggregated via a single physical port (e.g., ENNI), then the failure of that physical port impacts many vESes and triggers equally many 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 [RFC7432] for withdrawing many Ethernet A-D per ES 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 identifier: 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 B-MAC address assignment which is performed similar to what is described in section 7.2.1.1 of [RFC7623] with the following refinements:

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

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

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

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

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

### 4.1. EVPN DF Election for vES

The procedure for service carving for virtual Ethernet Segments is almost the same as the ones outlined in section 8.5 of [RFC7432] and [RFC8584] except for the fact that ES is replaced with vES.

For the sake of clarity and completeness, the default DF election procedure of [RFC7432] is repeated below with the necessary changes:

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 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 because of DF re-election, the PE, which ends up being the elected DF for the service, MUST trigger a MAC address flush notification towards the associated vES if the multi-homing device is a bridge or the multi-homing network is an Ethernet bridged network.

For LSP-based and PW-based vES, the non-DF PE SHOULD signal PW-status 'standby' to the Aggregation PE (e.g., AG1 and AG2 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 (EVPN instance) and the PW.

### 4.2. Grouping and Route Coloring for vES

Physical ports (e.g. ENNI) which aggregate many EVCs are 'colored' to enable the grouping schemes described below.

By default, the MAC address of the corresponding port (e.g. ENNI) is used to represent the 'color' of the port, and the EVPN Router's MAC Extended Community defined in [RFC9135] is used to signal this color.

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 may be advertised with the B-MAC route.

The following sections describe Grouping Ethernet A-D per ES and Grouping B-MAC, will become crucial for port failure handling as seen in Section 5.3, Section 5.4, and Section 5.5 below.

#### 4.2.1. EVPN Route Coloring for vES

When a PE discovers the vESI or is configured with the vESI associated with its attached vES, an Ethernet-Segment route and Ethernet A-D per ES route are generated using the vESI identifier.

These Ethernet-Segment and Ethernet A-D per ES routes specific to each vES are colored with an attribute representing their association to a physical port (e.g. ENNI).

The corresponding port 'color' is encoded in the EVPN Router's MAC Extended Community defined in [RFC9135] and advertised along with the Ethernet Segment and Ethernet A-D per ES routes for this vES.

The PE also constructs a special Grouping Ethernet A-D per ES route which represents all the vES associated with the port (e.g. ENNI). The corresponding port 'color' is encoded in the ESI field. For this encoding, Type 3 ESI (Section 5 of [RFC7432]) is used with the MAC field set to the color (MAC address) of the port and the 3-octet local discriminator field set to 0xFFFFFF.

The ESI label extended community (Section 7.5 of [RFC7432]) is not relevant to Grouping Ethernet A-D per ES route. The label value is not used for encapsulating BUM (Broadcast, Unknown-unicast, Multicast) packets for any split-horizon function. The ESI label extended community SHOULD NOT be added to Grouping Ethernet A-D per ES route and SHOULD be ignored on receiving PE.

This Grouping Ethernet A-D per ES route is advertised with a list of Route Targets corresponding to the impacted service instances. If the number of attached Route Targets exceeds the limit than can fit into a single route, then a set of Grouping Ethernet A-D per ES routes are advertised.

#### 4.2.2. PBB-EVPN Route Coloring for vES

For PBB-EVPN, especially where there are large number of service instances (i.e., I-SIDs) associated with each EVC the PE MAY color each vES B-MAC route with an attribute representing their association to a physical port (e.g. ENNI).

The corresponding port 'color' is encoded in the EVPN Router's MAC Extended Community defined in [RFC9135] and advertised along with the B-MAC for this vES in PBB-EVPN.

The PE MAY then also construct a special Grouping B-MAC route which represents all the vES associated with the port (e.g. ENNI). The corresponding port 'color' is encoded directly into this special Grouping B-MAC route.

## 5. Failure Handling and Recovery

There are several 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, individual 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.

[RFC7023] 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] to meet the above requirements. These extensions are described in the next section.

[RFC4377] and [RFC6310] are used for monitoring the status of LSPs and/or PWs associated to vES.

