Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2014

H. Chen, Ed.
Huawei Technologies
R. Torvi, Ed.
Juniper Networks

February 14, 2014

Extensions to RSVP-TE for LSP Ingress Local Protection
draft-chen-mpls-p2mp-ingress-protection-11.txt

Abstract

This document describes extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for locally protecting the ingress node of a Traffic Engineered (TE) Label Switched Path (LSP) in a Multi-Protocol Label Switching (MPLS) and Generalized MPLS (GMPLS) network.

Status of this Memo

This Internet-Draft is submitted to IETF 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 August 18, 2014.

Table of Contents

1. Co-authors

Ning So, Autumn Liu, Alia Atlas, Yimin Shen, Fengman Xu, Mehmet Toy, Lei Liu

2. Introduction

For MPLS LSPs it is important to have a fast-reroute method for protecting its ingress node as well as transit nodes. This is not covered either in the fast-reroute method defined in [RFC4090] or in the P2MP fast-reroute extensions to fast-reroute in [RFC4875].

An alternate approach to local protection (fast-reroute) is to use global protection and set up a second backup LSP (whether P2MP or P2P) from a backup ingress to the egresses. The main disadvantage of this is that the backup LSP may reserve additional network bandwidth.

This specification defines a simple extension to RSVP-TE for local protection of the ingress node of a P2MP or P2P LSP.

2.1. An Example of Ingress Local Protection

Figure 1 shows an example of using a backup P2MP LSP to locally protect the ingress of a primary P2MP LSP, which is from ingress R1 to three egresses: L1, L2 and L3. The backup LSP is from backup ingress Ra to the next hops R2 and R4 of ingress R1.

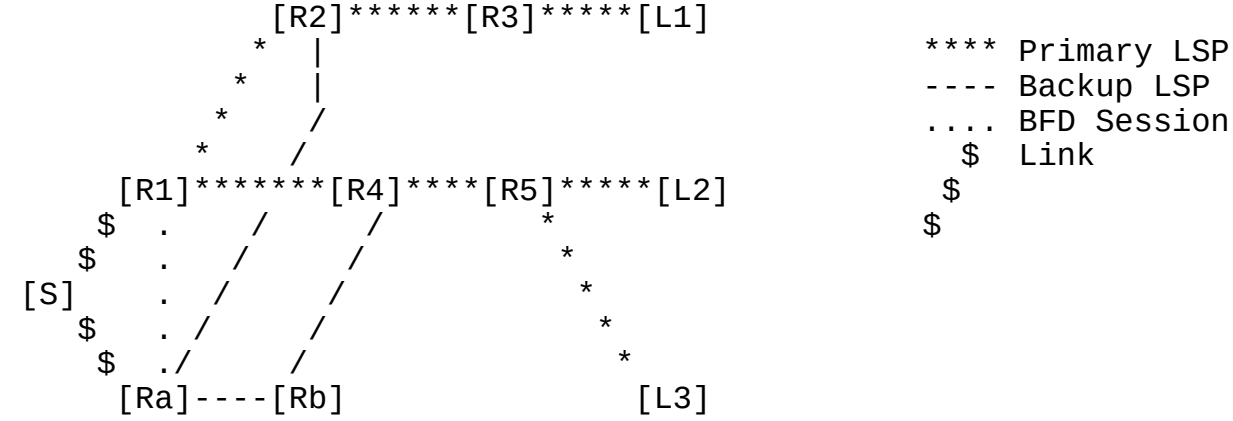


Figure 1: Backup P2MP LSP for Locally Protecting Ingress

Source S may send the traffic simultaneously to both primary ingress R1 and backup ingress Ra. R1 imports the traffic into the primary LSP. Ra normally does not put the traffic into the backup LSP.

Ra should be able to detect the failure of R1 and switch the traffic within 10s of ms. The exact method by which Ra does so is out of scope. Different options are discussed in this draft.

When Ra detects the failure of R1, it imports the traffic from S into the backup LSP to R1's next hops R2 and R4, where the traffic is merged into the primary LSP, and then sent to egresses L1, L2 and L3.

Note that the backup egress must be one logical hop away from the ingress. A logical hop is a direct link or a tunnel such as a GRE tunnel, over which RSVP-TE messages may be exchanged.

2.2. Ingress Local Protection with FRR

Through using the ingress local protection and the FRR, we can locally protect the ingress node, all the links and the intermediate nodes of an LSP. The traffic switchover time is within tens of milliseconds whenever the ingress, any of the links and the intermediate nodes of the LSP fails.

The ingress node of the LSP can be locally protected through using the ingress local protection. All the links and all the intermediate nodes of the LSP can be locally protected through using the FRR.

3. Ingress Failure Detection

Exactly how the failure of the ingress (e.g. R1 in Figure 1) is detected is out of scope for this document. However, it is necessary to discuss different modes for detecting the failure because they determine what must be signaled and what is the required behavior for the traffic source, backup ingress, and merge-points.

3.1. Backup and Source Detect Failure

Backup and Source Detect Failure or Backup-Source-Detect for short means that both the backup ingress and the source are concurrently responsible for detecting the failures of the primary ingress.

In normal operations, the source sends the traffic to the primary ingress. It switches the traffic to the backup ingress when it detects the failure of the primary ingress.

The backup ingress does not import any traffic from the source into the backup LSP in normal operations. When it detects the failure of the primary ingress, it imports the traffic from the source into the backup LSP to the next hops of the primary ingress, where the traffic is merged into the primary LSP.

Note that the source may locally distinguish between the failure of the primary ingress and that of the link between the source and the primary ingress. When the source detects the failure of the link, it may continue to send the traffic to the primary ingress via another link between the source and the primary ingress if there is one.

3.2. Backup Detects Failure

Backup Detects Failure or Backup-Detect means that the backup ingress is responsible for detecting the failure of the primary ingress of an LSP. The source SHOULD send the traffic simultaneously to both the primary ingress and backup ingress.

The backup ingress does not import any traffic from the source into the backup LSP in normal operations. When it detects the failure of the primary ingress, it imports the traffic from the source into the backup LSP to the next hops of the primary ingress, where the traffic is merged into the primary LSP.

