Internet Engineering Task Force (IETF)
Request for Comments: 8214
Category: Standards Track
ISSN: 2070-1721

S. Boutros
VMware
A. Sajassi
S. Salam
Cisco
J. Drake
Juniper Networks
J. Rabadan
Nokia

August 2017

Virtual Private Wire Service Support in Ethernet VPN

Abstract

This document describes how Ethernet VPN (EVPN) can be used to support the Virtual Private Wire Service (VPWS) in MPLS/IP networks. EVPN accomplishes the following for VPWS: provides Single-Active as well as All-Active multihoming with flow-based load-balancing, eliminates the need for Pseudowire (PW) signaling, and provides fast protection convergence upon node or link failure.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8214.

Table of Contents

1. Introduction

This document describes how EVPN can be used to support VPWS in MPLS/IP networks. The use of EVPN mechanisms for VPWS (EVPN-VPWS) brings the benefits of EVPN to Point-to-Point (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 the traditional way of PW signaling for P2P Ethernet services, as described in Section 4.

[RFC7432] provides the ability to forward customer traffic to/from a given customer Attachment Circuit (AC), without any Media Access Control (MAC) lookup. This capability is ideal in providing P2P services (aka VPWS services). [MEF] defines the Ethernet Virtual Private Line (EVPL) service as a P2P service between a pair of ACs (designated by VLANs) and the 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 Ethernet Segments ES(es). EVPL can be considered as a VPWS with only two ACs. In delivering an EVPL service, the traffic-forwarding capability of EVPN is based on the exchange of a pair of Ethernet Auto-Discovery (A-D) routes, whereas for more general VPWS as per [RFC4664], the traffic-forwarding capability of EVPN is based on the exchange of a group of Ethernet A-D routes (one Ethernet A-D route per AC/ES). 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-EVPN instance (EVI) Ethernet A-D route can be used in forwarding user traffic to the destination AC.

For both EPL and EVPL services, a specific VPWS service instance is identified by a pair of per-EVI Ethernet A-D routes that 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 Provider Edge (PE) node. 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.

For EVPN routes, the Ethernet Tag IDs are set to zero for port-based, VLAN-based, and VLAN bundle interface mode and set to non-zero Ethernet Tag IDs for VLAN-aware bundle mode. Conversely, for EVPN-VPWS, the Ethernet Tag ID in the Ethernet A-D route MUST be set to a non-zero value for all four service interface types.

In terms of route advertisement and MPLS label lookup behavior, EVPN-VPWS resembles the VLAN-aware bundle mode of [RFC7432] such that when a PE advertises a per-EVI Ethernet A-D route, the VPWS service instance serves as a 32-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 MPLS label and subsequently perform any required tag translation. For the EVPL service, the Ethernet frames transported over an MPLS/IP network SHOULD remain tagged with the originating VLAN ID (VID), and any VID translation MUST be performed at the disposition PE. For the 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 Virtual Extensible LAN (VXLAN) Network Identifier (VNI) for VXLAN encapsulation as per [RFC7348], and this VNI will have a local scope per PE and may also be equal to the VPWS service instance identifier set in the Ethernet A-D route. When using VXLAN encapsulation, the BGP Encapsulation extended community is included in the Ethernet A-D route as described in [EVPN-OVERLAY]. The VNI is like the MPLS label that will be set in the tunnel header used to tunnel Ethernet packets from all the service interface types defined in Section 2. The EVPN-VPWS techniques defined in this document have no dependency on the tunneling technology.

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 per the [RFC7432] baseline.

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

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

EVPN: Ethernet VPN.

MAC: Media Access Control.

MPLS: Multiprotocol Label Switching.

OAM: Operations, Administration, and Maintenance.

PE: Provider Edge Node.

AS: Autonomous System.

ASBR: Autonomous System Border Router.

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

EVPL: Ethernet Virtual Private Line.

EPL: Ethernet Private Line.

EP-LAN: Ethernet Private LAN.

EVP-LAN: Ethernet Virtual Private LAN.

S-VLAN: Service VLAN identifier.

C-VLAN: Customer VLAN identifier.

VID: VLAN ID.

VPWS: Virtual Private Wire Service.

EVI: EVPN Instance.

