Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 13, 2015

IJ. Wijnands, Ed.
K. Raza
Cisco Systems, Inc.
E. Rosen
A. Atlas
Juniper Networks, Inc.
J. Tantsura
Ericsson
Q. Zhao
Huawei Technology

February 9, 2015

mLDP Node Protection
draft-ietf-mpls-mldp-node-protection-05

Abstract

This document describes procedures to support node protection for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths (MP LSPs) that has been built by "Multipoint Label Distribution Protocol"(mLDP). In order to protect a node N, the Point of Local Repair (PLR) LSR of N must learn the Merge Point (MPT) LSR(s) of node N such that traffic can be redirected to them in case node N fails. Redirecting the traffic around the failed node N depends on existing P2P LSPs. The pre-established LSPs originate from the PLR LSR and terminate on the MPT LSRs while bypassing LSR N.

Table of Contents

1. Introduction

This document describes procedures to support node protection for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths (MP LSPs) that has been built by "Multipoint Label Distribution Protocol"(mLDP). In order to protect a node N, the Point of Local Repair (PLR) LSR of N must learn the Merge Point (MPT) LSR(s) of node N such that traffic can be redirected to them in case node N fails. Redirecting the traffic around the failed node N depends on existing P2P LSPs. The pre-established LSPs originate from the PLR LSR and terminate on the MPT LSRs while bypassing LSR N. The procedures to setup these P2P LSPs are outside the scope of this document, but one can imagine using RSVP-TE or LDP LFA based techniques to accomplish this.

The solution described in this document notifies the PLR(s) of the MPT LST(s) via signalling using a Targetted LDP (tLDP) session [RFC5036]. By having a tLDP session with the PLR, most of the (m)LDP features currently defined should just work, like Make-Before-Break (MBB), Graceful Restart (GR), Typed Wildcard FEC support, etc. All this is achieved at the expense of having additional tLDP sessions between each MPT and PLR LSR.

1.1. Conventions used in this document

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].

The terms "node" is used to refer to an LSR and used interchangeably. The terms "PLR" and "MPT" are used as shorthand to refer to "PLR LSR" and "MPT LSR" respectively.

1.2. Terminology

mLDP: Multipoint extensions to LDP.

PLR: Point of Local Repair (the LSR that redirects the traffic to one or more Merge Point LSRs).

MPT: Merge Point (the LSR that merges the backup LSP with primary LSP. Note, there can be multiple MPT LSRs for a single MP-LSP node protection).

tLDP: Targeted LDP.

MP LSP: Multi-Point LSP (either a P2MP or MP2MP LSP).

2. PLR Determination

In order for a MPT to establish a tLDP session with a PLR, it first has to learn the PLR for a particular MP LSP. It is the responsibility of the protected node N to advertise the address of the PLR to the MPT. The PLR address for a MP LSP on node N is the address of the upstream LDP peer, but only when node N is NOT the root node of the MP2MP LSP. If the upstream LDP peer is unable to function as PLR, the procedures in this document do not apply and are out of the scope. If node N is the root node, the procedures are slightly different as described in Section 2.2. The procedures that follow assume that all the participating nodes (N, PLRs, MPTs) are enabled (e.g. by a user configuration) to support and implement the PLR determination feature.

2.1. Transit node procedure

Below we are describing the procedures when the protected node is a transit node along the path to the root.

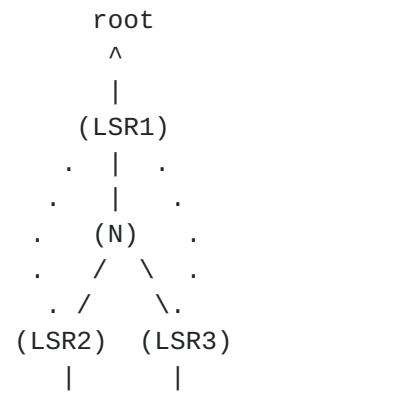


Figure 1.

N: The node being protected,
...: Backup LSPs from LSR1 to the LSR2 and LSR3.

Node N uses the root address of the MP LSP to determine the upstream LSR for a given MP LSP following the procedures as documented in [RFC6388] section 2.4.1.1. The upstream LSR in figure 1 is LSR1 because it is the first hop along the shortest path to reach the root address. After determining the upstream LSR, node N (which is feature enabled), MUST advertise the address of LSR1 as the PLR address to the downstream members of the MP LSP (i.e. LSR2 and LSR3) if the given downstream member has announced support for node

protection (see Section 5) for Capability negotiation). For the format and encoding of PLR address information, see Section 2.3.