Note that the backup ingress may locally distinguish between the failure of the primary ingress and that of the link between the backup ingress and the primary ingress through two BFDs between the backup ingress and the primary ingress. One is through the link, and the other is not. If the first BFD is down and the second is up, the link fails and the primary ingress does not.

3.3. Source Detects Failure

Source Detects Failure or Source-Detect means that the source is responsible for detecting the failure of the primary ingress of an LSP. The backup ingress is ready to import the traffic from the source into the backup LSP after the backup LSP is up.

In normal operations, the source sends the traffic to the primary ingress. When the source detects the failure of the primary ingress, it switches the traffic to the backup ingress, which delivers the traffic to the next hops of the primary ingress through the backup LSP, where the traffic is merged into the primary LSP.

3.4. Next Hops Detect Failure

Next Hops Detect Failure or Next-Hop-Detect means that each of the next hops of the primary ingress of an LSP is responsible for detecting the failure of the primary ingress.

In normal operations, the source sends the traffic to both the primary ingress and the backup ingress. Both ingresses deliver the traffic to the next hops of the primary ingress. Each of the next

hops selects the traffic from the primary ingress and sends the traffic to the destinations of the LSP.

When each of the next hops detects the failure of the primary ingress, it switches to receive the traffic from the backup ingress and then sends the traffic to the destinations.

3.5. Comparing Different Detection Modes

_Behavior / Detection\ Mode	Traffic Always Sent to Backup Ingress	Backup Ingress Activation of Forwarding Entry	Next-Hop Select Stream	Incorrect Failure Detection Cause Traffic Duplication (Ingress does FRR)
Backup-Source-Detect	No	Yes	No	No
Backup-Detect	Yes	Yes	No	Yes
Source-Detect	No	No (Always Active)	No	No
Next-Hop-Detect	Yes	No (Always Active)	Yes	(If Ingress-Next-Hop link fails, stream selection at Next-Next-Hops can mitigate)

A primary goal of failure detection and FRR protection is to avoid traffic duplication, particularly along the P2MP. A reasonable assumption when this ingress protection is in use is that the ingress is also trying to provide link and node protection. When the failure cannot be accurately identified as that of the ingress, this can lead to the ingress sending traffic on bypass to the next-next-hop(s) for node-protection while the backup ingress is sending traffic to its next-hop(s) if Next-Hop-Detect mode is used. RSVP Path messages from the bypass may help to eventually resolve this by removing the forwarding entry for receiving the traffic from the next-hop.

4. Backup Forwarding State

Before the primary ingress fails, the backup ingress is responsible

for creating the necessary backup LSPs to the next hops of the ingress. These LSPs might be multiple bypass P2P LSPs that avoid the ingress. Alternately, the backup ingress could choose to use a single backup P2MP LSP as a bypass or detour to protect the primary ingress of a primary P2MP LSP.

The backup ingress may be off-path or on-path of an LSP. When a backup ingress is not any node of the LSP, we call the backup ingress is off-path. When a backup ingress is a next-hop of the primary ingress of the LSP, we call it is on-path. If the backup ingress is on-path, the primary forwarding state associated with the primary LSP SHOULD be clearly separated from the backup LSP(s) state. Specifically in Backup-Detect mode, the backup ingress will receive traffic from the primary ingress and from the traffic source; only the former should be forwarded until failure is detected even if the backup ingress is the only next-hop.

4.1. Forwarding State for Backup LSP

A forwarding entry for a backup LSP is created on the backup ingress after the LSP is set up. Depending on the failure-detection mode (e.g., source-detect), it may be used to forward received traffic or simply be inactive (e.g., backup-detect) until required. In either case, when the primary ingress fails, this forwarding entry is used to import the traffic into the backup LSP to the next hops of the primary ingress, where the traffic is merged into the primary LSP.

The forwarding entry for a backup LSP is a local implementation issue. In one device, it may have an inactive flag. This inactive forwarding entry is not used to forward any traffic normally. When the primary ingress fails, it is changed to active, and thus the traffic from the source is imported into the backup LSP.

4.2. Forwarding State on Next Hops

When Next-Hop-Detect is used, a forwarding entry for a backup LSP is created on each of the next hops of the primary ingress of the LSP. This forwarding entry does not forward any traffic normally. When the primary ingress fails, it is used to import/select the traffic from the backup LSP into the primary LSP.

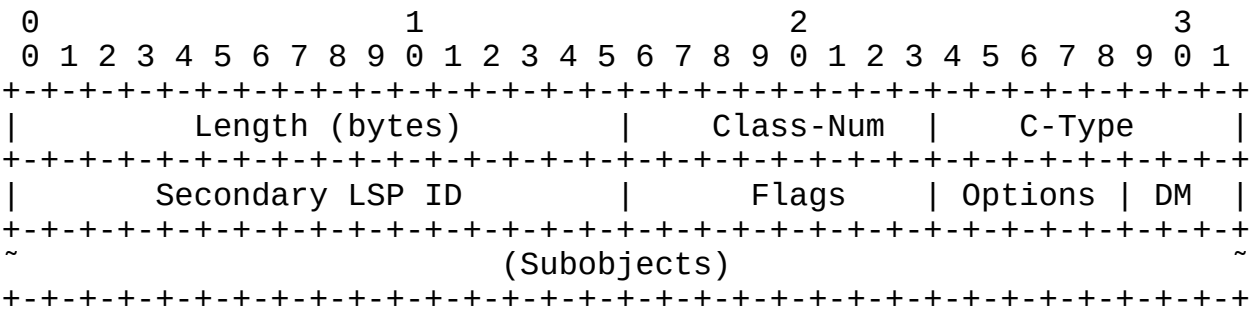
5. Protocol Extensions

A new object INGRESS_PROTECTION is defined for signaling ingress local protection. It is backward compatible.