P2P: Point to Point.

VXLAN: Virtual Extensible LAN.

DF: Designated Forwarder.

L2: Layer 2.

MTU: Maximum Transmission Unit.

eBGP: External Border Gateway Protocol.

iBGP: Internal Border Gateway Protocol.

ES: "Ethernet Segment" on a PE refers to the link attached to it. This link can be part of a set of links attached to different PEs in multihomed cases or could be a single link in single-homed cases.

ESI: Ethernet Segment Identifier.

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

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

VPWS Service Instance: A VPWS service instance is represented by a pair of EVPN service labels associated with a pair of endpoints. Each label is downstream-assigned and advertised by the disposition PE through an Ethernet A-D per-EVI route. The downstream label identifies the endpoint on the disposition PE. A VPWS service instance can be associated with only one VPWS service identifier.

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 an MPLS Label Switched Path (LSP) identified by the VPWS service instance identifier and a specific <port, VLAN>. If the VLAN is represented by different VIDs on different PEs and different ES(es) (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 the 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 that no VLAN translation is allowed. The receiving PE can direct the traffic, based on the 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 a VLAN-based service interface (defined in Section 2.1); 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, one would realize that it is the same as that of the VLAN-based service.

3. BGP Extensions

This document specifies the use of the per-EVI Ethernet A-D route to signal VPWS services. The ESI field is set to the customer ES, and the 32-bit Ethernet Tag ID field MUST be set to the VPWS service instance identifier value. The VPWS service instance identifier value MAY be set to a 24-bit value, and when a 24-bit value is used, it MUST be right-aligned. For both EPL and EVPL services using 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 multipoint services.

3.1. EVPN Layer 2 Attributes Extended Community

This document defines a new extended community [RFC4360], to be included with per-EVI Ethernet A-D routes. This attribute is mandatory if multihoming is enabled.

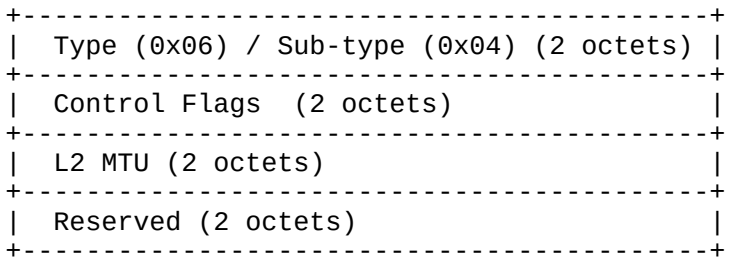

Figure 1: EVPN Layer 2 Attributes Extended Community

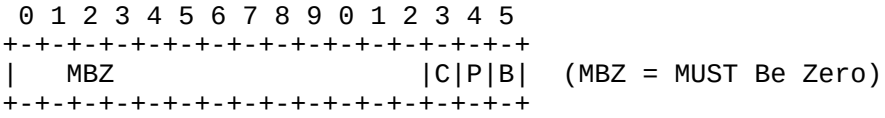

Figure 2: EVPN Layer 2 Attributes Control Flags

The following bits in 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, this flag indicates that the advertising PE is the primary PE. MUST be set to 1 for multihoming All-Active scenarios by all active PE(s).
B	If set to 1 in multihoming Single-Active scenarios, this flag indicates that the advertising PE is the backup PE.
C	If set to 1, a control word [RFC4448] MUST be present when sending EVPN packets to this PE. It is recommended that the control word be included in the absence of an entropy label [RFC6790].

L2 MTU is a 2-octet value indicating the MTU in bytes.

A received L2 MTU of zero means that no MTU checking against the local MTU is needed. A received non-zero MTU MUST be checked against the 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 A-D 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 Designated Forwarder (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 Flag. A remote PE will do per-flow load-balancing to the PEs that set the P Flag for the same Ethernet Tag and ESI. The B Flag in Control Flags SHOULD NOT be set in the multihoming All-Active scenario and MUST be ignored by receiving PE(s) if set.

