Fault Management for EVPN networks
draft-ietf-bess-evpn-bfd-00

Abstract

   This document specifies proactive, in-band network OAM mechanisms 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 Ethernet VPN (EVPN) network.
   The mechanisms specified in the draft are based on the widely adopted
   Bidirectional Forwarding Detection (BFD) protocol.

Status of this Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

   Distribution of this document is unlimited. Comments should be sent
   to the authors or the BESS working group mailing list: bess@ietf.org.

   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.

Table of Contents

1. Introduction

[ietf-bess-evpn-oam-req-frmwk] outlines the OAM requirements of
Ethernet VPN networks (EVPN [RFC7432]). This document specifies
mechanisms for proactive fault detection at the network (overlay)
layer of EVPN. The mechanisms proposed in the draft use the widely
adopted Bidirectional Forwarding Detection (BFD [RFC5880]) protocol.

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. Hence this document proposes different fault
detection mechanisms to suit each type, for unicast traffic using BFD
[RFC5880] and for BUM traffic using BFD or [RFC8563] depending on
whether an MP2P or P2MP tunnel is being used.

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.

   BFD - Bidirectional Forwarding Detection [RFC5880]

   BUM - Broadcast, Unknown Unicast, and Multicast

   CC - Continuity Check

   CV - Connectivity Verification

   EVI - EVPN Instance

   EVPN - Ethernet VPN [RFC7432]

   FEC - Forwarding Equivalency Class

   GAL - Generic Associated Channel Label [RFC5586]

   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

VXLAN - Virtual eXtesible Local Area Network (VXLAN) [RFC7348]

[2](#). Scope of this Document

This document specifies BFD based mechanisms for proactive fault
detection for EVPN both as specified in [RFC7432] and also for EVPN
using VXLAN encapsulation [ietf-vxlan-bfd]. It covers the following:

   o Unicast traffic.

   o BUM traffic using Multi-point-to-Point (MP2P) tunnels (ingress
      replication).

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

   o MPLS and VXLAN encapsulation.

This document does not discuss BFD mechanisms for:

   o EVPN variants like PBB-EVPN [RFC7623]. It is intended to
      address this in future versions.

   o Integrated Routing and Bridging (IRB) solution based on EVPN
      [ietf-bess-evpn-inter-subnet-forwarding]. It is intended to
      address this in future versions.

   o EVPN using other encapsulations such as NVGRE or MPLS over GRE
      [RFC8365].

   o BUM traffic using MP2MP tunnels.

This specification 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 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 [ietf-bess-evpn-oam-req-frmwk] was made after considering the
following:

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

o 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 entropy mechanisms such as entropy labels,
   when they are used, or VXLAN source ports. However, paths that
   cannot be realized by entropy variations cannot be monitored.
   Fault monitoring requirements outlined by
   [ietf-bess-evpn-oam-req-frmwk] are addressed by the mechanisms
   proposed by this draft.

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 to CE.) For example, an egress EVPN-PE could understand
EVPN labeling received but could switch data to an incorrect
interface. 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.

4. Fault Detection for Unicast Traffic

The mechanisms specified in BFD for MPLS LSPs [RFC5884] [RFC7726] are applied to test the handling of unicast EVPN traffic. The discriminators required for de-multiplexing the BFD sessions are advertised through BGP as specified in Section 7. This is needed for MPLS since the label stack does not contain enough information to disambiguate the sender of the packet.

The usage of MPLS entropy labels or various VXLAN source ports 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 such entropy is used. At least one path of multi-path connectivity between two PE routers MUST be tracked with BFD, but in that case the granularity of fault-detection will be coarser. To support unicast OAM, each PE node MUST allocate a BFD discriminator to be used for BFD messages to that PE and MUST advertise this discriminator with BGP as specified in Section 7. 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 bring down the session as specified in [RFC5884].

5. Fault Detection for BUM Traffic

   Section 5.1 below discusses fault detection for MP2P tunnels using
   ingress replication and Section 5.2 discusses fault detection for
   P2MP tunnels.

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 the head makes a unique copy 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. This contains the BUM EVPN 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.