2.2. MP2MP root node procedure

In this section we are describing the procedures for when the protected node is the root of a MP2MP LSP. Consider figure 2 below;

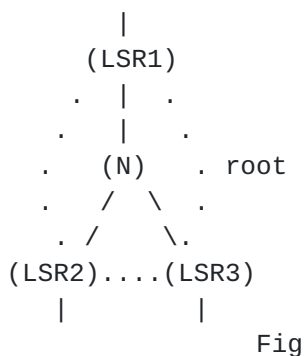


Figure 2.

N: The MP2MP root node being protected.
...: Backup LSPs between LSR1, LSR2 and LSR3.

Assume that LSR1, LSR2 and LSR3 are all members of a MP2MP LSP for which N is the root node. Since N is the root of the MP2MP LSP, there is no upstream LSR and no 'single' PLR LSR for protecting node N. In order to protect node N, all the directly connected members of the MP2MP must participate in protecting node N by acting both as PLR and MPT LSR. An LSR will act as MPT for traffic coming from the other LSR(s) and it will act as PLR for traffic it is sending to the other LSR(s). Since node N knows the members of the MP2MP LSP, it will advertise the member list to its directly connected members, excluding the member it is sending to. For example, node N will advertise {LSR3,LSR1} list to LSR2 excluding LSR2 from it. Instead of advertising a single PLR when node N is not the root, a list of PLRs is advertised using the procedures documented in Section 2.3.

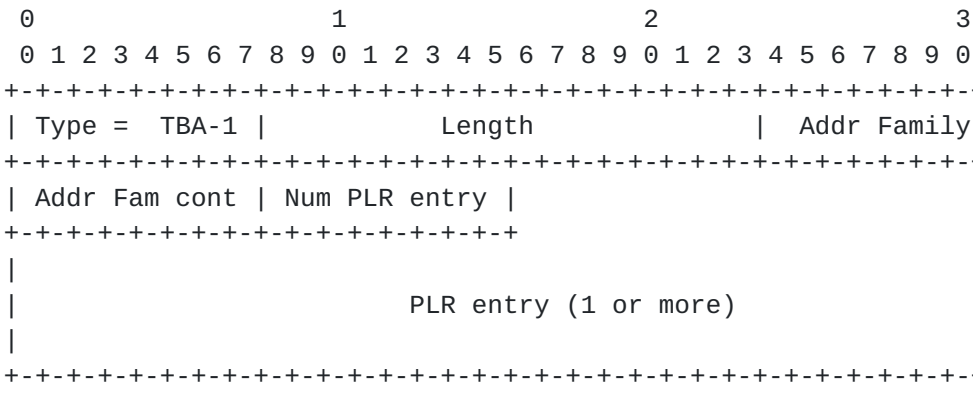
It should be noted that the MP2MP root node protection mechanism don't replace the Root Node Redundancy (RNR) procedures as described in [RFC6388] section 7. The node protection procedures in this draft will help restoring traffic for the existing MP2MP LSPs after node failure, but a new root node has to be elected eventually in order to allow new MP2MP LSPs to be created.

2.3. PLR information encoding

The upstream LSR address is conveyed via an LDP Notification message with MP Status TLV, where the MP status TLV contains a new "PLR Status Value Element" that specifies the address of the PLR.

The new "PLR Status Value Element" is encoded as follows;

PLR Status Element:



Where

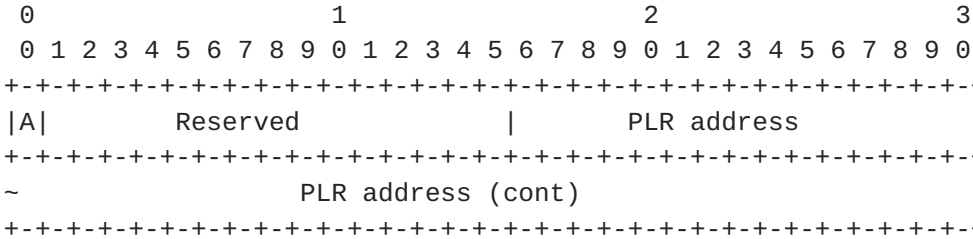
Type: PLR Status Value Element (Type TBA-1 to be assigned by IANA)

Length: The Length field encodes the length of the Status Value following the Length field. The encoded Length varies based on the Address Family and the number of PLR entries.

