BESS Workgroup
Internet Draft
Intended status: Informational

J. Rabadan, Ed.
S. Sathappan
K. Nagaraj
Nokia

J. Bueno
J. Crespo
Telefonica

Expires: January 4, 2018

July 3, 2017

Loop Protection in EVPN networks
draft-snr-bess-evpn-loop-protect-00

Abstract

Ethernet Virtual Private Networks (EVPN) is becoming the de-facto standard-based control plane solution for Data Center and layer-2 Service Provider applications. The risk of loops caused by backdoor paths accidentally created within the same broadcast domain, is a general common concern, especially among Service Providers in large Layer-2 networks. While other layer-2 Ethernet technologies use Spanning Tree based Protocols (xSTP) to provide a network-wide loop protection, EVPN has the right tools to detect and protect the network against loops in an efficient and effective way. This document describes a mechanism to provide global loop protection in EVPN networks.

Status of this Memo

This Internet-Draft is submitted 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/ietf/1id-abstracts.txt

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

This Internet-Draft will expire on January 4, 2018.

Table of Contents

1. Introduction

Ethernet Virtual Private Networks (EVPN) is becoming the de-facto standard-based control plane solution for Data Center and layer-2 Service Provider applications. The risk of loops caused by backdoor paths accidentally created within the same broadcast domain, is a general common concern, especially among Service Providers in large Layer-2 networks. While other layer-2 Ethernet technologies use Spanning Tree based Protocols (xSTP) to provide global loop protection, EVPN has the right tools to detect and protect the network against loops in an efficient and effective way. However, [RFC7432] only addresses the MAC duplication detection and protection at the control plane, and not all the possible loop scenarios.

In this document, backdoor path is defined as a layer-2 connection between two Attachment Circuits (ACs) that, along with the layer-2 connectivity in the EVI, creates a loop. We differentiate between a local and a global loop. A local loop is created by a backdoor path within the same physical port or between two Attachment Circuits (ACs) of the same MAC-VRF. A global loop is created by a backdoor path between two ACs of the same EVI but different PEs. This document addresses global loop protection, since it requires interoperability between PEs. Local loop protection is implementation specific and it is not addressed in this specification.

Figure 1 shows a typical example of a backdoor path that may be created by mistake in a Service Provider network that uses EVPN to provide E-LAN services. A backdoor path is accidentally created between AC4 and AC5.

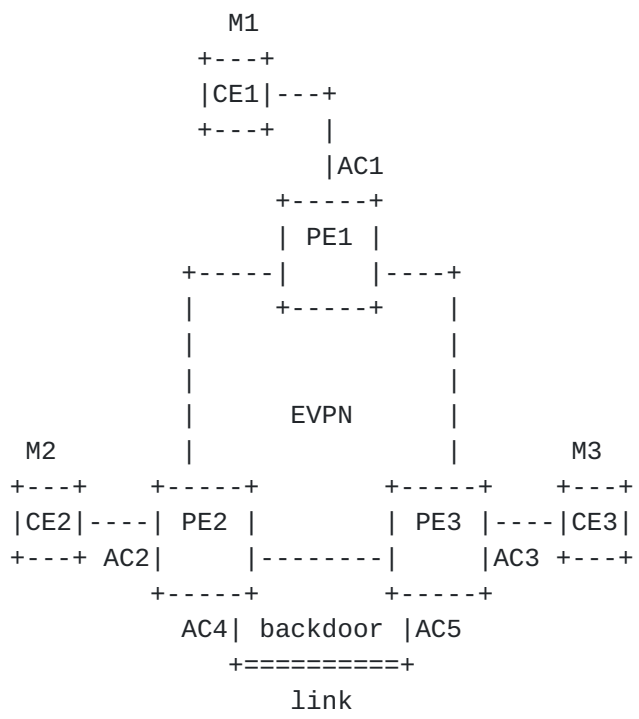

Figure 1 Backdoor link example in Service Provider EVPN networks

When, for instance, CE1 (in Figure 1) sends Broadcast, Unknown unicast or Multicast (BUM) traffic, the frames will be flooded to PE2 and PE3, looped to each other through the backdoor link and flooded back again in the EVPN network, creating an endless loop.

Figure 2 illustrates another example of backdoor path between NVEs in two remote Data Centers.