In a multihoming Single-Active scenario for a given VPWS service instance, the DF election should result in the primary-elected PE for the VPWS service instance advertising the P Flag set and the B Flag clear, the backup-elected PE should advertise the P Flag clear and the B Flag set, and the rest of the PEs in the same ES should signal both the P Flag and the B Flag clear. When the primary PE/ES fails, the primary PE will withdraw the associated Ethernet A-D routes for

the VPWS service instance from the remote PE, and the remote PE should then send traffic associated with the VPWS instance to the backup PE. DF re-election will happen between the PE(s) in the same ES, and there will be a newly elected primary PE and newly elected backup PE that will signal the P and B Flags as described. A remote PE SHOULD receive the P Flag set from only one primary PE and the B Flag set from only one backup PE. However, during transient situations, a remote PE receiving a P Flag set from more than one PE will select the last advertising PE as the primary PE when forwarding traffic. A remote PE receiving a B Flag set from more than one PE will select the last advertising PE as the backup PE. A remote PE MUST receive a P Flag set from at least one PE before forwarding traffic.

If a network uses entropy labels per [RFC6790], then the C Flag MUST NOT be set, and the control word MUST 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.

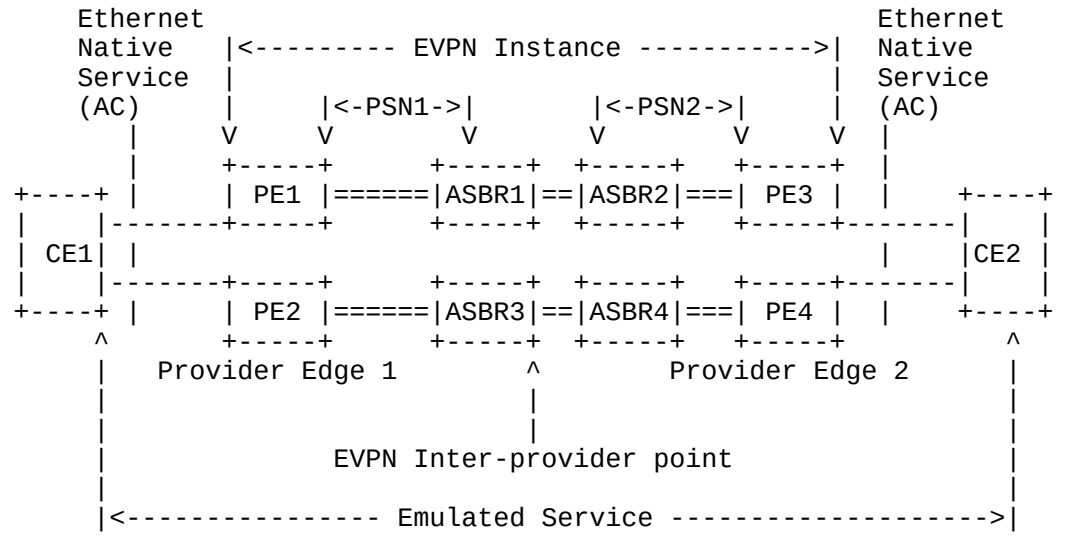

Figure 3: EVPN-VPWS Deployment Model

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

All PEs and ASBRs are enabled for the EVPN Subsequent Address Family Identifier (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 the NEXT_HOP attribute set to their IP addresses as per [RFC4271]. 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 a single-homed CE, in an advertised per-EVI Ethernet A-D route, the ESI field is set to zero and the Ethernet Tag ID is set to the VPWS service instance identifier that identifies the EVPL or EPL service.

For a multihomed CE, in an advertised per-EVI Ethernet A-D route, the ESI field is set to the CE's ESI and the Ethernet Tag ID 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 in a multihoming All-Active scenario 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.

Either (1) the VPWS service instance identifier encoded in the Ethernet Tag ID in an advertised per-EVI Ethernet A-D route MUST be unique across all ASes or (2) 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.

A per-ES Ethernet A-D route can be used for mass withdraw to withdraw all per-EVI Ethernet A-D routes associated with the multihomed 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 implementations of VPWS using PWs, redundancy is limited to Single-Active mode, while with EVPN implementations of VPWS, both Single-Active and All-Active redundancy modes can be supported.

In existing implementations with PWs, backup PWs are not used to carry traffic, while with EVPN, traffic can be load-balanced among different PEs multihomed 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 the 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 the backup PE.