Address Family: Two octet quantity containing a value from IANA's "Address Family Numbers" registry that encodes the address family for the PLR Address encoded in the PLR entry.

Num PLR entry: Number of "PLR entries" encoded in the Status Value Element, followed by "Num PLR entry" field (please see format of a PLR entry below).

The format of a "PLR Entry" is as follows:



Where

A bit: 0 = Withdraw, 1 = Add.

Reserved: 15 bits, must be zero on transmit and ignored on receipt

PLR address: PLR Address encoded according to Address Family field encoded in the PLR Status Value Element. Note, the length of the PLR address field is specific to the Address Family that is encoded.

The size of a "PLR Entry" is the 2 octets ("A bit + Reserved") + PLR address length. The length of the PLR address is depending on the Address Family as encoded in the PLR Status Value Element. The size of a "PLR entry" is 6 octets and 18 octets respectively for an IPv4 PLR address and an IPv6 PLR address.

If the PLR address on N changes for a given MP LSP, N needs to trigger a new PLR Status to update the MPT(s). A node N can advertise or withdraw a given PLR from its PLR set by setting "A bit" to 1 or 0 respectively in corresponding PLR entry. Removing a PLR address is likely due to a link failure, see the procedures as documented in Section 4.1. To remove all PLR addresses belonging to the encoded Address Family, an LSR N MUST encode PLR Status Value Element with no PLR entry and "Num PLR entry" field MUST be set to zero.

Along with the PLR MP Status a MP FEC TLV MUST be included in the LDP Notification message so that a receiver is able to associate the PLR Status with the MP LSP.

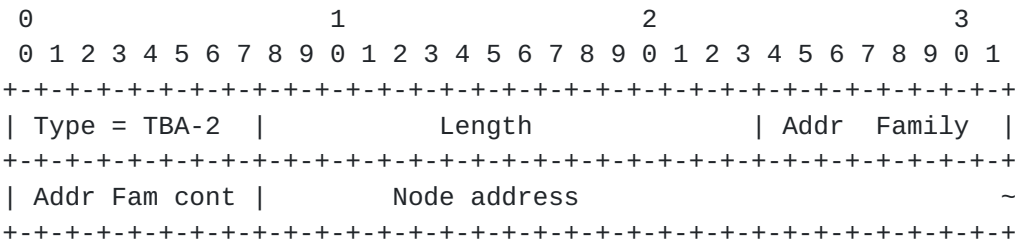
3. Using the tLDP session

The receipt of a PLR MP Status (with PLR addresses) for a MP LSP on a receiving LSR makes it an MPT for node protection. If not already established, the MPT LSR MUST establish a tLDP session with all of the learned PLR addresses using the procedures as documented in [RFC7060].

Using Figure 1 as the reference topology, let us assume that both LSR2 and LSR3 are MPTs and have established a tLDP session with the PLR being LSR1. Assume that both LSR2 and LSR3 have a FEC <R,X> with a upstream LSR N and label Ln assigned to FEC towards N. The MPTs will create a secondary upstream LSR (using the received PLR address) and assigned a Label Lpx to FEC <R,X> towards PLR for it. The MPTs will do that for each PLR address that was learned for the MP LSP. In this example, the MPTs will have a FEC <R,X> with two local labels associated with it. Ln that was assigned to N via the normal mLDP procedures, and Label Lpx that was assigned for PLR (LSR1) for the purpose of node protecting MP LSP via node N. Note, when the protected node is a MP2MP root node, there will be an upstream LSR

for each PLR address that was advertised along with a unique Label Lpx.

The receipt of a FEC Label Mapping alone over the tLDP session from MPT on a PLR conveys the label information but does not convey the node being protected. The information about a protected node is known to the MPT LSR and needs to be communicated to the PLR as well. For this reason, the FEC Label Mapping (FEC <R,X> : Lpx) sent by the MPT over the tLDP session to the PLR MUST include a Status TLV with MP Status including a new LDP MP status Value Element called the "Protected Node Status Value Element". This new value element is used to specify the address of the node being protected. The "Protected Node Status Value Element" has the following format;



Type : Protected Node Status Value Element (Type TBA-2 to be assigned by IANA)

Length: The Length field encodes the length of the Status Value following the Length field. The encoded Length varies based on the Address Family and is 6 octets (for Address Family + IPv4 address and 18 octets for Address Family + IPv6 address.