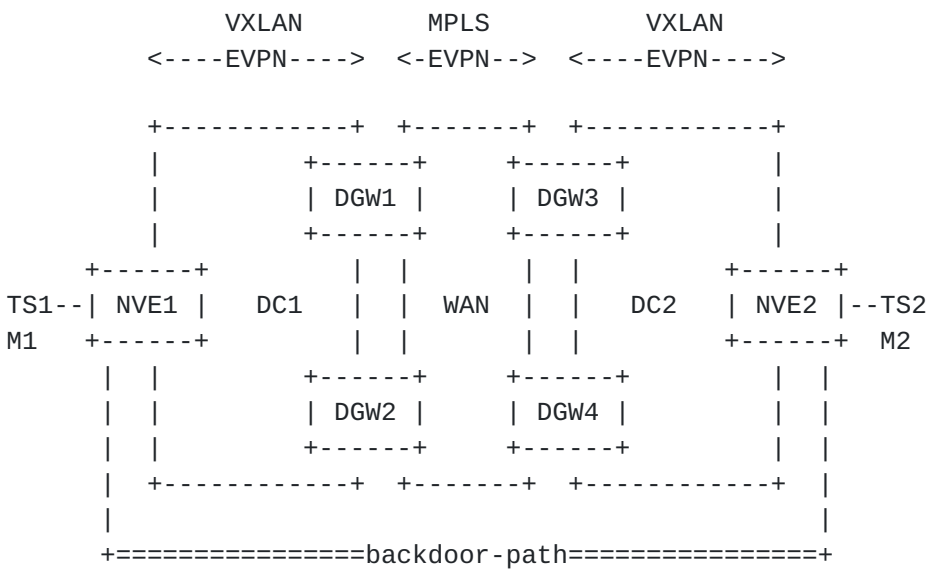

Figure 2 Backdoor path example in DCI EVPN networks

In Figure 2, a backdoor path is accidentally created between NVE1 and NVE2 in two remote Data Centers. BUM traffic generated by TS1 or TS2 will cause a layer-2 loop across DC1 and DC2.

2. Loop Protection Requirements in EVPN networks

The following requirements have been identified for loop protection in EVPN networks:

1- The EVPN PEs in a network MUST provide an automatic mechanism for detecting and resolving a loop within the same broadcast domain. In this document 'resolving a loop' refers to an automatic action executed by a PE or group of PEs that stops a frame from being endlessly forwarded back and forth between two PEs.

2- The Loop Protection mechanism MUST be compatible with all the procedures described in EVPN [RFC7432], in particular, it must not interfere with regular EVPN Multi-homing, MAC Mobility and MAC Protection procedures.

3- The Loop Resolution action SHOULD discard the looped flows without bringing down the Attachment Circuits (ACs) involved in the created loop. For example, when CE2 sends a broadcast frame (in Figure 1) the Loop Resolution action should discard the looped frames that are forwarded between PE2 and PE3 instead of bringing down any AC in the backdoor path.

4- The Loop Resolution action MAY bring down the ACs that are involved in the loop for a given flow instead of only discarding the identified looped frames. This action may impact some unicast flows that are not looped in the EVI, but provides an immediate solution to the loop situation. For example, when a loop (for BUM frames sent from CE1) is detected in PE3, the router may bring down the AC corresponding to the backdoor link.

5- A PE detecting a loop SHOULD log an event, warning the operator of the existence of a loop.

6- The operator SHOULD be able to configure whether the Loop Resolution action is manually or automatically cleared from a given PE, before the Loop Protection mechanism is restarted.

7- The solution MUST be compatible with other implementation-specific procedures that protect the PE against local loops.

3. Loop Protection Solution for EVPN networks

This document re-uses and enhances the MAC duplication solution specified in EVPN [RFC7432]. Section 3.1 clarifies this baseline EVPN MAC duplication mechanism and describes the required enhancements so that the EVPN network can protect the EVI user against loops.

3.1 The RFC7432 EVPN MAC Duplication Mechanism and Loop Protection

EVPN [RFC7432] describes a MAC duplication issue and how this anomaly is resolved. In this document, the terms VLAN and broadcast domain are used interchangeably. A VLAN is equivalent to an EVI in case of VLAN-based or VLAN Bundle services, and to a broadcast domain in case of VLAN-Aware Bundle services.

As per RFC7432, if a duplicate MAC situation exists in two or more hosts that are part of two different Ethernet Segments within the same VLAN, the traffic originating from these hosts would trigger continuous MAC moves among the PEs attached to them. If no action was made, the sequence number (in the MAC Mobility extended community attribute) would be incremented by the PEs to infinity.