5.1. INGRESS_PROTECTION Object

The INGRESS_PROTECTION object with the FAST_REROUTE object in a PATH message is used to control the backup for protecting the primary ingress of a primary LSP. The primary ingress MUST insert this object into the PATH message to be sent to the backup ingress for protecting the primary ingress. It has the following format:

Class-Num = TBD C-Type = TBD



Flags
- 0x01 Ingress local protection available
- 0x02 Ingress local protection in use
- 0x04 Bandwidth protection

Options
- 0x01 Revert to Ingress
- 0x02 Ingress-Proxy/Relay-Message
- 0x04 P2MP Backup

DM (Detection Mode)
- 0x00 Backup-Source-Detect
- 0x01 Backup-Detect
- 0x02 Source-Detect
- 0x03 Next-Hop-Detect

For backward compatible, the two high-order bits of the Class-Num in the object are set as follows:

- Class-Num = 0bbbbbbb for the object in a message not on LSP path. The entire message should be rejected and an "Unknown Object Class" error returned.

- Class-Num = 10bbbbbb for the object in a message on LSP path. The node should ignore the object, neither forwarding it nor sending an error message.

The Secondary LSP ID in the object is an LSP ID that the primary ingress has allocated for a protected LSP tunnel. The backup ingress will use this LSP ID to set up a new LSP from the backup ingress to the destinations of the protected LSP tunnel. This allows the new LSP to share resources with the old one.

The flags are used to communicate status information from the backup ingress to the primary ingress.

- Ingress local protection available: The backup ingress sets this flag after backup LSPs are up and ready for locally protecting the primary ingress. The backup ingress sends this to the primary ingress to indicate that the primary ingress is locally protected.

- Ingress local protection in use: The backup ingress sets this flag when it detects a failure in the primary ingress. The backup ingress keeps it and does not send it to the primary ingress since the primary ingress is down.

- Bandwidth protection: The backup ingress sets this flag if the backup LSPs guarantee to provide desired bandwidth for the protected LSP against the primary ingress failure.

The options are used by the primary ingress to specify the desired behavior to the backup ingress and next-hops.

- Revert to Ingress: The primary ingress sets this option indicating that the traffic for the primary LSP successfully re-signaled will be switched back to the primary ingress from the backup ingress when the primary ingress is restored.

- Ingress-Proxy/Relay-Message: This option is set to one indicating that Ingress-Proxy method is used. It is set to zero indicating that Relay-Message method is used.

- P2MP Backup: This option is set to ask for the backup ingress to use P2MP backup LSP to protect the primary ingress. Note that one spare bit of the flags in the FAST-REROUTE object can be used to indicate whether P2MP or P2P backup LSP is desired for protecting an ingress and intermediate node.

The DM (Detection Mode) is used by the primary ingress to specify a desired failure detection mode.

- Backup-Source-Detect (0x00): The backup ingress and the source are concurrently responsible for detecting the failure involving the primary ingress and redirecting the traffic.

- o Backup-Detect (0x01): The backup ingress is responsible for detecting the failure and redirecting the traffic.

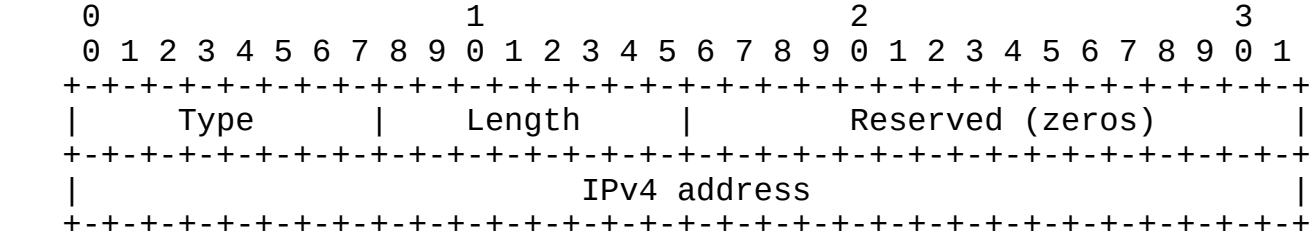
- o Source-Detect (0x02): The source is responsible for detecting the failure and redirecting the traffic.

- o Next-Hop-Detect (0x03): The next hops of the primary ingress are responsible for detecting the failure and selecting the traffic.

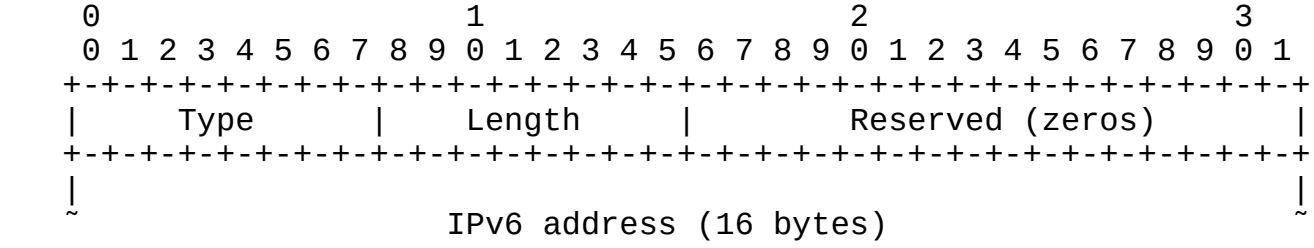
The INGRESS_PROTECTION object may contain some of the sub objects described below.

5.1.1. Subobject: Backup Ingress IPv4/IPv6 Address

When the primary ingress of a protected LSP sends a PATH message with an INGRESS_PROTECTION object to the backup ingress, the object may have a Backup Ingress IPv4/IPv6 Address sub object containing an IPv4/IPv6 address belonging to the backup ingress. The formats of the sub object for Backup Ingress IPv4/IPv6 Address is given below:



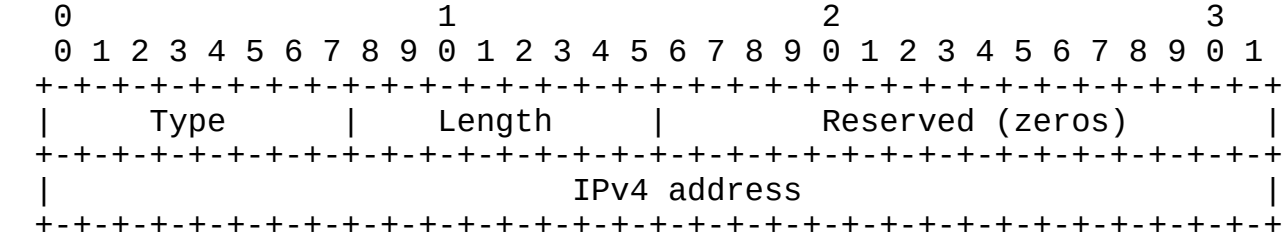
Type: TBD-1 Backup Ingress IPv4 Address
Length: Total length of the subobject in bytes, including the Type and Length fields. The Length is always 8.
Reserved: Reserved two bytes are set to zeros.
IPv4 address: A 32-bit unicast, host address.



