

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  
Alcatel-Lucent

Expires: September 17, 2016

March 16, 2016

VPWS support in EVPN  
draft-ietf-bess-evpn-vpws-03.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 single-segment and multi-segment PW signaling, and provides fast protection using data-plane prefix independent 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 DRAFT

VPWS support in EVPN

March 16, 2016

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

<a href="#">1</a>	Introduction . . . . .	<a href="#">4</a>
<a href="#">1.1</a>	Terminology . . . . .	<a href="#">5</a>
<a href="#">1.2</a>	Requirements . . . . .	<a href="#">6</a>
<a href="#">2</a>	Service interface . . . . .	<a href="#">6</a>
<a href="#">2.1</a>	VLAN-Based Service Interface . . . . .	<a href="#">6</a>
<a href="#">2.2</a>	VLAN Bundle Service Interface . . . . .	<a href="#">7</a>
<a href="#">2.2.1</a>	Port-Based Service Interface . . . . .	<a href="#">7</a>
<a href="#">2.3</a>	VLAN-Aware Bundle Service Interface . . . . .	<a href="#">7</a>
<a href="#">2.4</a>	Flexible CrossConnect Service . . . . .	<a href="#">7</a>
<a href="#">3</a>	BGP Extensions . . . . .	<a href="#">8</a>

<a href="#">3.1</a>	EVPN Layer 2 attributes extended community . . . . .	<a href="#">9</a>
<a href="#">4</a>	Operation . . . . .	<a href="#">10</a>
<a href="#">5</a>	EVPN Comparison to PW Signaling . . . . .	<a href="#">12</a>
<a href="#">6</a>	ESI Bandwidth . . . . .	<a href="#">12</a>
<a href="#">7</a>	Failure Scenarios . . . . .	<a href="#">12</a>

<a href="#">7.1</a>	Single-Homed CEs . . . . .	<a href="#">13</a>
<a href="#">7.2</a>	Multi-Homed CEs . . . . .	<a href="#">13</a>
<a href="#">8</a>	VPWS with multiple sites . . . . .	<a href="#">13</a>
<a href="#">9</a>	Acknowledgements . . . . .	<a href="#">13</a>
<a href="#">10</a>	Security Considerations . . . . .	<a href="#">13</a>
<a href="#">11</a>	IANA Considerations . . . . .	<a href="#">13</a>
<a href="#">12</a>	References . . . . .	<a href="#">13</a>
<a href="#">12.1</a>	Normative References . . . . .	<a href="#">13</a>
<a href="#">12.2</a>	Informative References . . . . .	<a href="#">14</a>
	Authors' Addresses . . . . .	<a href="#">14</a>

INTERNET DRAFT

VPWS support in EVPN

March 16, 2016

## 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 signaling single-segment and multi-segment PWs for p2p Ethernet services.

[EVPN] has the ability to forward customer traffic to/from a given customer Attachment Circuit (AC), aka Ethernet Segment in EVPN terminology, 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 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/segment) 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 Ethernet A-D per EVI 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 Ethernet A-D per EVI 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 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 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 Ethernet A-D per EVI route, the VPWS service instance serves as a 24-bit normalized Ethernet tag ID. 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 MUST 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 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.

VPWS: Virtual private wire service.

EVI: EVPN Instance.

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 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 access trunk could have hundreds of EVPL services, and 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. EPL-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 LSP 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 corss connect traffic from a group of VLANs on a specific port to the MPLS LSP. 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 6.2](#).

## [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 6.1](#)) 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.

## [2.4](#) Flexible CrossConnect Service

This service provides the ultimate flexibility at the expense of additional lookup. With this EVPN-VPWS service a large number of attachments circuits (ACs), each of which represented by either

single VLAN tag or double VLAN tags (QinQ) across multiple endpoints,



are multiplexed in a single EVPN-VPWS service instance. An endpoint can be a physical interface, VSI, an IP-VRF, a MAC-VRF, or any other endpoint where cross-connection of the associated AC is desired. Because in this service mode, aggregation is performed across multiple endpoints, besides MPLS label, an additional VLAN ID lookup (either single tag or double tag) needs to be performed at the disposition PE in order to identify the destination endpoint. One can think of this as, the EVPN label identifies a cross-connect table and then a single tag (or double tag) lookup is performed to identify the endpoint. Each cross-connect table has its own unique VLAN space which mean it can have upto 4K single-tag VLAN (or upto 16M double-tag VLANs). VLAN IDs can be overlap across different cross-connect tables but MUST be unique within a table.