   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 ports for an inclusive multicast FEC and (2) due to
   multiple MP2P tunnels indicated by different tail labels or IP
   addresses for MPLS or VXLAN. The BFD discriminator to be used is
   distributed by BGP as specified in Section 7. Once the BFD session
   for the EVPN label is UP, the BFD systems terminating the BFD session
   MUST NOT change the local discriminator values of the BFD Control
   packets they generate, unless they first bring down the session as
   specified in [RFC5884].

5.2 P2MP Tunnels (Label Switched Multicast)

   Fault detection for BUM traffic distributed using a P2MP tunnel uses
   active tail multipoint BFD [RFC8563] in one of the three scenarios
   providing head notification (see Section 5.2 of [RFC8563]).

   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.

6. BFD Packet Encapsulation

The sections below describe the MPLS and VXLAN encapsulations of BFD
for EVPN OAM use.

6.1 MPLS Encapsulation

This section describes use of the Generic Associated Channel Label
(GAL) for BFD encapsulation in MPLS based EVPN OAM.

6.1.1 Unicast

The packet initially contains the following labels: LSP label
(transport), 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 destination L2 header (in overlay space) followed by the
IP header that encapsulates the BFD packet. The MAC address of the
inner packet is used to validate the <EVI, MAC> in the receiving
node.

   - The destination MAC MUST be the dedicated MAC TBD-A (see Section
     9) or the MAC address of the destination PE.
   - The destination IP address MUST be in the 127.0.0.0/8 range for
     IPv4 or in the 0:0:0:0:0:FFFF:7F00:0/104 range for IPv6.
   - The destination IP port MUST be 3784 [RFC5881].
   - The source IP port MUST be in the range 49152 through 65535.
   - The discriminator values for BFD are obtained through BGP as
     specified in Section 7 or are exchanged out-of-band or through
     some other means outside the scope of this document.

```
       <---------- 4 bytes ---------->
      +------------------------------+  -----
      |          LSP Label           |     |
      +------------------------------+     |
      :      entropy label indicator :     |
      + (optional)                   +  MPLS Label Stack
      :          entropy label       :     |
      +------------------------------+     |
      |       EVPN Unicast label     |     |
      +------------------------------+     |
      | Generic Assoc. Channel Label |     |
      +------------------------------+  -----
      |   ACH word, Type TBD1 no TLVs |
      +------------------------------+  ---      -------
      |     Destination MAC Address  |   |          |
      +              +--------------+   |          |
      |   TBD-A      |               |   |          |
      +--------------+               +  L2 Header  |
      |       Source MAC Address     |   |          |
      +--------------+--------------+   |          |
      | VLAN Ethertype|    VLAN-ID   |   |          |
      +--------------+--------------+   |          |
      |IP4/6 Ethertype|               |   |          |
      +--------------+--------------+  ---         |
      /                              /       G-ACh Payload
      /...      IP4/6 Header    .../           |
      /                              /           |
      +------------------------------+           |
      |                              |           |
      +           UDP Header         +           |
      |                              |           |
      +------------------------------+           |
      |                              |           |
      +        BFD Control Packet    +           |
      /                              /           |
      /...                      .../  --------------
```
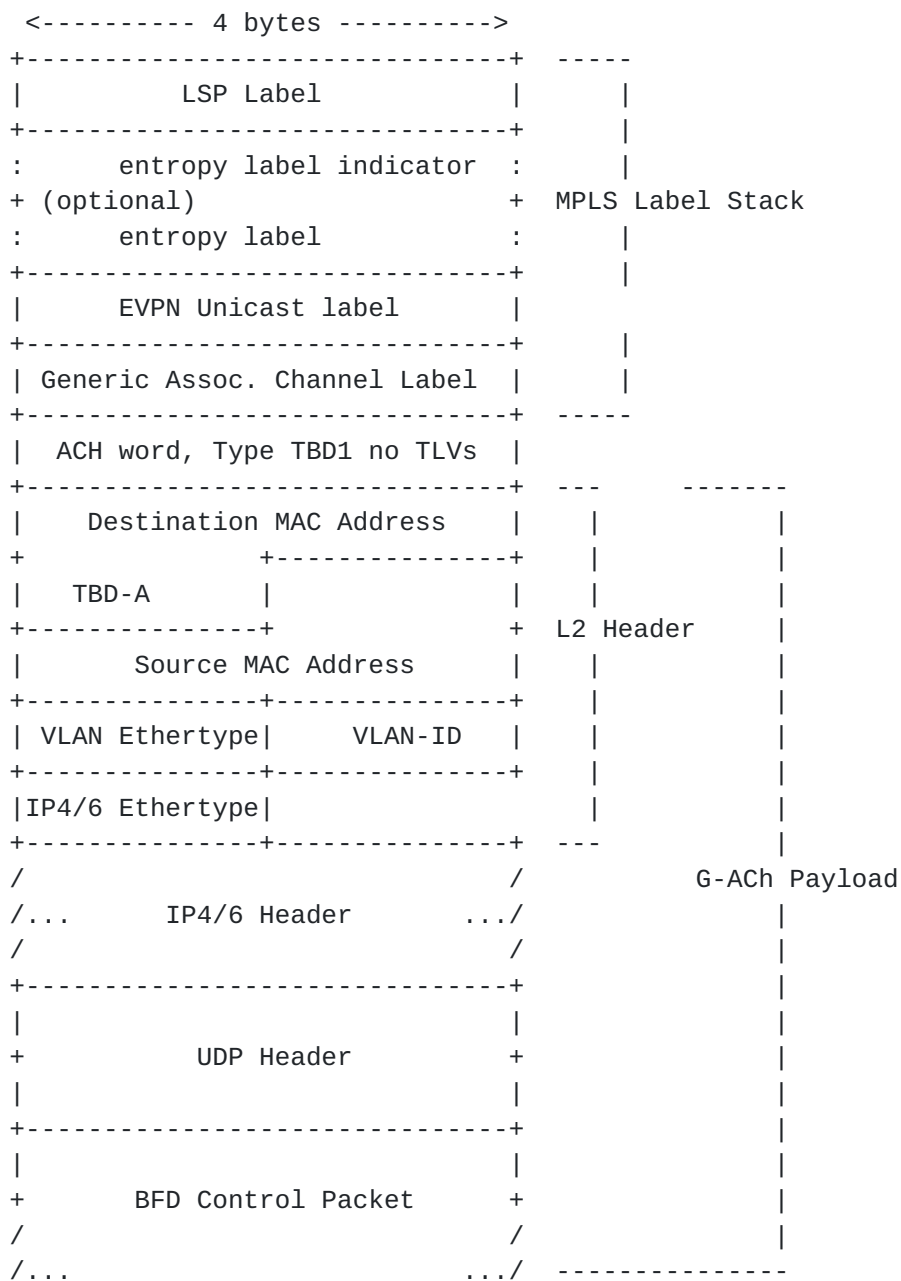
6.1.2 Ingress Replication