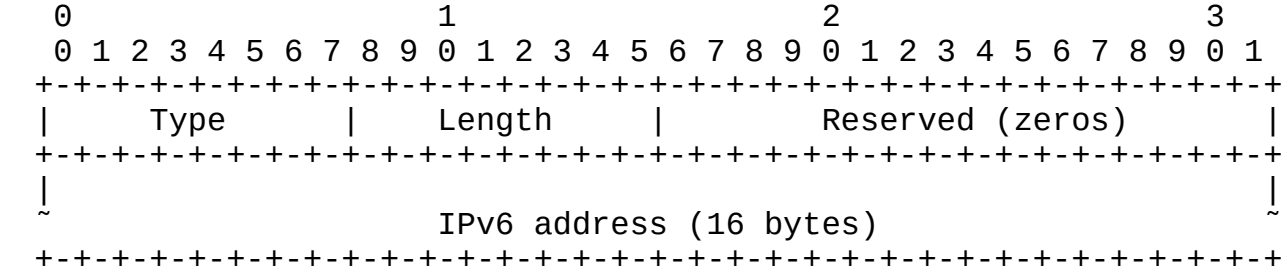
Type: TBD-2 Backup Ingress IPv6 Address
Length: Total length of the subobject in bytes, including the Type and Length fields. The Length is always 20.
Reserved: Reserved two bytes are set to zeros.
IPv6 address: A 128-bit unicast, host address.

5.1.2. Subobject: Ingress IPv4/IPv6 Address

The INGRESS_PROTECTION object in a PATH message from the primary ingress to the backup ingress may have an Ingress IPv4/IPv6 Address sub object containing an IPv4/IPv6 address belonging to the primary ingress. The sub object has the following format:



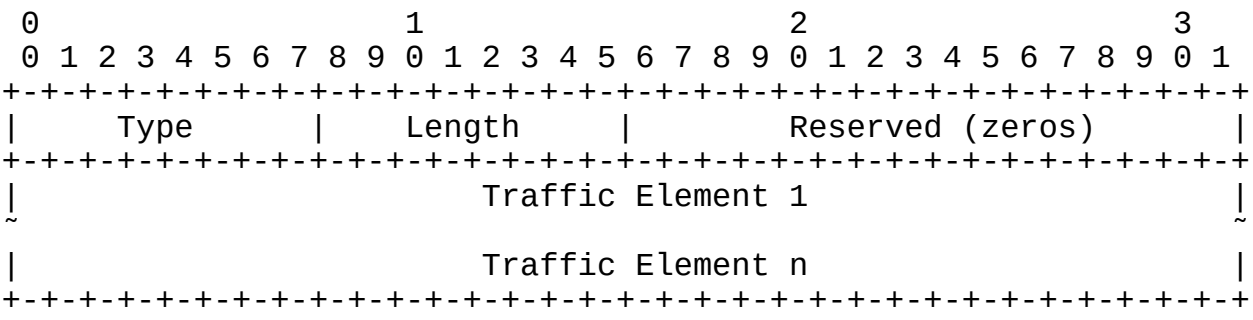
Type: TBD-3 Ingress IPv4 Address
Length: Total length of the subobject in bytes, including the Type and Length fields. The Length is always 8.
Reserved: Reserved two bytes are set to zeros.
IPv4 address: A 32-bit unicast, host address.



Type: TBD-4 Backup Ingress IPv6 Address
Length: Total length of the subobject in bytes, including the Type and Length fields. The Length is always 20.
Reserved: Reserved two bytes are set to zeros.
IPv6 address: A 128-bit unicast, host address.

5.1.3. Subobject: Traffic Descriptor

The INGRESS_PROTECTION object in a PATH message from the primary ingress to the backup ingress may have a Traffic Descriptor sub object describing the traffic to be mapped to the backup LSP on the backup ingress for locally protecting the primary ingress. The sub object has the following format:



Type: TBD-5/TBD-6/TBD-7 Interface/IPv4/6 Prefix
Length: Total length of the subobject in bytes, including the Type and Length fields.
Reserved: Reserved two bytes are set to zeros.

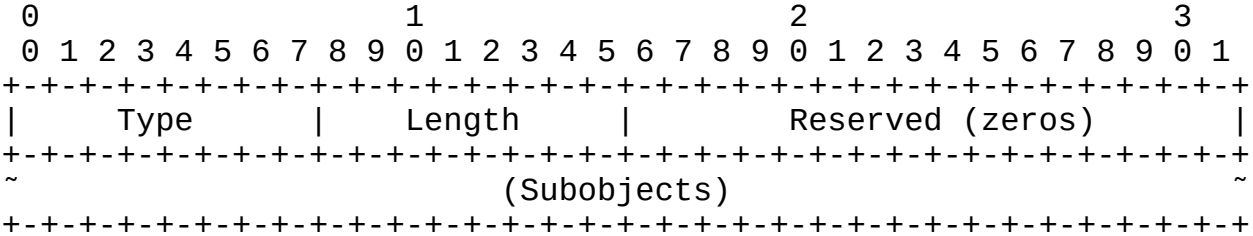
The Traffic Descriptor sub object may contain multiple Traffic Elements of same type as follows.

- Interface Traffic (Type TBD-5): Each of the Traffic Elements is a 32 bit index of an interface, from which the traffic is imported into the backup LSP.

- IPv4/6 Prefix Traffic (Type TBD-6/TBD-7): Each of the Traffic Elements is an IPv4/6 prefix, containing an 8-bit prefix length followed by an IPv4/6 address prefix, whose length, in bits, was specified by the prefix length, padded to a byte boundary.

