Internet Engineering Task Force   N. Akiya
Internet-Draft   G. Swallow
Updates: 4379,6424 (if approved)   Cisco Systems
Intended status: Standards Track   S. Litkowski
Expires: May 26, 2015   B. Decraene
                                                              Orange
                                                             J. Drake
                                                     Juniper Networks
                                                   November 22, 2014

          Label Switched Path (LSP) Ping/Trace Multipath Support for
                  Link Aggregation Group (LAG) Interfaces
                 draft-akiya-mpls-lsp-ping-lag-multipath-03

Abstract

   This document defines an extension to the Multiprotocol Label
   Switching (MPLS) Label Switched Path (LSP) Ping and Traceroute to
   describe Multipath Information for Link Aggregation (LAG) member
   links separately, thus allowing MPLS LSP Ping and Traceroute to
   discover and exercise specific paths of layer 2 (L2) Equal-Cost
   Multipath (ECMP) over LAG interfaces.

   This document updates RFC4379 and RFC6424.

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].

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

1.1. Terminology

   The following acronyms/terminologies are used in this document:

   o MPLS - Multiprotocol Label Switching.

   o LSP - Label Switched Path.

   o LSR - Label Switching Router.

   o ECMP - Equal-Cost Multipath.

   o LAG - Link Aggregation.

   o Initiating LSR - LSR which sends MPLS echo request.

   o Responder LSR - LSR which receives MPLS echo request and sends
      MPLS echo reply.

1.2. Background

   The Multiprotocol Label Switching (MPLS) Label Switched Path (LSP)
   Ping and Traceroute [RFC4379] are powerful tools designed to diagnose
   all available layer 3 (L3) paths of LSPs, i.e. provides diagnostic
   coverage of L3 Equal-Cost Multipath (ECMP). In many MPLS networks,
   Link Aggregation (LAG) as defined in [IEEE802.1AX], which provide
   layer 2 (L2) ECMP, are often used for various reasons. MPLS LSP Ping
   and Traceroute tools were not designed to discover and exercise
   specific paths of L2 ECMP. Result raises a limitation for following
   scenario when LSP X traverses over LAG Y:

   o MPLS switching of LSP X over one or more member links of LAG Y is
      succeeding.

   o MPLS switching of LSP X over one or more member links of LAG Y is
      failing.

   o MPLS echo request for LSP X over LAG Y is load balanced over a
      member link which is MPLS switching successfully.

   With above scenario, MPLS LSP Ping and Traceroute will not be able to
   detect the MPLS switching failure of problematic member link(s) of
   the LAG. In other words, lack of L2 ECMP discovery and exercise
   capability can produce an outcome where MPLS LSP Ping and Traceroute
   can be blind to MPLS switching failures over LAG interface that are
   impacting MPLS traffic. It is, thus, desirable to extend the MPLS

LSP Ping and Traceroute to have deterministic diagnostic coverage of
LAG interfaces.

2. Overview

This document defines an extension to the MPLS LSP Ping and
Traceroute to describe Multipath Information for LAG member links
separately, thus allowing MPLS LSP Ping and Traceroute to discover
and exercise specific paths of L2 ECMP over LAG interfaces. Reader
is expected to be familiar with mechanics of the MPLS LSP Ping and
Traceroute described in Section 3.3 of [RFC4379] and Downstream
Detailed Mapping TLV (DDMAP) described in Section 3.3 of [RFC6424].

MPLS echo request carries a DDMAP and an optional TLV to indicate
that separate load balancing information for each L2 nexthop over LAG
is desired in MPLS echo reply. Responder LSR places the same
optional TLV in the MPLS echo reply to provide acknowledgement back
to the initiator. It also adds, for each downstream LAG member, a
load balance information (i.e. multipath information and interface
index). For example:
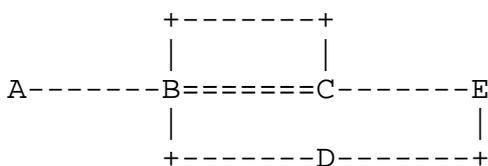
```
   <----- LDP Network ----->


          +-------+
          |       |
  A-------B=======C-------E
          |       |
          +-------D-------+


   ---- Non-LAG
   ==== LAG comprising of two member links
```

        Figure 1: Example LDP Network

When node A is initiating LSP Traceroute to node E, node B will
return to node A load balance information for following entries.