Address Family: Two octet quantity containing a value from IANA's "Address Family Numbers" registry that encodes the address family for the Node Address.

Node address: Protected node address encoded according to Address Family field.

When a PLR receives a Label Mapping for FEC <R,X> that includes a Protected Node Status, it will only use that label binding once the Node advertised in the Status value becomes unreachable. If the LSP is a MP2MP LSP, the PLR would have assigned a Label Mapping for the upstream MP2MP FEC Element to the MPT ([RFC6388] section 3) for FEC <R,X>. This label binding on the MPT MUST only be used once node N becomes unreachable.

The procedures to determine if a node is unreachable is a local

decision and not spelled out in this draft. Typical link failure or Bidirectional Forwarding Detection (BFD) can be used to determine and detect node unreachability.

4. Link or node failure

Consider the following topology;

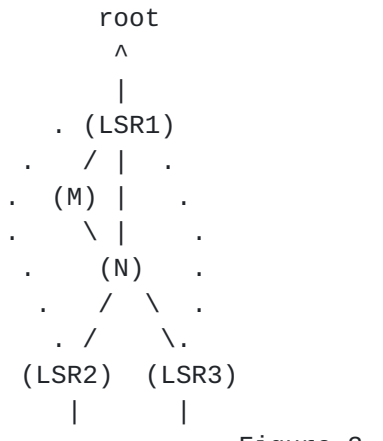


Figure 3.

N: The node being protected
M: The backup node to protect link LSR1 - N
...; Backup LSPs from LSR1 to LSR2 and LSR3.

Assume that LSR1 is the PLR for protected node N, LSR2 and LSR3 are MPTs for node N. When LSR1 discovered that node N is unreachable, it can't determine whether it is the 'LSR1 - N' link or node N that failed. In Figure 3, the link between LSR1 and N is also protected using Fast ReRoute (FRR) [RFC4090] link protection via node M. LSR1 MAY potentially invoke 2 protection mechanisms at the same time, redirection the traffic due to link protection via node M to N, and for node protection directly to LSR1 and LSR2. If only the link failed, LSR2 and LSR3 will receive the packets twice due to the two protection mechanisms. To prevent duplicate packets to be forwarded to the receivers on the tree, LSR2 and LSR3 need to determin which upstream node to accept the packets from. So, either from the primary upstream LSR N or from the secondary upstream LSR1, but never both at the same time. The selection between the primary upstream LSR or (one or more) secondary upstream LSRs (on LSR2 and LSR3) is based on the reachability of the protected node N. As long as N is reachable, N is the primary upstream LSR who is accepting the MPLS packets and forwarding them. Once N becomes unreachable, the secondary upstream LSRs (LSR1 in our example) are activated. Note

that detecting if N is unreachable is a local decision and not spelled out in this draft. Typical link failure or Bidirectional Forwarding Detection (BFD) can be used to determine and detect node unreachability.

4.1. Re-convergence after node/link failure

Consider the following topology;

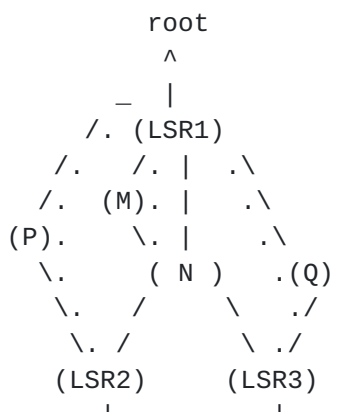


Figure 4.

N: The node being protected.
M: The backup node to protect link 'LSR1 - N'.
P and Q: The nodes on the new primary path after N failure.
...: P2P backup LSPs.

Assume that LSR1 has detected that Node N is unreachable and invoked both the Link Protection and Node Protection procedures as described in this draft. LSR1 is acting as PLR and sending traffic over both the backup P2P LSP to node N (via M) and the P2P LSPs directly to LSR2 and LSR3, acting as MPT LSRs. The sequence of events are depending on whether the link 'LSR1 - N' has failed or node N itself. The node's downstream from the protected node (and participating in node protection) MUST have the capability to determin that the protected node became unreachable. Otherwise the procedures below can not be applied.

