BESS Working Group
Internet-Draft
Intended status: Standards Track
Expires: 26 March 2026

V. Govindan
A. Sajassi
Cisco Systems
M. Mudigonda
Celestica Inc.
G. Mirsky
Ericsson
D. Eastlake
Independent

22 September 2025

EVPN Network Layer Fault Management
draft-ietf-bess-evpn-bfd-12

Abstract

This document specifies proactive, in-band network layer OAM (RFC 9062) mechanisms to detect loss of continuity faults that affect unicast and multi-destination paths (used by Broadcast, Unknown Unicast, and Multicast traffic) in an Ethernet VPN (EVPN, RFC 7432bis) network. The mechanisms specified in this document use the widely adopted Bidirectional Forwarding Detection (RFC 5880) protocol.

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). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on 26 March 2026.

Copyright Notice

Table of Contents

1. Introduction

[RFC9062] outlines the OAM requirements of Ethernet VPN (EVPN) [rfc7432bis]. This document specifies mechanisms for proactive fault detection at the network (overlay) layer of EVPN, that is to say between Provider Edge (PE) nodes, as shown in Figure 1 taken from [RFC9062] and described in Section 2.3 of [RFC9062]. The mechanisms specified in this document use the widely adopted Bidirectional Forwarding Detection (BFD, [RFC5880] [RFC5881] [RFC5882] [RFC5883] [RFC5884]) protocol, which is a lightweight in-band protocol using fixed length messages suitable for implementation in hardware. All

these mechanisms are applied as active in-band OAM methods, i.e., specially constructed OAM packets traverse the same set of links and interfaces receiving the same forwarding behavior as the monitored EVPN flow. EVPN service restoration mechanisms (redundancy and recovery/convergence) are the most logical clients, in the [RFC5882] sense, for BFD sessions specified herein.

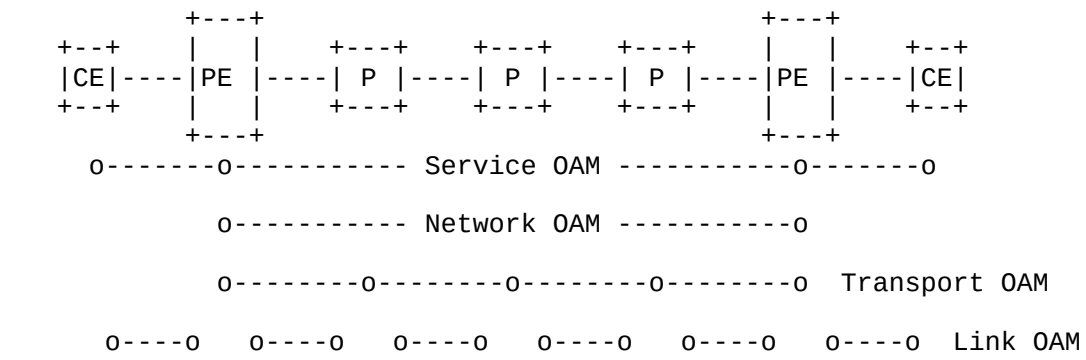

Figure 1: OAM Layering

EVPN Network Layer OAM monitors connectivity between peer Bridge-table instances. In the case of VLAN-based and VLAN Bundle service interfaces, this is equivalent to MAC-VRF-to-MAC-VRF connectivity because each MAC-VRF contains just one BT which is identified by the application label/VNI in the DP. In the case of VLAN-aware Bundle service interface:

* Each MAC-VRF contains multiple BTs.

* In EVPN-VxLAN each BT is still identified by its own VNI in the DP.

* In EVPN-MPLS a BT can be identified in the DP by its own application label, or by the combination of the application label and the "normalized VLAN tag" in the customer Ethernet frame. This document does not cover the case where the BT is identified using the VLAN tag but relies on the identification of the BT solely by using the application label.

EVPN fault detection mechanisms need to consider unicast traffic separately from Broadcast, Unknown Unicast, and Multicast (BUM) traffic since they map to different Forwarding Equivalency Classes (FECs) in EVPN so such traffic may follow different paths. Hence this document specifies different continuity fault detection mechanisms, depending on the type of traffic and the type of tunnel used, as follows (see also Section 2.3 of [RFC9062]):

* Using BFD [RFC5880] for unicast traffic and BUM traffic via Point to Point (P2P) and Multipoint to Point (MP2P) tunnels.

* Using BFD Multipoint [RFC8562] or BFD Multipoint Active Tails [RFC8563] [ietf-mpls-p2mp-bfd] for BUM traffic via a Point to Multipoint (P2MP) tunnels.

Packet loss and packet delay measurement are out of scope for this document. See [ietf-bmwg-evpntest] for EVPN benchmarking guidance.

At least three types of mis-programming of the application label can occur:

* It is not known in the data plane of the egress EVPN PE.

* It is known and is programmed with action "swap and forward".

* It is known and is programmed as "pop and forward" but to a different BT.