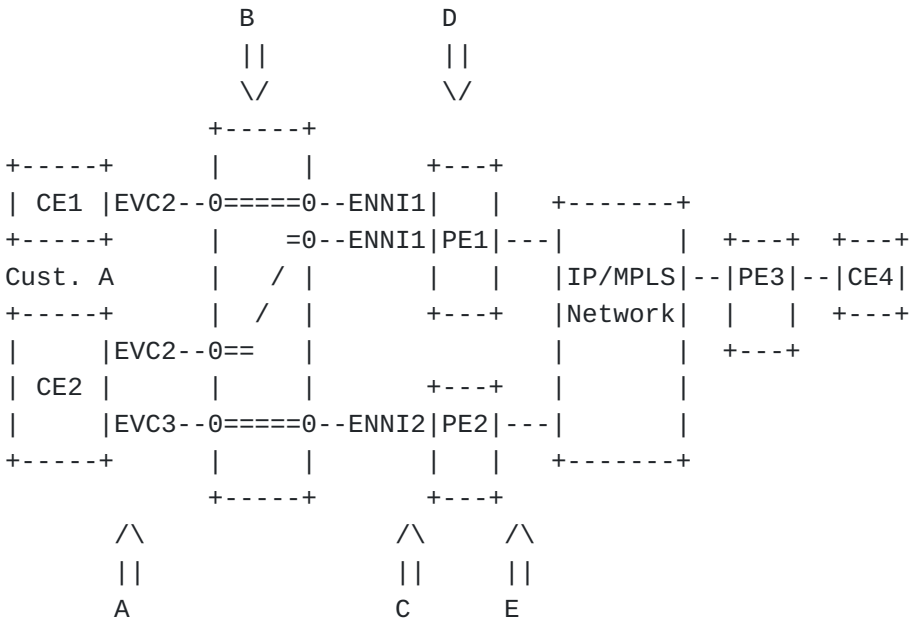


Figure 4: Failure Scenarios A,B,C,D and E

### 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 invalidate all MAC addresses associated with that Ethernet Segment. This is done by means of a mass-withdraw message, by withdrawing the 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 invalidating can be avoided by the remote PEs as they can update their next hop 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, this same EVPN procedure is used. In this case, the mass-withdraw is conveyed by withdrawing the Ethernet A-D per vES route carrying the vESI representing 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 C-MAC addresses associated with that Ethernet Segment. This is done by updating the advertised a B-MAC route's 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 C-MAC flushing because a single physical port can support large number of EVCs (and their associated vESes) and thus updating the advertised B-MAC corresponding to the physical port, with MAC mobility Extended community, will result in flushing C-MAC addresses not just for the impacted EVC but for all other EVCs on that port.

To reduce the scope of C-MAC flushing to only the impacted service instances (the service instance(s) impacted by the EVC failure), the PBB-EVPN C-MAC 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 B-MAC/I-SID route where existing PBB-EVPN B-MAC 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 C-MAC addresses associated with that I-SID for that B-MAC. This C-MAC 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 B-MAC/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 B-MAC/I-SID routes - one for each VLAN (broadcast domain) - i.e., one for each I-SID. Each B-MAC/I-SID route basically instructs the remote PEs to perform flushing for C-MACs corresponding to the advertised B-MAC only for the advertised I-SID.

The C-MAC flushing based on B-MAC/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 B-MAC per vES and upon EVC failure, the affected PE simply withdraws this B-MAC message to other PEs.

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

When many 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 propagation of failure in BGP needs to address all these impacted service instances. In order to achieve this, the following extensions are added to the baseline EVPN mechanism:

1. The PE MAY color each Ethernet A-D per ES route for a given vES, as described in Section 4.2.1. PE SHOULD use the physical port MAC by default. The receiving PEs take note of this color and create a list of vESes for this color.

2. The PE MAY advertises a special Grouping Ethernet A-D per ES route for that color, which represents all the vES associated with the port.

3. Upon a port failure (e.g., ENNI failure), the PE MAY send a mass-withdraw message by withdrawing the Grouping Ethernet A-D per ES route.

4. When this message is received, the remote PE MAY detect the special vES mass-withdraw message by identifying the Grouping Ethernet A-D per ES route. The remote PEs MAY then access the list created in (1) of the vESes for the specified color, and initiate locally MAC address invalidating procedures for each of the vESes in the list.

