BESS Working Group
Internet Draft
Category: Standard Track

A. Sajassi
S. Salam
P. Brissette
Cisco

L. Jalil
Verizon

Expires: January 2, 2018

July 2, 2017

(PBB-)EVPN Integration with (PBB-)VPLS in All-Active Mode
draft-sajassi-bess-evpn-vpls-all-active-00

Abstract

This draft discusses the backward compatibility of the (PBB-)EVPN solution with (PBB-)VPLS in all-active redundancy mode.

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/1id-abstracts.html

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

Copyright and License Notice

Table of Contents

1 Introduction

VPLS and PBB-VPLS are widely-deployed L2VPN technologies. Many SPs who are looking at adopting EVPN and PBB-EVPN want to preserve their investment in the (PBB-)VPLS networks. Hence, it is required to provide mechanisms by which (PBB-)EVPN technology can be introduced into existing L2VPN networks without requiring a fork-lift upgrade. [EVPN-VPLS] discusses mechanisms for the seamless integration of the two technologies in the same MPLS/IP network, however, operation is limited to single-active redundancy mode. In this document, we extend the solution to support all-active redundancy.

Section 2 provides the limitations in the current (PBB-)EVPN/(PBB-)VPLS interoperability solution. Section 3 discusses the solution for addressing those limitations. Section 4 describes the required control and data plane changes to support all-active redundancy. Section 5 covers the failure handling.

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].

2 Limitations

[EVPN-VPLS] defines mechanisms for (PBB-)EVPN seamless interoperability with (PBB-)VPLS. The solution defined in [EVPN-VPLS] suffers from a major limitation that hinders brown-field deployment of EVPN solution: It provides support for all-active redundancy only for VPN instances confined to (PBB-)EVPN PEs. For VPN instances that span both (PBB-)EVPN as well as (PBB-)VPLS PEs only single-active redundancy mode is supported. This eliminates one of the key value propositions of inserting EVPN solution in existing networks.

The reason why this capability is not currently supported is due to the issue of MAC address flip-flopping on the VPLS PEs. This is best explained with an example: Consider the example network of Figure 1 below. Assume that CE1 is connected over an all-active link aggregation group (LAG) to EVPN-capable PEs (PE2 and PE3). For traffic destined from CE1 to CE2, different flows from the same source MAC address MAC-A will be load-balanced over the LAG to PE2 and PE3. PE2 will forward the traffic over its own pseudowire (call it PW-Blue) to PE5, whereas PE3 will forward the traffic over its own pseudowire (call it PW-Red) to PE5. As such, VPLS PE (PE5) will learn the same MAC address (MAC-A) over both PW-Red and PW-Blue, depending on the load-balancing. This MAC flip-flopping will continue

indefinitely depending on traffic patterns.

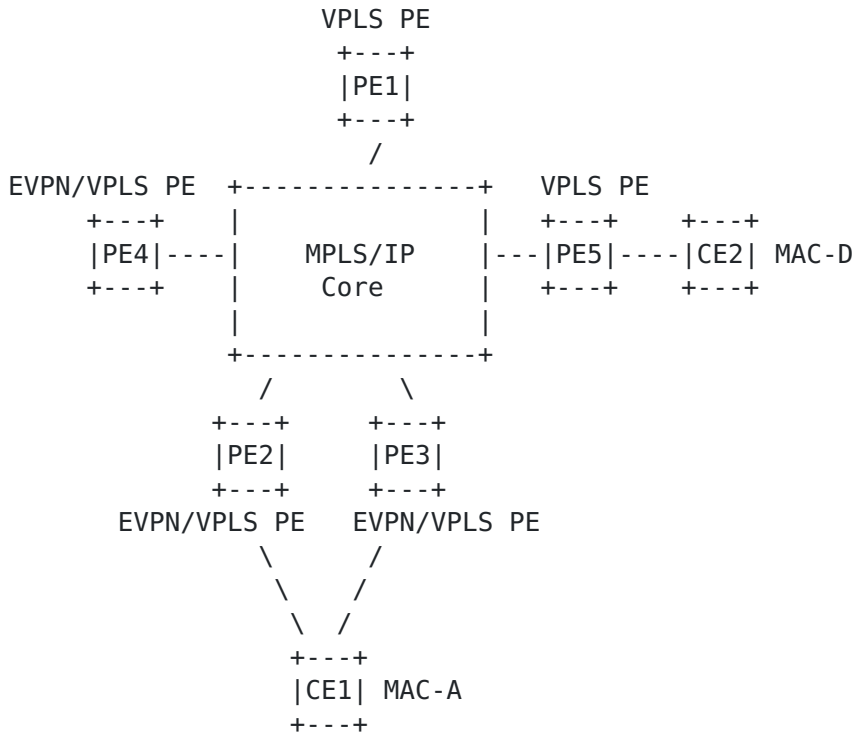

Figure 1: Seamless Integration of (PBB-)EVPN PEs & (PBB-)VPLS

The focus of this draft is on providing a solution that addresses the above limitation, thereby enabling the support of all-active redundancy in mixed (PBB-)EVPN/(PBB-)VPLS deployments.