The packet initially contains the following labels: LSP label
(transport), the optional entropy label, the BUM label, and the split
horizon label [RFC7432] (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.

6.1.3 LSM (Label Switched Multicast, P2MP)

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.

6.2 VXLAN Encapsulation

This section describes the use of the VXLAN [RFC7348] for BFD
encapsulation in VXLAN based EVPN OAM. This specification conforms to
[ietf-bfd-vxlan].

6.2.1 Unicast

The outer and inner IP headers have a unicast source IP address of
the BFD message source and a destination IP address of the BFD
message destination

The destination UDP port MUST be 3784 [RFC5881]. The source port MUST
be in the range 49152 through 65535. If the BFD source has multiple
IP addresses, entropy MAY be further obtained by using any of those
addresses assuming the source is prepared for responses directed to
the IP address used.

The Your BFD discriminator is the value distributed for this unicast
OAM purpose by the destination using BGP as specified in Section 7 or
is exchanged out-of-band or through some other means outside the
scope of this document.

```
          <---------- 4 bytes ---------->
        +------------------------------+  ---
        |    Destination MAC Address    |    |
        +              +---------------+    |
        |              |               |    |
        +---------------+               +  L2 Header
        |      Source MAC Address       |    |
        +------------------------------+    |
        |            VLAN Tag           |    |
        +--------------+---------------+    |
        |IP4/6 Ethertype|               |
        +--------------+---------------+  ---
        /                              /
        /...     IP4/6 Header      .../
        /                              /
        +------------------------------+
        |                              |
        +          UDP Header          +
        |                              |
        +------------------------------+
        |                              |
        +          VXLAN Header        +
        |                              |
        +------------------------------+  ---
        |    Destination MAC Address    |    |
        +              +---------------+    |
        |              |               |    |
        +---------------+               +  L2 Header
        |      Source MAC Address       |    |
        +--------------+---------------+    |
        | IP4 Ethertype |               |
        +--------------+---------------+  ---
        /                              /
        /...      IP4 Header       .../
        /                              /
        +------------------------------+
        |                              |
        +          UDP Header          +
        |                              |
        +--------------+---------------+
        |                              |
        +       BFD Control Packet      +
        |                              |
        /...                       .../
```

6.2.2 **Ingress Replication**

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 specified in Section 7 for the MP2P
    tunnel in question or is exchanged out-of-band or through some
    other means outside the scope of this document.