4.1.1. Node failure

If node N failed, both LSR2 and LSR3 will have changed the primary upstream LSR to the secondary upstream LSR (LSR1) due to node N being unreachable. With that, the label bindings previously assigned to LSR1 will be activated on the MPTs (LSR2 and LSR3) and the label binding to N will be disabled. Traffic is now switched over the label bindings that where installed for node protection.

4.1.2. Link failure

If the link 'LSR1 - N' has failed, both LSR2 and LSR3 will not change the primary upstream LSR because node N is still reachable. LSR2 and LSR3 will receive traffic over two different bindings, the primary label binding assigned to node N (due to link protection via node M) as well as over the binding assigned to LSR1 for the node protection. Since the secondary upstream LSRs have not been activated, the traffic received due to node protection will be dropped. Node N will re-converge and update LSR2 and LSR3 (Section 2.3) with the information that the PLR address (LSR1) is no longer applicable and must be removed. In response, LSR2 and LSR3 MUST sent a Label Withdraw to LSR1 to withdraw the label binding. This will stop the traffic being forwarded over the backup P2P LSPs for node protection. LSR1 will respond back with a Label Release as soon as the binding has been removed.

4.1.3. Switching to new primary path

The network will eventually re-converge and a new best path to the root will be found by LSR2 and LSR3. LSR2 will find that P is its new primary upstream LSR to reach the Root and LSR3 will find Q. Note that although the current active upstream LSR can either be node N or LSR1 (depending on link or node failure), it does not matter for the following procedures. Both LSR2 and LSR3 SHOULD use the Make-Before-Break (MBB) procedures as described in [RFC6388] section 8 to switch to the new primary upstream node. As soon as the new primary upstream LSRs P and Q are activated, a Label Withdraw message MUST be sent to the old upstream LSR. Note that an upstream LSR switchover from a tLDP neighbor to a directly connected LDP neighbor is no different compared to switching between two directly connected neighbors. After the Label Withdraw message has been received by LSR1 or node N, forwarding will stop and a Label Release will be sent.

When it is determined that after re-convergence there is no more interest in the tLDP session between the MPT and the PLR, the tLDP session MAY be taken down. It is possible that having no more interest in the tLDP session is temporarily due to link flapping. In order to avoid the tLDP session from flapping, it is RECOMMENDED to apply a delay before tearing down the session. Determining the delay is a local implementation matter.

5. mLDP Capabilities for Node Protection

In order to describe the capabilities of the participating LSRs , we are organizing it per role in the network i.e., Point of Local Repair

(PLR), Merge Point (MPT), and Protected Node (as depicted in Fig 1).

5.1. PLR capability

A PLR node should handle the following conditions;

1. Accept an incoming tLDP session from the MPT LSR.

2. Support the receipt of a "Protected Node Status Value Element" status in a MP Status TLV over tLDP session.

3. Upon node failure detection, capable of switching traffic towards one or more MPT(s) over P2P LSP (bypassing N) using the labels previously advertised for MP LSPs over the tLDP session.

An LSR capable of performing these actions will advertise it self as PLR capable in the Node Protection capability (see Section 5.4). This is a unidirectional capability announced from PLR to the protected LSR.

5.2. MPT capability

An MPT node should handle the following conditions;

1. Support the receipt of "PLR Status Value Element" in a MP Status TLV from a protected node N.

2. Support to transmit "Protected Node Status Value Element" in a MP Status TLV to a PLR.

A LSR capable of performing these actions will advertise itself as the MPT capable in the Node Protection capability (see Section 5.4). This is a unidirectional capability from MPT to the protected LSR.

5.3. The Protected LSR

A protected node should handle the following conditions;

1. Determine the PLR and MPT capability for directly connected upstream and downstream LSRs for a given MP FEC.

2. Support transmitting of "PLR Status Value Element" in a MP Status TLV to one or more downstream MPT LSRs.

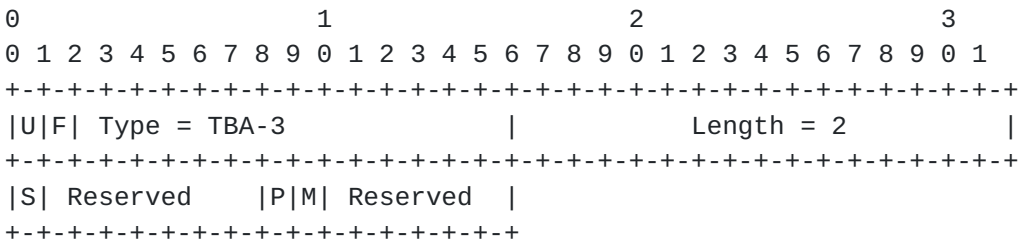
The protected LSR does not advertise any capability for mLDP Node Protection because it does not need to receive any of the defined MP Status values as described above. However, the protected node does play an important role in the signaling and setup of the node