1. Downstream C over Non-LAG (upper path).

2. First Downstream C over LAG (middle path).

3. Second Downstream C over LAG (middle path).

4. Downstream D over Non-LAG (lower path).

This document defines:

o In Section 3, a mechanism to discover L2 ECMP multipath
   information;

o In Section 4, a mechanism to validate L2 ECMP traversal in some
   LAG provisioning models;

o In Section 5, the LAG Interface Info TLV;

o In Section 6, the LAG Description Indicator flag;

o In Section 7, the Interface Index Sub-TLV;

o In Section 8, the Detailed Interface and Label Stack TLV;

o In Appendix A, issues with LAG having an L2 Switch.

Note that the mechanism described in this document does not impose
any changes to scenarios where an LSP is pinned down to a particular
LAG member (i.e. the LAG is not treated as one logical interface by
the LSP).

3. Mechanism to Discover L2 ECMP Multipath

   The MPLS echo request carries a DDMAP and the LAG Interface Info TLV
   (described in Section 5) to indicate that separate load balancing
   information for each L2 nexthop over LAG is desired in MPLS echo
   reply. Responder LSRs that understand the LAG Interface Info TLV but
   unable to describe outgoing LAG member links separately MUST add the
   LAG Interface Info TLV in the MPLS echo reply to provide
   acknowledgement back to the initiating LSR. The Downstream LAG Info
   Accommodation flag MUST NOT be set in LAG Interface Info Flags. The
   responder LSRs that understands the LAG Interface Info TLV and able
   to describe outgoing LAG member links separately MUST use the follow
   procedures, regardless of whether or not outgoing interfaces include
   LAG interfaces:

   o MUST add the LAG Interface Info TLV in the MPLS echo reply to
      provide acknowledgement back to the initiator. The Downstream LAG
      Info Accommodation flag MUST be set in the LAG Interface Info
      Flags field.

   o For each downstream that is a LAG interface:

      * MUST add DDMAP in the MPLS echo reply.

      * MUST set the LAG Description Indicator flag in the DS Flags
         field (described in Section 6) of the DDMAP.

* In the DDMAP, Interface Index Sub-TLV and Multipath Data Sub-
  TLV are to describe each LAG member link. All other fields of
  the DDMAP are to describe the LAG interface.

* For each LAG member link of this LAG interface:

  + MUST add an Interface Index Sub-TLV (described in Section 7)
    with the LAG Member Link Indicator flag set in the Interface
    Index Flags field, describing this LAG member link.

  + MUST add an Multipath Data Sub-TLV for this LAG member link,
    if received DDMAP requested multipath information.

When both the Interface Index Sub-TLV and the Multipath Data Sub-TLV
is placed in the DDMAP to describe a LAG member link, Interface Index
Sub-TLV MUST be added first with Multipath Data Sub-TLV immediately
following.

For example, a responder LSR possessing a LAG interface with two
member links would send the following DDMAP for this LAG interface:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   DDMAP fields describing LAG interface with DS Flags G set   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Interface Index Sub-TLV of LAG member link #1         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Multipath Data Sub-TLV LAG member link #1           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Interface Index Sub-TLV of LAG member link #2         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Multipath Data Sub-TLV LAG member link #2           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Label Stack Sub-TLV                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
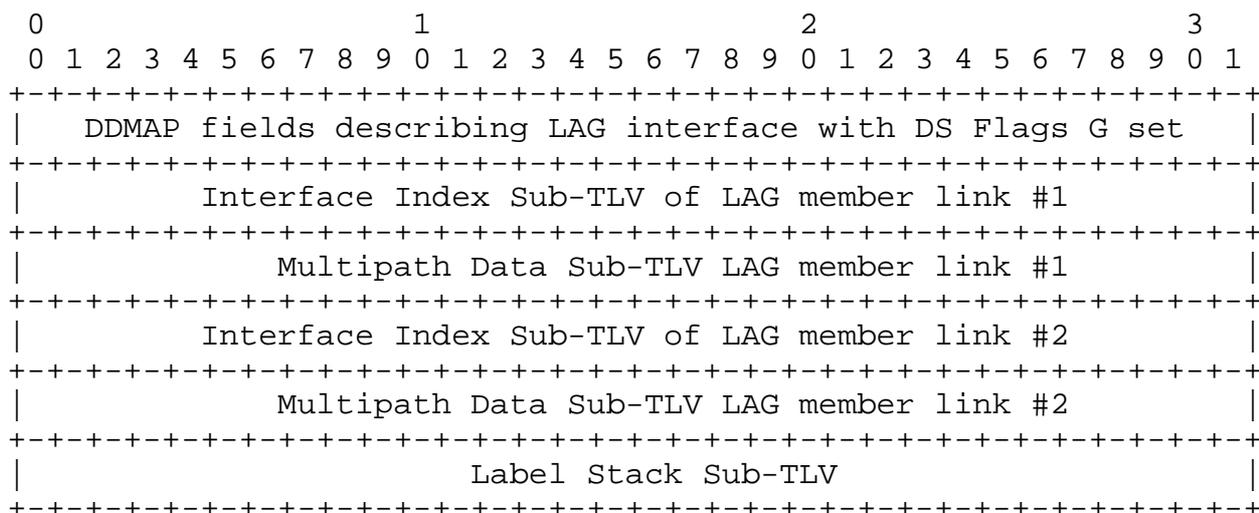
        Figure 2: Example of DDMAP in MPLS Echo Reply

