            Label Switched Path (LSP) and Pseudowire (PW) Ping/Trace
                 over MPLS Networks Using Entropy Labels (ELs)

Abstract

   Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) ping
   and traceroute are methods used to test Equal-Cost Multipath (ECMP)
   paths. Ping is known as a connectivity-verification method and
   traceroute is known as a fault-isolation method, as described in RFC
   4379. When an LSP is signaled using the Entropy Label (EL) described
   in RFC 6790, the ability for LSP ping and traceroute operations to
   discover and exercise ECMP paths is lost for scenarios where Label
   Switching Routers (LSRs) apply different load-balancing techniques.
   One such scenario is when some LSRs apply EL-based load balancing
   while other LSRs apply load balancing that is not EL based (e.g.,
   IP). Another scenario is when an EL-based LSP is stitched with
   another LSP that can be EL based or not EL based.

   This document extends the MPLS LSP ping and traceroute multipath
   mechanisms in RFC 6424 to allow the ability of exercising LSPs that
   make use of the EL. This document updates RFC 6790.

Copyright Notice

Table of Contents

1. Introduction

   [RFC4379] describes LSP traceroute as an operation where the
   initiating LSR sends a series of MPLS echo requests towards the same
   destination. The first packet in the series has the TTL set to 1.
   When the echo reply is received from the LSR one hop away, the second
   echo request in the series is sent with the TTL set to 2. For each
   additional echo request, the TTL is incremented by one until a
   response is received from the intended destination. The initiating
   LSR discovers and exercises ECMP by obtaining Multipath Information
   from each transit LSR and using a specific destination IP address or
   specific entropy label.

   From here on, the notation {x, y, z} refers to Multipath Information
   Types x, y, or z. Multipath Information Types are defined in
   Section 3.3 of [RFC4379] .

   The LSR initiating LSP ping sends an MPLS echo request with the
   Multipath Information. This Multipath Information is described in
   the echo request's DDMAP TLV and may contain a set of IP addresses or
   a set of labels. Multipath Information Types {2, 4, 8} carry a set
   of IP addresses, and the Multipath Information Type {9} carries a set
   of labels. The responder LSR (the receiver of the MPLS echo request)
   will determine the subset of initiator-specified Multipath
   Information, which load balances to each downstream (outgoing)
   interface. The responder LSR sends an MPLS echo reply with the
   resulting Multipath Information per downstream (outgoing interface)
   back to the initiating LSR. The initiating LSR is then able to use a
   specific IP destination address or a specific label to exercise a
   specific ECMP path on the responder LSR.

   The current behavior is problematic in the following scenarios:

   o The initiating LSR sends the IP Multipath Information, but the
      responder LSR load balances on labels.

   o The initiating LSR sends the Label Multipath Information, but the
      responder LSR load balances on IP addresses.

   o The initiating LSR sends the existing Multipath Information to an
      LSR that pushes ELI/EL in the label stack, but the initiating LSR
      can only continue to discover and exercise specific paths of the
      ECMP if the LSR that pushes ELI/EL responds with both IP addresses
      and the associated EL corresponding to each IP address. This is
      because:

      * An ELI/EL-pushing LSR that is a stitching point will load
         balance based on the IP address.

      * Downstream LSR(s) of an ELI/EL-pushing LSR may load balance
         based on ELs.

   o The initiating LSR sends existing Multipath Information to an ELI/
      EL-pushing LSR, but the initiating LSR can only continue to
      discover and exercise specific paths of ECMP if the ELI/EL-pushing
      LSR responds with both labels and the associated EL corresponding
      to the label. This is because:

      * An ELI/EL-pushing LSR that is a stitching point will load
         balance based on the EL from the previous LSP and push a new
         EL.

      * Downstream LSR(s) of ELI/EL-pushing LSR may load balance based
         on new ELs.

   The above scenarios demonstrate that the existing Multipath
   Information is insufficient when LSP traceroute is used on an LSP
   with entropy labels [RFC6790]. This document defines a new Multipath
   Information Type to be used in the DDMAP of MPLS echo request/reply
   packets for [RFC6790] LSPs.

   The responder LSR can reply with empty Multipath Information if no IP
   address set or if no label set is received with the Multipath
   Information. An empty return is also possible if an initiating LSR
   sends Multipath Information of one type, IP Address or Label, but the
   responder LSR load balances on the other type. To disambiguate
   between the two results, this document introduces new flags in the
   DDMAP TLV to allow the responder LSR to describe the load-balancing
   technique being used.

   To use this enhanced method end-to-end, all LSRs along the LSP need
   to be able to understand the new flags and the new Multipath
   Information Type. Mechanisms to verify this condition are outside of
   the scope of this document. The rest of the requirements are
   detailed in the initiating LSR and responder LSR procedures. Two
   additional DS Flags are defined for the DDMAP TLV in Section 6.
   These two flags are used by the responder LSR to describe its load-
   balancing behavior on a received MPLS echo request.