While the mechanisms of this document can cover the first two cases, the last case is not be covered by this document since that may require changes to rules for association of a received BFD packets as specified in [RFC5880] especially when EVPN encapsulated BFD packets are handled by OAM processors in off-the-shelf forwarding hardware.

The primary motivation of using BFD for detecting liveness is to report failures to an operator who may initiate corrective action. This document does not provide any mechanisms for repairing faults.

2. 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.

The following acronyms are used in this document.

BFD - Bidirectional Forwarding Detection [RFC5880]

BUM - Broadcast, Unknown Unicast, and Multicast

CC - Continuity Check

CE - Customer Edge

EVI - EVPN Instance

EVPN - Ethernet VPN [rfc7432bis]

FEC - Forwarding Equivalency Class

LSM - Label Switched Multicast (P2MP)

LSP - Label Switched Path

MP2MP - Multipoint to Multipoint

MP2P - Multipoint to Point

MPLS - MultiProtocol Label Switching

OAM - Operations, Administration, and Maintenance

P2MP - Point to Multipoint (LSM)

P2P - Point to Point

PE - Provider Edge

VXLAN - Virtual eXtensible Local Area Network [RFC7348]

3. Scope of this Document

This document specifies BFD-based mechanisms for proactive fault detection at the Network Layer (as specified in Section 2.3 of [RFC9062]) for MPLS based EVPN (as specified in [rfc7432bis]) and also for EVPN using VXLAN encapsulation [RFC8365]. Specifically, this document covers the following:

* Unicast traffic using Point to Point (P2P) and Multipoint to Point (MP2P) tunnels.

* BUM traffic using ingress replication via Point to Point (P2P) and Multipoint to Point (MP2P) tunnels.

* BUM traffic using Point to Multipoint (P2MP) tunnels (Label Switched Multicast (LSM)).

* MPLS and VXLAN encapsulation.

This document does not discuss BFD mechanisms for:

* The PBB-EVPN [RFC7623] EVPN variant. It is intended to address this in a future document.

* Monitoring EVPN E-Tree services [RFC8317] is outside the scope of this document since lack of connectivity between two BTs that have just Leaf ACs would have no impact on the EVPN BFD session.

* EVPN using other encapsulations such as NVGRE or MPLS over GRE (see Section 5 of [RFC8365]).

* BUM traffic using MP2MP tunnels.

This document specifies procedures for BFD asynchronous mode. BFD demand mode is outside the scope of this specification except as it is used in [RFC8563]. The use of the BFD Echo function is outside the scope of this specification.

4. Running BFD at the EVPN Network Layer

The following considerations motivated the use of BFD at the network layer of the OAM model for EVPN (Section 2.3 of [RFC9062]):

* In addition to detecting network failures in an EVPN network, BFD sessions at the network layer can be used to monitor the successful setup, such as label programming, of MP2P and P2MP EVPN tunnels transporting Unicast and BUM traffic. The scope of reachability detection covers the ingress and the egress EVPN PE (Provider Edge) nodes and the network connecting them.

* Monitoring a representative set of paths or a particular path among multiple paths available between two EVPN PE nodes could be done by exercising entropy mechanisms such as entropy labels, when they are used [RFC6790], or VXLAN source port numbers [RFC7348]. However, paths that cannot be realized by entropy variations cannot be monitored. The fault monitoring requirements outlined by Section 3.1.1.1 of [RFC9062] are addressed by the mechanisms specified in this document.

BFD sessions, as described herein, are requested when an EVPN route is established and the information necessary for the BFD session or sessions, as specified in Section 5 and Section 6, is available. Data is not sent over the EVPN route until the BFD session or sessions are in the UP state.

BFD testing between EVPN PE nodes does not guarantee that the EVPN service is functioning. This can be monitored at the service level, that is CE (Customer Edge) to CE (Section 2.2 of [RFC9062]) as shown in Figure 1. For example, an egress EVPN PE could recognize EVPN

labeling received and correctly process BFD packets but switch data to incorrect interfaces. However, BFD testing in the EVPN Network Layer does provide additional confidence that data transported using those tunnels will reach the expected egress node.

When BFD testing in the EVPN overlay fails, that can be used as an indication of a Loss-of-Connectivity defect in the EVPN underlay that would cause EVPN service failure; however, normally continuity checking at lower layers (the transport / link layers) SHOULD be done to detect underlay failures as this permits better localization of failures. Continuity testing at each layer SHOULD, if possible, be configured to detect failures more rapidly than at higher layers, for example with shorter BFD timers, for the following reasons:

* To facilitate repair, it is best to localize the area of a failure to as small an area as practical, which is best done by lower layers. If the failure is detected at a broader, higher layer area, while that detection can be reported, it may mask more localized lower layer failure detections.

* If lower layer failure detection is coupled with automatic lower layer path repair, premature detection at a higher layer may be a false detection in that the lower layer fault was in the process of being repaired.