protection. For a given FEC, the protected node can only send PLR information to a downstream LSR if the PLR has signaled the PLR capability and the downstream LSR has signaled the MPT capability. When the downstream LSR (acting as MPT) receives the PLR status, it can implicitly infer that the advertised LSR(s) are PLR capable. The MPT LSR can now proceed with setting up a tLDP session with the PLR(s) and MP LSP node protection signaling.

5.4. The Node Protection Capability

We define a single capability "MP Node Protection Capability" to announce the PLR and MPT capability.

The format of the capability parameter TLV is as follows:



Where

U/F bits: MUST be set to 1 and 0 respectively (as per [RFC5561])

Type: MP Node Protection Capability (Type = TBA-3 to be assigned by IANA)

Length: MUST be set to 2.

S bit: Set to 1 to announce and 0 to withdraw the capability (as per [RFC5561])

P bit: PLR capable for MP LSP node protection

M bit: MPT capable for MP LSP node protection

Reserved: Must be zero on transmit and ignored on receipt

The above capability can be sent in an LDP Initialization message to announce capability at the session establishment time, or it can be sent in LDP Capability message to dynamically update (announce or withdraw) its capability towards its peer using procedures specified in [RFC5561].

An LSR that supports the PLR functionality LSR MAY send this capability to its downstream MP peers with "P" bit set; whereas, an LSR that supports an the MPT functionality MAY send this capability to its upstream peer with "M" bit set. Moreover, an LSR that supports both the PLR and MPT functionality MAY sent this capability to its peers with both "P" and "M" bit set.

6. Security Considerations

The same security considerations apply as those for the base mLDP specification, as described in [RFC6388] and [RFC5920].

7. IANA considerations

IANA is requested to allocate two new code points from the "LDP MP Status Value Element type" registry within the Label Distribution Protocol (LDP) Parameters;

Value	Name	Reference
TBA-1	PLR Status Value Element	this doc
TBA-2	Protected Node Status Value Element	this doc

IANA is requested to assign a new code points for a new Capability Parameter TLV. The code point should be assigned from the IETF Consensus range of the "TLV Type Name Space" registry within the LDP Parameters. The lowest available new code point after 0x0970 should be used.

Value	Description	Reference	Notes/Reg Date
TBA-3	MP Node Protection Capability	This doc	

8. Acknowledgments

The authors like to thank Nagendra Kumar, Duan Hong, Martin Vigoureux, Kenji Fujihira and Loa Andersson for their comments on this draft.

9. References

9.1. Normative References

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

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

[RFC6388] Wijnands, IJ., Minei, I., Kompella, K., and B. Thomas, "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", RFC 6388, November 2011.

[RFC5561] Thomas, B., Raza, K., Aggarwal, S., Aggarwal, R., and JL. Le Roux, "LDP Capabilities", RFC 5561, July 2009.

[RFC7060] Napierala, M., Rosen, E., and IJ. Wijnands, "Using LDP Multipoint Extensions on Targeted LDP Sessions", RFC 7060, November 2013.

9.2. Informative References

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

[RFC5920] Fang, L., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

Authors' Addresses

IJsbrand Wijnands (editor)
Cisco Systems, Inc.
De kleetlaan 6a
Diegem 1831
Belgium

Email: ice@cisco.com

Kamran Raza
Cisco Systems, Inc.
2000 Innovation Drive
Ottawa Ontario K2K-3E8
Canada

Email: skraza@cisco.com

Eric Rosen
Juniper Networks, Inc.
10 Technology Park Drive
Westford MA 01886
USA

Email: erosen@juniper.net

Alia Atlas
Juniper Networks, Inc.
10 Technology Park Drive
Westford MA 01886
USA

Email: akatlas@juniper.net

Jeff Tantsura
Ericsson
300 Holger Way
San Jose CA 95134
USA

Email: jeff.tantsura@ericsson.com

Quintin Zhao
Huawei Technology
125 Nagog Technology Park
Acton MA 01719
USA

Email: quintin.zhao@huawei.com