These procedures allow initiating LSR to:

o Mandate the responder LSR to always add the LAG Interface Info TLV
   in the MPLS echo reply. This allows the initiating LSR to
   identify whether or not the responder LSR understands the LAG
   Interface Info TLV and can describe outgoing LAG member links
   separately.

   o Utilize the value of the LAG Description Indicator flag in DS
      Flags to identify whether each DDMAP describes a LAG interface or
      a non-LAG interface.

   o Obtain multipath information which is expected to traverse the
      specific LAG member link described by corresponding interface
      index.

   When an initiating LSR receives a DDMAP containing LAG member
   information from a downstream LSR with TTL=n, then the subsequent
   DDMAP sent by the initiating LSR to the downstream LSR with TTL=n+1
   through a particular LAG member link MUST be updated with following
   procedures:

   o The Interface Index Sub-TLVs MUST NOT be present in the sending
      DDMAP.

   o The Multipath Data Sub-TLVs SHOULD be updated to include just the
      one corresponding to the LAG member link being traversed. The
      initiating LSR MAY combine the Multipath Data Sub-TLVs for all LAG
      member links into a single Multipath Data Sub-TLV, but there MUST
      be only one Multipath Data Sub-TLV in the sending DDMAP.

   o All other fields of the DDMAP are to comply with procedures
      described in [RFC6424].

   Using the DDMAP example described in the Figure 2, the DDMAP being
   sent by the initiating LSR through LAG member link #1 to the next
   downstream LSR should be:

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |    DDMAP fields describing LAG interface with DS Flags G set   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Multipath Data Sub-TLV LAG member link #1           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                      Label Stack Sub-TLV                       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
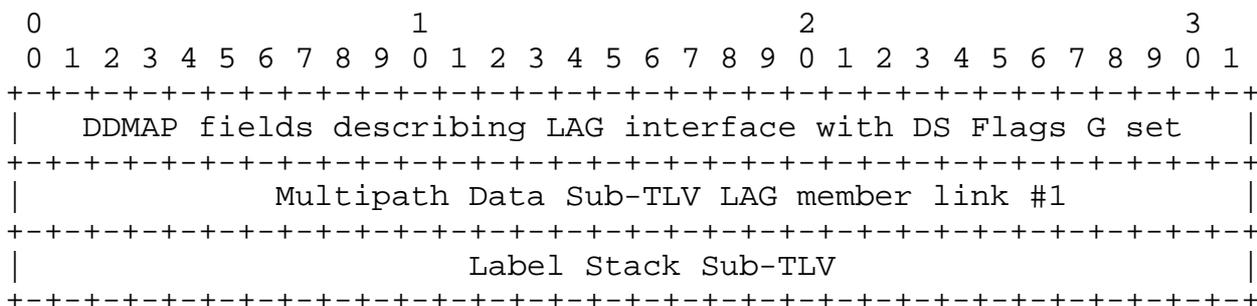
           Figure 3: Example of DDMAP in MPLS Echo Request

4. Mechanism to Validate L2 ECMP Traversal

   This document does not update the FEC validation procedures nor the
   DDMAP validation procedures. Rather this document provides the
   mechanism for the initiating LSR to obtain additional information
   from the downstream LSRs when incoming and/or outgoing interfaces are

LAGs. With this additional information, it is the responsibility of
the initiating LSR to validate the L2 ECMP traversal.