5.1.4. Subobject: Label-Routes

The INGRESS_PROTECTION object in a PATH message from the primary ingress to the backup ingress will have a Label-Routes sub object containing the labels and routes that the next hops of the ingress use. The sub object has the following format:



Type: TBD-8 Label-Routes
Length: Total length of the subobject in bytes, including the Type and Length fields.
Reserved: Reserved two bytes are set to zeros.

The Subobjects in the Label-Routes are copied from the Subobjects in the RECORD_ROUTE objects contained in the RESV messages that the primary ingress receives from its next hops for the protected LSP. They MUST contain the first hops of the LSP, each of which is paired with its label.

6. Behavior of Ingress Protection

6.1. Overview

There are four parts of ingress protection: 1) setting up the necessary backup LSP forwarding state; 2) identifying the failure and providing the fast repair (as discussed in Sections 2 and 3); 3) maintaining the RSVP-TE control plane state until a global repair can be done; and 4) performing the global repair(see Section 5.5).

There are two different proposed signaling approaches to obtain ingress protection. They both use the same new INGRESS-PROTECTION object. The object is sent in both PATH and RESV messages.

6.1.1. Relay-Message Method

The primary ingress relays the information for ingress protection of an LSP to the backup ingress via PATH messages. Once the LSP is created, the ingress of the LSP sends the backup ingress a PATH message with an INGRESS-PROTECTION object with Label-Routes subobject, which is populated with the next-hops and labels. This provides sufficient information for the backup ingress to create the appropriate forwarding state and backup LSP(s).

The ingress also sends the backup ingress all the other PATH messages for the LSP with an empty INGRESS-PROTECTION object. Thus, the backup ingress has access to all the PATH messages needed for modification to be sent to refresh control-plane state after a failure.

The advantages of this method include: 1) the primary LSP is independent of the backup ingress; 2) simple; 3) less configuration; and 4) less control traffic.

6.1.2. Proxy-Ingress Method

Conceptually, a proxy ingress is created that starts the RSVP signaling. The explicit path of the LSP goes from the proxy ingress to the backup ingress and then to the real ingress. The behavior and signaling for the proxy ingress is done by the real ingress; the use of a proxy ingress address avoids problems with loop detection.

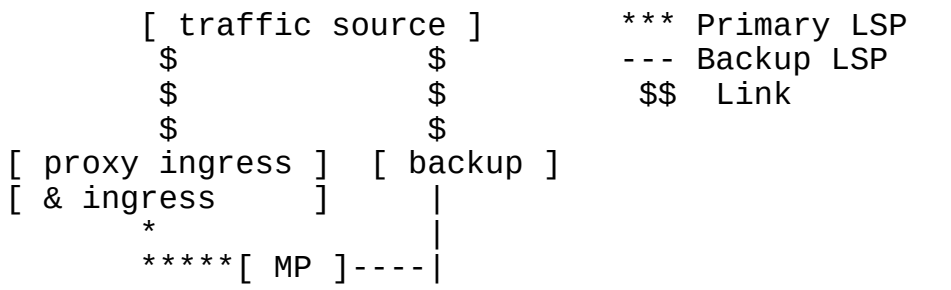


Figure 2: Example Protected LSP with Proxy Ingress Node

The backup ingress must know the merge points or next-hops and their associated labels. This is accomplished by having the RSVP PATH and RESV messages go through the backup ingress, although the forwarding path need not go through the backup ingress. If the backup ingress fails, the ingress simply removes the INGRESS-PROTECTION object and forwards the PATH messages to the LSP's next-hop(s). If the ingress has its LSP configured for ingress protection, then the ingress can add the backup ingress and itself to the ERO and start forwarding the PATH messages to the backup ingress.

Slightly different behavior can apply for the on-path and off-path cases. In the on-path case, the backup ingress is a next hop node after the ingress for the LSP. In the off-path, the backup ingress is not any next-hop node after the ingress for all associated sub-LSPs.

The key advantage of this approach is that it minimizes the special handling code requires. Because the backup ingress is on the signaling path, it can receive various notifications. It easily has access to all the PATH messages needed for modification to be sent to refresh control-plane state after a failure.

6.1.3. Comparing Two Methods

Method	Primary LSP Depends on Backup Ingress	Simple	Config Proxy-Ingress-ID	PATH Msg from Backup to primary RESV Msg from Primary to backup	Reuse Some of Existing Functions
Relay-Message	No	Yes	No	No	Yes-
Proxy-Ingress	Yes	Yes-	Yes	Yes	Yes

6.2. Ingress Behavior

The primary ingress must be configured with four pieces of information for ingress protection.

- Backup Ingress Address: The primary ingress must know an IP address for it to be included in the INGRESS-PROTECTION object.

- Failure Detection Mode: The primary ingress must know what failure detection mode is to be used: Backup-Source-Detect, Backup-Detect, Source-Detect, or Next-Hop-Detect.

- Proxy-Ingress-Id (only needed for Proxy-Ingress Method): The Proxy-Ingress-Id is only used in the Record Route Object for recording the proxy-ingress. If no proxy-ingress-id is specified, then a local interface address that will not otherwise be included in the Record Route Object can be used. A similar technique is used in [RFC4090 Sec 6.1.1].

- Application Traffic Identifier: The primary ingress and backup ingress must both know what application traffic should be directed into the LSP. If a list of prefixes in the Traffic Descriptor sub-object will not suffice, then a commonly understood Application Traffic Identifier can be sent between the primary ingress and backup ingress. The exact meaning of the identifier should be configured similarly at both the primary ingress and backup ingress. The Application Traffic Identifier is understood within the unique context of the primary ingress and backup ingress.

With this additional information, the primary ingress can create and signal the necessary RSVP extensions to support ingress protection.

6.2.1. Relay-Message Method

To protect the ingress of an LSP, the ingress does the following after the LSP is up.

1. Select a PATH message.

2. If the backup ingress is off-path, then send the backup ingress a PATH message with the content from the selected PATH message and an INGRESS-PROTECTION object; else (the backup ingress is a next hop, i.e., on-path case) add an INGRESS-PROTECTION object into the existing PATH message to the backup ingress (i.e., the next hop). The INGRESS-PROTECTION object contains the Traffic-Descriptor sub-object, the Backup Ingress Address sub-object and the Label-Routes sub-object. The DM (Detection Mode) in the

object is set to indicate the failure detection mode desired. The flags is set to indicate whether a Backup P2MP LSP is desired. If not yet allocated, allocate a second LSP-ID to be used in the INGRESS-PROTECTION object. The Label-Routes sub-object contains the next-hops of the ingress and their labels.