In scenarios where a logical ENNI is used the above procedure equally applies. The logical ENNI is represented by a Grouping Ethernet A-D per ES where the Type 3 ESI and the 6 bytes used in the ENNI's ESI MAC address field is used as a color for vESes as described above and in Section 4.2.1.

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

When many 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 B-MAC 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 B-MAC 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.

As noted above, the advertisement of the extended community along with B-MAC route for coloring purposes 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 number of I-SIDs).

If there are large number of service instances (i.e., I-SIDs) associated with each EVC, and if there is a B-MAC assigned per vES as recommended in the above section, then to handle port failure efficiently, the following extensions are added to the baseline PBB-EVPN mechanism:

1. Each vES MAY be colored with a MAC address representing the physical port like the coloring mechanism for EVPN. In other words, each B-MAC representing a vES is advertised with the 'color' of the physical port per Section 4.2.2. The receiving PEs take note of this color being advertised along with the B-MAC route and for each such color, create a list of vESes associated with this color.

2. The PE MAY advertise a special Grouping B-MAC route for that color (consisting by default of port MAC address), which represents all the vES associated with the port.

3. Upon a port failure (e.g., ENNI failure), the PE MAY send a mass-withdraw message by withdrawing the Grouping B-MAC route.

4. When this message is received, the remote PE MAY detect the special vES mass-withdraw message by identifying the Grouping B-MAC route. The remote PEs MAY then access the list created in (1) for the specified color, and flush all C-MACs associated with the failed physical port.

### 5.5. Fast Convergence in (PBB-)EVPN

As described above, when many 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:

*For EVPN initiate mass-withdraw procedure for all vESes associated with the failed port to invalidate MACs and for PBB-EVPN flush all C-MACs associated with the failed port across all vESes and the impacted I-SIDs

*DF election for all impacted vESes associated with the failed port

Section 5.3 already describes how to perform mass-withdraw for all affected vESes and invalidating MACs using a single BGP withdrawal of the Grouping Ethernet A-D per ES route. Section 5.4 describes how to only flush C-MAC address associated with the failed physical port (e.g., optimum C-MAC flushing) as well as, optionally, the withdrawal of a Grouping B-MAC route.

This section describes how to perform DF election in the most optimal 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 is used and route coloring is used.

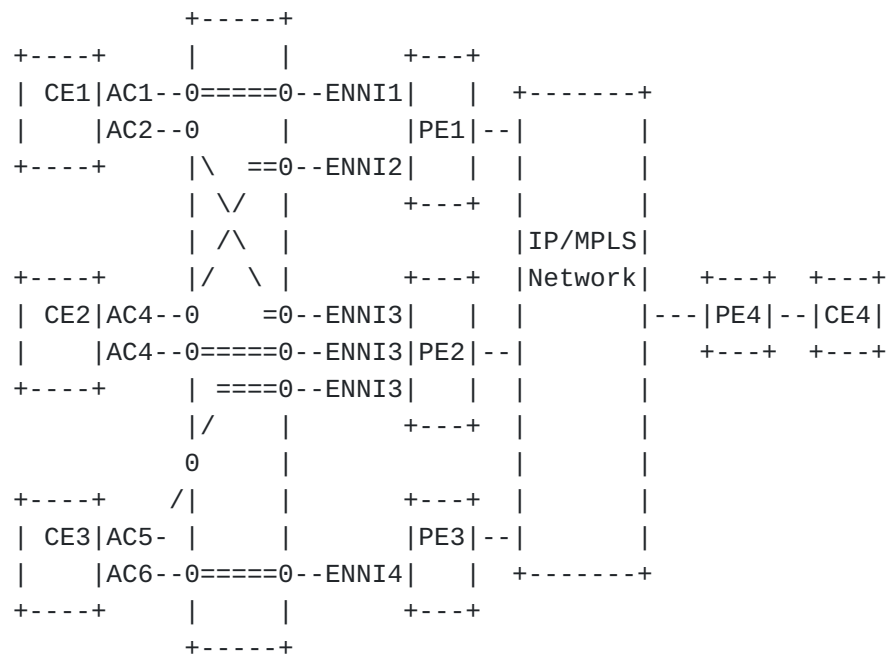


Figure 5: Fast Convergence Upon ENNI Failure

The procedure for coloring vES Ethernet Segment routes is described in Section 4.2. The following describes the procedure for fast convergence for DF election using these colored routes:

1. When a vES is configured, the PE SHOULD advertise the Ethernet Segment route for this vES with a color corresponding to the physical port.