In order to remedy such a situation, a PE that detects a MAC mobility event via local learning:

o Starts an M-second timer. M is configurable, with a default value of M = 180.

o If it detects N MAC moves before the timer expires, it concludes

that a duplicate-MAC situation has occurred and adds the MAC to a duplicate-MAC list. N is configurable with a default value of N = 5.

- The PE MUST alert the operator and stop sending and processing any BGP MAC/IP Advertisement routes for that MAC address until a corrective action is taken by the operator.

- While a MAC address is on the duplicate-MAC list for the VLAN, the other PEs in the EVI will forward the traffic for the duplicate-MAC address to one of the PEs that advertised it.

In the example of Figure 1, when CE1 sends BUM traffic to the EVI, the EVPN MAC Duplication Mechanism prevents an endless MAC/IP route exchange for M1 between PE1, PE2 and PE3. For instance, when MAC M1 moves N times in PE2 within the M-second timer period, PE2 will add M1 to the duplicate-MAC list for the broadcast domain and will stop advertising a MAC/IP route for M1. While this helps the control plane settle, Broadcast frames being sent by CE1 are still endlessly looped within the broadcast domain through the backdoor link. This may cause unpredictable issues in the CEs connected to the affected EVI.

3.2 Loop Protection Solution

This document enhances the EVPN MAC Duplication Mechanism by extending it with an optional Loop-protection action that is applied on the duplicate-MAC addresses. This additional mechanism resolves loops created by accidental backdoor links and SHOULD be enabled in all the PEs in the EVI.

Figure 3 outlines the Loop Protection solution when a backdoor link exists between two PEs (PE2 and PE3) in the same EVI and broadcast domain. The following assumptions are made:

- Loop Protection (this document) is enabled on (at least) PE3.
- PEs in the EVI are configured with window M-timer = M seconds and number of moves = N.
- PEs are also configured with a R-timer (retry-timer) = R seconds. This timer is explained later.
- In this document, a MAC-move refers to a relearn event in the same MAC-VRF, where the same MAC is first learned on an AC and later learned from BGP EVPN. Vice versa is also considered a MAC-move. Relearn events between two ACs in the same PE (i.e. local loops) or between two different EVPN endpoints are not considered. To protect the network against local loops, this procedure should be combined with local loop protection mechanisms.

