                   Bidirectional Forwarding Detection (BFD)
                    for MPLS Label Switched Paths (LSPs)

Abstract

   One desirable application of Bidirectional Forwarding Detection (BFD)
   is to detect a Multiprotocol Label Switching (MPLS) Label Switched
   Path (LSP) data plane failure. LSP Ping is an existing mechanism for
   detecting MPLS data plane failures and for verifying the MPLS LSP
   data plane against the control plane. BFD can be used for the
   former, but not for the latter. However, the control plane
   processing required for BFD Control packets is relatively smaller
   than the processing required for LSP Ping messages. A combination of
   LSP Ping and BFD can be used to provide faster data plane failure
   detection and/or make it possible to provide such detection on a
   greater number of LSPs. This document describes the applicability of
   BFD in relation to LSP Ping for this application. It also describes
   procedures for using BFD in this environment.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   One desirable application of Bidirectional Forwarding Detection (BFD)
   is to track the liveness of a Multiprotocol Label Switching (MPLS)
   Label Switched Path (LSP). In particular, BFD can be used to detect
   a data plane failure in the forwarding path of an MPLS LSP. LSP Ping
   [RFC4379] is an existing mechanism for detecting MPLS LSP data plane
   failures and for verifying the MPLS LSP data plane against the
   control plane. This document describes the applicability of BFD in
   relation to LSP Ping for detecting MPLS LSP data plane failures. It
   also describes procedures for using BFD for detecting MPLS LSP data
   plane failures.

2. Specification of Requirements

   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 [RFC2119].

3. Applicability

   In the event of an MPLS LSP failing to deliver data traffic, it may
   not always be possible to detect the failure using the MPLS control
   plane. For instance, the control plane of the MPLS LSP may be
   functional while the data plane may be mis-forwarding or dropping
   data. Hence, there is a need for a mechanism to detect a data plane
   failure in the MPLS LSP path [RFC4377].

3.1. BFD for MPLS LSPs: Motivation

   LSP Ping described in [RFC4379] is an existing mechanism for
   detecting an MPLS LSP data plane failure. In addition, LSP Ping also
   provides a mechanism for verifying the MPLS control plane against the
   data plane. This is done by ensuring that the LSP is mapped to the
   same Forwarding Equivalence Class (FEC), at the egress, as the
   ingress.

   BFD cannot be used for verifying the MPLS control plane against the
   data plane. However, BFD can be used to detect a data plane failure
   in the forwarding path of an MPLS LSP. The LSP may be associated
   with any of the following FECs:

      a) Resource Reservation Protocol (RSVP) LSP_Tunnel IPv4/IPv6
         Session [RFC3209]

      b) Label Distribution Protocol (LDP) IPv4/IPv6 prefix [RFC5036]

      c) Virtual Private Network (VPN) IPv4/IPv6 prefix [RFC4364]

    d) Layer 2 VPN [L2-VPN]

    e) Pseudowires based on PWid FEC and Generalized PWid FEC
       [RFC4447]

    f) Border Gateway Protocol (BGP) labeled prefixes [RFC3107]

LSP Ping includes extensive control plane verification. BFD, on the
other hand, was designed as a lightweight means of testing only the
data plane. As a result, LSP Ping is computationally more expensive
than BFD for detecting MPLS LSP data plane faults. BFD is also more
suitable for being implemented in hardware or firmware due to its
fixed packet format. Thus, the use of BFD for detecting MPLS LSP
data plane faults has the following advantages:

    a) Support for fault detection for greater number of LSPs.

    b) Fast detection. Detection with sub-second granularity is
       considered as fast detection. LSP Ping is intended to be used
       in an environment where fault detection messages are exchanged,
       either for diagnostic purposes or for infrequent periodic fault
       detection, in the order of tens of seconds or minutes. Hence,
       it is not appropriate for fast detection. BFD, on the other
       hand, is designed for sub-second fault detection intervals.
       Following are some potential cases when fast detection may be
       desirable for MPLS LSPs:

       1. In the case of a bypass LSP used for a facility-based link
          or node protection [RFC4090]. In this case, the bypass LSP
          is essentially being used as an alternate link to protect
          one or more LSPs. It represents an aggregate and is used to
          carry data traffic belonging to one or more LSPs, when the
          link or the node being protected fails. Hence, fast failure
          detection of the bypass LSP may be desirable particularly in
          the event of link or node failure when the data traffic is
          moved to the bypass LSP.

       2. MPLS Pseudowires (PWs). Fast detection may be desired for
          MPLS PWs depending on i) the model used to layer the MPLS
          network with the Layer 2 network, and ii) the service that
          the PW is emulating. For a non-overlay model between the
          Layer 2 network and the MPLS network, the provider may rely
          on PW fault detection to provide service status to the end-
          systems. Also, in that case, interworking scenarios such as
          ATM/Frame Relay interworking may force periodic PW fault
          detection messages. Depending on the requirements of the
          service that the MPLS PW is emulating, fast failure
          detection may be desirable.