   Note that the terms "IP-Based Load Balancer" and "Label-Based Load
   Balancer" are in context of how a received MPLS echo request is
   handled by the responder LSR.

1.1. Terminology

   The following abbreviations and terms are used in this document:

   o MPLS: Multiprotocol Label Switching.

   o LSP: Label Switched Path.

   o Stitched LSP: Stitched Label Switched Paths combine several LSPs
      such that a single end-to-end LSP is realized. [RFC6424]
      describes LSP ping for Stitched LSPs.

   o LSR: Label Switching Router.

   o FEC: Forwarding Equivalence Class.

   o ECMP: Equal-Cost Multipath.

   o EL: Entropy Label.

   o ELI: Entropy Label Indicator.

   o GAL: Generic Associated Channel Label.

   o MS-PW: Multi-Segment Pseudowire.

   o Initiating LSR: An LSR that sends an MPLS echo request.

   o Responder LSR: An LSR that receives an MPLS echo request and sends
      an MPLS echo reply.

   o IP-Based Load Balancer: An LSR that load balances on fields from
      an IP header (and possibly fields from upper layers) and does not
      consider an entropy label from an MPLS label stack (i.e., flow
      label [RFC6391] or entropy label [RFC6790]) for load-balancing
      purposes.

   o Label-Based Load Balancer: An LSR that load balances on an entropy
      label from an MPLS label stack (i.e., flow label or entropy label)
      and does not consider fields from an IP header (and possibly
      fields from upper layers) for load-balancing purposes.

   o Label and IP-Based Load Balancer: An LSR that load balances on
      both entropy labels from an MPLS label stack and fields from an IP
      header (and possibly fields from upper layers).

1.1.1. Requirements Language

   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. Background

   MPLS implementations employ a wide variety of load-balancing
   techniques in terms of fields used for hash "keys". The mechanisms
   in [RFC4379] and updated by [RFC6424] are designed to provide
   multipath support for a subset of techniques. The intent of this
   document is to provide multipath support for the supported techniques
   that are compromised by the use of ELs [RFC6790]. Section 9
   describes supported and unsupported cases, and it may be useful for
   the reader to first review this section.

   The Downstream Detailed Mapping (DDMAP) TLV [RFC6424] provides
   Multipath Information, which can be used by an LSP ping initiator to
   trace and validate ECMP paths between an ingress and egress. The
   Multipath Information encodings defined by [RFC6424] are sufficient
   when all the LSRs along the path(s), between ingress and egress,
   consider the same set of "keys" as input for load-balancing
   algorithms, e.g., either all IP based or all label based.

   With the introduction of [RFC6790], some LSRs may perform load
   balancing based on labels while others may be IP based. This results
   in an LSP ping initiator that is unable to trace and validate all the
   ECMP paths in the following scenarios:

   o One or more transit LSRs along an LSP with ELI/EL in the label
      stack do not perform ECMP load balancing based on EL (hashes based
      on "keys" including the IP destination address). This scenario is
      not only possible but quite common due to transit LSRs not
      implementing [RFC6790] or transit LSRs implementing [RFC6790] but
      not implementing the suggested transit LSR behavior in Section 4.3
      of [RFC6790].

   o Two or more LSPs stitched together with at least one of these LSPs
      pushing ELI/EL into the label stack.

   These scenarios can be quite common because deployments of [RFC6790]
   typically have a mixture of nodes that support ELI/EL and nodes that
   do not. There will also typically be a mixture of areas that support
   ELI/EL and areas that do not.

As pointed out in [RFC6790], the procedures of [RFC4379] (and
consequently of [RFC6424]) with respect to Multipath Information Type
{9} are incomplete. However, [RFC6790] does not actually update
[RFC4379]. Further, the specific EL location is not clearly defined,
particularly in the case of Flow-Aware Pseudowires [RFC6391]. This
document defines a new FEC Stack sub-TLV for the entropy label.
Section 2 of this document updates the procedures for the Multipath
Information Type {9} that are described in [RFC4379] and that are
applicable to [RFC6424]. The rest of this document describes
extensions required to restore ECMP discovery and tracing
capabilities for the scenarios described.

[RFC4379], [RFC6424], and this document will support IP-based load
balancers and label-based load balancers that limit their hash to the
first (top-most) or only entropy label in the label stack. Other use
cases (refer to Section 9) are out of scope.

2. Multipath Type {9}

[RFC4379] defined Multipath Type {9} for the tracing of LSPs where
label-based load balancing is used. However, as pointed out in
[RFC6790], the procedures for using this type are incomplete as the
specific location of the label was not defined. It was assumed that
the presence of Multipath Type {9} implied that the value of the
bottom-of-stack label should be varied by the values indicated by the
multipath to determine the respective outgoing interfaces.