The MPLS echo request is sent with a DDMAP with DS Flags I set and
the optional LAG Interface Info TLV to indicate the request for
Detailed Interface and Label Stack TLV with additional LAG member
link information (i.e. interface index) in the MPLS echo reply.
Responder LSRs that understands the LAG Interface Info TLV but unable
to describe incoming LAG member link MUST add the LAG Interface Info
TLV in the MPLS echo reply to provide acknowledgement back to the
initiator. The Upstream LAG Info Accommodation flag MUST NOT be set
in LAG Interface Info Flags. The responder LSRs that understands the
LAG Interface Info TLV and able to describe incoming LAG member link
MUST use the following procedures, regardless of whether or not
incoming interface was a LAG interface:

o Add the LAG Interface Info TLV in the MPLS echo reply to provide
   acknowledgement back to the initiator. The Upstream LAG Info
   Accommodation flag MUST be set in the LAG Interface Info Flags
   field.

o When the received DDMAP had DS Flags I set, add the Detailed
   Interface and Label Stack TLV (described in Section 8) in the MPLS
   echo reply.

o When the received DDMAP had DS Flags I set and incoming interface
   was a LAG, add the Incoming Interface Index Sub-TLV (described in
   Section 8.1.2). The LAG Member Link Indicator flag MUST be set in
   the Interface Index Flags field, and the Interface Index field set
   to the LAG member link which received the MPLS echo request.

These procedures allow initiating LSR to:

o Identify whether or not the responder LSR understands the LAG
   Interface Info TLV and can describe the incoming LAG member links
   (the responder LSR is mandated to always add the LAG Interface
   Info TLV in the MPLS echo reply).

Along with procedures described in Section 3, described procedures in
this section will allow an initiating LSR to know:

o The expected load balance information of every LAG member link, at
   LSR with TTL=n.

o With specific entropy, the expected interface index of the
   outgoing LAG member link at TTL=n.

   o With specific entropy, the interface index of the incoming LAG
      member link at TTL=n+1.

   Expectation is that there's a relationship between the interface
   index of the outgoing LAG member link at TTL=n and the interface
   index of the incoming LAG member link at TTL=n+1 for all discovered
   entropies. In other words, set of entropies that load balances to
   outgoing LAG member link X at TTL=n should all reach the nexthop on
   same incoming LAG member link Y at TTL=n+1.

   With additional logics added in the initiating LSR, following checks
   can be performed:

   o Success case:

      * Traversing LAG member=1 at TTL=n results in LAG member=1' as
         the incoming interface at TTL=n+1.

      * Traversing LAG member=2 at TTL=n results in LAG member=2' as
         the incoming interface at TTL=n+1.

   o Error case:

      * Traversing LAG member=1 at TTL=n results in LAG member=1' as
         the incoming interface at TTL=n+1.

      * Traversing LAG member=2 at TTL=n results in LAG member=1' as
         the incoming interface at TTL=n+1.

   Note that defined procedures will provide a deterministic result for
   LAG interfaces that are back-to-back connected between routers (i.e.
   no L2 switch in between). If there is a L2 switch between LSR at
   TTL=n and LSR at TTL=n+1, there is no guarantee that traversal of
   every LAG member link at TTL=n will result in reaching different
   interface index at TTL=n+1. Issues resulting from LAG with L2 switch
   in between are further described in Appendix A. LAG provisioning
   models in operated network should be considered when analyzing the
   output of LSP Traceroute exercising L2 ECMPs.

5. LAG Interface Info TLV

   The LAG Interface Info object is a new TLV that MAY be included in
   the MPLS echo request message. An MPLS echo request MUST NOT include
   more than one LAG Interface Info object. Presence of LAG Interface
   Info object is a request that responder LSR describes upstream and
   downstream LAG interfaces according to procedures defined in this
   document. If the responder LSR is able to accommodate this request,

then the LAG Interface Info object MUST be included in the MPLS echo
reply message.