5. Fault Detection for Unicast Traffic

The mechanisms specified in BFD for MPLS LSPs [RFC5884] [RFC7726] and BFD for VXLAN [RFC8971] are, except as otherwise provided herein, applied to test loss of continuity for unicast EVPN traffic. This includes the following provision of [RFC5884]:

> Note that once the BFD session for the MPLS LSP is UP, either end of the BFD session MUST NOT change the source IP address and the local discriminator values of the BFD Control packets it generates, unless it first brings down the session.

The MPLS control plane can be verified against the data plane as specified in [RFC8029]. When the discriminators required for de-multiplexing the BFD sessions are not otherwise available, for example by configuration, they can be advertised through BGP using the BFD Discriminator Attribute [RFC9026]. Discriminators are needed for MPLS since the label stack does not contain enough information to identify the sender of the packet.

The usage of different MPLS entropy labels [RFC6790] or different VXLAN source ports takes care of the requirement to monitor various paths of the multi-path provider network. Each unique realizable

path between the participating PE nodes MAY be monitored separately when such entropy is used. At least one path of multi-path connectivity between two PE nodes MUST be tracked with BFD, but in that case the granularity of fault-detection will be coarser.

To support unicast fault management with BFD packets sent to a PE node, that PE node MUST allocate or be configured with a BFD discriminator to be used as Your Discriminator (Section 4.1 of [RFC5880]) in the BFD messages to it. By default, a PE node advertises this discriminator with BGP using the BFD Discriminator Attribute [RFC9026] with BFD Mode TBD2 in an EVPN Ethernet Autodiscovery Route [rfc7432bis] or MAC/IP Advertisement Route as long as it advertises it in at least one route. It extracts its peer's discriminator from such an attribute. However, these discriminators MAY be exchanged out-of-band or through some other mechanism outside the scope of this document.

Once a PE node knows a unicast route and discriminator for another PE node and is configured to do so, it endeavors to bring UP and maintain a BFD session to that other PE node. The BFD session is brought down if a PE node is no longer configured to maintain it or if a route and discriminator are no longer available.

6. Fault Detection for BUM Traffic

Section 5.1 below discusses BUM traffic fault detection for P2P and MP2P tunnels using ingress replication and Section 5.2 discusses such fault detection for P2MP tunnels. In both cases the following provision of [RFC5884] applies:

> | Note that once the BFD session for the MPLS LSP is UP, either end
> | of the BFD session MUST NOT change the source IP address and the
> | local discriminator values of the BFD Control packets it
> | generates, unless it first brings down the session.

6.1. Ingress Replication

Ingress replication (see Section 11 of [rfc7432bis]) uses separate P2P or MP2P tunnels for transporting BUM traffic from the ingress PE (head) to a set of one or more egress PEs (tails). The fault detection mechanism specified by this document takes advantage of the fact that the head makes a unique copy for each tail.

Another key aspect to be considered in EVPN is the advertisement of the Inclusive Multicast Ethernet Tag Route (see Section 7.3 of [rfc7432bis]). The BUM traffic flows from a head node to a particular tail only after the head receives such an inclusive multicast route from the tail. This route contains the BUM EVPN MPLS

label (downstream allocated) corresponding to the MP2P tunnel for MPLS encapsulation and contains the IP address of the PE originating the inclusive multicast route for use in VXLAN encapsulation. It also contains a BFD Discriminator Attribute [RFC9026] with BFD Mode TBD2 giving the BFD discriminator that will be used by the tail unless this information has been otherwise distributed. This is the P2P mode BFD Discriminator Attribute since a P2P BFD session is used in both the P2P and MP2P cases with ingress replication.

There MAY exist multiple BFD sessions between a head PE and an individual tail due to (1) the usage of MPLS entropy labels [RFC6790] or VXLAN source port numbers for an inclusive multicast FEC and (2) due to multiple MP2P tunnels indicated by different tail labels for MPLS or different IP addresses for VXLAN encapsulation. If a PE node is configured to do so, once it knows a multicast route and discriminator for another PE mode it endeavors to bring UP and maintain a BFD session to that other PE node. The BFD session is brought down if a PE node is no longer configured to maintain it or if a route and discriminator are no longer available.

6.2. P2MP Tunnels (Label Switched Multicast)

Fault detection for BUM traffic distributed using a P2MP tunnel uses BFD Multipoint Active Tails [RFC8563] in one of the three methods providing head notification. Which method is used depends on the local configuration. Sections 5.2.2 and 5.2.3 of [RFC8563] describe two of these methods ("Head Notification and Tail Solicitation with Multipoint Polling" and "Head Notification with Composite Polling"). The third method ("Head Notification without Polling") is touched on in Section 5.2.1 of [RFC8563] and fully specified in [ietf-mpls-p2mp-bfd]. All these three modes assume the existence of a unicast return path from each tail to the head. In addition, Head Notification with Composite Polling assumes a head to tail unicast path disjoint from the path used by the P2MP tunnel.