Section 4 defines a new FEC-Stack sub-TLV to indicate an entropy
label. These labels MAY appear anywhere in a label stack.

Multipath Type {9} applies to the first label in the label stack that
corresponds to an EL-FEC. If no such label is found, it applies to
the label at the bottom of the label stack.

3. Pseudowire Tracing

This section defines procedures for tracing Pseudowires. These
procedures pertain to the use of Multipath Information Type {9} as
well as Type {10}. In all cases below, when a control word is in
use, the N flag in the DDMAP MUST be set. Note that when a control
word is not in use, the returned DDMAPs may not be accurate.

In order to trace a Pseudowire that is not flow aware, the initiator
includes an EL-FEC instead of the appropriate PW FEC at the bottom of
the FEC Stack. Tracing in this way will cause compliant routers to
return the proper outgoing interface. Note that this procedure only
traces to the end of the MPLS LSP that is under test and will not
verify the PW FEC. To actually verify the PW FEC or in the case of a

MS-PW, to determine the next Pseudowire label value, the initiator
MUST repeat that step of the trace (i.e., repeating the TTL value
used) but with the FEC Stack modified to contain the appropriate PW
FEC. Note that these procedures are applicable to scenarios where an
initiator is able to vary the bottom label (i.e., Pseudowire label).
Possible scenarios are tracing multiple Pseudowires that are not flow
aware on the same endpoints or tracing a Pseudowire that is not flow-
aware provisioned with multiple Pseudowire labels.

In order to trace a flow-aware Pseudowire [RFC6391], the initiator
includes an EL FEC at the bottom of the FEC Stack and pushes the
appropriate PW FEC onto the FEC Stack.

In order to trace through routers that are not compliant, the
initiator forms an MPLS echo request message and includes a DDMAP
with the Multipath Type {9}. For a Pseudowire that is not flow
aware, it includes the appropriate PW FEC in the FEC Stack. For a
flow- aware Pseudowire, the initiator includes a Nil FEC at the
bottom of the FEC Stack and pushes the appropriate PW FEC onto the
FEC Stack.

4. Entropy Label FEC

The ELI is a reserved label that has no associated explicit FEC, and
has the label value 7 assigned from the reserved range. Use the Nil
FEC as the Target FEC Stack sub-TLV to account for ELI in a Target
FEC Stack TLV.

The EL is a special-purpose label with the label value being
discretionary (i.e., the label value is not from the reserved range).
For LSP verification mechanics to perform its purpose, it is
necessary for a Target FEC Stack sub-TLV to clearly describe the EL,
particularly in the scenario where the label stack does not carry ELI
(e.g., flow-aware Pseudowire [RFC6391]). Therefore, this document
defines an EL FEC sub-TLV (33, see Section 11.1) that allows a Target
FEC Stack sub-TLV to be added to the Target FEC Stack to account for
EL.

   The Length is 4. Labels are 20-bit values treated as numbers.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Label                |          MBZ          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
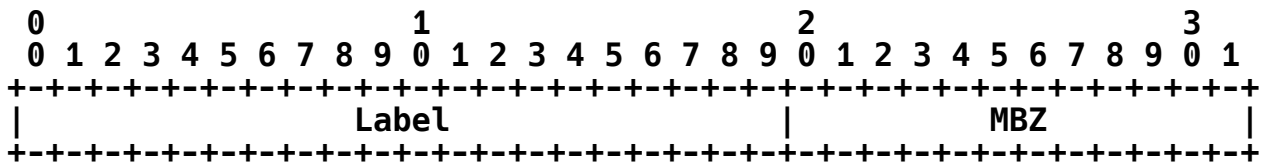
                     Figure 1: Entropy Label FEC

   "Label" is the actual label value inserted in the label stack; the
   "MBZ" field MUST be zero when sent and ignored on receipt.

5. DS Flags: L and E

   Two flags, L and E, are added to the DS Flags field of the DDMAP TLV.
   Both flags MUST NOT be set in the echo request packets when sending
   and SHOULD be ignored when received. Zero, one, or both new flags
   MUST be set in the echo reply packets.

   DS Flags

```
     0 1 2 3 4 5 6 7
    +-+-+-+-+-+-+-+-+
    |  MBZ  |L|E|I|N|
    +-+-+-+-+-+-+-+-+
```

   Flag   Name and Meaning
   ----   ----------------
     L   Label-based load balance indicator
          This flag MUST be cleared in the echo request. An LSR
          that performs load balancing on a label MUST set this
          flag in the echo reply. An LSR that performs load
          balancing on IP MUST NOT set this flag in the echo
          reply.