3 Solution for MAC Flip-Flopping

In order to address the MAC flip-flopping problem on the VPLS PEs, these PEs must learn the traffic originating from a given source MAC address over the same pseudowire consistently, regardless of which remote EVPN-capable PE forwarded the traffic in a given multi-homed setup. To that end, every multi-homed EVPN-capable PE must maintain, in addition to its own pseudowires, a set of shadow or "alias" pseudowires for each of its peers in a given Redundancy Group (RG). For instance, in the example network of Figure 1, PE2 maintains its own pseudowire towards PE5 in addition to an "alias" pseudowire corresponding to the pseudowire between PE3 and PE5.

When traffic arrives from a multi-homed CE over a multi-chassis LAG, the EVPN-capable PE then examines whether or not it is the Designated Forwarder (DF) for the Ethernet Segment (ES) in question. In the case

where the PE is the DF for the ES, it would use its own pseudowire label to forward traffic towards a remote VPLS PE. However, in the case where the PE is not the DF for the ES, it would then use the "alias" pseudowire label associated with the DF PE in order to forward traffic towards the remote VPLS PE. To illustrate this using the example of Figure 1, consider that PE3 is the DF for the ES associated with CE1. Furthermore, assume that the pseudowire labels from PE2 and PE3 to PE5 are Label-Blue and Label-Red, respectively. When CE1 load-balances traffic destined to CE2 towards PE3, the latter will use its own pseudowire label (Label-Red) to forward traffic to PE5. Whereas, when CE1 forwards traffic destined to CE2 towards PE2, it will use the alias pseudowire label (Label-Red) instead of its own pseudowire label to forward traffic towards PE5. This is because PE2 is not the DF for the Ethernet Segment associated with CE1.

3.1 Load-Balancing

For traffic flowing from the EVPN-capable PEs towards the MPLS network, the load-balancing is on a per-flow granularity, regardless of whether the traffic is destined towards remote EVPN or VPLS PEs.

For traffic flowing from the VPLS PEs towards the EVPN-capable PEs, the load-balancing is on a per-VLAN per destination site granularity. That is, the traffic for a given VLAN in a destination site is sent to only one of the multi-homed EVPN-capable PEs. This is because all the EVPN-capable PEs in a given redundancy group will use the the pseudowire label associated with the DF to forward traffic towards remote VPLS PEs (recall, also, that EVPN DF election is per VLAN per ES).

4 Changes on EVPN PEs

The changes to support the mechanisms of this draft are confined to the EVPN-capable PEs. In the following two sub-sections we cover both the control plane as well as data plane changes required.

4.1 Control Plane Changes

In order for the EVPN-capable PEs to maintain the alias pseudowires, it is required to synchronize the VPLS pseudowire labels among the PEs in the same Redundancy Group. For VPLS-BGP [RFC4761], this is straight-forward to achieve because the VE-IDs and label blocks associated with all PEs are advertised in BGP. Hence, a PE in an EVPN RG can easily extract the alias pseudowire labels associated with its peers in the same RG. For VPLS-LDP [RFC4762], protocol message extensions are required but are outside the scope of the current document.

Another control plane extension that is required is to synchronize the MAC addresses learnt over the active pseudowire at DF EVPN PEs to the non-DF EVPN PEs with alias pseudowire using BGP. This can be done using the existing EVPN MAC Advertisement route. The identity of the pseudowire over which the address was learnt is encoded in the ESI field. This can be done using a Type 4 ESI, where the Router ID holds the IP address of the remote pseudowire endpoint IP address (i.e. VPLS PE address) and the high-order 2 octets of the Local Discriminator encode the VE-ID of the remote pseudowire endpoint (i.e. EVPN-capable PE that is the DF).

4.2 Data Plane Changes

4.2.1 Known Unicast Traffic

After DF election is complete, the EVPN-capable PE programs its data plane based on the outcome of DF election as follows:

If known unicast traffic is received by the PE from an Ethernet Segment for which it is the DF, then it uses its own pseudowire label in the label stack when forwarding traffic to remote VPLS PEs.