The BUM traffic flows from a head node to the tails after the head transmits an Inclusive Multicast Tag Route [rfc7432bis] if local configuration so directs. This route contains the BUM EVPN MPLS label (upstream allocated) corresponding to the P2MP tunnel for MPLS encapsulation. The route also includes a BFD Discriminator Attribute [RFC9026] with the BFD Mode set to 1 and a Source IP Address TLV, which gives the address associated with the MultiPoint Head of the P2MP session. This BFD discriminator advertised by the head in the Inclusive Multicast route or otherwise configured at or communicated to a tail MUST be used in any reverse BFD control message as Your Discriminator so the head can determine the tail of which P2MP BFD session is responding. If a PE node is configured to do so, once a PE knows a P2MP multicast route and the needed discriminators, it

brings UP and maintains a P2MP BFD active tails session to the tails. The BFD session is brought down if a PE node is no longer configured to maintain it or the multicast route and discriminators are no longer available.

For MPLS encapsulation of the head to tails BFD, Label Switched Multicast is used. For VXLAN encapsulation, BFD is delivered to the tails through underlay multicast using an outer multicast IP address.

7. BFD Packet Encapsulation

The following subsections describe the MPLS and VXLAN encapsulations of BFD for EVPN network layer fault management:

7.1. MPLS Encapsulation

This section describes use of the Generic Associated Channel Label (GAL, [RFC5586]) for BFD encapsulation in MPLS-based EVPN network layer fault management. Since the use of BFD specified in this document is encapsulated between PEs, it is treated as single hop and uses the single hop BFD port number [RFC5881].

7.1.1. Unicast MPLS Encapsulation

As shown in Figure 2, the packet contains the following labels in the order given: LSP label (transport), optionally an entropy label, the EVPN Unicast label, and then the Generic Associated Channel label with the G-ACh type set to TBD1. The G-ACh payload of the packet MUST contain the destination L2 header (in overlay space) followed by the IP header that encapsulates the BFD packet. The source MAC address of the inner packet can be used to validate the <EVI, MAC> in the receiving node.

* The destination MAC address MUST be the dedicated unicast MAC TBD4 (see Section 8) or the MAC address of the destination PE node.

* The destination IP address MUST be 127.0.0.1/32 for IPv4 or ::1/128 for IPv6 [RFC6890].

* The destination UDP port number MUST be 3784 [RFC5881].

* The source UDP port number MUST be in the range 49152 through 65535.

* The discriminator values for BFD are obtained as discussed in Section 4.

* IPv4 TTL or IPv6 Hop Limit MUST be set to 255 according to [RFC5082].