     E   ELI/EL push indicator
          This flag MUST be cleared in the echo request. An LSR
          that pushes ELI/EL MUST set this flag in the echo
          reply. An LSR that does not push ELI/EL MUST NOT set
          this flag in the echo reply.

The two flags result in four load-balancing techniques, which the
echo reply generating LSR can indicate:

o {L=0, E=0} LSR load balances based on IP and does not push ELI/EL.

o {L=0, E=1} LSR load balances based on IP and pushes ELI/EL.

o {L=1, E=0} LSR load balances based on labels and does not push
   ELI/EL.

o {L=1, E=1} LSR load balances based on labels and pushes ELI/EL.

6. New Multipath Information Type {10}

   One new Multipath Information Type is added to be used in DDMAP TLV.
   This new Multipath Type has the value of 10.

```
   Key   Type                   Multipath Information
   ---   ----------------       ---------------------
   10    IP and Label set       IP addresses and label prefixes
```

   Multipath Information Type {10} is comprised of three sections. The
   first section describes the IP address set. The second section
   describes the label set. The third section describes another label
   set, which associates to either the IP address set or the label set
   specified in the other sections.

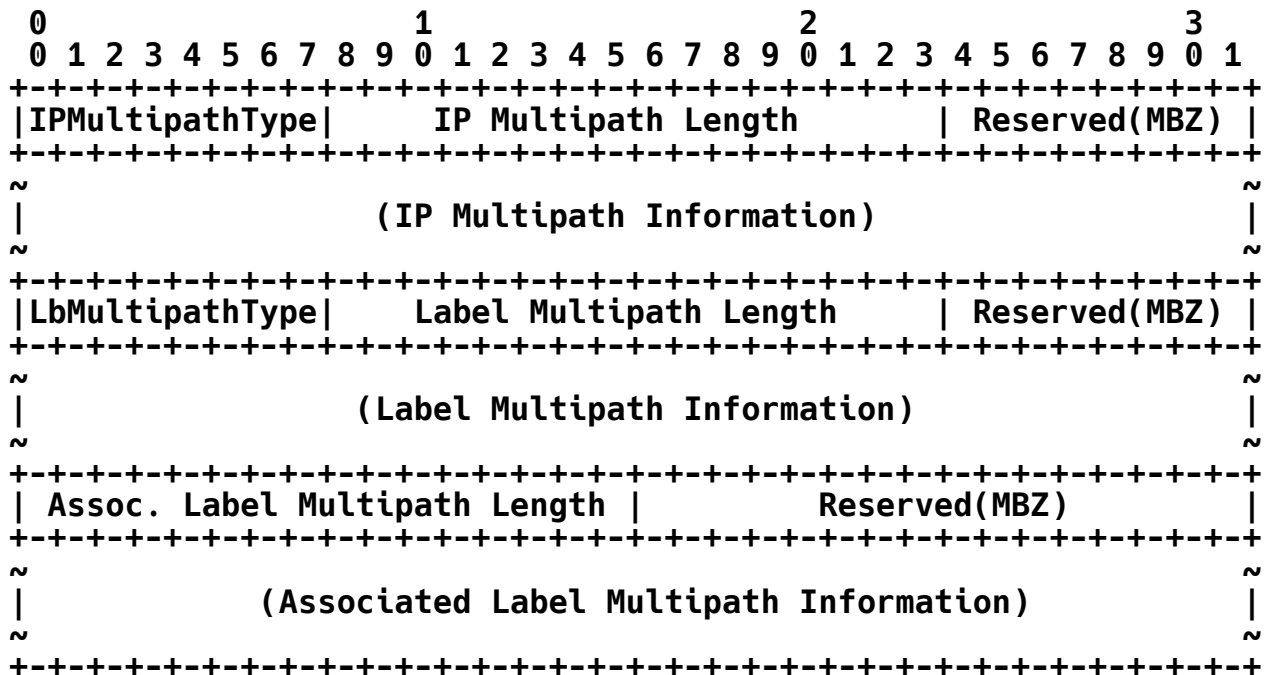   Multipath Information Type {10} has the following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|IPMultipathType|       IP Multipath Length       | Reserved(MBZ) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
|                 (IP Multipath Information)                    |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|LbMultipathType|    Label Multipath Length      | Reserved(MBZ) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
|               (Label Multipath Information)                   |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Assoc. Label Multipath Length |        Reserved(MBZ)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
|          (Associated Label Multipath Information)            |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

              Figure 2: Multipath Information Type {10}

   o IPMultipathType

      * 0 when "IP Multipath Information" is omitted. Otherwise, one
         of the IP Multipath Information values: {2, 4, 8}.