LAG Interface Info TLV Type is TBD1. Length is 4. The Value field
of LAG Interface TLV has 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   LAG Interface Info Flags    |           Must Be Zero        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

        Figure 4: LAG Interface Info TLV

LAG Interface Info Flags

   LAG Interface Info Flags field is a bit vector with following
   format.

```
 0                   1
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Must Be Zero (Reserved) |U|D|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   Two flags are defined: U and D. The remaining flags MUST be set
   to zero when sending and ignored on receipt. Both U and D flags
   MUST be cleared in MPLS echo request message when sending, and
   ignored on receipt. Either or both U and D flags MAY be set in
   MPLS echo reply message.

   Flag Name and Meaning
   ---- ----------------

      U   Upstream LAG Info Accommodation

          When this flag is set, LSR is capable of placing Incoming
          Interface Index Sub-TLV, describing LAG member link, in
          the Detailed Interface and Label Stack TLV.

      D   Downstream LAG Info Accommodation

          When this flag is set, LSR is capable of placing Interface
          Index Sub-TLV and Multipath Data Sub-TLV, describing LAG
          member link, in the Downstream Detailed Mapping TLV.

6. DDMAP TLV DS Flags: G

   One flag, G, is added in DS Flags field of the DDMAP TLV. In the
   MPLS echo request message, G flag MUST be cleared when sending, and
   ignored on receipt. In the MPLS echo reply message, G flag MUST be
   set if the DDMAP TLV describes a LAG interface. It MUST be cleared
   otherwise.

   DS Flags

      DS Flags G is added, in Bit Number TBD4, in DS Flags bit vector.

       0 1 2 3 4 5 6 7
      +-+-+-+-+-+-+-+-+
      | MBZ |G|MBZ|I|N|
      +-+-+-+-+-+-+-+-+
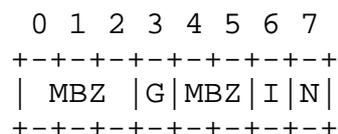
      RFC-Editor-Note: Please update above figure to place the flag G in
      the bit number TBD4.

      Flag Name and Meaning
      ---- ----------------

         G LAG Description Indicator

            When this flag is set, DDMAP describes a LAG interface.

7. Interface Index Sub-TLV

   The Interface Index object is a Sub-TLV that MAY be included in a
   DDMAP TLV. Zero or more Interface Index object MAY appear in a DDMAP
   TLV. The Interface Index Sub-TLV describes the index assigned by
   local LSR to the egress interface.

   Interface Index Sub-TLV Type is TBD2. Length is 8, and the Value
   field has 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
      +-+
      | Interface Index Flags | Must Be Zero |
      +-+
      | Interface Index |
      +-+
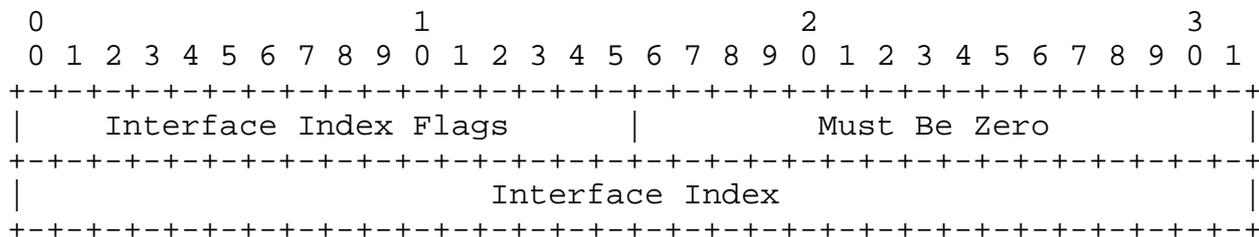
           Figure 5: Interface Index Sub-TLV

   Interface Index Flags

Interface Index Flags field is a bit vector with following format.

```
 0                   1
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Must Be Zero (Reserved)  |M|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

One flag is defined: M. The remaining flags MUST be set to zero
when sending and ignored on receipt.

Flag   Name and Meaning
----   ----------------

  M   LAG Member Link Indicator

      When this flag is set, interface index described in
      this sub-TLV is member of a LAG.

Interface Index

   Index assigned by the LSR to this interface.

8. Detailed Interface and Label Stack TLV

   The Detailed Interface and Label Stack object is a TLV that MAY be
   included in a MPLS echo reply message to report the interface on
   which the MPLS echo request message was received and the label stack
   that was on the packet when it was received. A responder LSR MUST
   NOT insert more than one instance of this TLV. This TLV allows the
   initiating LSR to obtain the exact interface and label stack
   information as it appears at the responder LSR.

