

INTERNET-DRAFT
Intended Status: Standard Track

Sami Boutros
VMware

Ali Sajassi
Samer Salam
Cisco Systems

John Drake
Juniper Networks

Jeff Tantsura
Ericsson

Dirk Steinberg
Steinberg Consulting

Thomas Beckhaus
Deutsche Telecom

J. Rabadan
Nokia

Expires: December 8, 2016

June 6, 2016

VPWS support in EVPN
draft-ietf-bess-evpn-vpws-04.txt

Abstract

This document describes how EVPN can be used to support Virtual Private Wire Service (VPWS) in MPLS/IP networks. EVPN enables the following characteristics for VPWS: single-active as well as all-active multi-homing with flow-based load-balancing, eliminates the need for traditional way of PW signaling, and provides fast protection convergence upon node or link failure.

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) 2016 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.

Table of Contents

1	Introduction	4
1.1	Terminology	5
1.2	Requirements	6
2	Service interface	7
2.1	VLAN-Based Service Interface	7
2.2	VLAN Bundle Service Interface	7
2.2.1	Port-Based Service Interface	7
2.3	VLAN-Aware Bundle Service Interface	8
3	BGP Extensions	8
3.1	EVPN Layer 2 attributes extended community	8
4	Operation	10
5	EVPN Comparison to PW Signaling	11
6	Failure Scenarios	12
6.1	Single-Homed CEs	12
6.2	Multi-Homed CEs	12
7	Acknowledgements	12
8	Security Considerations	12

9	IANA Considerations	12
10	References	13
10.1	Normative References	13
10.2	Informative References	13
	Authors' Addresses	13

1 Introduction

This document describes how EVPN can be used to support Virtual Private Wire Service (VPWS) in MPLS/IP networks. The use of EVPN mechanisms for VPWS brings the benefits of EVPN to p2p services. These benefits include single-active redundancy as well as all-active redundancy with flow-based load-balancing. Furthermore, the use of EVPN for VPWS eliminates the need for traditional way of PW signaling for p2p Ethernet services, as described in [section 4](#).

[EVPN] has the ability to forward customer traffic to/from a given customer Attachment Circuit (AC), without any MAC lookup. This capability is ideal in providing p2p services (aka VPWS services). [MEF] defines Ethernet Virtual Private Line (EVPL) service as p2p service between a pair of ACs (designated by VLANs) and Ethernet Private Line (EPL) service, in which all traffic flows are between a single pair of ports, that in EVPN terminology would mean a single pair of ESes. EVPL can be considered as a VPWS with only two ACs. In delivering an EVPL service, the traffic forwarding capability of EVPN based on the exchange of a pair of Ethernet AD routes is used; whereas, for more general VPWS, traffic forwarding capability of EVPN based on the exchange of a group of Ethernet AD routes (one Ethernet AD route per AC/ES) is used. In a VPWS service, the traffic from an originating Ethernet Segment can be forwarded only to a single destination Ethernet Segment; hence, no MAC lookup is needed and the MPLS label associated with the per-EVI Ethernet AD route can be used in forwarding user traffic to the destination AC.

Both services are supported by using the per EVI Ethernet A-D route which contains an Ethernet Segment Identifier, in which the customer ES is encoded, and an Ethernet Tag, in which the VPWS service instance identifier is encoded. I.e., for both EPL and EVPL services, a specific VPWS service instance is identified by a pair of per EVI Ethernet A-D routes which together identify the VPWS service instance endpoints and the VPWS service instance. In the control plane the VPWS service instance is identified using the VPWS service instance identifiers advertised by each PE and in the data plane the value of the MPLS label advertised by one PE is used by the other PE to send traffic for that VPWS service instance. As with the Ethernet Tag in standard EVPN, the VPWS service instance identifier has uniqueness within an EVPN instance.

Unlike EVPN where Ethernet Tag ID in EVPN routes are set to zero for Port-based, vlan-based, and vlan-bundle interface mode and it is set to non-zero Ethernet tag ID for vlan-aware bundle mode, in EVPN-VPWS, for all the four interface modes, Ethernet tag ID in the Ethernet A-D route MUST be set to a valid value in all the service interface types.

