

Internet Engineering Task Force
Internet-Draft
Updates: 4379,6790 (if approved)
Intended status: Standards Track
Expires: April 24, 2014

N. Akiya
G. Swallow
C. Pignataro
Cisco Systems
October 21, 2013

Label Switched Path (LSP) Ping/Trace over MPLS Network
using Entropy Labels (EL)
draft-akiya-mpls-entropy-lsp-ping-00

Abstract

The Multiprotocol Label Switching (MPLS) Label Switched Path (LSP) Ping and Traceroute are used to exercise specific paths of Equal Cost Multipath (ECMP). This ability has been lost on some scenarios which makes use of [RFC6790]: Entropy Labels (EL).

This document extends the MPLS LSP Ping and Traceroute mechanisms to restore the ability of exercising specific paths of ECMP over LSP which make use of Entropy Label. This document updates [RFC4379] and [RFC6790].

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

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 24, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

1. Introduction	2
2. Overview	3
3. Multipath Type 9	5
4. Initiating LSR Procedures	6
5. Responder LSR Procedures	7
5.1. IP Based Load Balancer & Not Imposing ELI/EL	7
5.2. IP Based Load Balancer & Imposing ELI/EL	8
5.3. Label Based Load Balancer & Not Imposing ELI/EL	8
5.4. Label Based Load Balancer & Imposing ELI/EL	9
5.5. FAT MS-PW Stitching LSR	9
6. Entropy Label FEC	10
7. DS Flags: L and E	10
8. New Multipath Information Type: 10	11
9. Unsupported Cases	13
10. Security Considerations	13
11. IANA Considerations	13
11.1. DS Flags	13
11.2. Multipath Information Types	13
11.3. Entropy Label FEC	13
12. Acknowledgements	14
13. Contributing Authors	14
14. References	14
14.1. Normative References	14
14.2. Informative References	14
Authors' Addresses	14

1. Introduction

Section 3.3.1 of [RFC4379] specifies multipath information encoding which can be used by LSP Ping initiator to trace and validate all ECMP paths between ingress and egress. These encodings are sufficient when all the LSRs along the path(s), between ingress and egress, consider same set of "keys" as input for load balancing algorithm: all IP based or all label based.

With introduction of [RFC6790], it is quite normal to see set of LSRs performing load balancing based on EL/ELI while others still follow the traditional way (IP based). This results in LSP Ping initiator not be able to trace and validate all ECMP paths in following scenarios:

- o One or more transit LSRs along ELI/EL imposed LSP do not perform ECMP load balancing based on EL (hashes based on "keys" including IP destination address). This scenario is not only possible but quite common due transit LSRs not implementing [RFC6790] or transit LSRs implementing [RFC6790] but not implementing suggested transit LSR behavior in Section 4.3 of [RFC6790].
- o Two or more LSPs stitched together with at least one LSP being ELI/EL imposing LSP. Such scenarios are described in [I-D.ravisingh-mpls-el-for-seamless-mpls].

These scenarios will be quite common because every deployment of [RFC6790] will invariably end up with nodes that support ELI/EL and nodes that do not. There will typically be areas that support ELI/EL and areas that do not.

As pointed out in [RFC6790] the procedures of [RFC4379] 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 FAT Pseudowires [RFC6391]. Herein is defined a new FEC Stack sub-TLV for the Entropy Label. Section 3 of this document updates the procedures for multipath information type {9}.

2. Overview

[RFC4379] describes LSP traceroute as an operation performed through initiating LSR sending LSP Ping packet (LSP echo request) with incrementing TTL, starting with TTL of one. Initiating LSR discovers and exercises ECMP by obtaining multipath information from each transit LSR and using specific destination IP address or specific entropy label.

LSP Ping initiating LSR sends LSP echo request with multipath information. This multipath information is described in DSMAP/DDMAP

TLV of echo request, and can contain set of IP addresses or set of labels today. Multipath information types {2, 4, 8} carry set of IP addresses and multipath information type {9} carries set of labels. Responder LSR (receiver of LSP echo request) is to determine subset of initiator specified multipath information which load balances to each downstream (outgoing interface). Responder LSR sends LSP echo reply with resulting multipath information per downstream (outgoing interface) back to the initiating LSR. Initiating LSR is then able to use specific IP destination address or specific label to exercise specific ECMP path on the responder LSR.