3. For each of the other PATH messages, if the node to which the message is sent is not the backup ingress, then send the backup ingress a PATH message with the content copied from the message to the node and an empty INGRESS-PROTECTION object; else send the node the message with an empty INGRESS-PROTECTION object.

6.2.2. Proxy-Ingress Method

The primary ingress is responsible for starting the RSVP signaling for the proxy-ingress node. To do this, the following is done for the RSVP PATH message.

1. Compute the EROs for the LSP as normal for the ingress.

2. If the selected backup ingress node is not the first node on the path (for all sub-LSPs), then insert at the beginning of the ERO first the backup ingress node and then the ingress node.

3. In the PATH RRO, instead of recording the ingress node’s address, replace it with the Proxy-Ingress-Id.

4. Leave the HOP object populated as usual with information for the ingress-node.

5. Add the INGRESS-PROTECTION object to the PATH message. Allocate a second LSP-ID to be used in the INGRESS-PROTECTION object. Include the Backup Ingress Address (IPv4 or IPv6) sub-object and the Traffic-Descriptor sub-object. Set the control-options to indicate the failure detection mode desired. Set or clear the flag indicating that a Backup P2MP LSP is desired.

6. Optionally, add the FAST-REROUTE object [RFC4090] to the Path message. Indicate whether one-to-one backup is desired. Indicate whether facility backup is desired.

7. The RSVP PATH message is sent to the backup node as normal.

If the ingress detects that it can’t communicate with the backup ingress, then the ingress should instead send the PATH message to the next-hop indicated in the ERO computed in step 1. Once the ingress detects that it can communicate with the backup ingress, the ingress SHOULD follow the steps 1-7 to obtain ingress failure protection.

When the ingress node receives an RSVP PATH message with an INGRESS-PROTECTION object and the object specifies that node as the ingress node and the PHOP as the backup ingress node, the ingress node SHOULD check the Failure Scenario specified in the INGRESS-PROTECTION object and, if it is not the Next-Hop-Detect, then the ingress node SHOULD remove the INGRESS-PROTECTION object from the PATH message before sending it out. Additionally, the ingress node must store that it will install ingress forwarding state for the LSP rather than midpoint forwarding.

When an RSVP RESV message is received by the ingress, it uses the NHOP to determine whether the message is received from the backup ingress or from a different node. The stored associated PATH message contains an INGRESS-PROTECTION object that identifies the backup ingress node. If the RESV message is not from the backup node, then ingress forwarding state should be set up, and the INGRESS-PROTECTION object MUST be added to the RESV before it is sent to the NHOP, which should be the backup node. If the RESV message is from the backup node, then the LSP should be considered available for use.

If the backup ingress node is on the forwarding path, then a RESV is received with an INGRESS-PROTECTION object and an NHOP that matches the backup ingress. In this case, the ingress node’s address will not appear after the backup ingress in the RRO. The ingress node should set up ingress forwarding state, just as is done if the LSP weren’t ingress-node protected.

6.3. Backup Ingress Behavior

An LER determines that the ingress local protection is requested for an LSP if the INGRESS_PROTECTION object is included in the PATH message it receives for the LSP. The LER can further determine that it is the backup ingress if one of its addresses is in the Backup Ingress Address sub-object of the INGRESS-PROTECTION object. The LER as the backup ingress will assume full responsibility of the ingress after the primary ingress fails. In addition, the LER determines that it is off-path if it is not a next hop of the primary ingress.

6.3.1. Backup Ingress Behavior in Off-path Case

The backup ingress considers itself as a PLR and the primary ingress as its next hop and provides a local protection for the primary ingress. It behaves very similarly to a PLR providing fast-reroute where the primary ingress is considered as the failure-point to protect. Where not otherwise specified, the behavior given in [RFC4090] for a PLR should apply.

The backup ingress SHOULD follow the control-options specified in the

INGRESS-PROTECTION object and the flags and specifications in the FAST-REROUTE object. This applies to providing a P2MP backup if the "P2MP backup" is set, a one-to-one backup if "one-to-one desired" is set, facility backup if the "facility backup desired" is set, and backup paths that support the desired bandwidth, and administrative-colors that are requested.

If multiple INGRESS-PROTECTION objects have been received via multiple PATH messages for the same LSP, then the most recent one that specified a Traffic-Descriptor sub-object MUST be the one used.

The backup ingress creates the appropriate forwarding state based on failure detection mode specified. For the Source-Detect and Next-Hop-Detect, this means that the backup ingress forwards any received identified traffic into the backup LSP tunnel(s) to the merge point(s). For the Backup-Detect and Backup-Source-Detect, this means that the backup ingress creates state to quickly determine the primary ingress has failed and switch to sending any received identified traffic into the backup LSP tunnel(s) to the merge point(s).

When the backup ingress sends a RESV message to the primary ingress, it should add an INGRESS-PROTECTION object into the message. It SHOULD set or clear the flags in the object to report "Ingress local protection available", "Ingress local protection in use", and "bandwidth protection".

If the backup ingress doesn’t have a backup LSP tunnel to all the merge points, it SHOULD clear "Ingress local protection available". [Editor Note: It is possible to indicate the number or which are unprotected via a sub-object if desired.]

When the primary ingress fails, the backup ingress redirects the traffic from a source into the backup P2P LSPs or the backup P2MP LSP transmitting the traffic to the next hops of the primary ingress, where the traffic is merged into the protected LSP.

In this case, the backup ingress keeps the PATH message with the INGRESS_PROTECTION object received from the primary ingress and the RESV message with the INGRESS_PROTECTION object to be sent to the primary ingress. The backup ingress sets the "local protection in use" flag in the RESV message, indicating that the backup ingress is actively redirecting the traffic into the backup P2P LSPs or the backup P2MP LSP for locally protecting the primary ingress failure.

Note that the RESV message with this piece of information will not be sent to the primary ingress because the primary ingress has failed.