   Detailed Interface and Label Stack TLV Type is TBD3. Length is K +
   Sub-TLV Length, and the Value field has 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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Address Type  |                 Must Be Zero                  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 IP Address (4 or 16 octets)                   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  Interface (4 or 16 octets)                   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          Must Be Zero          |         Sub-TLV Length        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   .                                                               .
   .                       List of Sub-TLVs                        .
   .                                                               .
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
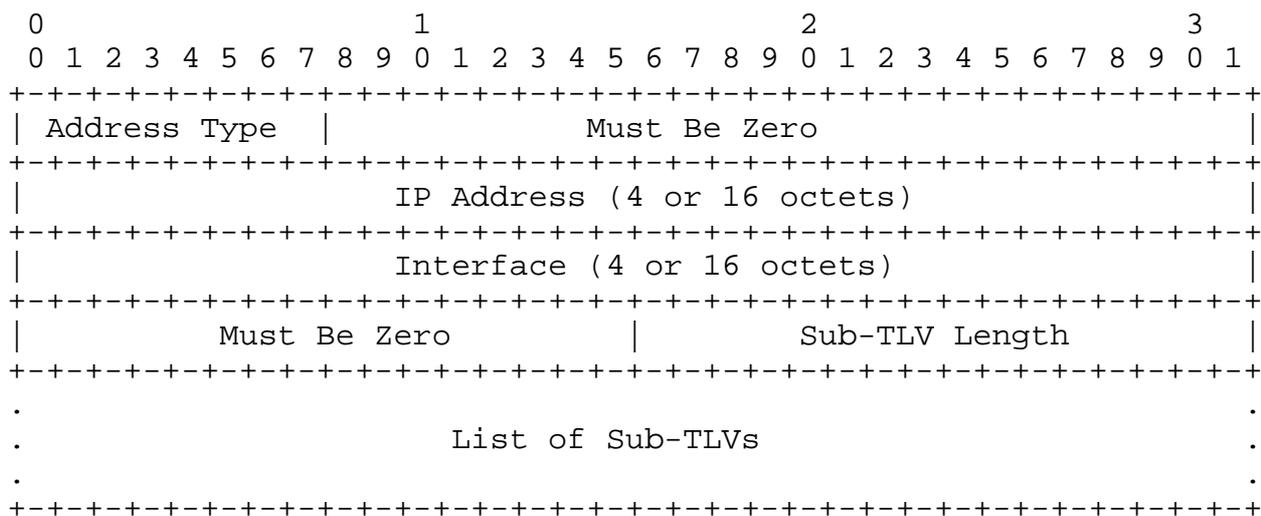
              Figure 6: Detailed Interface and Label Stack TLV

The Detailed Interface and Label Stack TLV format is derived from the
Interface and Label Stack TLV format (from [RFC4379]). Two changes
are introduced. First is that label stack, which is of variable
length, is converted into a sub-TLV. Second is that a new sub-TLV is
added to describe an interface index. The fields of Detailed
Interface and Label Stack TLV have the same use and meaning as in
[RFC4379]. A summary of the fields taken from the Interface and
Label Stack TLV is as below:

   Address Type

      The Address Type indicates if the interface is numbered or
      unnumbered. It also determines the length of the IP Address
      and Interface fields. The resulting total for the initial part
      of the TLV is listed in the table below as "K Octets". The
      Address Type is set to one of the following values:

            Type #   Address Type   K Octets
            ------   ------------   --------
                 1   IPv4 Numbered   16
                 2   IPv4 Unnumbered   16
                 3   IPv6 Numbered   40
                 4   IPv6 Unnumbered   28

   IP Address and Interface

      IPv4 addresses and interface indices are encoded in 4 octets;
      IPv6 addresses are encoded in 16 octets.

If the interface upon which the echo request message was received is numbered, then the Address Type MUST be set to IPv4 Numbered or IPv6 Numbered, the IP Address MUST be set to either the LSR's Router ID or the interface address, and the Interface MUST be set to the interface address.

If the interface is unnumbered, the Address Type MUST be either IPv4 Unnumbered or IPv6 Unnumbered, the IP Address MUST be the LSR's Router ID, and the Interface MUST be set to the index assigned to the interface.