In terms of route advertisement and MPLS label lookup behavior, EVPN-VPWS resembles the vlan-aware bundle mode of [\[RFC 7432\]](#) such that when a PE advertises per EVI Ethernet A-D route, the VPWS service instance serves as a 24-bit normalized Ethernet tag ID. The value of the MPLS label in this route represents both the EVI and the VPWS service instance, so that upon receiving an MPLS encapsulated packet, the disposition PE can identify the egress AC from the lookup of the MPLS label alone and perform any required tag translation. For EVPL service, the Ethernet frames transported over an MPLS/IP network SHOULD remain tagged with the originating VID and any VID translation is performed at the disposition PE. For EPL service, the Ethernet frames are transported as is and the tags are not altered.

The MPLS label value in the Ethernet A-D route can be set to the VNI for VxLAN encap, and this VNI may have a global scope or local scope per PE and may also be made equal to the VPWS service instance identifier set in the Ethernet A-D route.

The Ethernet Segment identifier encoded in the Ethernet A-D per EVI route is not used to identify the service, however it can be used for flow-based load-balancing and mass withdraw functions.

As with standard EVPN, the Ethernet A-D per ES route is used for fast convergence upon link or node failure and the Ethernet Segment route is used for auto-discovery of the PEs attached to a given multi-homed CE and to synchronize state between them.

1.1 Terminology

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](#)].

MAC: Media Access Control

MPLS: Multi Protocol Label Switching.

OAM: Operations, Administration and Maintenance.

PE: Provide Edge Node.

CE: Customer Edge device e.g., host or router or switch.

EVPL: Ethernet Virtual Private Line.

EPL: Ethernet Private Line.

EP-LAN: Ethernet Private LAN.

EVP-LAN: Ethernet Virtual Private LAN.

VPWS: Virtual Private Wire Service.

EVI: EVPN Instance.

ES: Ethernet Segment on a PE refer to the link attached to it, this link can be part of a set of links attached to different PEs in multi home cases, or could be a single link in single home cases.

Single-Active Mode: When a device or a network is multi-homed to two or more PEs and when only a single PE in such redundancy group can forward traffic to/from the multi-homed device or network for a given VLAN, then such multi-homing or redundancy is referred to as "Single-Active".

All-Active: When a device is multi-homed to two or more PEs and when all PEs in such redundancy group can forward traffic to/from the multi-homed device for a given VLAN, then such multi-homing or redundancy is referred to as "All-Active".

1.2 Requirements

1. EPL service access circuit maps to the whole Ethernet port.
2. EVPL service access circuits are VLANs on single or double tagged trunk ports. Each VLAN individually (or <S-VLAN,C-VLAN> combination) will be considered to be an endpoint for an EVPL service, without any direct dependency on any other VLANs on the trunk. Other VLANs on the same trunk could also be used for EVPL services, but could also be associated with other services.
3. If multiple VLANs on the same trunk are associated with EVPL services, the respective remote endpoints of these EVPLs could be dispersed across any number of PEs, i.e. different VLANs may lead to different destinations.
4. The VLAN tag on the access trunk only has PE-local significance. The VLAN tag on the remote end could be different, and could also be double tagged when the other side is single tagged.
5. Also, multiple EVPL service VLANs on the same trunk could belong to the same EVPN instance (EVI), or they could belong to different EVIs. This should be purely an administrative choice of the network operator.

6. A given PE could have thousands of EVPLs configured. It must be possible to configure multiple EVPL services within the same EVI.
7. Local access circuits configured to belong to a given EVPN instance could also belong to different physical access trunks.
8. EP-LAN and EVP-LAN are possible on the same system and also ESIs can be shared between EVPL and EVP-LANs.

2 Service interface

2.1 VLAN-Based Service Interface

With this service interface, a VPWS instance identifier corresponds to only a single VLAN on a specific interface. Therefore, there is a one-to-one mapping between a VID on this interface and the VPWS service instance identifier. The PE provides the cross-connect functionality between MPLS LSP identified by the VPWS service instance identifier and a specific <port,VLAN>. If the VLAN is represented by different VIDs on different PEs. (e.g., a different VID per Ethernet segment per PE), then each PE needs to perform VID translation for frames destined to its Ethernet segment. In such scenarios, the Ethernet frames transported over an MPLS/IP network SHOULD remain tagged with the originating VID, and a VID translation MUST be supported in the data path and MUST be performed on the disposition PE.