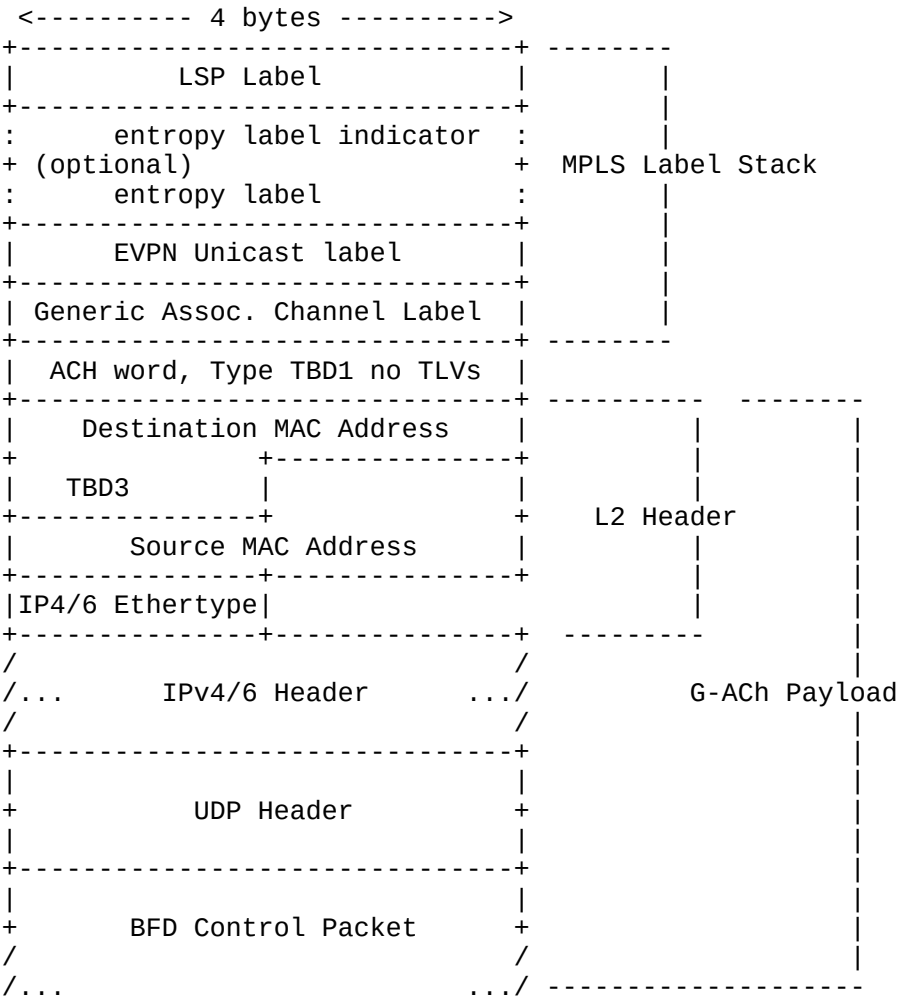

Figure 2: MPLS Unicast Encapsulation

7.1.2. MPLS Ingress Replication

When ingress replication is used for BUM traffic, a packet contains the following labels in the order given: LSP label (transport), optionally an entropy label, the BUM label, and the split horizon label [rfc7432bis] (where applicable). The G-ACh type is set to TBD1. The G-ACh payload of the packet is as described in Section 6.1.1 except that the destination MAC address, if not that of

the destination PE node, is the dedicated multicast MAC TBD3.

7.1.3. MPLS LSM (Label Switched Multicast, P2MP)

When Label Switched Multicast is used for BUM traffic, the encapsulation is the same as in Section 6.1.2 for ingress replication except that the transport label identifies the P2MP tunnel, in effect the set of tail PEs, rather than identifying a single destination PE at the end of an MP2P tunnel.

7.2. VXLAN Encapsulation

This section describes the use of the VXLAN [RFC7348] [RFC8365] for BFD encapsulation in VXLAN based EVPN fault management.

7.2.1. Unicast VXLAN Encapsulation

Figure 3 below shows the unicast VXLAN encapsulation on the wire on an Ethernet link. The outer and inner IP headers have a unicast source and destination IP address, both IPv4 or both IPv6 in each header, that are the addresses of the PE nodes that are the BFD message source and destination. The source port number MAY be varied as a source of entropy. If the BFD source has multiple IP addresses, whether multiple IPv4 addresses, multiple IPv6 addresses, or a mixture thereof, entropy MAY be further obtained by using any of those addresses assuming the destination has a same version IP address and the source is prepared for responses directed to the IP address used.

* The outer destination UDP port number MUST be 4789 [RFC7348].

* The inner destination UDP port number MUST be 3784 [RFC5881].

* The outer and inner source UDP port numbers MUST each be in the range 49152 through 65535.

* The inner destination MAC number MUST be the MAC address of the destination PE or the dedicated unicast BFD over VXLAN address 00-00-5E-00-52-02 (see [RFC8971]).