Current behavior is problematic in following scenarios:

- o Initiating LSR sends IP multipath information, but responder LSR load balances on labels.
- o Initiating LSR sends label multipath information, but responder LSR load balances on IP addresses.
- o Initiating LSR sends any of existing multipath information to ELI/EL imposing LSR, but initiating LSR can only continue to discover and exercise specific path of ECMP if ELI/EL imposing LSR responds with both IP addresses and associated EL corresponding to each IP address. This is because:
 - * ELI/EL imposing LSR that is a stitching point will load balance based on IP address.
 - * Downstream LSR(s) of ELI/EL imposing LSR may load balance based on ELs.
- o Initiating LSR sends any of existing multipath information to ELI/EL imposing LSR, but initiating LSR can only continue to discover and exercise specific path of ECMP if ELI/EL imposing LSR responds with both labels and associated EL corresponding to label. This is because:
 - * ELI/EL imposing LSR that is a stitching point will load balance based on EL from previous LSP and imposes new EL.
 - * Downstream LSR(s) of ELI/EL imposing LSR may load balance based on new ELs.

The above scenarios point to how the existing multipath information is insufficient when LSP traceroute is operated on an LSP with Entropy Labels described by [RFC6790]. Therefore, this document defines a multipath information type to be used in the DSMAP/DDMAP of LSP echo request/reply packets in Section 8.

In addition, responder LSR can reply with empty multipath information if no IP address set or label set from received multipath information matched load balancing to a downstream. Empty return is also possible if initiating LSR sends multipath information of one type, IP address or label, but responder LSR load balances on the other type. To disambiguate between the two results, this document introduces new flags in the DSMAP/DDMAP TLV to allow responder LSR to describe the load balance technique being used.

It is required that all LSRs along the LSP understand new flags as well as new multipath information type. It is also required that initiating LSR can select both IP destination address and label to use on transmitting LSP echo request packets. Two additional DS Flags are defined for the DSMAP and DDMAP TLVs in Section 7.

3. Multipath Type 9

[RFC4379] defined multipath type {9} for tracing of LSPs where label based load-balancing is used. However, as pointed out in [RFC6790], the procedures for using this type are incomplete. First, the specific location of the label was not defined. What was assumed, but not spelled out, was that the presence of multipath type {9} meant the responder should act as if the payload of the received packet were non-IP and that the bottom-of-stack label should be replaced by the values indicated by multipath type {9} to determine their respective out-going interfaces.

Further, with the introduction of [RFC6790], entropy labels may now appear anywhere in a label stack.

This section defines to which labels multipath type {9} can apply. Additionally it defines procedures for tracing pseudowires and flow-aware pseudowires. These procedures pertain to the use of multipath information type {9} as well as type {10}.

Section 6 defines a new FEC-Stack sub-TLV to indicate an entropy label. Multipath type {9} applies exclusively to this sub-TLV. Any LSP Ping message containing a DD-MAP or DS-MAP with multipath type {9} MUST include an EL_FEC at the bottom of the FEC-Stack.

When an MPLS echo request message is received containing a FEC-Stack with an EL-FEC at the bottom of the FEC stack and is not preceded by an entropy label, the responder must behave (for load balancing purposes) as if the first word of the message were a Pseudowire Control Word.

In order to trace a non-FAT pseudowire, instead of including the appropriate PW-FEC in the FEC-Stack, an EL-FEC is included. 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 transport LSP (e.g. LDP and/or RSVP). 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.

In order to trace a FAT pseudowire, the initiator includes an EL-FEC at the bottom of the FEC-Stack and pushes the appropriate PW-FEC onto the FEC-Stack.

4. Initiating LSR Procedures

In order to facilitate the flow of the following text we speak in terms of a boolean called EL_LSP maintained by the initiating LSR. This value controls the multipath information type to be used in transmitted echo request packets. When the initiating LSR is transmitting an echo request packet with DSMAP/DDMAP with a non-zero multipath information type, then EL_LSP boolean MUST be consulted to determine the multipath information type to use.

In addition to procedures described in [RFC4379] as updated by Section 3 and [RFC6424], initiating LSR MUST operate with following procedures.

- o When initiating LSR is IP based load balancer (not imposing ELI/EL), initialize EL_LSP=False.
- o When initiating LSR imposes ELI/EL, initialize EL_LSP=True.
- o When initiating LSR is transmitting non-zero multipath information type:
 - If (EL_LSP) initiating LSR MUST use multipath information type {10}.
 - Else initiating LSR MUST use multipath information type {2, 4, 8, 9}.
- o When initiating LSR is transmitting multipath information type {10}, both "IP Multipath Information" and "Label Multipath Information" MUST be included, and "IP Associated Label Multipath Information" MUST be omitted (NULL).
- o When initiating LSR receives echo reply with {L=0, E=1} in DS flags with valid contents, set EL_LSP=True.