2.2 VLAN Bundle Service Interface

With this service interface, a VPWS service instance identifier corresponds to multiple VLANs on a specific interface. The PE provides the cross-connect functionality between MPLS label identified by the VPWS service instance identifier and a group of VLANs on a specific interface. For this service interface, each VLAN is presented by a single VID which means no VLAN translation is allowed. The receiving PE, can direct the traffic based on EVPN label alone to a specific port. The transmitting PE can cross connect traffic from a group of VLANs on a specific port to the MPLS label. The MPLS-encapsulated frames MUST remain tagged with the originating VID.

2.2.1 Port-Based Service Interface

This service interface is a special case of the VLAN bundle service interface, where all of the VLANs on the port are mapped to the same VPWS service instance identifier. The procedures are identical to those described in [Section 2.2](#).

2.3 VLAN-Aware Bundle Service Interface

Contrary to EVPN, in EVPN-VPWS this service interface maps to VLAN-based service interface (defined in [section 2.1](#)) and thus this service interface is not used in EVPN-VPWS. In other words, if one tries to define data-plane and control plane behavior for this service interface, he would realize that it is the same as that of VLAN-based service.

3. BGP Extensions

This document proposes the use of the per EVI Ethernet A-D route to signal VPWS services. The Ethernet Segment Identifier field is set to the customer ES and the Ethernet Tag ID 32-bit field is set to the 24-bit VPWS service instance identifier. For both EPL and EVPL services, for a given VPWS service instance the pair of PEs instantiating that VPWS service instance will each advertise a per EVI Ethernet A-D route with its VPWS service instance identifier and will each be configured with the other PE's VPWS service instance identifier. When each PE has received the other PE's per EVI Ethernet A-D route the VPWS service instance is instantiated. It should be noted that the same VPWS service instance identifier may be configured on both PEs.

The Route-Target (RT) extended community with which the per EVI Ethernet A-D route is tagged identifies the EVPN instance in which the VPWS service instance is configured. It is the operator's choice as to how many and which VPWS service instances are configured in a given EVPN instance. However, a given EVPN instance **MUST NOT** be configured with both VPWS service instances and standard EVPN multi-point services.

3.1 EVPN Layer 2 attributes extended community

This draft proposes a new extended community, defined below, to be included with the per EVI Ethernet A-D route. This attribute is mandatory if multihoming is enabled.

```

+-----+
|  Type(0x06)/Sub-type(0x04)(2 octet)|
+-----+
|  Control Flags (2 octets)           |
+-----+
|  L2 MTU (2 octets)                  |
+-----+
|  Reserved (2 octets)                 |
+-----+

```



```

 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+--+--+--+--+--+--+--+--+--+--+
|   MBZ                               |C|P|B|   (MBZ = MUST Be Zero)
+--+--+--+--+--+--+--+--+--+--+

```

The following bits in the Control Flags are defined; the remaining bits MUST be set to zero when sending and MUST be ignored when receiving this community.

Name	Meaning
------	---------

- | | |
|---|--|
| P | If set to 1 in multihoming single-active scenarios, it indicates that the advertising PE is the Primary PE. SHOULD be set to 1 for multihoming all-active scenarios. |
| B | If set to 1 in multihoming single-active scenarios, it indicates that the advertising PE is the Backup PE. |
| C | If set to 1, a Control word [RFC 4448] MUST be present when sending EVPN packets to this PE. |

A received L2 MTU=0 means no MTU checking against local MTU is needed. A received non-zero MTU SHOULD be checked against local MTU and if there is a mismatch, the local PE MUST not add the remote PE as the EVPN destination for the corresponding VPWS service instance.

The usage of the Per ES Ethernet AD route is unchanged from its usage in [[RFC7432](#)], i.e. the "Single-Active" bit in the flags of the ESI Label extended community will indicate if single-active or all-active redundancy is used for this ES.

In a multihoming all-active scenario, there is no DF election, and all the PEs in the ES that are active and ready to forward traffic to/from the CE will set the P bit to 1. A remote PE will do per-flow load balancing to the PEs that send P=1 for the same Ethernet Tag and ESI.

In multihoming single-active scenario, the DF election will determine who the primary and the backup PEs are, and only those PEs will set the P bit and B bit respectively. A remote PE will forward the traffic to the primary PE and switch over to the backup PE as soon as it receives an Ethernet A-D route withdrawal from the primary PE in the Ethernet Segment.

