Fault Management for EVPN networks
draft-gmsm-bess-evpn-bfd-01

Abstract

This document specifies a proactive, in-band network OAM mechanism to detect loss of continuity and miss-connection faults that affect unicast and multi-destination paths, used by Broadcast, unknown Unicast and Multicast traffic, in an EVPN network. The mechanisms proposed in the draft use the widely adopted Bidirectional Forwarding Detection (BFD) protocol.
Table of Contents

1. Introduction ............................................ 3
   1.1 Terminology ........................................ 3
2. Scope of this Document ................................. 4
3. Motivation for Running BFD at the EVPN Network Layer... 4
4. Fault Detection of Unicast Traffic ...................... 6
5. Fault Detection for BUM Traffic .......................... 7
   5.1 Ingress Replication .................................... 7
   5.2 Label Switched Multicast .............................. 7
6. BFD Packet Encapsulation ................................ 8
   6.1 Using GAL/G-ACh Encapsulation Without IP Headers.... 8
   6.1.1 Ingress Replication ................................ 8
   6.1.1.1 Alternative Encapsulation Format ................. 8
   6.1.2 LSM (Label Switched Multicast) ....................... 9
   6.1.3 Unicast .............................................. 9
   6.1.3.1 Alternative Encapsulation Format ................ 9
   6.2 Using IP Headers ....................................... 10
7. Scalability Considerations .............................. 11
8. IANA Considerations ...................................... 12
9. Security Considerations .................................. 13

Normative References ...................................... 14
Informative References .................................... 15

Authors’ Addresses ........................................ 17
1. Introduction

[I-D.eastlake-bess-evpn-oam-req-frmwk] and [I-D.ooamdt-rtgwg-ooam-requirement] outline the OAM requirements of Ethernet VPN networks [RFC7432]. This document proposes mechanisms for proactive fault detection at the network (overlay) OAM layer of EVPN. EVPN fault detection mechanisms need to consider unicast traffic separately from Broadcast, unknown Unicast, and Multicast (BUM) traffic since they map to different FECs in EVPN, hence this document proposes different fault detection mechanisms to suit each type using the principles of [RFC5880], [RFC5884] and Point-to-multipoint BFD [I-D.ietf-bfd-multipoint] and [I-D.ietf-bfd-multipoint-active-tail]. Packet loss and packet delay measurement are out of scope for this document.

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.

The following acronyms are used in this document.

BUM - Broadcast, Unknown Unicast, and Multicast
CC - Continuity Check
CV - Connectivity Verification
FEC - Forwarding Equivalency Class
GAL - Generic Associated Channel Label
LSM - Label Switched Multicast (P2MP)
LSP - Label Switched Path
MP2P - Multi-Point to Point
OAM - Operations Administration, and Maintenance
P2MP - Point to Multi-Point (LSM)
PE - Provider Edge
PHP - Penultimate Hop Popping
2. Scope of this Document

This document specifies proactive fault detection for EVPN [RFC7432] using BFD mechanisms for:

- Unicast traffic.
- BUM traffic using Multi-point-to-Point (MP2P) tunnels (ingress replication).
- BUM traffic using Point-to-Multipoint (P2MP) tunnels (LSM).

This document does not discuss BFD mechanisms for:

- EVPN variants like PBB-EVPN [RFC7623]. This will be addressed in future versions.
- Integrated Routing and Bridging (IRB) solution based on EVPN [I-D.ietf-bess-evpn-inter-subnet-forwarding]. This will be addressed in future versions.
- EVPN using other encapsulations like VxLAN, NVGRE and MPLS over GRE [RFC8365].
- BUM traffic using MP2MP tunnels will also be addressed in a future version of this document.

This specification describes procedures only for BFD asynchronous mode. BFD demand mode is outside the scope of this specification. Further, the use of the Echo function is outside the scope of this specification.