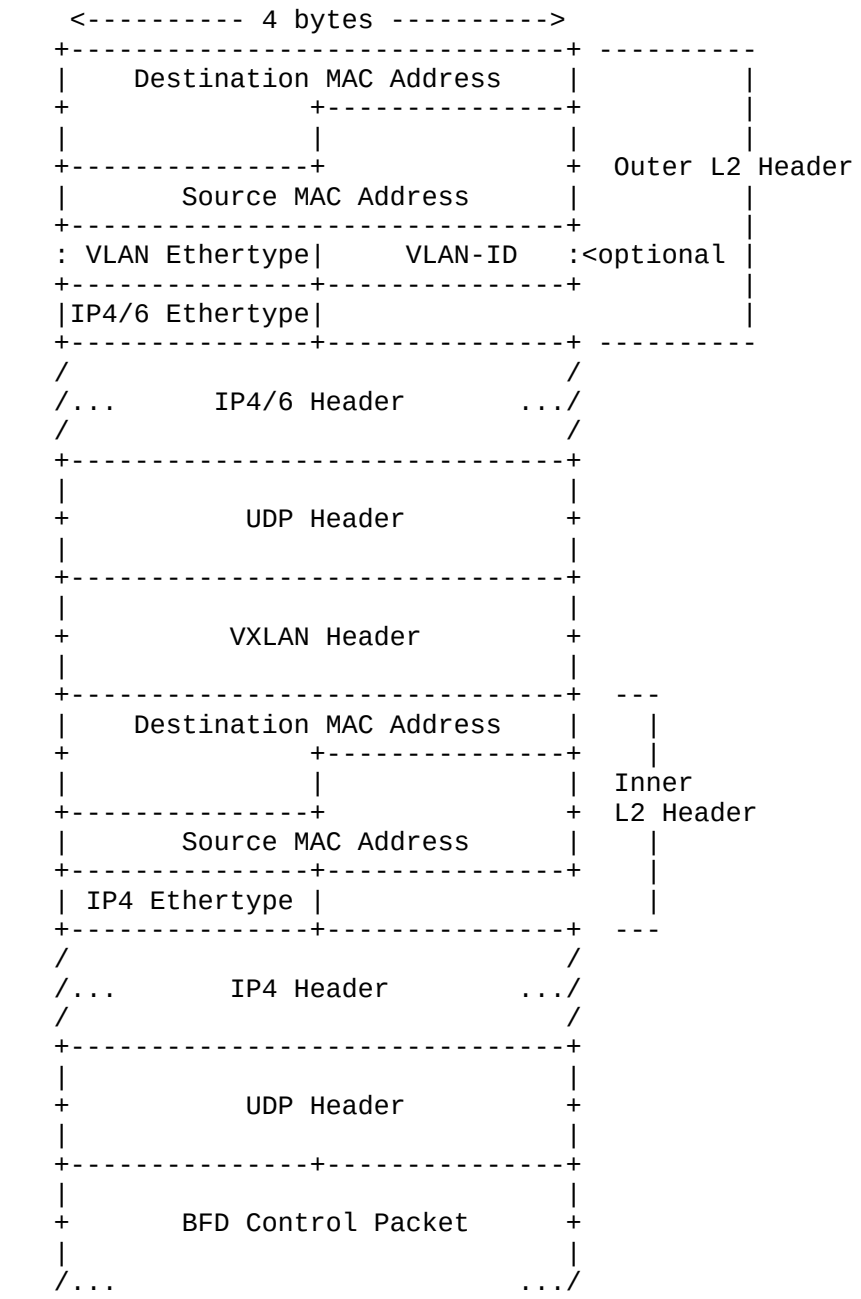

Figure 3: VXLAN Unicast Encapsulation

7.2.2. VXLAN Ingress Replication

When VXLAN encapsulated ingress replication is used, the BFD packet construction is as given in Section 6.2.1 except as follows:

1. The destination IP address used by the BFD message source is that advertised by the destination PE in its Inclusive Multicast EVPN route for the MP2P tunnel in question; and

2. The Your BFD discriminator used is the one advertised by the BFD destination using BGP as discussed in Section 5.1 for the MP2P tunnel.

3. The destination MAC address, if not that of the destination PE node, is the dedicated multicast BFD over VXLAN MAC address TBD5.

7.2.3. VXLAN P2MP

When VXLAN head-to-tails (P2MP) is used, the encapsulation is as given in Section 6.2.2 except as follows:

1. The multicast destination IP address used is that corresponding to the VXLAN VNI.

2. The Your BFD discriminator is the value distributed for this multicast fault management purpose as discussed in Section 5.2.

8. Scalability Considerations

The mechanisms specified by this document could affect the packet load on the network and its elements especially when supporting configurations involving a large number of EVIs. The option of decreasing or increasing BFD timer values can be used by an administrator or a network management entity to maintain the overhead incurred due to fault monitoring at an acceptable level.

9. IANA Considerations

The following IANA Actions are requested.

IANA is requested to assign a channel type from the "Pseudowire Associated Channel Types" registry in [RFC4385] as follows.

Value	Description	Reference
TBD1	BFD-EVPN OAM	[this document]

IANA is requested to assign a value from the IETF Review range in the BFD Mode sub-registry on the Border Gateway Protocol Parameters Registry web page as follows:

Value	Description	Reference
TBD2	P2P BFD Session	[this document]

IANA is requested to assign parallel multicast and unicast MAC addresses under the IANA OUI [0x01005E900101 and 0x00005E900101 suggested] as follows:

IANA Multicast 48-bit MAC Addresses

Address	Usage	Reference
TBD3	EVPN Network Layer OAM	[this document]

IANA Unicast 48-bit MAC Addresses

Address	Usage	Reference
TBD4	EVPN Network Layer OAM	[this document]

IANA is requested to assign a multicast MAC address under the IANA OUI [00-00-0E-90-00-04 suggested] as follows:

IANA Multicast 48-bit MAC Addresses

Address	Usage	Reference
TBD5	Multicast BFD over VXLAN	[this document]

10. Security Considerations

Security considerations discussed in [RFC5880], [RFC5883], and [RFC8029] apply.

MPLS security considerations [RFC5920] apply to BFD Control packets encapsulated in a MPLS label stack. When BFD Control packets are routed, the authentication considerations discussed in [RFC5883] should be followed.

VXLAN BFD security considerations in [RFC8971] apply to BFD packets encapsulated in VXLAN.

11. Acknowledgements

The authors place on record very special thanks to Alexander Vainshtein for his thoughtful comments and follow-up in shaping this document.

The authors wish to thank the following for their comments and suggestions: Mach Chen, Jorge Rabadan, and Mohammed Boucadair

12. Normative References

[ietf-mpls-p2mp-bfd]
Mirsky, G., Mishra, G., and D. Eastlake, "BFD for Multipoint Networks over Point-to-Multi-Point MPLS LSP", December 2022, <https://datatracker.ietf.org/doc/draft-ietf-mpls-p2mp-bfd/>.