If known unicast traffic is received by the PE from an Ethernet Segment for which is non-DF, then it uses the alias pseudowire label (associated with the DF) instead of its own pseudowire label in the label stack when forwarding traffic to remote VPLS PEs.

In other words, the EVPN-capable PE must use the DF/non-DF status of the incoming attachment circuit interface in order to choose the correct label stack for VPLS forwarding.

4.2.2 BUM Traffic

The EVPN-capable PEs must maintain two replication lists: one that uses their own pseudowires, and another that uses the alias pseudowires. When BUM traffic is received from the attachment circuit, the PE examines the DF status of the incoming interface to identify which of the two replication lists to use: If the PE is the DF, then it uses the replication list which encompasses its own pseudowires. Whereas, if the PE is non-DF, then it uses the replication list encompassing the alias pseudowires.

BUM traffic received over a VPLS pseudowire is handled as follows:

Broadcast and multicast traffic is identified as such by inspecting the destination MAC address, and is handled as usual per EVPN MPLS ingress flooding mechanisms. At egress to the attachment circuit, all broadcast and multicast VPLS traffic is subjected to DF filtering

procedures per existing EVPN procedures.

Unknown unicast traffic cannot be identified as such by the disposition PE on egress from the pseudowire, since nothing in the Ethernet frame or the MPLS label stack (unlike EVPN) distinguishes this traffic from known unicast. Furthermore, the disposition PE cannot rely on its own MAC forwarding table to infer whether the frame was flooded or not - i.e., an unknown MAC address on the imposition PE cannot be known to the disposition PE. Due to this, the egress (disposition) PE will treat unicast MAC addresses based on its own local forwarding state - i.e., if the MAC address is known locally, then it is treated as such and if the MAC address is unknown locally, then it is treated as BUM traffic and will apply DF filtering. This can lead to a side-effect for a very specific scenario where the MAC-DA is unknown at the ingress PE but it is known to the egress multi-homing PEs (i.e., there is no issue when MAC-DA is known at the ingress and unknown at the egress, or MAC-DA is unknown at both the ingress and egress PEs). In such a specific scenario, a multi-homed CE will experience duplicate packets for an interim period of time until the remote VPLS PE learns the MAC address from reverse traffic. The CE's application layer will handle the discard of transient duplicate frames. While it is acknowledged that this behavior deviates from classical Ethernet, which guarantees the absence of packet duplication, the side-effect occurs in very specific scenario and it is both short-lived and confined in scope to the PE/CE links. Hence, it is a reasonable trade-off to accept in favor of enabling all-active redundancy in the solution.

5 Failure Handling

Failure handling follows standard EVPN and VPLS procedures:

For link failure on DF EVPN-capable PE, the PE sends a mass withdraw indication using per ES Ethernet A-D route to other EVPN PEs, causing them to update their forwarding entries to point to only the non-DF PE. The DF PE also sends VPLS MAC address flush message to remote VPLS PEs, causing them to flush their entries. The non-DF EVPN PE takes over and assumes the DF role. It uses its own VPLS pseudowire labels for sending traffic towards the VPLS PEs.

For link failure on non-DF EVPN PE, the PE sends mass withdraw per ES Ethernet A-D route to other EVPN PEs, causing them to update their forwarding entries to point to only the DF PE. Nothing is done with respect to the VPLS PEs, as this failure is transparent to them.

6 EVPN-VPWS termination onto multi-homing EVPN PEs This section will be added in the future revision to describe how the MAC synchroniation

mechanism over PW described above can be used for this scenario.

7 Security Considerations

No new security considerations beyond those for VPLS and EVPN.

8 IANA Considerations

This document has no actions for IANA.

9 References

9.1 Normative References

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

[RFC4761] Kompella, K., Ed., and Y. Rekhter, Ed., "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007.

[RFC4762] Lasserre, M., Ed., and V. Kompella, Ed., "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

[EVPN-VPLS] Sajassi, A., Salam, S., Del Regno, N., and Rabadan, J., "(PBB-)EVPN Seamless Integration with (PBB-)VPLS", draft-ietf-bess-evpn-vpls-seamless-integ-00, work in progress, February 2015, <https://datatracker.ietf.org/doc/html/draft-sajassi-bess-evpn-vpls-seamless-integ>.

9.2 Informative References

Authors' Addresses

Ali Sajassi
Cisco
170 West Tasman Drive

San Jose, CA 95134, US
Email: sajassi@cisco.com

Samer Salam
Cisco
595 Burrard Street, Suite 2123
Vancouver, BC V7X 1J1, Canada
Email: ssalam@cisco.com

Patrice Brissette
Cisco
Email: pbrisset@cisco.com

Luay Jalil
Cisco
Email: luay.jalil@verizon.com