6.2.3 **LSM (Label Switched Multicast, P2MP)**

The VXLAN encapsulation for the head-to-tails BFD packets uses the
multicast destination IP corresponding to the VXLAN VNI.

The destination port MUST be 3784. For entropy purposes, the source
port can vary but MUST be in the range 49152 through 65535 [RFC5881].
If the head PE has multiple IP addresses, entropy MAY be further
obtained by using any of those addresses.

The Your BFD discriminator is the value distributed for this unicast
OAM purpose by the BFD message using BGP as specified in Section 7 or
is exchanged out-of-band or through some other means outside the
scope of this document.

7. BGP Distribution of BFD Discriminators

   BGP is used to distribute BFD discriminators for use in EVPN OAM as
   follows using the BGP-BFD Attribute as specified in
   [ietf-bess-mvpn-fast-failover]. This attribute is included with
   appropriate EVPN routes as follows:

      Unicast: MAC/IP Advertisement Route [RFC7432].

      MP2P Tunnel: Inclusive Multicast Ethernet Tag Route [RFC7432].

      P2MP: TBD

    [Need more text on BFD sessions reacting to the new advertisement
   and withdrawal of the BGP-BFD Attribute.]

8. 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.

9. IANA Considerations

The following IANA Actions are requested.

9.1 Pseudowire Associated Channel Type

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]
```

9.2 MAC Address

IANA is requested to assign a multicast MAC address under the IANA OUI [0x01005E900004 suggested] as follows:

```
    Address    Usage       Reference
    -------    --------    ---------------
    TBD-A      EVPN OAM    [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 BPD Control packets are routed, the authentication considerations discussed in [RFC5883] should be followed.

VXLAN BFD security considerations in [ietf-vxlan-bfd] apply to BFD packets encapsulate in VXLAN.

Acknowledgement

The authors wish to thank the following for their comments and suggestions:

    Mach Chen

Normative References

    [ietf-bess-evpn-inter-subnet-forwarding] Sajassi, A., Salam, S.,
            Thoria, S., Rekhter, Y., Drake, J., Yong, L., and L.
            Dunbar, "Integrated Routing and Bridging in EVPN",
            draft-ietf-bess-evpn-inter-subnet-forwarding-08, work in
            progress, March 2019.

    [ietf-bess-mvpn-fast-failover] Morin, T., Kebler, R., Mirsky, G.,
            "Multicast VPN fast upstream failover",
            draft-ietf-bess-mvpn-fast-failover-05 (work in progress),
            February 2019.

    [ietf-bfd-vxlan] Pallagatti, S., Paragiri, S., Govindan, V.,
            Mudigonda, M., G. Mirsky, "BFD for VXLAN",
            draft-ietf-bfd-vxlan-07 (work in progress), May 2019.

    [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
            Requirement Levels", BCP 14, RFC 2119, DOI
            10.17487/RFC2119, March 1997, <http://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, <http://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,
            <http://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>.

    [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, <http://www.rfc-
           editor.org/info/rfc6790>.

[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>.

[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>.

[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, <http://www.rfc-editor.org/info/rfc7623>.

[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>.

[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>.

Informative References

   [ietf-bess-evpn-oam-req-frmwk] Salam, S., Sajassi, A., Aldrin, S., J.
              Drake, and D. Eastlake, "EVPN Operations, Administration
              and Maintenance Requirements and Framework",
              draft-ietf-bess-evpn-oam-req-frmwk-00, work in progress,
              February 2019.

   [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>.

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
        Futurewei Technologies
        2386 Panoramic Circle
        Apopka, FL 32703 USA

        Phone: +1-508-333-2270
        Email: d3e3e3@gmail.com

Copyright, Disclaimer, and Additional IPR Provisions