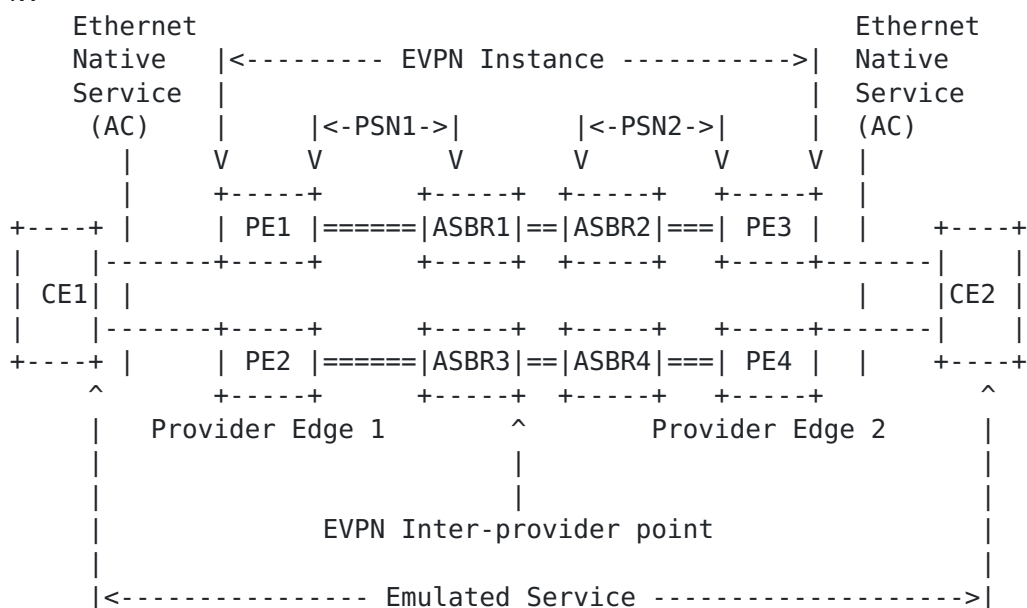
In multihoming single-active scenario, during transient situations, a remote PE receiving P=1 from more than one PE will select the last advertising PE as the primary PE when forwarding traffic. A remote PE

receiving B=1 from more than one PE will select only one backup PE. A remote PE MUST receive P=1 from at least one PE before forwarding traffic.

As per [RFC6790], if a network uses entropy labels then the control word (C bit set) SHOULD not be used when sending EVPN-encapsulated packets over a P2P LSP.

4 Operation

The following figure shows an example of a P2P service deployed with EVPN.



iBGP sessions are established between PE1, PE2, ASBR1 and ASBR3, possibly via a BGP route-reflector. Similarly, iBGP sessions are established between PE3, PE4, ASBR2 and ASBR4. eBGP sessions are established among ASBR1, ASBR2, ASBR3, and ASBR4.

All PEs and ASBRs are enabled for the EVPN SAFI and exchange per EVI Ethernet A-D routes, one route per VPWS service instance. For inter-AS option B, the ASBRs re-advertise these routes with Next Hop attribute set to their IP addresses. The link between the CE and the PE is either a C-tagged or S-tagged interface, as described in [802.1Q], that can carry a single VLAN tag or two nested VLAN tags and it is configured as a trunk with multiple VLANs, one per VPWS service instance. It should be noted that the VLAN ID used by the customer at either end of a VPWS service instance to identify that service instance may be different and EVPN doesn't perform that translation between the two values. Rather, the MPLS label will

identify the VPWS service instance and if translation is needed, it should be done by the Ethernet interface for each service.

For single-homed CE, in an advertised per EVI Ethernet A-D route the ESI field is set to 0 and the Ethernet Tag field is set to the VPWS service instance identifier that identifies the EVPL or EPL service.

For a multi-homed CE, in an advertised per EVI Ethernet A-D route the ESI field is set to the CE's ESI and the Ethernet Tag field is set to the VPWS service instance identifier, which MUST have the same value on all PEs attached to that ES. This allows an ingress PE to perform flow-based load-balancing of traffic flows to all of the PEs attached to that ES. In all cases traffic follows the transport paths, which may be asymmetric.