   o IP Multipath Information

      * This section is omitted when "IPMultipathType" is 0.
         Otherwise, this section reuses the IP Multipath Information
         from [RFC4379]. Specifically, Multipath Information for values
         {2, 4, 8} can be used.

   o LbMultipathType

      * 0 when the "Label Multipath Information" is omitted.
         Otherwise, the Label Multipath Information value {9}.

   o Label Multipath Information

      * This section is omitted when the "LbMultipathType" is 0.
         Otherwise, this section reuses the Label Multipath Information
         from [RFC4379]. Specifically, the Multipath Information for
         value {9} can be used.

   o Associated Label Multipath Information

      * "Associated Label Multipath Length" is a 16-bit field of
         Multipath Information that indicates the length in octets of
         the Associated Label Multipath Information.

      * "Associated Label Multipath Information" is a list of labels
         with each label described in 24 bits. This section MUST be
         omitted in an MPLS echo request message. A midpoint that
         pushes ELI/EL labels SHOULD include "Associated Label Multipath
         Information" in its MPLS echo reply message, along with either
         "IP Multipath Information" or "Label Multipath Information".
         Each specified associated label described in this section maps
         to a specific IP address OR label described in the "IP
         Multipath Information" section or the "Label Multipath
         Information" section. For example, if three IP addresses are
         specified in the "IP Multipath Information" section, then there
         MUST be three labels described in this section. The first
         label maps to the first IP address specified, the second label
         maps to the second IP address specified, and the third label
         maps to the third IP address specified.

   When a section is omitted, the length for that section MUST be set to
   zero.

7. Initiating LSR Procedures

   The following procedure is described in terms of an EL_LSP boolean
   maintained by the initiating LSR. This value controls the Multipath
   Information Type to be used in the transmitted echo request packets.
   When the initiating LSR is transmitting an echo request packet with
   DDMAP with a non-zero Multipath Information Type, then the EL_LSP
   boolean MUST be consulted to determine the Multipath Information Type
   to use.

   In addition to the procedures described in [RFC4379], as updated by
   Section 2 and [RFC6424], the initiating LSR MUST operate with the
   following procedures:

   o When the initiating LSR pushes ELI/EL, initialize EL_LSP=True.
      Else, set EL_LSP=False.

   o When the initiating LSR is transmitting a non-zero Multipath
      Information Type:

      * If (EL_LSP), the initiating LSR MUST use the Multipath
         Information Type {10} unless the responder LSR cannot handle
         Type {10}. When the initiating LSR is transmitting the
         Multipath Information Type {10}, both "IP Multipath
         Information" and "Label Multipath Information" MUST be
         included, and "Associated Label Multipath Information" MUST be
         omitted (NULL).

      * Else, the initiating LSR MAY use the Multipath Information Type
         {2, 4, 8, 9, 10}. When the initiating LSR is transmitting the
         Multipath Information Type {10} in this case, "IP Multipath
         Information" MUST be included, and "Label Multipath
         Information" and "Associated Label Multipath Information" MUST
         be omitted (NULL).

   o When the initiating LSR receives an echo reply with {L=0, E=1} in
      the DS Flags with valid contents, set EL_LSP=True.

   In the following conditions, the initiating LSR may have lost the
   ability to exercise specific ECMP paths. The initiating LSR MAY
   continue with "best effort" in the following cases:

   o Received echo reply contains empty Multipath Information.

   o Received echo reply contains {L=0, E=<any>} DS Flags, but does not
      contain IP Multipath Information.

   o Received echo reply contains {L=1, E=<any>} DS Flags, but does not
      contain Label Multipath Information.

   o Received echo reply contains {L=<any>, E=1} DS Flags, but does not
      contain Associated Label Multipath Information.

   o IP Multipath Information Types {2, 4, 8} sent, and received echo
      reply with {L=1, E=0} in DS Flags.

   o Multipath Information Type {10} sent, and received echo reply with
      Multipath Information Type other than {10}.

8. Responder LSR Procedures

    Common Procedures:

    o The responder LSR receiving an MPLS echo request packet MUST first
       determine whether or not the initiating LSR supports this LSP ping
       and traceroute extension for entropy labels. If either of the
       following conditions are met, the responder LSR SHOULD determine
       that the initiating LSR supports this LSP ping and traceroute
       extension for entropy labels.

       1. Received MPLS echo request contains the Multipath Information
           Type {10}.

       2. Received MPLS echo request contains a Target FEC Stack TLV
           that includes the entropy label FEC.

       If the initiating LSR is determined not to support this LSP ping
       and traceroute extension for entropy labels, then the responder
       LSR MUST NOT follow further procedures described in this section.
       Specifically, MPLS echo reply packets:

       * MUST have the following DS Flags cleared (i.e., not set): "ELI/
          EL push indicator" and "Label-based load balance indicator".

