Abstract

This document proposes a proactive, in-band network OAM mechanism to detect connectivity faults that affect unicast and multi-destination paths in an EVPN network. The multi-destination paths are used by Broadcast, unknown Unicast and Multicast (BUM) traffic. The mechanisms proposed in the draft use the principles of the widely adopted Bidirectional Forwarding Detection (BFD) protocol.

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

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal
Provisions Relating to IETF Documents
(http://trustee.ietf.org/license-info) in effect on the date of
publication of this document. Please review these documents
carefully, as they describe your rights and restrictions with respect
to this document. Code Components extracted from this document must
include Simplified BSD License text as described in Section 4.e of
the Trust Legal Provisions and are provided without warranty as
described in the Simplified BSD License.

Table of Contents

1 Introduction ........................................... 2
  1.1 Terminology ........................................ 3
  1.2 Motivation for running BFD at the network layer of EVPN .. 3
2. Scope of fault detection mechanisms proposed in this
document .............................................. 4
  2.1 Fault Detection of BUM traffic using ingress replication
    (MP2P) .............................................. 4
    2.1.1 Bootstrapping BFD sessions at the head of the MP2P
tunnel .................................................. 4
    2.1.2 Bootstrapping BFD sessions at the tail nodes of the
         MP2P tunnel ....................................... 5
  2.2 Fault Detection of BUM traffic using P2MP tunnels (LSM) ... 5
    2.2.1 Bootstrapping BFD sessions at the root of the P2MP
tunnel .................................................. 5
    2.2.2 Bootstrapping BFD sessions at the tail nodes of the
         P2MP tunnel ....................................... 6
  2.3 Fault Detection of unicast traffic .......................... 6
3 BFD packet encapsulation .................................. 6
  3.1 Using GAL/G-ACh encapsulation without IP headers ........ 6
    3.1.1 Ingress replication ................................. 6
    3.1.2 LSM ............................................. 6
    3.1.3 Unicast .......................................... 7
  3.2 Using IP headers ...................................... 7
4. Scalability Considerations ................................. 7
5. Security Considerations .................................. 7
6 IANA Considerations ..................................... 7
7. Acknowledgments ........................................ 8
8 References ............................................. 8
  8.1 Normative References .................................. 8
  8.2 Informative References ................................ 8
Authors’ Addresses ...................................... 9

1 Introduction

[EVPNOAM] outlines the OAM requirements of Ethernet VPN networks
This document proposes mechanisms for proactive fault detection at the network OAM layer of EVPN. These mechanisms could either be deployed for periodic and proactive monitoring, or be triggered by specific events to aid troubleshooting. EVPN fault detection mechanisms need to consider unicast and BUM traffic separately since they map to different FECs in EVPN. Since BUM traffic can be transported using MP2P or P2MP tunnels, this document proposes slightly different fault detection mechanisms to suit each type using the principles of BFD over MPLS LSPs [BFD-MPLS] and Point-to-multipoint BFD (P2MPBFD). Please note that this document uses the term EVPN loosely to include [EVPN], [PBB-EVPN] as well as [TRILL-EVPN].

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].

1.2 Motivation for running BFD at the network layer of EVPN

The choice of running BFD at the network layer of the OAM model for EVPN [EVPNOAM] 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 [EVPNOAM] 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 BFD sessions in the EVPN Network Layer exits UP state, it provides additional confidence that data transported using those tunnels will not reach the expected egress node.
2. Scope of fault detection mechanisms proposed in this document

This section proposes proactive fault detection using BFD mechanisms for:

a. BUM traffic using MP2P tunnels (ingress replication).
b. BUM traffic using P2MP tunnels (LSM).
c. Unicast traffic

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.

The approach takes advantage of the inclusive multicast route used in EVPN to advertise the multi-destination FEC for bootstrapping the BFD sessions. Earlier approaches for P2MP BFD [MPLSCVBFD] have used periodic MPLS ping requests to bootstrap P2MP BFD sessions over MPLS.

2.1 Fault Detection of BUM traffic using ingress replication (MP2P)

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 proposed 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.

Note that once the BFD session for the EVPN BUM label is UP, either end of the BFD session MUST NOT change the local discriminator values of the BFD Control packets it generates, unless it first brings down the session as specified in [BFD-MPLS].

2.1.1 Bootstrapping BFD sessions at the head of the MP2P tunnel

To simplify BFD session de-multiplexing, we take advantage of the fact that the head replicates a BUM packet for each tail by using unique sets of discriminators in each copy of the (replicated) BFD packet. These discriminators MUST be exchanged out-of-band using MPLS ping [BFD-MPLS] before the start of the BFD session between the head and the tail node(s). The head PE performing ingress replication MUST initiate an LSP ping using the inclusive multicast FEC [EVPNPING] upon receiving an inclusive multicast route from a
tail to bootstrap the BFD session. This MPLS ping MUST include the BFD TLV specified in [BFD-MPLS].

There could exist multiple BFD sessions between a head of the multi-destination tunnel and an individual tail due to the usage of entropy labels [MPLSEL] for an inclusive multicast FEC. For fine-grained fault detection, a BFD session MAY be bootstrapped to monitor all unique path(s) that can be realized using entropy labels between a head and a given tail. However, the path(s) MUST be monitored using at least one or more number of representative BFD session(s) to satisfy the fault monitoring requirements [EVPNNOAM].