2. All receiving PEs (in the redundancy group) SHOULD take note of this color and create a list of vESes for this color.

3. Recall, that the PE SHOULD also advertise a Grouping Ethernet A-D per ES (for EVPN) and a Grouping B-MAC (for PBB-EVPN) representing this color and vES grouping.

4. Upon a port failure (e.g., ENNI failure), the PE SHOULD withdraw this previously advertised Grouping Ethernet A-D per ES or Grouping B-MAC associated with the failed port. The PE SHOULD prioritize sending these Grouping routes withdraw message over individual vES route withdrawal messages of impacted vESes. For example, in Figure 5, when the physical port associated with

ENNI3 fails on PE2, it withdraws the previously advertised Grouping Ethernet A-D per ES route. When other multi-homing PEs(i.e., PE1 and PE3) receives this withdrawal message, they know that the vESes associated with CE1 and CE3 are impacted (because of the associated color), and thus to initiate DF election procedure for these vESes. Furthermore, the remote PEs (i.e., PE4) upon receiving this withdrawal message, it intiates failover procedure for vESes associated with CE1, CE3, and switches over to the other PE for each vES redundancy group.

5. On reception of Grouping Ethernet A-D per ES or Grouping B-MAC route withdrawal, other PEs in the redundancy group SHOULD initiate DF election procedures across all their affected vESes.

6. The PE with the physical port failure (ENNI failure), SHOULD send vES route withdrawal for every impacted vES. 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 when Grouping Ethernet A-D per ES or Grouping B-MAC withdrawal has been previously received.

## 6. Acknowledgements

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

## 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

This document requests no actions from IANA.

## 9. References

### 9.1. Normative References

**[I-D.ietf-bess-pbb-evpn-isid-cmacflush]**
Rabadan, J., Sathappan, S., Nagaraj, K., Miyake, M., and T. Matsuda, "PBB-EVPN ISID-based C-MAC-Flush", Work in Progress, Internet-Draft, draft-ietf-bess-pbb-evpn-isid-

cmacflush-08, 5 July 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-pbb-evpn-isid-cmacflush-08>.

[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>.

[RFC9135] Sajassi, A., Salam, S., Thoria, S., Drake, J., and J. Rabadan, "Integrated Routing and Bridging in Ethernet VPN (EVPN)", RFC 9135, DOI 10.17487/RFC9135, October 2021, <https://www.rfc-editor.org/info/rfc9135>.

### 9.2. Informative References

[MEF63] Metro Ethernet Forum, MEF., "[MEF6.3]: Subscriber Ethernet Services Definitions", 2019.

[RFC4377] Nadeau, T., Morrow, M., Swallow, G., Allan, D., and S. Matsushima, "Operations and Management (OAM) Requirements for Multi-Protocol Label Switched (MPLS) Networks", RFC 4377, DOI 10.17487/RFC4377, February 2006, <https://www.rfc-editor.org/info/rfc4377>.

[RFC6310] Aissaoui, M., Busschbach, P., Martini, L., Morrow, M., Nadeau, T., and Y. Stein, "Pseudowire (PW) Operations, Administration, and Maintenance (OAM) Message Mapping",

RFC 6310, DOI 10.17487/RFC6310, July 2011, <https://www.rfc-editor.org/info/rfc6310>.

**[RFC7023]** Mohan, D., Ed., Bitar, N., Ed., Sajassi, A., Ed., DeLord, S., Niger, P., and R. Qiu, "MPLS and Ethernet Operations, Administration, and Maintenance (OAM) Interworking", RFC 7023, DOI 10.17487/RFC7023, October 2013, <https://www.rfc-editor.org/info/rfc7023>.

**[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>.

**[RFC8584]** Rabadan, J., Ed., Mohanty, S., Ed., Sajassi, A., Drake, J., Nagaraj, K., and S. Sathappan, "Framework for Ethernet VPN Designated Forwarder Election Extensibility", RFC 8584, DOI 10.17487/RFC8584, April 2019, <https://www.rfc-editor.org/info/rfc8584>.

## 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

Jorge Rabadan
Nokia

Email: jorge.rabadan@nokia.com