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

- 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}.

5. Responder LSR Procedures

Common Procedures: Responder LSR receiving LSP echo request packet with multipath information type {10} MUST validate following contents. Any deviation MUST result in responder LSR to consider the packet as malformed and return code 1 (Malformed echo request received) in LSP echo reply packet.

- o IP multipath information MUST be included.
- o Label multipath information MUST be included.
- o IP associated label multipath information MUST be omitted (NULL).

Following subsections describe expected responder LSR procedures when echo reply is to include DSMAP/DDMAP TLVs, based on local load balance technique being employed. In case responder LSR performs deviating load balance techniques per downstream basis, appropriate procedures matching to each downstream load balance technique MUST be operated.

5.1. IP Based Load Balancer & Not Imposing ELI/EL

- o Responder MUST set {L=0, E=0} in DS flags.
- o If multipath information type {2, 4, 8} is received, responder MUST comply with [RFC4379]/[RFC6424].

- o If multipath information type {9} is received, responder MUST reply with multipath type {0}.
- o If multipath information type {10} is received, responder MUST reply with multipath information type {10}. "Label Multipath Information" and "Associated Label Multipath Information" sections MUST be omitted (NULL). If no matching IP address is found, then "IPMultipathType" field MUST be set to multipath information type {0} and "IP Multipath Information" section MUST also be omitted (NULL). If at least one matching IP address is found, then "IPMultipathType" field MUST be set to appropriate multipath information type {2, 4, 8} and "IP Multipath Information" section MUST be included.

5.2. IP Based Load Balancer & Imposing ELI/EL

- o Responder MUST set {L=0, E=1} in DS flags.
- o If multipath information type {9} is received, responder MUST reply with multipath type {0}.
- o If multipath type {2, 4, 8, 10} is received, responder MUST respond with multipath type {10}. "Label Multipath Information" section MUST be omitted (NULL). IP address set specified in received IP multipath information MUST be used to determine the returning IP/Label pairs. If received multipath information type was {10}, received "Label Multipath Information" sections MUST NOT be used to determine the associated label portion of returning IP/Label pairs. If no matching IP address is found, then "IPMultipathType" field MUST be set to multipath information type {0} and "IP Multipath Information" section MUST be omitted (NULL). In addition, "Assoc Label Multipath Length" MUST be set to 0, and "Associated Label Multipath Information" section MUST also be omitted (NULL). If at least one matching IP address is found, then "IPMultipathType" field MUST be set to appropriate multipath information type {2, 4, 8} and "IP Multipath Information" section MUST be included. In addition, "Associated Label Multipath Information" section MUST be populated with list of labels corresponding to each IP address specified in "IP Multipath Information" section. "Assoc Label Multipath Length" MUST be set to appropriate value.

5.3. Label Based Load Balancer & Not Imposing ELI/EL

- o Responder MUST set {L=1, E=0} in DS flags.
- o If multipath information type {2, 4, 8} is received, responder MUST reply with multipath type {0}.

- o If multipath information type {9} is received, responder MUST comply with [RFC4379] / [RFC6424] as updated by Section 3.
- o If multipath information type {10} is received, responder MUST reply with multipath information type {10}. "IP Multipath Information" and "Associated Label Multipath Information" sections MUST be omitted (NULL). If no matching label is found, then "LbMultipathType" field MUST be set to multipath information type {0} and "Label Multipath Information" section MUST also be omitted (NULL). If at least one matching label is found, then "LbMultipathType" field MUST be set to appropriate multipath information type {9} and "Label Multipath Information" section MUST be included.

5.4. Label Based Load Balancer & Imposing ELI/EL

- o Responder MUST set {L=1, E=1} in DS flags.
- o If multipath information type {2, 4, 8} is received, responder MUST reply with multipath type {0}.
- o If multipath type {9, 10} is received, responder MUST respond with multipath type {10}. "IP Multipath Information" section MUST be omitted (NULL). Label set specified in received label multipath information MUST be used to determine the returning Label/Label pairs. If received multipath information type was {10}, received "Label Multipath Information" sections MUST NOT be used to determine the associated label portion of returning Label/Label pairs. If no matching label is found, then "LbMultipathType" field MUST be set to multipath information type {0} and "Label Multipath Information" section MUST be omitted (NULL). In addition, "Assoc Label Multipath Length" MUST be set to 0, and "Associated Label Multipath Information" section MUST also be omitted (NULL). If at least one matching label is found, then "LbMultipathType" field MUST be set to appropriate multipath information type {9} and "Label Multipath Information" section MUST be included. In addition, "Associated Label Multipath Information" section MUST be populated with list of labels corresponding to each label specified in "Label Multipath Information" section. "Assoc Label Multipath Length" MUST be set to appropriate value.