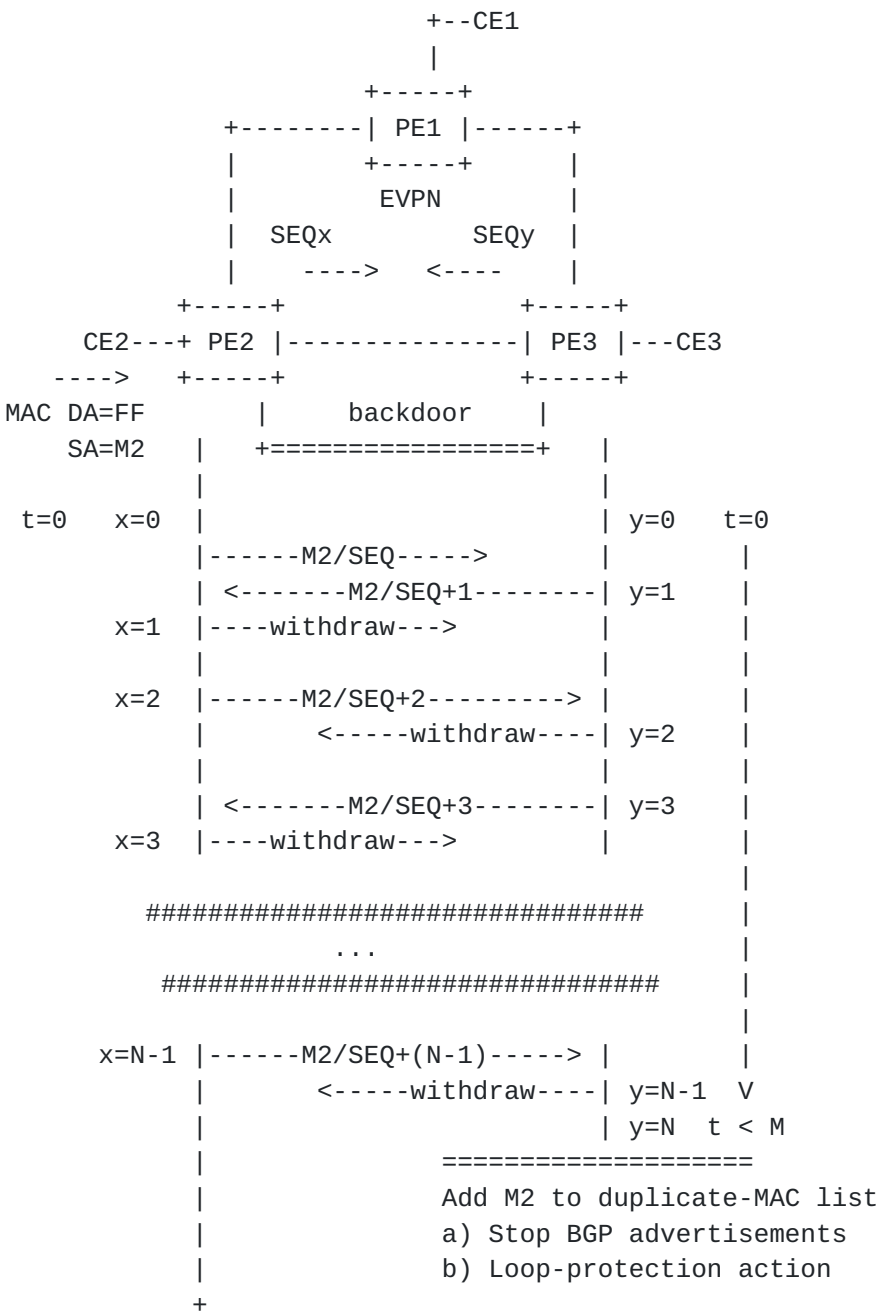

Figure 3 MAC Duplication and Loop Protection process

In the example of Figure 3, we assume CE2 sends a broadcast frame with MAC SA (Source Address) M2. We also assume PE3 learns M2 via BGP first, and via data path later. Although that is unlikely since data path learning is normally faster than BGP-based learning, it helps understand and generalize the procedure. The procedure will work as long as the PE detects N MAC-moves within M seconds for a given MAC.

The following process takes place:

T0 - PE2 receives the frame, learns M2 (if not learned before) and initializes counter x and timer t. Counter x stores the number of MAC moves, while t stores the delta time since the first MAC move for M2 occurred. PE2 advertises M2 with the currently stored Sequence Number (SEQ). Also, PE2 does a MAC DA (Destination Address) lookup and, since the MAC DA is a broadcast address, it floods the frame to PE1, PE3 and the AC on the backdoor link. This causes a loop between PE2 and PE3.

T1 - PE3 receives the BGP update and learns M2. Counter y and timer t are initialized. Counter y stores the number of moves for M2 and t stores the delta time since y was initialized. PE3 now advertises M2 with SEQ+1. M2/SEQ+1 route arrives at PE2 and it is installed in the MAC-VRF. The advertisement makes PE2 withdraw the MAC/IP route for M2 and increment x. Immediately after, PE2 receives the frame again through the backdoor link, relearns M2 locally, increments x and advertises M2 with SEQ+2.

T2 - M2/SEQ+2 route arrives at PE3 and it is installed in the MAC-VRF. The advertisement makes PE3 withdraw the MAC/IP route for M2 and increment y. PE3 receives the frame again through the backdoor link, relearns M2 locally, increments y and advertises M2 with SEQ+3. PE2 receives the route, relearns M2 and increments x. PE2 also withdraws the route for M2. Immediately after, PE2 receives the frame through the backdoor link and repeats the process (updates y and withdraws the route).

Since the frame (with MAC SA=M2) keeps being learned locally on the backdoor link ACs on PE2 and PE3, the above process is repeated until y reaches number of moves = N.

Tr - When y=N, PE3 compares t against the configured window M, and in case t<M, PE3 adds M2 to the duplicate-MAC list for the broadcast domain. Declaring M2 as duplicate triggers three actions:
  a) PE3 stops advertising M2 and logs a duplicate event.
  b) PE3 initializes a retry-timer.
  c) Since Loop Protection is enabled in PE3, PE3 executes the Loop Protection action, which we will refer to as "Black-holing" M2. When P3 programs M2 as a Black-Hole MAC in the MAC-VRF, M2 is no longer associated to the backdoor AC, but to a Black-Hole destination.

Ts - At this point and while M2 is in Black-Hole state:
  a) If a new frame is received at PE3 (from the EVPN core or the backdoor AC) with MAC SA = M2, PE3 will identify M2 as Black-

Hole and discard the frame, ending the loop.
b) Optionally, instead of simply discarding the frame with MAC SA = M2, PE3 MAY bring down the AC on which the offending frame is seen last. In this example, PE3 would bring down the backdoor AC, ending in that way the loop not only for frames from CE2, but for any traffic.
c) Optionally, any frame that arrives at PE3 with MAC DA = M2 SHOULD be discarded too.