There may be other potential cases where fast failure detection is
desired for MPLS LSPs.

3.2. Using BFD in Conjunction with LSP Ping

BFD can be used for MPLS LSP data plane fault detection. However, it
does not have all the functionality of LSP Ping. In particular, it
cannot be used for verifying the control plane against the data
plane. LSP Ping performs the following functions that are outside
the scope of BFD:

    a) Association of an LSP Ping Echo request message with a FEC. In
       the case of Penultimate Hop Popping (PHP) or when the egress
       Label Switching Router (LSR) distributes an explicit null label
       to the penultimate hop router, for a single label stack LSP,
       the only way to associate a fault detection message with a FEC
       is by carrying the FEC in the message. LSP Ping provides this
       functionality. Next-hop label allocation also makes it
       necessary to carry the FEC in the fault detection message as
       the label alone is not sufficient to identify the LSP being
       verified. In addition, presence of the FEC in the Echo request
       message makes it possible to verify the control plane against
       the data plane at the egress LSR.

    b) Equal Cost Multi-Path (ECMP) considerations. LSP Ping
       traceroute makes it possible to probe multiple alternate paths
       for LDP IP FECs.

    c) Traceroute. LSP Ping supports traceroute for a FEC and it can
       be used for fault isolation.

Hence, BFD is used in conjunction with LSP Ping for MPLS LSP fault
detection:

    i) LSP Ping is used for bootstrapping the BFD session as described
       later in this document.

   ii) BFD is used to exchange fault detection (i.e., BFD session)
       packets at the required detection interval.

  iii) LSP Ping is used to periodically verify the control plane
       against the data plane by ensuring that the LSP is mapped to
       the same FEC, at the egress, as the ingress.

4. Theory of Operation

   To use BFD for fault detection on an MPLS LSP, a BFD session MUST be
   established for that particular MPLS LSP. BFD Control packets MUST
   be sent along the same data path as the LSP being verified and are
   processed by the BFD processing module of the egress LSR. If the LSP
   is associated with multiple FECs, a BFD session SHOULD be established
   for each FEC. For instance, this may happen in the case of next-hop
   label allocation. Hence, the operation is conceptually similar to
   the data plane fault detection procedures of LSP Ping.

   If MPLS fast-reroute is being used for the MPLS LSP, the use of BFD
   for fault detection can result in false fault detections if the BFD
   fault detection interval is less than the MPLS fast-reroute
   switchover time. When MPLS fast-reroute is triggered because of a
   link or node failure, BFD Control packets will be dropped until
   traffic is switched on to the backup LSP. If the time taken to
   perform the switchover exceeds the BFD fault detection interval, a
   fault will be declared even though the MPLS LSP is being locally
   repaired. To avoid this, the BFD fault detection interval should be
   greater than the fast-reroute switchover time. An implementation
   SHOULD provide configuration options to control the BFD fault
   detection interval.

   If there are multiple alternate paths from an ingress LSR to an
   egress LSR for an LDP IP FEC, LSP Ping traceroute MAY be used to
   determine each of these alternate paths. A BFD session SHOULD be
   established for each alternate path that is discovered.

   Periodic LSP Ping Echo request messages SHOULD be sent by the ingress
   LSR to the egress LSR along the same data path as the LSP. This is
   to periodically verify the control plane against the data plane by
   ensuring that the LSP is mapped to the same FEC, at the egress, as
   the ingress. The rate of generation of these LSP Ping Echo request
   messages SHOULD be significantly less than the rate of generation of
   the BFD Control packets. An implementation MAY provide configuration
   options to control the rate of generation of the periodic LSP Ping
   Echo request messages.

   To enable fault detection procedures specified in this document, for
   a particular MPLS LSP, this document requires the ingress and egress
   LSRs to be configured. This includes configuration for supporting
   BFD and LSP Ping as specified in this document. It also includes
   configuration that enables the ingress LSR to determine the method
   used by the egress LSR to identify Operations, Administration, and
   Maintenance (OAM) packets, e.g., whether the Time to Live (TTL) of
   the innermost MPLS label needs to be set to 1 to enable the egress

LSR to identify the OAM packet. For fault detection for MPLS PWs,
this document assumes that the PW control channel type [RFC5085] is
configured and the support of LSP Ping is also configured.

5. Initialization and Demultiplexing

A BFD session may be established for a FEC associated with an MPLS
LSP. As described above, in the case of PHP or when the egress LSR
distributes an explicit null label to the penultimate hop router, or
next-hop label allocation, the BFD Control packet received by the
egress LSR does not contain sufficient information to associate it
with a BFD session. Hence, the demultiplexing MUST be done using the
remote discriminator field in the received BFD Control packet. The
exchange of BFD discriminators for this purpose is described in the
next section.