The VPWS service instance identifier encoded in the Ethernet Tag field in an advertised per EVI Ethernet A-D route MUST either be unique across all ASs, or an ASBR needs to perform a translation when the per EVI Ethernet A-D route is re-advertised by the ASBR from one AS to the other AS.

Per ES Ethernet A-D route can be used for mass withdraw to withdraw all per EVI Ethernet A-D routes associated with the multi-home site on a given PE.

5 EVPN Comparison to PW Signaling

In EVPN, service endpoint discovery and label signaling are done concurrently using BGP. Whereas, with VPWS based on [\[RFC4448\]](#), label signaling is done via LDP and service endpoint discovery is either through manual provisioning or through BGP.

In existing implementation of VPWS using pseudowires(PWs), redundancy is limited to single-active mode, while with EVPN implementation of VPWS both single-active and all-active redundancy modes can be supported.

In existing implementation with PWs, backup PWs are not used to carry traffic, while with EVPN, traffic can be load-balanced among different PEs multi-homed to a single CE.

Upon link or node failure, EVPN can trigger failover with the withdrawal of a single BGP route per EVPL service or multiple EVPL services, whereas with VPWS PW redundancy, the failover sequence requires exchange of two control plane messages: one message to deactivate the group of primary PWs and a second message to activate the group of backup PWs associated with the access link.

Finally, EVPN may employ data plane egress link protection mechanisms not available in VPWS. This can be done by the primary PE (on local AC down) using the label advertised in the per EVI Ethernet A-D route by the backup PE to encapsulate the traffic and direct it to backup PE.

6 Failure Scenarios

On a link or port failure between the CE and the PE for both single and multi-homed CEs, unlike [\[EVPN\]](#) the PE must withdraw all the associated Ethernet AD routes for the VPWS service instances on the failed port or link.

6.1 Single-Homed CEs

Unlike [\[EVPN\]](#), EVPN-VPWS uses Ethernet AD route advertisements for single-homed Ethernet Segments. Therefore, upon a link/port failure of this single-homed Ethernet Segment, the PE MUST withdraw the associated per EVI Ethernet A-D routes.

6.2 Multi-Homed CEs

For a faster convergence in multi-homed scenarios with either Single-Active Redundancy or All-active redundancy, mass withdraw technique as per [\[EVPN\]](#) baseline is used. A PE previously advertising a per ES Ethernet A-D route, can withdraw this route signaling to the remote PEs to switch all the VPWS service instances associated with this multi-homed ES to the backup PE

7 Acknowledgements

The authors would like to acknowledge Jeffrey Zhang, Wen Lin, Nitin Singh, Senthil Sathappan and Vinod Prabhu for their feedback and contributions to this document.

8 Security Considerations

The mechanisms in this document use EVPN control plane as defined in [\[RFC7432\]](#). Security considerations described in [\[RFC7432\]](#) are equally applicable.

This document uses MPLS and IP-based tunnel technologies to support data plane transport. Security considerations described in [\[RFC7432\]](#) and in [\[ietf-evpn-overlay\]](#) are equally applicable.

9 IANA Considerations

IANA has allocated the following EVPN Extended Community sub-type in

[[RFC7153](#)].

0x04 EVPN Layer 2 attributes [RFCXXXX]

10 References

10.1 Normative References

[KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

[RFC7432] A. Sajassi, R. Aggarwal et. al., "BGP MPLS Based Ethernet VPN".

10.2 Informative References

[RFC7209] A. Sajassi, R. Aggarwal et. al., "Requirements for Ethernet VPN".

[RFC7623] A. Sajassi et. al., "PBB-EVPN", "Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)".

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", [RFC4761](#), January 2007.

Authors' Addresses

Sami Boutros
VMware, Inc.
Email: sboutros@vmware.com

Ali Sajassi
Cisco
Email: sajassi@cisco.com

Samer Salam
Cisco
Email: ssalam@cisco.com

John Drake
Juniper Networks
Email: jdrake@juniper.net

Jeff Tantsura
Ericsson
Email: jeff.tantsura@ericsson.com

Dirk Steinberg
Steinberg Consulting
Email: dws@steinbergnet.net

Patrice Brissette
Cisco
Email: pbrisset@cisco.com

Thomas Beckhaus
Deutsche Telecom
Email: Thomas.Beckhaus@telekom.de>

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

Ryan Bickhart
Juniper Networks
Email: rbickhart@juniper.net