       * MUST NOT use the Multipath Information Type {10}.

    o The responder LSR receiving an MPLS echo request packet with the
       Multipath Information Type {10} MUST validate the following
       contents. Any deviation MUST result in the responder LSR
       considering the packet to be malformed and returning code 1
       ("Malformed echo request received") in the MPLS echo reply packet.

       * IP Multipath Information MUST be included.

       * Label Multipath Information MAY be included.

       * Associated Label Multipath Information MUST be omitted (NULL).

    The following subsections describe expected responder LSR procedures
    when the echo reply is to include DDMAP TLVs, based on the local load
    balance technique being employed. In case the responder LSR performs
    deviating load balance techniques on a per-downstream basis,
    appropriate procedures matched to each downstream load balance
    technique MUST be followed.

8.1. IP-Based Load Balancer That Does Not Push ELI/EL

   o The responder MUST set {L=0, E=0} in DS Flags.

   o If the Multipath Information Type {2, 4, 8} is received, the
      responder MUST comply with [RFC4379] and [RFC6424].

   o If the Multipath Information Type {9} is received, the responder
      MUST reply with Multipath Type {0}.

   o If the Multipath Information Type {10} is received, the following
      procedures are to be used:

      * The responder MUST reply with the Multipath Information Type
         {10}.

      * The "Label Multipath Information" and "Associated Label
         Multipath Information" sections MUST be omitted (NULL).

      * If no matching IP address is found, then the "IPMultipathType"
         field MUST be set to the Multipath Information Type {0} and the
         "IP Multipath Information" section MUST also be omitted (NULL).

      * If at least one matching IP address is found, then the
         "IPMultipathType" field MUST be set to the appropriate
         Multipath Information Type {2, 4, 8} and the "IP Multipath
         Information" section MUST be included.

8.2. IP-Based Load Balancer That Pushes ELI/EL

   o The responder MUST set {L=0, E=1} in DS Flags.

   o If the Multipath Information Type {9} is received, the responder
      MUST reply with Multipath Type {0}.

   o If the Multipath Type {2, 4, 8, 10} is received, the following
      procedures are to be used:

      * The responder MUST respond with Multipath Type {10}. See
         Section 6 for details of Multipath Type {10}.

      * The "Label Multipath Information" section MUST be omitted
         (i.e., it is not there).

      * The IP address set specified in the received IP Multipath
         Information MUST be used to determine the returned IP/Label
         pairs.

*   If the received Multipath Information Type was {10}, the
    received "Label Multipath Information" sections MUST NOT be
    used to determine the associated label portion of the returned
    IP/Label pairs.

*   If no matching IP address is found, then the "IPMultipathType"
    field MUST be set to the Multipath Information Type {0} and the
    "IP Multipath Information" section MUST be omitted. In
    addition, the "Associated Label Multipath Length" MUST be set
    to 0, and the "Associated Label Multipath Information" section
    MUST also be omitted.

*   If at least one matching IP address is found, then the
    "IPMultipathType" field MUST be set to the appropriate
    Multipath Information Type {2, 4, 8} and the "IP Multipath
    Information" section MUST be included. In addition, the
    "Associated Label Multipath Information" section MUST be
    populated with a list of labels corresponding to each IP
    address specified in the "IP Multipath Information" section.
    "Associated Label Multipath Length" MUST be set to a value
    representing the length in octets of the "Associated Label
    Multipath Information" field.

8.3. Label-Based Load Balancer That Does Not Push ELI/EL

   o The responder MUST set {L=1, E=0} in DS Flags.

   o If the Multipath Information Type {2, 4, 8} is received, the
      responder MUST reply with Multipath Type {0}.

   o If the Multipath Information Type {9} is received, the responder
      MUST comply with [RFC4379] and [RFC6424] as updated by Section 2.

   o If the Multipath Information Type {10} is received, the following
      procedures are to be used:

      *   The responder MUST reply with the Multipath Information Type
          {10}.

      *   The "IP Multipath Information" and "Associated Label Multipath
          Information" sections MUST be omitted (NULL).

      *   If no matching label is found, then the "LbMultipathType" field
          MUST be set to the Multipath Information Type {0} and the
          "Label Multipath Information" section MUST also be omitted
          (NULL).

   * If at least one matching label is found, then the
      "LbMultipathType" field MUST be set to the appropriate
      Multipath Information Type {9} and the "Label Multipath
      Information" section MUST be included.

8.4. Label-Based Load Balancer That Pushes ELI/EL

   o The responder MUST set {L=1, E=1} in DS Flags.

   o If the Multipath Information Type {2, 4, 8} is received, the
      responder MUST reply with Multipath Type {0}.

   o If the Multipath Type {9, 10} is received, the following
      procedures are to be used:

      * The responder MUST respond with the Multipath Type {10}.