Note: Usage of IPv6 Unnumbered has the same issue as [RFC4379], described in Section 3.4.2 of [I-D.ietf-mpls-ipv6-only-gap]. A solution should be considered an applied to both [RFC4379] and this document.

Sub-TLV Length

Total length in octets of the sub-TLVs associated with this TLV.

8.1. Sub-TLVs

This section defines the sub-TLVs that MAY be included as part of the Detailed Interface and Label Stack TLV.

```
        Sub-Type      Value Field
        ---------     ------------
           1          Incoming Label stack
           2          Incoming Interface Index
```

8.1.1. Incoming Label Stack Sub-TLV

The Incoming Label Stack sub-TLV contains the label stack as received by the LSR. If any TTL values have been changed by this LSR, they SHOULD be restored.

Incoming Label Stack Sub-TLV Type is 1. Length is variable, and the Value field has 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Label                  | TC  |S|     TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
.                                                               .
.                                                               .
.                                                               .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Label                  | TC  |S|     TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

          Figure 7: Incoming Label Stack Sub-TLV

8.1.2. Incoming Interface Index Sub-TLV

   The Incoming Interface Index object is a Sub-TLV that MAY be included
   in a Detailed Interface and Label Stack TLV. The Incoming Interface
   Index Sub-TLV describes the index assigned by this LSR to the
   interface which received the MPLS echo request message.

   Incoming Interface Index Sub-TLV Type is 2. Length is 8, and the
   Value field has 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Interface Index Flags       |          Must Be Zero         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Interface Index                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

        Figure 8: Incoming Interface Index Sub-TLV

  Interface Index Flags

     Interface Index Flags field is a bit vector with following format.

```
  0                   1
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     Must Be Zero (Reserved)   |M|
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

     One flag is defined: M. The remaining flags MUST be set to zero
     when sent and ignored on receipt.

```
   Flag   Name and Meaning
   ----   ----------------

     M   LAG Member Link Indicator

         When this flag is set, the interface index described in
         this sub-TLV is a member of a LAG.
```

Interface Index

   Index assigned by the LSR to this interface.

9. Security Considerations

   This document extends LSP Traceroute mechanism to discover and
   exercise L2 ECMP paths. Additional processing are required for
   initiating LSR and responder LSR, especially to compute and handle
   increasing number of multipath information. Due to additional
   processing, it is critical that proper security measures described in
   [RFC4379] and [RFC6424] are followed.

10. IANA Considerations

10.1. LAG Interface Info TLV

   The IANA is requested to assign new value TBD1 for LAG Interface Info
   TLV from the "Multiprotocol Label Switching Architecture (MPLS) Label
   Switched Paths (LSPs) Ping Parameters - TLVs" registry.

```
   Value    Meaning                                    Reference
   -----    -------                                    ---------
   TBD1     LAG Interface Info TLV                     this document
```

10.2. Interface Index Sub-TLV

   The IANA is requested to assign new value TBD2 for Interface Index
   Sub-TLV from the "Multiprotocol Label Switching Architecture (MPLS)
   Label Switched Paths (LSPs) Ping Parameters - TLVs" registry, "Sub-
   TLVs for TLV Types 20" sub-registry.

```
   Value    Meaning                                    Reference
   -----    -------                                    ---------
   TBD2     Interface Index Sub-TLV                    this document
```

10.3. Detailed Interface and Label Stack TLV

The IANA is requested to assign new value TBD3 for Detailed Interface
and Label Stack TLV from the "Multiprotocol Label Switching
Architecture (MPLS) Label Switched Paths (LSPs) Ping Parameters -
TLVs" registry ([IANA-MPLS-LSP-PING]).

```
   Value   Meaning                                    Reference
   -----   -------                                    ---------
   TBD3    Detailed Interface and Label Stack TLV     this document
```

10.4. DS Flags

The IANA is requested to assign a new bit number from the "DS flags"
sub-registry from the "Multi-Protocol Label Switching (MPLS) Label
Switched Paths (LSPs) Ping Parameters - TLVs" registry
([IANA-MPLS-LSP-PING]).

Note: the "DS flags" sub-registry is created by
[I-D.ietf-mpls-lsp-ping-registry].