[rfc7432bis]
Sajassi, A., Burdet, LA., Drake, J., and J. Rabadan, "BGP MPLS-Based Ethernet VPN", 13 March 2023, <https://datatracker.ietf.org/doc/draft-ietf-bess-rfc7432bis/>.

[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>.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, DOI 10.17487/RFC4385, February 2006, <https://www.rfc-editor.org/info/rfc4385>.

[RFC5586] Bocci, M., Ed., Vigoureux, M., Ed., and S. Bryant, Ed., "MPLS Generic Associated Channel", RFC 5586, DOI 10.17487/RFC5586, June 2009, <https://www.rfc-editor.org/info/rfc5586>.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <https://www.rfc-editor.org/info/rfc5880>.

[RFC5881] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)", RFC 5881, DOI 10.17487/RFC5881, June 2010, <https://www.rfc-editor.org/info/rfc5881>.

[RFC5882] Katz, D. and D. Ward, "Generic Application of Bidirectional Forwarding Detection (BFD)", RFC 5882, DOI 10.17487/RFC5882, June 2010, <https://www.rfc-editor.org/info/rfc5882>.

[RFC5883] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for Multihop Paths", RFC 5883, DOI 10.17487/RFC5883, June 2010, <https://www.rfc-editor.org/info/rfc5883>.

[RFC5884] Aggarwal, R., Kompella, K., Nadeau, T., and G. Swallow, "Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)", RFC 5884, DOI 10.17487/RFC5884, June 2010, <https://www.rfc-editor.org/info/rfc5884>.

[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>.

[RFC6890] Cotton, M., Vegoda, L., Bonica, R., Ed., and B. Haberman, "Special-Purpose IP Address Registries", BCP 153, RFC 6890, DOI 10.17487/RFC6890, April 2013, <https://www.rfc-editor.org/info/rfc6890>.

[RFC7726] Govindan, V., Rajaraman, K., Mirsky, G., Akiya, N., and S. Aldrin, "Clarifying Procedures for Establishing BFD Sessions for MPLS Label Switched Paths (LSPs)", RFC 7726, DOI 10.17487/RFC7726, January 2016, <https://www.rfc-editor.org/info/rfc7726>.

[RFC8029] Kompella, K., Swallow, G., Pignataro, C., Ed., Kumar, N., Aldrin, S., and M. Chen, "Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures", RFC 8029, DOI 10.17487/RFC8029, March 2017, <https://www.rfc-editor.org/info/rfc8029>.

[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>.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <https://www.rfc-editor.org/info/rfc8365>.

[RFC8562] Katz, D., Ward, D., Pallagatti, S., Ed., and G. Mirsky, Ed., "Bidirectional Forwarding Detection (BFD) for Multipoint Networks", RFC 8562, DOI 10.17487/RFC8562, April 2019, <https://www.rfc-editor.org/info/rfc8562>.

[RFC8563] Katz, D., Ward, D., Pallagatti, S., Ed., and G. Mirsky, Ed., "Bidirectional Forwarding Detection (BFD) Multipoint Active Tails", RFC 8563, DOI 10.17487/RFC8563, April 2019, <https://www.rfc-editor.org/info/rfc8563>.

[RFC9026] Morin, T., Ed., Kebler, R., Ed., and G. Mirsky, Ed., "Multicast VPN Fast Upstream Failover", RFC 9026, DOI 10.17487/RFC9026, April 2021, <https://www.rfc-editor.org/info/rfc9026>.

13. Informative References

[ietf-bmwg-evpntest] Jacob, S. and K. Tiruveedhula, "Benchmarking Methodology for EVPN and PBB-EVPN", June 2021, <https://datatracker.ietf.org/doc/draft-ietf-bmwg-evpntest/>.

[RFC5082] Gill, V., Heasley, J., Meyer, D., Savola, P., Ed., and C. Pignataro, "The Generalized TTL Security Mechanism (GTSM)", RFC 5082, DOI 10.17487/RFC5082, October 2007, <https://www.rfc-editor.org/info/rfc5082>.

[RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, DOI 10.17487/RFC5920, July 2010, <https://www.rfc-editor.org/info/rfc5920>.

[RFC8317] Sajassi, A., Ed., Salam, S., Drake, J., Uttaro, J., Boutros, S., and J. Rabadan, "Ethernet-Tree (E-Tree) Support in Ethernet VPN (EVPN) and Provider Backbone Bridging EVPN (PBB-EVPN)", RFC 8317, DOI 10.17487/RFC8317, January 2018, <https://www.rfc-editor.org/info/rfc8317>.

[RFC7623] Sajassi, A., Ed., Salam, S., Bitar, N., Isaac, A., and W. Henderickx, "Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)", RFC 7623, DOI 10.17487/RFC7623, September 2015, <https://www.rfc-editor.org/info/rfc7623>.

[RFC8971] Pallagatti, S., Ed., Mirsky, G., Ed., Paragiri, S., Govindan, V., and M. Mudigonda, "Bidirectional Forwarding Detection (BFD) for Virtual eXtensible Local Area Network (VXLAN)", RFC 8971, DOI 10.17487/RFC8971, December 2020, <https://www.rfc-editor.org/info/rfc8971>.