6. Failure Scenarios

On a link or port failure between the CE and the PE for both single-homed and multihomed CEs, unlike [RFC7432], the PE MUST withdraw all the associated Ethernet A-D routes for the VPWS service instances on the failed port or link.

6.1. Single-Homed CEs

Unlike [RFC7432], EVPN-VPWS uses Ethernet A-D route advertisements for single-homed Ethernet Segments. Therefore, upon a link/port failure of a given single-homed Ethernet Segment, the PE MUST withdraw the associated per-EVI Ethernet A-D routes.

6.2. Multihomed CEs

For a faster convergence in multihomed scenarios with either Single-Active redundancy or All-Active redundancy, a mass withdraw technique is used. A PE previously advertising a per-ES Ethernet A-D route can withdraw this route by signaling to the remote PEs to switch all the VPWS service instances associated with this multihomed ES to the backup PE.

Just like RFC 7432, the Ethernet A-D per-EVI route MUST NOT be used for traffic forwarding by a remote PE until it also receives the associated set of Ethernet A-D per-ES routes.

7. Security Considerations

The mechanisms in this document use the EVPN control plane as defined in [RFC7432]. The security considerations described in [RFC7432] are equally applicable.

This document uses MPLS and IP-based tunnel technologies to support data-plane transport. The security considerations described in [RFC7432] and in [EVPN-OVERLAY] are equally applicable.

8. IANA Considerations

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

Sub-Type Value	Name	Reference
0x04	EVPN Layer 2 Attributes	RFC 8214

This document creates a registry called "EVPN Layer 2 Attributes Control Flags". New registrations will be made through the "RFC Required" procedure defined in [RFC8126].

Initial registrations are as follows:

P	Advertising PE is the primary PE.
B	Advertising PE is the backup PE.
C	Control word [RFC4448] MUST be present.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[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>.

[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>.

[RFC4448] Martini, L., Ed., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks", RFC 4448, DOI 10.17487/RFC4448, April 2006, <https://www.rfc-editor.org/info/rfc4448>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <https://www.rfc-editor.org/info/rfc6790>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <https://www.rfc-editor.org/info/rfc4360>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <https://www.rfc-editor.org/info/rfc7348>.

9.2. Informative References

[MEF] Metro Ethernet Forum, "EVC Ethernet Services Definitions Phase 3", Technical Specification MEF 6.2, August 2014, <https://www.mef.net/Assets/Technical_Specifications/PDF/MEF_6.2.pdf>.

[RFC4664] Andersson, L., Ed., and E. Rosen, Ed., "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, DOI 10.17487/RFC4664, September 2006, <https://www.rfc-editor.org/info/rfc4664>.

[EVPN-OVERLAY]
Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution using EVPN", Work in Progress, draft-ietf-bess-evpn-overlay-08, March 2017.

[802.1Q] IEEE, "IEEE Standard for Local and metropolitan area networks -- Media Access Control (MAC) Bridges and Virtual Bridge Local Area Networks", IEEE Std 802.1Q-2011, DOI 10.1109/IEEESTD.2011.6009146.

Acknowledgements

The authors would like to acknowledge Jeffrey Zhang, Wen Lin, Nitin Singh, Senthil Sathappan, Vinod Prabhu, Himanshu Shah, Iftekhar Hussain, Alvaro Retana, and Acee Lindem for their feedback and contributions to this document.

Contributors

In addition to the authors listed on the front page, the following coauthors have also contributed to this document:

Jeff Tantsura
Individual
Email: jefftant@gmail.com

Dirk Steinberg
Steinberg Consulting
Email: dws@steinbergnet.net

Patrice Brissette
Cisco Systems
Email: pbrisset@cisco.com

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

Ryan Bickhart
Juniper Networks
Email: rbickhart@juniper.net

Daniel Voyer
Bell Canada

Authors' Addresses

Sami Boutros
VMware, Inc.

Email: sboutros@vmware.com

Ali Sajassi
Cisco Systems

Email: sajassi@cisco.com

Samer Salam
Cisco Systems

Email: ssalam@cisco.com

John Drake
Juniper Networks

Email: jdrake@juniper.net

Jorge Rabadan
Nokia

Email: jorge.rabadan@nokia.com