If the backup ingress has not received any PATH message from the primary ingress for an extended period of time (e.g., a cleanup timeout interval) and a confirmed primary ingress failure did not occur, then the standard RSVP soft-state removal SHOULD occur. The backup ingress SHALL remove the state for the PATH message from the primary ingress, and tear down the one-to-one backup LSPs for protecting the primary ingress if one-to-one backup is used or unbind the facility backup LSPs if facility backup is used.

When the backup ingress receives a PATH message from the primary ingress for locally protecting the primary ingress of a protected LSP, it checks to see if any critical information has been changed. If the next hops of the primary ingress are changed, the backup ingress SHALL update its backup LSP(s).

6.3.1.1. Relay-Message Method

When the backup ingress receives a PATH message with the INGRESS-PROTECTION object, it examines the object to learn what traffic associated with the LSP and what ingress failure detection mode is being used. It determines the next-hops to be merged to by examining the Label-Routes sub-object in the object. If the Traffic-Descriptor sub-object isn’t included, this object is considered "empty".

The backup ingress stores the PATH message received from the primary ingress, but does NOT forward it.

The backup ingress MUST respond with a RESV to the PATH message received from the primary ingress. If the INGRESS-PROTECTION object is not "empty", the backup ingress SHALL send the RESV message with the state indicating protection is available after the backup LSP(s) are successfully established.

6.3.1.2. Proxy-Ingress Method

The backup ingress determines the next-hops to be merged to by collecting the set of the pair of (IPv4/IPv6 sub-object, Label sub-object) from the Record Route Object of each RESV that are closest to the top and not the Ingress router; this should be the second to the top pair. If a Label-Routes sub-object is included in the INGRESS-PROTECTION object, the included IPv4/IPv6 sub-objects are used to filter the set down to the specific next-hops where protection is desired. A RESV message must have been received before the Backup Ingress can create or select the appropriate backup LSP.

When the backup ingress receives a PATH message with the INGRESS-PROTECTION object, the backup ingress examines the object to learn what traffic associated with the LSP and what ingress failure

detection mode is being used. The backup ingress forwards the PATH message to the ingress node with the normal RSVP changes.

When the backup ingress receives a RESV message with the INGRESS-PROTECTION object, the backup ingress records an IMPLICIT-NULL label in the RRO. Then the backup ingress forwards the RESV message to the ingress node, which is acting for the proxy ingress.

6.3.2. Backup Ingress Behavior in On-path Case

An LER as the backup ingress determines that it is on-path if one of its addresses is a next hop of the primary ingress and the primary ingress is not its next hop via checking the PATH message with the INGRESS_PROTECTION object received from the primary ingress. The LER on-path sends the corresponding PATH messages without any INGRESS_PROTECTION object to its next hops. It creates a number of backup P2P LSPs or a backup P2MP LSP from itself to the other next hops (i.e., the next hops other than the backup ingress) of the primary ingress. The other next hops are from the Label-Routes sub object.

It also creates a forwarding entry, which sends/multicasts the traffic from the source to the next hops of the backup ingress along the protected LSP when the primary ingress fails. The traffic is described by the Traffic-Descriptor.

After the forwarding entry is created, all the backup P2P LSPs or the backup P2MP LSP is up and associated with the protected LSP, the backup ingress sends the primary ingress the RESV message with the INGRESS_PROTECTION object containing the state of the local protection such as "local protection available" flag set to one, which indicates that the primary ingress is locally protected.

When the primary ingress fails, the backup ingress sends/multicasts the traffic from the source to its next hops along the protected LSP and imports the traffic into each of the backup P2P LSPs or the backup P2MP LSP transmitting the traffic to the other next hops of the primary ingress, where the traffic is merged into protected LSP.

During the local repair, the backup ingress continues to send the PATH messages to its next hops as before, keeps the PATH message with the INGRESS_PROTECTION object received from the primary ingress and the RESV message with the INGRESS_PROTECTION object to be sent to the primary ingress. It sets the "local protection in use" flag in the RESV message.

6.3.3. Failure Detection

Failure detection happens much faster than RSVP, whether via a link-level notification or BFD. As discussed, there are different modes for detecting it. The backup ingress MUST have properly set up its forwarding state to either always forward the specified traffic into the backup LSP(s) for the Source-Detect and Next-Hop-Detect modes or to swap from discarding to forwarding when a failure is detected for the Backup-Source-Detect and Backup-Detect modes.

For facility backup LSPs, the correct inner MPLS label to use must be determined. For the ingress-proxy method, that MPLS label comes directly from the RRO of the RESV. For the relay-message method, that MPLS label comes from the Label-Routes sub-object in the non-empty INGRESS-PROTECTION object.

As described in [RFC4090], it is necessary to refresh the PATH messages via the backup LSP(s). The Backup Ingress MUST wait to refresh the backup PATH messages until it can accurately detect that the ingress node has failed. An example of such an accurate detection would be that the IGP has no bi-directional links to the ingress node and the last change was long enough in the past that changes should have been received (i.e., an IGP network convergence time or approximately 2-3 seconds) or a BFD session to the primary ingress' loopback address has failed and stayed failed after the network has reconverged.

As described in [RFC4090 Section 6.4.3], the backup ingress, acting as PLR, SHOULD modify - including removing any INGRESS-PROTECTION and FAST-REROUTE objects - and send any saved PATH messages associated with the primary LSP.

6.4. Merge Point Behavior

An LSR that is serving as a Merge Point may need to support the INGRESS-PROTECTION object and functionality defined in this specification if the LSP is ingress-protected where the failure scenario is Next-Hop-Detect. An LSR can determine that it must be a merge point if it is not the ingress, it is not the backup ingress (determined by examining the Backup Ingress Address (IPv4 or IPv6) sub-object in the INGRESS-PROTECTION object), and the PHOP is the ingress node.

In that case, when the LSR receives a PATH message with an INGRESS-PROTECTION object, the LSR MUST remove the INGRESS-PROTECTION object before forwarding on the PATH message. If the failure scenario specified is Next-Hop-Detect, the MP must connect up the fast-failure detection (as configured) to accepting backup traffic received from

the backup node. There are a number of different ways that the MP can enforce not forwarding traffic normally received from the backup node. For instance, first, any LSPs set up from the backup node should not be signaled with an IMPLICIT NULL label and second, the associated label for the ingress- protected LSP could be set to normally discard inside that context.