[RFC9062] Salam, S., Sajassi, A., Aldrin, S., Drake, J., and D. Eastlake 3rd, "Framework and Requirements for Ethernet VPN (EVPN) Operations, Administration, and Maintenance (OAM)", RFC 9062, DOI 10.17487/RFC9062, June 2021, <https://www.rfc-editor.org/info/rfc9062>.

[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>.

Appendix A. Special considerations for EVPN BFD

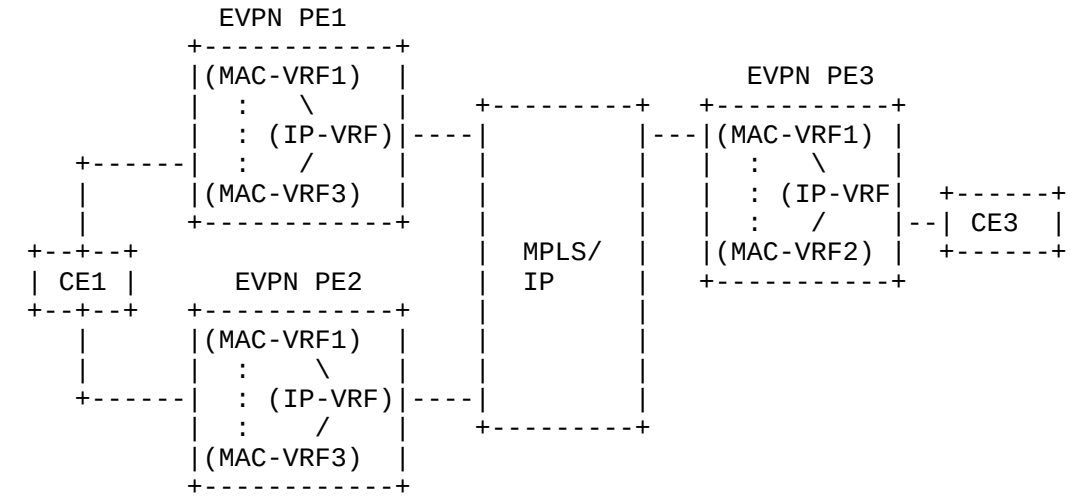

Figure 4: EVPN & EVPN PE scenario for BFD

The following are the topological aspects considered based on the above figure:

* A group of EVI implementing VLAN-based service interface (for simplicity) is instantiated in PE1, PE2 and PE3.

* PE1 and PE2 are attached to a multi-homed Ethernet Segment in Single-Active redundancy mode.

* PE3 is not attached to this multi-Homed Ethernet Segment.

* Some of the EVI (say, EVI-1 and EVI-2) in this group elects PE1 as the Designated Forwarder for this Multihomed Etherent Segment.

* At least one other EVI (say, EVI-3) selects PE2 as the DF for this MH ES.

* An EVPN BFD session has been set up for monitoring unicast network layer connectivity between a pair MAC-VRFs that locally EVI-1 in PE1 and PE3 and between PE2 and PE3.

When due to mis-programming of the application label used by MAC-VRF that represents EVI-1 in PE1:

* The BFD session endpoint in PE1 stops receiving EVPN-BFD packets and goes to DOWN state due to a local timeout.

* This BFD session endpoint indicates its DOWN state in the BFD Control packets it sends, so that the BFD session endpoint goes DOWN also in PE3. It must be noted that the BFD session at PE3 goes down because PE1 signals the BFD Down State using Diagnostic code "Neighbor Signaled Session Down" [RFC5880].

* No other failures will be detected (e.g. in PE2).

What actions can be taken at the individual PE Nodes are described below:

* There would be no actions required in PE2 since it is least impacted due to the failure.

* At PE3 the action could limited to reporting the fault to the operator. using the Diagnostic Code received from PE1 [RFC5880]

* At PE1 there are various actions that could be considered:

  - Notifying the operator about the fault is the basic operation that can be carried out in PE1 as well. This is mandatory.

  - Additionally, there could be more optional actions considered. For example, if the fault at PE1 causes BFD failures greater than a certain number of sessions (minimum threshold) then the EVPN PE1 can decide to force a DF election by withdrawing the Ethernet Segment Route. How such procedures are done are outside the scope of this specification.

Authors' Addresses

Vengada Prasad Govindan
Cisco Systems
Email: venggovi@cisco.com

Ali Sajassi
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
United States of America
Email: sajassi@cisco.com

Mudigonda Mallik
Celestica Inc.
Email: mallik.mj@celestica.com

Gregory Mirsky
Ericsson
Email: gregmirsky@gmail.com

Donald E. Eastlake 3rd
Independent
2386 Panoramic Circle
Apopka, FL 32703
United States of America
Phone: +1-508-333-2270
Email: d3e3e3@gmail.com