```
  Bit number Name                                      Reference
  ---------- ----------------------------------------- ---------
       TBD4 G: LAG Description Indicator                this document
```

10.5. New Sub-Registry

10.5.1. Sub-TLVs for TLV Type TBD3

The IANA is requested to make a new "Sub-TLVs for TLV Type TBD3" sub-
registry under "Multiprotocol Label Switching Architecture (MPLS)
Label Switched Paths (LSPs) Ping Parameters - TLVs" registry.

Initial values for this sub-registry, "Sub-TLVs for TLV Types TBD3",
are described below.

```
  Sub-Type   Name                                      Reference
  ---------  ----------------------------------------- ---------
     1       Incoming Label Stack                      this document
     2       Incoming Interface Index                  this document
   4-65535   Unassigned
```

Assignments of Sub-Types are via Standards Action [RFC5226].

11. Acknowledgements

   Authors would like to thank Nagendra Kumar and Sam Aldrin for
   providing useful comments and suggestions.

12. References

12.1. Normative References

   [I-D.ietf-mpls-lsp-ping-registry]
            Decraene, B., Akiya, N., Pignataro, C., Andersson, L., and
            S. Aldrin, "IANA registries for LSP ping Code Points",
            draft-ietf-mpls-lsp-ping-registry-00 (work in progress),
            November 2014.

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

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

   [RFC6424] Bahadur, N., Kompella, K., and G. Swallow, "Mechanism for
            Performing Label Switched Path Ping (LSP Ping) over MPLS
            Tunnels", RFC 6424, November 2011.

12.2. Informative References

   [I-D.ietf-mpls-ipv6-only-gap]
            George, W. and C. Pignataro, "Gap Analysis for Operating
            IPv6-only MPLS Networks", draft-ietf-mpls-ipv6-only-gap-03
            (work in progress), October 2014.

   [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/
            mpls-lsp-ping-parameters.xhtml>.

   [IEEE802.1AX]
            IEEE Std. 802.1AX, "IEEE Standard for Local and
            metropolitan area networks - Link Aggregation", November
            2008.

   [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an
            IANA Considerations Section in RFCs", BCP 26, RFC 5226,
            May 2008.

Appendix A. LAG with L2 Switch Issues

   Several flavors of "LAG with L2 switch" provisioning models are
   described in this section, with MPLS data plane ECMP traversal
   validation issues with each.

A.1. Equal Numbers of LAG Members

   R1 ==== S1 ==== R2

   The issue with this LAG provisioning model is that packets traversing
   a LAG member from R1 to S1 can get load balanced by S1 towards R2.
   Therefore, MPLS echo request messages traversing specific LAG member
   from R1 to S1 can actually reach R2 via any LAG members, and sender
   of MPLS echo request messages have no knowledge of this nor no way to
   control this traversal. In the worst case, MPLS echo request
   messages with specific entropies to exercise every LAG members from
   R1 to S1 can all reach R2 via same LAG member. Thus it is impossible
   for MPLS echo request sender to verify that packets intended to
   traverse specific LAG member from R1 to S1 did actually traverse that
   LAG member, and to deterministically exercise "receive" processing of
   every LAG member on R2.

A.2. Deviating Numbers of LAG Members

        ‾‾‾‾
   R1 ==== S1 ==== R2

   There are deviating number of LAG members on the two sides of the L2
   switch. The issue with this LAG provisioning model is the same as
   previous model, sender of MPLS echo request messages have no
   knowledge of L2 load balance algorithm nor entropy values to control
   the traversal.

A.3. LAG Only on Right

   R1 ---- S1 ==== R2

   The issue with this LAG provisioning model is that there is no way
   for MPLS echo request sender to deterministically exercise both LAG
   members from S1 to R2. And without such, "receive" processing of R2
   on each LAG member cannot be verified.

A.4. LAG Only on Left

   R1 ==== S1 ---- R2

MPLS echo request sender has knowledge of how to traverse both LAG
members from R1 to S1. However, both types of packets will terminate
on the non-LAG interface at R2. It becomes impossible for MPLS echo
request sender to know that MPLS echo request messages intended to
traverse a specific LAG member from R1 to S1 did indeed traverse that
LAG member.

Authors' Addresses

Nobo Akiya
Cisco Systems

Email: nobo@cisco.com

George Swallow
Cisco Systems

Email: swallow@cisco.com

Stephane Litkowski
Orange

Email: stephane.litkowski@orange.com

Bruno Decraene
Orange

Email: bruno.decraene@orange.com

John E. Drake
Juniper Networks

Email: jdrake@juniper.net