5.5. FAT MS-PW Stitching LSR

MS-PW stitching LSR that xconnects flow-aware pseudowires behaves in one of two ways:

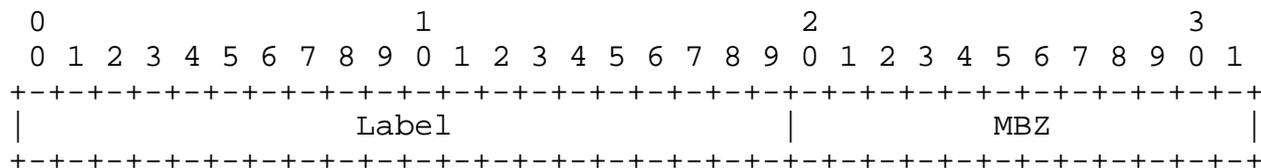
- o Load balances on previous flow label, and carries over same flow label. For this case, stitching LSR is to behave as procedures described in Section 5.3.
- o Load balances on previous flow label, and replaces flow label with newly computed. For this case, stitching LSR is to behave as procedures described in Section 5.4.

6. Entropy Label FEC

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

Entropy Label (EL) is a special purpose label with label value being discretionary (i.e. label value may not be 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 EL, particularly in the scenario where label stack does not carry ELI (ex: FAT-PW [RFC6391]). Therefore, this document defines a EL FEC to allow 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.

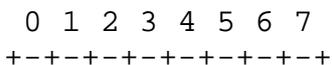


Label is the actual label value inserted in the label stack; the MBZ fields MUST be zero when sent and ignored on receipt.

7. DS Flags: L and E

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

DS Flags



```
|  MBZ  |L|E|I|N|
+--+--+--+--+--+--+--+
```

Flag Name and Meaning

- L Label based load balance indicator
This flag MUST be set to zero in the echo request. LSR which performs load balancing on a label MUST set this flag in the echo reply. LSR which performs load balancing on IP MUST NOT set this flag in the echo reply.
- E ELI/EL imposer indicator
This flag MUST be set to zero in the echo request. LSR which imposes ELI/EL MUST set this flag in the echo reply. LSR which does not impose ELI/EL MUST NOT set this flag in the echo reply.

Two flags result in four load balancing techniques which echo reply generating LSR can indicate:

- o {L=0, E=0} LSR load balances based on IP and does not impose ELI/EL.
- o {L=0, E=1} LSR load balances based on IP and imposes ELI/EL.
- o {L=1, E=0} LSR load balances based on label and does not impose ELI/EL.
- o {L=1, E=1} LSR load balances based on label and imposes ELI/EL.