      * The "IP Multipath Information" section MUST be omitted.

      * The label set specified in the received Label Multipath
         Information MUST be used to determine the returned Label/Label
         pairs.

      * If the received Multipath Information Type was {10} received,
         the "Label Multipath Information" sections MUST NOT be used to
         determine the associated label portion of the returned Label/
         Label pairs.

      * If no matching label is found, then the "LbMultipathType" field
         MUST be set to the Multipath Information Type {0} and the
         "Label Multipath Information" section MUST be omitted. In
         addition, the "Associated Label Multipath Length" MUST be set
         to 0, and the "Associated Label Multipath Information" section
         MUST also be omitted.

      * If at least one matching label is found, then the
         "LbMultipathType" field MUST be set to the appropriate
         Multipath Information Type {9} and the "Label Multipath
         Information" section MUST be included. In addition, the
         "Associated Label Multipath Information" section MUST be
         populated with a list of labels corresponding to each label
         specified in the "Label Multipath Information" section. The
         "Associated Label Multipath Length" MUST be set to a value
         representing the length in octets of the "Associated Label
         Multipath Information" field.

8.5. Flow-Aware MS-PW Stitching LSR

   A stitching LSR that cross-connects flow-aware Pseudowires behaves in
   one of two ways:

   o Load balances on the previous flow label and carries over the same
      flow label. For this case, the stitching LSR is to behave as
      described in Section 8.3.

   o Load balances on the previous flow label and replaces the flow
      label with a newly computed label. For this case, the stitching
      LSR is to behave as described in Section 8.4.

9. Supported and Unsupported Cases

   The MPLS architecture does not define strict rules on how
   implementations are to identify hash "keys" for load-balancing
   purposes. As a result, implementations may be of the following load
   balancer types:

   1. IP-based load balancer.
   2. Label-based load balancer.
   3. Label- and IP-based load balancer.

   For cases (2) and (3), an implementation can include different sets
   of labels from the label stack for load-balancing purpose. Thus, the
   following sub-cases are possible:

   a. Entire label stack.
   b. Top N labels from label stack where the number of labels in label
       stack is > N.
   c. Bottom N labels from label stack where the number of labels in
       label stack is > N.

   In a scenario where there is one flow label or entropy label present
   in the label stack, the following further cases are possible for
   (2b), (2c), (3b), and (3c):

   1. N labels from label stack include flow label or entropy label.
   2. N labels from label stack do not include flow label or entropy
       label.

Also, in a scenario where there are multiple entropy labels present
in the label stack, it is possible for implementations to employ
deviating techniques:

o Search for entropy stops at the first entropy label.

o Search for entropy includes any entropy label found plus continues
   to search for entropy in the label stack.

Furthermore, handling of reserved (i.e., special) labels varies among
implementations:

o Reserved labels are used in the hash as any other label would be
   (not a recommended practice).

o Reserved labels are skipped over and, for implementations limited
   to N labels, the reserved labels do not count towards the limit of
   N.

o Reserved labels are skipped over and, for implementations limited
   to N labels, the reserved labels count towards the limit of N.

It is important to point this out since the presence of GAL will
affect those implementations that include reserved labels for load-
balancing purposes.

As can be seen from the above, there are many types of potential
load-balancing implementations. Attempting to get any Operations,
Administration, and Maintenance (OAM) tools to support ECMP discovery
and traversal over all types would require fairly complex procedures.
Complexities in OAM tools have minimal benefit if the majority of
implementations are expected to employ only a small subset of the
cases described above.

o Section 4.3 of [RFC6790] states that in implementations, for load-
   balancing purposes, parsing beyond the label stack after finding
   an entropy label has "limited incremental value". Therefore, it
   is expected that most implementations will be of types "IP-based
   load balancer" or "Label-based load balancer".

o Section 2.4.5.1 of [RFC7325] recommends that searching for entropy
   labels in the label stack should terminate upon finding the first
   entropy label. Therefore, it is expected that implementations
   will only include the first (top-most) entropy label when there
   are multiple entropy labels in the label stack.

   o It is expected that, in most cases, the number of labels in the
      label stack will not exceed the number of labels (N) that
      implementations can include for load-balancing purposes.

   o It is expected that labels in the label stack, besides the flow
      label and entropy label, are constant for the lifetime of a single
      LSP multipath traceroute operation. Therefore, deviating load-
      balancing implementations with respect to reserved labels should
      not affect this tool.

   Thus, [RFC4379], [RFC6424], and this document support cases (1) and
   (2a1), where only the first (top-most) entropy label is included when
   there are multiple entropy labels in the label stack.