6. Session Establishment

A BFD session is bootstrapped using LSP Ping. 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
initiation of fault detection for a particular <MPLS LSP, FEC>
combination results in the exchange of LSP Ping Echo request and Echo
reply packets, in the ping mode, between the ingress and egress LSRs
for that <MPLS LSP, FEC>. To establish a BFD session, an LSP Ping
Echo request message MUST carry the local discriminator assigned by
the ingress LSR for the BFD session. This MUST subsequently be used
as the My Discriminator field in the BFD session packets sent by the
ingress LSR.

On receipt of the LSP Ping Echo request message, the egress LSR MUST
send a BFD Control packet to the ingress LSR, if the validation of
the FEC in the LSP Ping Echo request message succeeds. This BFD
Control packet MUST set the Your Discriminator field to the
discriminator received from the ingress LSR in the LSP Ping Echo
request message. The egress LSR MAY respond with an LSP Ping Echo
reply message that carries the local discriminator assigned by it for
the BFD session. The local discriminator assigned by the egress LSR
MUST be used as the My Discriminator field in the BFD session packets
sent by the egress LSR.

The ingress LSR follows the procedures in [BFD] to send BFD Control
packets to the egress LSR in response to the BFD Control packets
received from the egress LSR. The BFD Control packets from the
ingress to the egress LSR MUST set the local discriminator of the
egress LSR, in the Your Discriminator field. The egress LSR
demultiplexes the BFD session based on the received Your

Discriminator field. As mentioned above, the egress LSR MUST send
Control packets to the ingress LSR with the Your Discriminator field
set to the local discriminator of the ingress LSR. The ingress LSR
uses this to demultiplex the BFD session.

6.1. BFD Discriminator TLV in LSP Ping

LSP Ping Echo request and Echo reply messages carry a BFD
discriminator TLV for the purpose of session establishment as
described above. IANA has assigned a type value of 15 to this TLV.
This TLV has a length of 4. The value contains the 4-byte local
discriminator that the LSR, sending the LSP Ping message, associates
with the BFD session.

If the BFD session is not in UP state, the periodic LSP Ping Echo
request messages MUST include the BFD Discriminator TLV.

7. Encapsulation

BFD Control packets sent by the ingress LSR MUST be encapsulated in
the MPLS label stack that corresponds to the FEC for which fault
detection is being performed. If the label stack has a depth greater
than one, the TTL of the inner MPLS label MAY be set to 1. This may
be necessary for certain FECs to enable the egress LSR's control
plane to receive the packet [RFC4379]. For MPLS PWs, alternatively,
the presence of a fault detection message may be indicated by setting
a bit in the control word [RFC5085].

The BFD Control packet sent by the ingress LSR MUST be a UDP packet
with a well-known destination port 3784 [BFD-IP] and a source port
assigned by the sender as per the procedures in [BFD-IP]. The source
IP address is a routable address of the sender. The destination IP
address MUST be randomly chosen from the 127/8 range for IPv4 and
from the 0:0:0:0:0:FFFF:7F00/104 range for IPv6 with the following
exception. If the FEC is an LDP IP FEC, the ingress LSR may discover
multiple alternate paths to the egress LSR for this FEC using LSP
Ping traceroute. In this case, the destination IP address, used in a
BFD session established for one such alternate path, is the address
in the 127/8 range for IPv4 or 0:0:0:0:0:FFFF:7F00/104 range for IPv6
discovered by LSP Ping traceroute [RFC4379] to exercise that
particular alternate path.

The motivation for using the address range 127/8 is the same as
specified in Section 2.1 of [RFC4379]. This is an exception to the
behavior defined in [RFC1122].

The IP TTL or hop limit MUST be set to 1 [RFC4379].

BFD Control packets sent by the egress LSR are UDP packets. The
source IP address is a routable address of the replier.

The BFD Control packet sent by the egress LSR to the ingress LSR MAY
be routed based on the destination IP address as per the procedures
in [BFD-MHOP]. If this is the case, the destination IP address MUST
be set to the source IP address of the LSP Ping Echo request message,
received by the egress LSR from the ingress LSR.

Or the BFD Control packet sent by the egress LSR to the ingress LSR
MAY be encapsulated in an MPLS label stack. In this case, the
presence of the fault detection message is indicated as described
above. This may be the case if the FEC for which the fault detection
is being performed corresponds to a bidirectional LSP or an MPLS PW.
This may also be the case when there is a return LSP from the egress
LSR to the ingress LSR. In this case, the destination IP address
MUST be randomly chosen from the 127/8 range for IPv4 and from the
0:0:0:0:0:FFFF:7F00/104 range for IPv6.