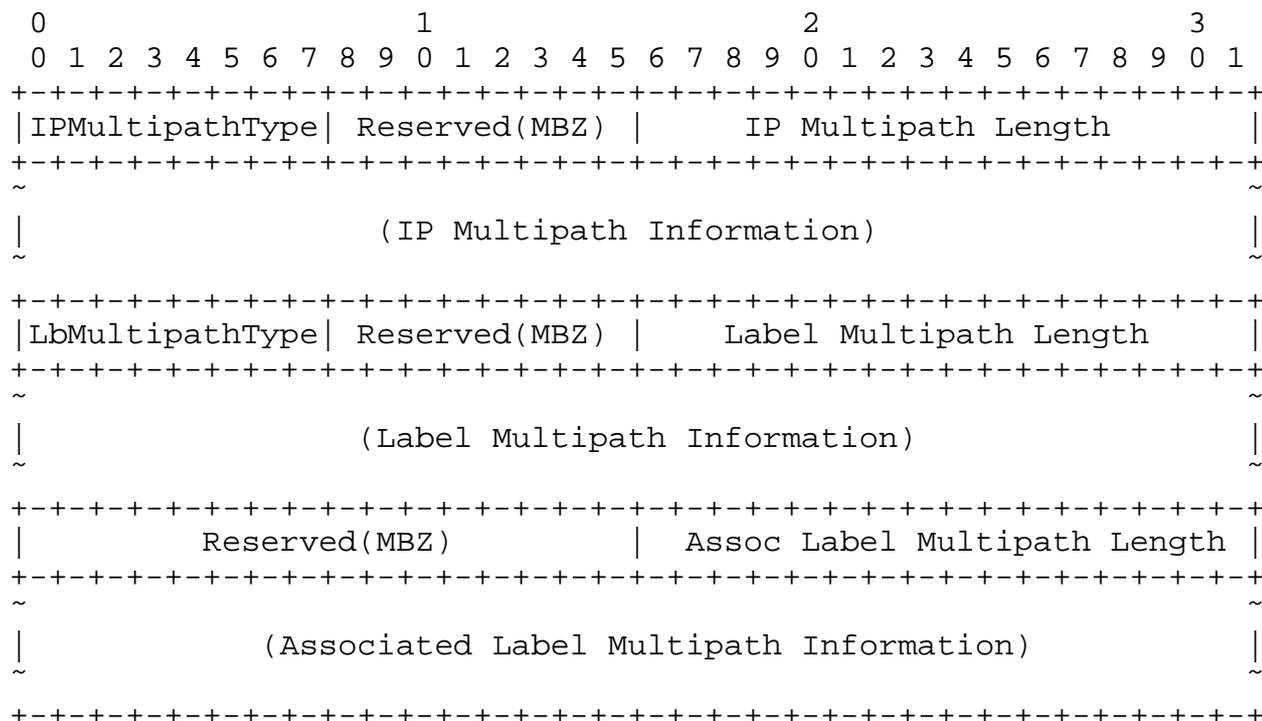
8. New Multipath Information Type: 10

One new multipath information type is added to be used in DSMAP/DDMAP TLVs. New multipath type has value of 10.

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

Multipath type 10 is comprised of three sections. One section to describe IP address set. One section to describe label set. One section to describe another label set which associates to either IP address set or label set specified in the other section.

Multipath information type 10 has following format:



- o IP Multipath Information

This section reuses IP multipath information from [RFC4379]. Specifically, values {0, 2, 4, 8} can be used.

- o Label Multipath Information

This section reuses label multipath information from [RFC4379]. Specifically, values {0, 9} can be used.

- o Associated Label Multipath Information

"Assoc Label Multipath Length" is a 16 bit field of multipath information which indicates 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 (NULL) in an MPLS Echo Request message. A midpoint which imposes ELI/EL labels SHOULD include "Assoc 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 specific IP address OR label described in the "IP Multipath Information" section or "Label Multipath Information" section. For example, if 3 IP addresses are specified in the "IP Multipath Information" section, then there MUST be 3 labels described in this section. First label maps to the lowest IP address specified, second label maps to the second lowest IP address specified and third label maps to the third lowest IP address specified.

9. Unsupported Cases

There are couple of scenarios where LSP path tracing mechanics are not supported in this draft revision.

- o When one or more LSP transit node(s) performs label based load balancing on a label that is not bottom-of-stack label when Entropy Label Indicator is not included.
- o When one or more LSP transit node(s) performs label based load balancing on a label other than Entropy Label when Entropy Label Indicator and Entropy Label pair is included.

10. Security Considerations

Beyond those specified in [RFC4379], [RFC6424] and [RFC6790], there are no further security measured required.

11. IANA Considerations

11.1. DS Flags

DS flags ... not maintained by IANA. Should it be?

11.2. Multipath Information Types

Multipath information types ... not maintained by IANA. Should it be?

11.3. Entropy Label FEC

IANA is requested to assign a new sub-TLV from the "Sub-TLVs for TLV Types 1 and 16" section from "TLVs" sub-registry within the "Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters" registry.

Following value appears to be next available sub-TLV value. Requesting IANA to allow specified value as early allocation.

Value	Meaning	Reference
-----	-----	-----
26	Entropy Label FEC	this document

12. Acknowledgements

TBD

13. Contributing Authors

Nagendra Kumar
Cisco Systems
Email: naikumar@cisco.com

14. References

14.1. Normative References

- [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.
- [RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, November 2012.

14.2. Informative References

- [I-D.ravisingh-mpls-el-for-seamless-mpls] Singh, R., Shen, Y., and J. Drake, "Entropy label for seamless MPLS", draft-ravisingh-mpls-el-for-seamless-mpls-00 (work in progress), February 2013.
- [RFC6391] Bryant, S., Filsfils, C., Drafz, U., Kompella, V., Regan, J., and S. Amante, "Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network", RFC 6391, November 2011.
- [RFC6424] Bahadur, N., Kompella, K., and G. Swallow, "Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels", RFC 6424, November 2011.

Authors' Addresses

Nobo Akiya
Cisco Systems

Email: nobo@cisco.com

George Swallow
Cisco Systems

Email: swallow@cisco.com

Carlos Pignataro
Cisco Systems

Email: cpignata@cisco.com