Tt - When the retry-timer for M2 reaches R seconds, PE3 will flush M2 from the MAC-VRF and the process will be restarted.

Section 3.3 provides more details about the Black-Hole MAC in the context of this document.

3.3 The Black-Hole MAC concept for Loop Protection

As discussed in section 3.2, this document enhances the EVPN MAC Duplication mechanism by converting the detected duplicate-MAC addresses into Black-Hole MAC addresses and ending the forwarding plane loop. A Black-hole MAC is modeled as a special MAC-VRF record that has the following characteristics:

a) A Black-Hole MAC M is automatically installed in the MAC-VRF when M is detected as duplicate-MAC address.

b) When M is installed as Black-Hole MAC, for any ingress frame and irrespective of the frame arriving at an AC or network port:
i) If MAC SA = M the ingress frame MUST be discarded, without any further action.
ii) If MAC DA = M the ingress frame SHOULD be discarded, without any further action.

c) Optionally, any ingress frame with MAC SA = M arriving at an access AC, MAY trigger the PE to bring down the AC. Note that this approach cuts off the backdoor path that created the loop, preventing traffic from other MAC addresses from being forwarded, even if they are not identified as duplicate-MAC addresses yet.

d) A Black-Hole MAC M can be flushed from the MAC-VRF if any of the following events occur:
o Retry-timer R for duplicate-MAC M expires. R is initialized when M is detected as duplicate-MAC. Its value is configurable and SHOULD be at least three times the EVPN MAC Duplication M-timer window. According to EVPN [RFC7432], M's default value is 180 seconds, hence R's default value SHOULD be 540 seconds.
o The operator manually flushes a Black-Hole MAC M. This should be done only if the conditions under which M was identified as

duplicate have been cleared.
- The remote PE withdraws the MAC/IP route for M and there are no other remote MAC/IP routes for M.
- The remote PE sends a MAC/IP route update for M with the sticky-bit set (in the MAC Mobility extended community).

e) When a Black-Hole MAC is flushed from the MAC-VRF, the actions described in (b) and (c) are naturally reverted and the EVPN MAC Duplication and Loop Protection process will be restarted.

4. Conclusions

As EVPN is deployed in large layer-2 networks to deliver E-LAN or E-Tree services, it is important that the technology provides a solid protection against loops accidentally created by backdoor links. These backdoors can exist between CEs that can be connected anywhere in the EVI.

The EVPN [RFC7432] MAC Duplication Detection mechanism solves a situation where the same MAC has been accidentally configured on two or more hosts connected to different EVPN Ethernet Segments in the same broadcast domain. However, that mechanism does not provide a solution to resolve loops in those cases where the MAC duplication is caused by backdoor links between CEs.

This document leverages and extends the EVPN [RFC7432] MAC Duplication Detection mechanism by providing additional Loop Protection actions for the duplicate-MAC addresses.

6. Conventions used in this document

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].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

In this document, the characters ">>" preceding an indented line(s) indicates a compliance requirement statement using the key words listed above. This convention aids reviewers in quickly identifying or finding the explicit compliance requirements of this RFC.

7. Security Considerations

This section will be added in future versions.

8. IANA Considerations

This document does not require new codepoints.

9. Terminology

EVI: EVPN Instance.
E-LAN: MEF-based Ethernet Local Area Network service.
E-Tree: MEF-based Ethernet Tree service.
BUM: Broadcast, Unknown unicast and Multicast traffic.
AC: Attachment Circuit.
MAC-VRF: MAC Virtual Routing and Forwarding instance. Instantiation of an EVI in a PE.
xSTP: Any Spanning Tree based Protocol, e.g. STP, RSTP, MSTP.

9. References

9.1 Normative References

[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, <http://www.rfc-editor.org/info/rfc7432>.

9.2 Informative References

10. Acknowledgments

11. Contributors

17. Authors' Addresses

Jorge Rabadan
Nokia
777 E. Middlefield Road
Mountain View, CA 94043 USA
Email: jorge.rabadan@nokia.com

Senthil Sathappan
Nokia
Email: senthil.sathappan@nokia.com

Kiran Nagaraj
Nokia
Email: kiran.nagaraj@nokia.com

Julio Bueno
Telefonica
Email: julio.buenohernandez@telefonica.com

Jose Manuel Crespo
Telefonica
Email: josemanuel.crespogarcia@telefonica.com