The BFD Control packet sent by the egress LSR MUST have a well-known
destination port 4784, if it is routed [BFD-MHOP], or it MUST have a
well-known destination port 3784 [BFD-IP] if it is encapsulated in a
MPLS label stack. The source port MUST be assigned by the egress LSR
as per the procedures in [BFD-IP].

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. This implies that an LSR MUST
ignore BFD packets for a given session, demultiplexed using the
received Your Discriminator field, if the session is in UP state and
if the My Discriminator or the Source IP address fields of the
received packet do not match the values associated with the session.

8. Security Considerations

   Security considerations discussed in [BFD], [BFD-MHOP], and [RFC4379]
   apply to this document. For BFD Control packets sent by the ingress
   LSR or when the BFD Control packet sent by the egress LSR are
   encapsulated in an MPLS label stack, MPLS security considerations
   apply. These are discussed in [MPLS-SEC]. When BFD Control packets
   sent by the egress LSR are routed, the authentication considerations
   discussed in [BFD-MHOP] should be followed.

9. IANA Considerations

   This document introduces a BFD discriminator TLV in LSP Ping. The
   BFD Discriminator has been assigned a value of 15 from the LSP Ping
   TLVs and sub-TLVs registry maintained by IANA.

10. Acknowledgments

   We would like to thank Yakov Rekhter, Dave Katz, and Ina Minei for
   contributing to the discussions that formed the basis of this
   document and for their comments. Thanks to Dimitri Papadimitriou for
   his comments and review. Thanks to Carlos Pignataro for his comments
   and review.

11. References

11.1. Normative References

   [BFD]   Katz, D. and D. Ward, "Bidirectional Forwarding
               Detection", RFC 5880, June 2010.

   [BFD-IP]   Katz, D. and D. Ward, "Bidirectional Forwarding Detection
               (BFD) for IPv4 and IPv6 (Single Hop)", RFC 5881, June
               2010.

   [RFC4379]   Kompella, K. and G. Swallow, "Detecting Multi-Protocol
               Label Switched (MPLS) Data Plane Failures", RFC 4379,
               February 2006.

   [RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
               Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC1122]   Braden, R., Ed., "Requirements for Internet Hosts -
               Communication Layers", STD 3, RFC 1122, October 1989.

11.2. Informative References

   [BFD-MHOP] Katz, D. and D. Ward, "Bidirectional Forwarding Detection
               (BFD) for Multihop Paths", RFC 5883, June 2010.

   [RFC5085]   Nadeau, T., Ed., and C. Pignataro, Ed., "Pseudowire
               Virtual Circuit Connectivity Verification (VCCV): A
               Control Channel for Pseudowires", RFC 5085, December 2007.

   [RFC3209]   Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V.,
               and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP
               Tunnels", RFC 3209, December 2001.

   [RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast
              Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090,
              May 2005.

   [RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed.,
              "LDP Specification", RFC 5036, October 2007.

   [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private
              Networks (VPNs)", RFC 4364, February 2006.

   [L2-VPN]   Kompella, K., Leelanivas, M., Vohra, Q., Achirica, J.,
              Bonica, R., Cooper, D., Liljenstolpe, C., Metz, E., Ould-
              Brahim, H., Sargor, C., Shah, H., Srinivasan, and Z.
              Zhang, "Layer 2 VPNs Over Tunnels", Work in Progress,
              February 2003.

   [RFC4447] Martini, L., Ed., Rosen, E., El-Aawar, N., Smith, T., and
              G. Heron, "Pseudowire Setup and Maintenance Using the
              Label Distribution Protocol (LDP)", RFC 4447, April 2006.

   [RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in
              BGP-4", RFC 3107, May 2001.

   [RFC4377] Nadeau, T., Morrow, M., Swallow, G., Allan, D., and S.
              Matsushima, "Operations and Management (OAM) Requirements
              for Multi-Protocol Label Switched (MPLS) Networks", RFC
              4377, February 2006.

   [MPLS-SEC] Fang, L., Ed., "Security Framework for MPLS and GMPLS
              Networks", Work in Progress, October 2009.

Authors' Addresses

   Rahul Aggarwal
   Juniper Networks
   1194 N. Mathilda Ave.
   Sunnyvale, CA 94089
   USA

   EMail: rahul@juniper.net

   Kireeti Kompella
   Juniper Networks
   1194 N. Mathilda Ave.
   Sunnyvale, CA 94089
   USA

   EMail: kireeti@juniper.net

   Thomas D. Nadeau
   BT
   BT Centre
   81 Newgate Street
   London EC1A 7AJ
   UK

   EMail: tom.nadeau@bt.com

   George Swallow
   Cisco Systems, Inc.
   300 Beaver Brook Road
   Boxborough, MA 01719
   USA

   EMail: swallow@cisco.com