10. Security Considerations

   While [RFC4379] and [RFC6424] already allow for the discovery and
   exercise of ECMP paths, this document extends the LSP ping and
   traceroute mechanisms to more precisely discover and exercise ECMP
   paths when an LSP uses ELI/EL in the label stack. Sourcing or
   inspecting LSP ping packets can be used for network reconnaissance.

   The extended capability defined in this document requires minor
   additional processing for the responder and initiator nodes. The
   responder node that pushes ELI/EL will need to compute and return
   multipath data including associated EL. The initiator node will need
   to store and handle both IP Multipath and Label Multipath
   Information, and include destination IP addresses and/or ELs in MPLS
   echo request packets as well as in the Multipath Information sent to
   downstream nodes. The security considerations of [RFC4379] already
   cover Denial-of-Service attacks by regulating LSP ping traffic going
   to the control plane.

   Finally, the security measures described in [RFC4379], [RFC6424], and
   [RFC6790] are applicable. [RFC6424] provides guidelines if a network
   operator wants to prevent tracing or does not want to expose details
   of the tunnel and [RFC6790] provides guidance on the use of the EL.

11. IANA Considerations

11.1. Entropy Label FEC

   IANA has assigned a new sub-TLV from the "Sub-TLVs for TLV Types 1,
   16, and 21" section from the "Multi-Protocol Label Switching (MPLS)
   Label Switched Paths (LSPs) Ping Parameters" registry under "TLVs"
   ([IANA-MPLS-LSP-PING]).

Sub-Type	Sub-TLV Name	Reference
33	Entropy label FEC	this document

11.2. DS Flags

   IANA has assigned new bit numbers from the "DS Flags" subregistry
   from the "TLVs" section of the "Multi-Protocol Label Switching (MPLS)
   Label Switched Paths (LSPs) Ping Parameters" registry
   ([IANA-MPLS-LSP-PING]).

   Note: The "DS Flags" subregistry was created by [RFC7537].

Bit number	Name	Reference
5	E: ELI/EL push indicator	this document
4	L: Label-based load balance indicator	this document

11.3. Multipath Type

   IANA has assigned a new value from the "Multipath Type" subregistry
   from the "TLVs" section of the "Multi-Protocol Label Switching (MPLS)
   Label Switched Paths (LSPs) Ping Parameters" registry
   ([IANA-MPLS-LSP-PING]).

   Note: The "Multipath Type" subregistry was created by [RFC7537].

Value	Meaning	Reference
10	IP and label set	this document

12. References

12.1. Normative References

   [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>.

   [RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol
              Label Switched (MPLS) Data Plane Failures", RFC 4379,
              DOI 10.17487/RFC4379, February 2006,
              <http://www.rfc-editor.org/info/rfc4379>.

   [RFC6424] Bahadur, N., Kompella, K., and G. Swallow, "Mechanism for
              Performing Label Switched Path Ping (LSP Ping) over MPLS
              Tunnels", RFC 6424, DOI 10.17487/RFC6424, November 2011,
              <http://www.rfc-editor.org/info/rfc6424>.

   [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>.

   [RFC7537] Decraene, B., Akiya, N., Pignataro, C., Andersson, L., and
              S. Aldrin, "IANA Registries for LSP Ping Code Points",
              RFC 7537, DOI 10.17487/RFC7537, May 2015,
              <http://www.rfc-editor.org/info/rfc7537>.

12.2. Informative References

   [IANA-MPLS-LSP-PING]
              IANA, "Multi-Protocol Label Switching (MPLS) Label
              Switched Paths (LSPs) Ping Parameters",
              <http://www.iana.org/assignments/mpls-lsp-ping-parameters>.

   [RFC6391] Bryant, S., Ed., Filsfils, C., Drafz, U., Kompella, V.,
              Regan, J., and S. Amante, "Flow-Aware Transport of
              Pseudowires over an MPLS Packet Switched Network",
              RFC 6391, DOI 10.17487/RFC6391, November 2011,
              <http://www.rfc-editor.org/info/rfc6391>.

   [RFC7325] Villamizar, C., Ed., Kompella, K., Amante, S., Malis, A.,
              and C. Pignataro, "MPLS Forwarding Compliance and
              Performance Requirements", RFC 7325, DOI 10.17487/RFC7325,
              August 2014, <http://www.rfc-editor.org/info/rfc7325>.

Authors' Addresses

   Nobo Akiya
   Big Switch Networks
   Email: nobo.akiya.dev@gmail.com

   George Swallow
   Cisco Systems, Inc.
   Email: swallow@cisco.com

   Carlos Pignataro
   Cisco Systems, Inc.
   Email: cpignata@cisco.com

   Andrew G. Malis
   Huawei Technologies
   Email: agmalis@gmail.com

   Sam Aldrin
   Google
   Email: aldrin.ietf@gmail.com