2.1.2 Bootstrapping BFD sessions at the tail nodes of the MP2P tunnel

The tail nodes MUST bootstrap a BFD session based on the incoming MPLS ping initiated by the head [P2MPBFD]. At the tail node, a new BFD discriminator MUST be allocated for each unique combination of the source IP and the attributes of the <inclusive multicast FEC, BUM label> when a MPLS ping initiated from the head is received. A tail node SHOULD include the entropy label, if present to support monitoring of specific paths or all realizable paths.

2.2 Fault Detection of BUM traffic using P2MP tunnels (LSM)

The case of using P2MP tunnels for distributing BUM traffic presents a different challenge for using BFD. Clearly, the yourDisc of the BFD packet MUST be zero [P2MPBFD] as the packet is multicast from the root unlike ingress replication where individual copies are made from the head. However the MPLS label that identifies the P-Tunnel [EVPN] used for forwarding the multi-destination traffic provides a convenient method of identifying the source and the FEC (multi-destination tree) being tracked by the BFD session. The tails of the multi-destination tree MUST use the MPLS label identifying the P-tunnel to de-multiplex the BFD packet. In the case of Aggregate Inclusive trees, where the root of the multi-destination tree reuses the same LSP label for traffic of various EVIs, the tail node MUST use the MPLS labels of the P-Tunnel and the upstream assigned label which the PE has bound uniquely to the EVI. The myDisc of the BFD packet is filled with an unique value allocated by the root to identify the multi-path session.

2.2.1 Bootstrapping BFD sessions at the root of the P2MP tunnel

The P2MP BFD sessions MUST be bootstrapped at the head [P2MPBFD] as soon as there is one receiver for the MDT traffic.
2.2.2 Bootstrapping BFD sessions at the tail nodes of the P2MP tunnel

The P2MP BFD sessions MUST be bootstrapped at the tail upon reception of the P2MP BFD packets from the head. The tail MUST use the P2MP MDT label to de-multiplex the incoming BFD packet. The BFD session MAY be destroyed immediately upon leaving Up state.

2.3 Fault Detection of unicast traffic

The mechanisms specified in BFD for MPLS LSPs [BFD-MPLS] 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 MPLS ping specifying the Unicast EVPN FEC [EVPNPING] before starting 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 take care of addressing the requirement of monitoring faults of the various paths of the multi-path server layer network [MPLSE]. Each unique realizable path between the participating PE routers MAY be monitored separately when entropy labels are used. Alternately, all paths MUST be tracked by at least one or a fewer number of representative BFD session(s) in which case the granularity of fault-detection would be coarser. The PE node receiving the MPLS ping MUST allocate one BFD discriminator for every unique combination of the source IP address and the tuple of <unicast FEC, EVPN label>. A node SHOULD include the entropy label, if present to satisfy the requirement of fault monitoring of specific paths or all realizable paths. Note that once the BFD session for the EVPN label is UP, either end of the BFD session MUST NOT change the local discriminator values of the BFD Control packets it generates, unless it first brings down the session as specified in [BFD-MPLS].

3 BFD packet encapsulation

3.1 Using GAL/G-ACh encapsulation without IP headers

3.1.1 Ingress replication

The packet contains the following labels: LSP label (transport) when not using PHP, the optional entropy label, the BUM label and the SH label[EVPN] (where applicable). The G-ACh type is set to TBD. The discriminator values of BFD are obtained through negotiation through the out-of-band MPLS ping.

3.1.2 LSM
The packet contains the following labels: label identifying the P-Tunnel, upstream label which the PE has bound uniquely to the EVI (for aggregate inclusive trees only). The G-ACh type is set to TBD. The yourDisc value is set to 0 and the myDisc value is uniquely generated by the root.

3.1.3 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 TBD. The discriminator values of BFD are obtained through negotiation through the out-of-band MPLS ping.

3.2 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.

4. 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.

5. Security Considerations

This document does not introduce any new security issues, the security considerations defined in [BFD] and [P2MPBFD] apply in this document.

6 IANA Considerations

A new G-ACh Type is requested for GAL encapsulated BFD as the existing type [VCCV-BFD] specifically applies to PW-ACH.
encapsulation.

7. Acknowledgments

We thank Nobo Akiya, Tina Lam, Jose Liste, Mudigonda Mallik and Gregory Mirsky for their valuable input, discussions and comments.

8. References

8.1. Normative References


8.2. Informative References

[PBB-EVPN] Sajassi et al, "PBB-EVPN", draft-ietf-l2vpn-pbb-evpn-05,
July 2013.

[TRILL-EVPN]
Sajassi et al., "TRILL-EVPN",
draft-ietf-l2vpn-trill-evpn-00.txt,
work in progress, June 2012.

[MPLSEL] Kompella et al,
"The Use of Entropy Labels in MPLS Forwarding",
RFC 6790, November 2012.

Authors’ Addresses
Vengada Prasad Govindan
EMail: venggovi@cisco.com

Samer Salam
EMail: ssalam@cisco.com

Ali Sajassi
EMail:sajassi@cisco.com