3. Motivation for Running BFD at the EVPN Network Layer

The choice of running BFD at the network layer of the OAM model for EVPN [I-D.eastlake-bess-evpn-oam-req-frmwk] and [I-D.oamdt-rtgwg-oam-requirement] was made after considering the following:

- In addition to detecting link failures in the EVPN network, BFD sessions at the network layer can be used to monitor the successful programming of labels used for setting up MP2P and P2MP EVPN tunnels transporting Unicast and BUM traffic. The scope of reachability detection covers the ingress and the egress EVPN PE nodes and the network connecting them.

- Monitoring a representative set of path(s) or a particular path among the multiple paths available between two EVPN PE nodes could be done by exercising the entropy labels when they are used.
However paths that cannot be realized by entropy variations cannot be monitored. Fault monitoring requirements outlined by [I-D.eastlake-bess-evpn-oam-req-frmwk] are addressed by the mechanisms proposed by this draft.

Successful establishment and maintenance of BFD sessions between EVPN PE nodes does not fully guarantee that the EVPN service is functioning. For example, an egress EVPN-PE can understand the EVPN label but could switch data to incorrect interface. However, once BFD sessions in the EVPN Network Layer reach UP state, it does provide additional confidence that data transported using those tunnels will reach the expected egress node. When the BFD session in EVPN overlay goes down that can be used as an indication of a Loss-of-Connectivity defect in the EVPN underlay that would cause EVPN service failure.
4. Fault Detection of Unicast Traffic

The mechanisms specified in BFD for MPLS LSPs [RFC5884] [RFC7726] can be applied to bootstrap and maintain BFD sessions for unicast EVPN traffic. The discriminators required for de-multiplexing the BFD sessions MUST be exchanged using EVPN LSP ping specifying the Unicast EVPN FEC [I-D.jain-bess-evpn-lsp-ping] before establishing the BFD session. This is needed since the MPLS label stack does not contain enough information to disambiguate the sender of the packet.

The usage of MPLS entropy labels takes care of the requirement to monitor various paths of the multi-path server layer network [RFC6790]. Each unique realizable path between the participating PE routers MAY be monitored separately when entropy labels are used. The multi-path connectivity between two PE routers MUST be tracked by at least one representative BFD session, but in that case the granularity of fault-detection would be coarser. The PE node receiving the EVPN LSP ping MUST allocate BFD discriminators using the procedures defined in [RFC7726]. Once the BFD session for the EVPN label is UP, the ends of the BFD session MUST NOT change the local discriminator values of the BFD Control packets they generate, unless they first brings down the session as specified in [RFC5884].
5. Fault Detection for BUM Traffic

5.1 Ingress Replication

Ingress replication uses separate 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 a unique copy is made by the head for each tail. Another key aspect to be considered in EVPN is the advertisement of the inclusive multicast route. The BUM traffic flows from a head node to a particular tail only after the head receives the inclusive multicast route containing the BUM EVPN label (downstream allocated) corresponding to the MP2P tunnel.

The head-end PE performing ingress replication MUST initiate an EVPN LSP ping using the inclusive multicast FEC [I-D.jain-bess-evpn-lsp-ping] upon receiving an inclusive multicast route from a tail to bootstrap the BFD session. There MAY exist multiple BFD sessions between a head PE and an individual tail due to the usage of entropy labels [RFC6790] for an inclusive multicast FEC. The PE node receiving the EVPN LSP ping MUST allocate BFD discriminators using the procedures defined in [RFC7726]. Once the BFD session for the EVPN label is UP, the ends of the BFD session MUST NOT change the local discriminator values of the BFD Control packets they generate, unless they first brings down the session as specified in [RFC5884].

5.2 Label Switched Multicast

Fault detection for BUM traffic distributed by a Label Switched Multicast (LSM) using a P2MP tunnel is done with active tail multipoint BFD in the reliable head notification scenario (see [I-D.ietf-bfd-multipoint] and [I-D.ietf-bfd-multipoint-active-tail] particularly Section 3.4).
6. BFD Packet Encapsulation

6.1 Using GAL/G-ACh Encapsulation Without IP Headers

This section describes use of the Generic Associated Channel Label (GAL/G-ACh).