The EVPN label besides identifying the cross-connect table, also identifies the following types of VID look-ups: Single VID lookup: The disposition PE MUST support single VID lookup where upon outer-VID lookup, the destination end-point is identified. Double VID lookup: The disposition PE MUST support double VID lookup where upon outer most two VIDs lookup, the destination end-point is identified.

Wildcard VID Lookup: The disposition PE MAY support special double VID lookup where the first VID is outer most VID and the 2nd VID is the wild card (\*).

If no entry is found upon the lookup, a counter per cross-connect table is incremented. Upon finding an entry and identifying the destination endpoint, the packet is forwarded to that destination endpoint. Any further tag manipulation such as re-write (single or double), addition, deletion (single or double) will be performed at the endpoint.

On the imposition PE, by associating an attachment circuit to an EVPN-VPWS service instance ID, we basically associate that attachment circuit with the corresponding cross-connect table.

Since VID lookup (single or double) needs to be performed at the disposition PE, then VID normalization MUST be performed prior to the MPLS encapsulation on the ingress PE. This requires that both imposition and disposition PE devices be capable of VLAN tag manipulation, such as re-write (single or double), addition, deletion (single or double), at their endpoints (e.g., their physical interfaces).

### [3. BGP Extensions](#)

This document proposes the use of the Ethernet A-D per EVI 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 an Ethernet A-D per EVI 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 Ethernet A-D per EVI 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 Ethernet A-D per EVI 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 Ethernet A-D per EVI route. This attribute is mandatory if multihoming is enabled.

```

+-----+
|  Type(0x06)/Sub-type(TBD)(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 [ <a href="#">RFC 4448</a> ] 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 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 multihoming single active scenario, a remote PE receiving P=1 from more than one PE will select only one 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.





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 Ethernet A-D per EVI route MUST either be unique across all ASs, or an ASBR needs to perform a translation when the Ethernet A-D per EVI route is re-advertised by the ASBR from one AS to the other AS.

Ethernet A-D per ES route can be used for mass withdraw to withdraw all Ethernet A-D per EVI 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 local repair mechanisms not available in VPWS.

## [6](#) ESI Bandwidth

The ESI Bandwidth will be encoded using the Link Bandwidth Extended community defined in [[draft-ietf-idr-link-bandwidth](#)] and associated with the Ethernet AD route used to realize the EVPL services.

When a PE receives this attribute for a given EVPL it MUST request the required bandwidth from the PSN towards the other EVPL service destination PE originating the message. When resources are allocated from the PSN for a given EVPL service, then the PSN SHOULD account for the Bandwidth requested by this EVPL service.

In the case where PSN resources are not available, the PE receiving this attribute MUST re-send its local Ethernet AD routes for this EVPL service with the ESI Bandwidth = All FFs to declare that the "PSN Resources Unavailable".

The scope of the ESI Bandwidth is limited to only one Autonomous System.

## [7](#) Failure Scenarios

On a link or port failure between the CE and the PE for both single

and multi-homed CEs, the PE must withdraw all the associated Ethernet AD routes for the VPWS service instances on the failed port or link.

### [7.1](#) 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 Ethernet A-D routes.

### [7.2](#) 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 an Ethernet A-D per ES 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

## [8](#) VPWS with multiple sites

The VPWS among multiple sites (full mesh of P2P connections - one per pair of sites) that can be setup automatically without any explicit provisioning of P2P connections among the sites is outside the scope of this document.

## [9](#) Acknowledgements

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

## [10](#) Security Considerations

This document does not introduce any additional security constraints.

## [11](#) IANA Considerations

Allocation of Extended Community Type and Sub-Type for EVPN L2 attributes.

## [12](#) References

### [12.1](#) Normative References

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

### [12.2](#) Informative References

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

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

[PBB-EVPN] A. Sajassi et. al., "PBB-EVPN", [draft-ietf-l2vpn-pbb-evpn-08.txt](#).

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

[[draft-ietf-idr-link-bandwidth](#)] P. Mohapatra, R. Fernando, "BGP Link Bandwidth Extended Community", [draft-ietf-idr-link-bandwidth-06.txt](#)

#### 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



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