6.1.1 Ingress Replication

The packet contains the following labels: LSP label (transport) when not using PHP (Penultimate Hop Popping), the optional entropy label, the BUM label and the SH label [RFC7432] (where applicable). The G-ACh type is set to TBD1. The G-Ach payload of the packet MUST contain the L2 header (in overlay space) followed by the IP header encapsulating the BFD packet. The MAC address of the inner packet is used to validate the <EVI, MAC> in the receiving node. The discriminator values of BFD are obtained through negotiation through the out-of-band EVPN LSP ping.

6.1.1.1 Alternative Encapsulation Format

A new TLV can be defined as proposed in Sec 3 of [RFC6428] to include the EVPN FEC information as a TLV following the BFD Control packet.

The format of the TLV can be reused from the EVPN Inclusive Multicast sub-TLV proposed by Fig 2 of [I-D.jain-bess-evpn-lsp-ping].

A new type (TBD3) to indicate the EVPN Inclusive Multicast SubTLV is requested from the "CC/ CV MEP-ID TLV" registry [RFC6428].
6.1.2 LSM (Label Switched Multicast)

TBD...

6.1.3 Unicast

The packet contains the following labels: LSP label (transport) when not using PHP, the optional entropy label and the EVPN Unicast label. The G-ACh type is set to TBD1. The G-Ach payload of the packet MUST contain the L2 header (in overlay space) followed by the IP header encapsulating the BFD packet. The MAC address of the inner packet is used to validate the <EVI, MAC> in the receiving node. The discriminator values for BFD are obtained through negotiation using the out-of-band EVPN ping.

6.1.3.1 Alternative Encapsulation Format

A new TLV can be defined as proposed in Sec 3 of [RFC6428] to include the EVPN FEC information as a TLV following the BFD Control packet. The format of the TLV can be reused from the EVPN MAC sub-TLV proposed by Figure 1 of [I-D.jain-bess-evpn-lsp-ping]. A new type (TBD4) to indicate the EVPN MAC SubTLV is requested from the "CC/ CV MEP-ID TLV" registry [RFC6428].
6.2 Using IP Headers

The encapsulation option using IP headers will not be suited for EVPN, as using different values in the destination IP address for data and OAM (BFD) packets could cause the BFD packets to follow a different path than that of data packets. Hence this option MUST NOT be used for EVPN.
7. Scalability Considerations

The mechanisms proposed by this draft could affect the packet load on the network and its elements especially when supporting configurations involving a large number of EVIs. The option of slowing down or speeding up 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.
8. IANA Considerations

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

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD1</td>
<td>EFD-EVPN CC</td>
<td>[this document]</td>
</tr>
<tr>
<td>TBD2</td>
<td>BFD-EVPN CV</td>
<td>[this document]</td>
</tr>
</tbody>
</table>

Ed Note: Do we need a CC code point?  TBD

IANA is requested to assign the following code-points from the "CC/CV MEP-ID TLV" registry [RFC6428].

<table>
<thead>
<tr>
<th>Value</th>
<th>Name</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD3</td>
<td>EVPN inclusive multicast</td>
<td>[this document]</td>
</tr>
<tr>
<td>TBD4</td>
<td>EVPN unicast</td>
<td>[this document]</td>
</tr>
</tbody>
</table>
9. 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.
Normative References


Informative References

[I-D.ooamdt-rtgwg-ooam-requirement] Kumar, N., Pignataro, C., Kumar, D., Mirsky, G., Chen, M., Nordmark, E., Networks, J., and D. Mozes, "Overlay OAM Requirements", draft-ooamdt-rtgwg-
oam-requirement-02 (work in progress), March 2016.


Authors’ Addresses

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

Mudigonda Mallik
Cisco Systems
Email: mmudigon@cisco.com

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

Gregory Mirsky
ZTE Corp.
Email: gregimirsky@gmail.com

Donald Eastlake, 3rd
Huawei Technologies
155 Beaver Street
Milford, MA 01757 USA
Phone: +1-508-333-2270
Email: d3e3e3@gmail.com