When the MP receives a RESV message whose matching PATH state had an INGRESS-PROTECTION object, the MP SHOULD add the INGRESS-PROTECTION object to the RESV message before forwarding it. The Backup PATH handling is as described in [RFC4090] and [RFC4875].

6.5. Revertive Behavior

Upon a failure event in the (primary) ingress of a protected LSP, the protected LSP is locally repaired by the backup ingress. There are a couple of basic strategies for restoring the LSP to a full working path.

- Revert to Primary Ingress: When the primary ingress is restored, it re-signals each of the LSPs that start from the primary ingress. The traffic for every LSP successfully re-signaled is switched back to the primary ingress from the backup ingress.

- Global Repair by Backup Ingress: After determining that the primary ingress of an LSP has failed, the backup ingress computes a new optimal path, signals a new LSP along the new path, and switches the traffic to the new LSP.

6.5.1. Revert to Primary Ingress

If "Revert to Primary Ingress" is desired for a protected LSP, the (primary) ingress of the LSP re-signals the LSP that starts from the primary ingress after the primary ingress restores. When the LSP is re-signaled successfully, the traffic is switched back to the primary ingress from the backup ingress and redirected into the LSP starting from the primary ingress.

It is possible that the Ingress failure was inaccurately detected, that the Ingress recovers before the Backup Ingress does Global Repair, or that the Ingress has the ability to take over an LSP based on receiving the associated RESVs.

If the ingress can resignal the PATH messages for the LSP, then the ingress can specify the "Revert to Ingress" control-option in the INGRESS-PROTECTION object. Doing so may cause a duplication of traffic while the Ingress starts sending traffic again before the Backup Ingress stops; the alternative is to drop traffic for a short

period of time.

Additionally, the Backup Ingress can set the "Revert To Ingress" control-option as a request for the Ingress to take over.

6.5.2. Global Repair by Backup Ingress

When the backup ingress has determined that the primary ingress of the protected LSP has failed (e.g., via the IGP), it can compute a new path and signal a new LSP along the new path so that it no longer relies upon local repair. To do this, the backup ingress uses the same tunnel sender address in the Sender Template Object and uses the previously allocated second LSP-ID in the INGRESS-PROTECTION object of the PATH message as the LSP-ID of the new LSP. This allows the new LSP to share resources with the old LSP.

When the backup ingress has determined that the primary ingress of the protected LSP has failed (e.g., via the IGP), it can compute a new path and signal a new LSP along the new path so that it no longer relies upon local repair. To do this, the backup ingress uses the same tunnel sender address in the Sender Template Object and uses the previously allocated second LSP-ID in the INGRESS-PROTECTION object of the PATH message as the LSP-ID of the new LSP. This allows the new LSP to share resources with the old LSP. In addition, if the Ingress recovers, the Backup Ingress SHOULD send it RESVs with the INGRESS-PROTECTION object where either the "Force to Backup" or "Revert to Ingress" is specified. The Secondary LSP ID should be the unused LSP ID - while the LSP ID signaled in the RESV will be that currently active. The Ingress can learn from the RESVs what to signal. Even if the Ingress does not take over, the RESVs notify it that the particular LSP IDs are in use. The Backup Ingress can reoptimize the new LSP as necessary until the Ingress recovers. Alternately, the Backup Ingress can create a new LSP with no bandwidth reservation that duplicates the path(s) of the protected LSP, move traffic to the new LSP, delete the protected LSP, and then resignal the new LSP with bandwidth.

7. Security Considerations

In principle this document does not introduce new security issues. The security considerations pertaining to RFC 4090, RFC 4875 and other RSVP protocols remain relevant.

8. IANA Considerations

TBD

9. Contributors

Renwei Li
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA
Email: renwei.li@huawei.com

Quintin Zhao
Huawei Technologies
Boston, MA
USA
Email: quintin.zhao@huawei.com

Zhenbin Li
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA
Email: zhenbin.li@huawei.com

Boris Zhang
Telus Communications
200 Consilium Pl Floor 15
Toronto, ON M1H 3J3
Canada
Email: Boris.Zhang@telus.com

Markus Jork
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
USA
Email: mjork@juniper.net

10. Acknowledgement

The authors would like to thank Rahul Aggarwal, Eric Osborne, Ross Callon, Loa Andersson, Michael Yue, Olufemi Komolafe, Rob Rennison, Neil Harrison, Kannan Sampath, and Ronhazli Adam for their valuable comments and suggestions on this draft.

11. References

11.1. Normative References

[RFC1700] Reynolds, J. and J. Postel, "Assigned Numbers", RFC 1700, October 1994.

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

[RFC3692] Narten, T., "Assigning Experimental and Testing Numbers Considered Useful", BCP 82, RFC 3692, January 2004.

[RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

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

[RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

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

[RFC4461] Yasukawa, S., "Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)", RFC 4461, April 2006.

[RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.

[P2MP-FRR]
Le Roux, J., Aggarwal, R., Vasseur, J., and M. Vigoureux, "P2MP MPLS-TE Fast Reroute with P2MP Bypass Tunnels", draft-leroux-mpls-p2mp-te-bypass , March 1997.

11.2. Informative References

[RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, January 2001.

Appendix A. Authors' Addresses

Huaimo Chen
Huawei Technologies
Boston, MA
USA
Email: huaimo.chen@huawei.com

Ning So
Tata Communications
2613 Fairbourne Cir.
Plano, TX 75082
USA
Email: ning.so@tatacommunications.com

Autumn Liu
Ericsson
300 Holger Way
San Jose, CA 95134
USA
Email: autumn.liu@ericsson.com

Raveendra Torvi
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
USA
Email: rtorvi@juniper.net

Alia Atlas
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
USA
Email: akatlas@juniper.net

Yimin Shen
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
USA
Email: yshen@juniper.net

Fengman Xu
Verizon
2400 N. Glenville Dr
Richardson, TX 75082
USA
Email: fengman.xu@verizon.com

Mehmet Toy
Comcast
1800 Bishops Gate Blvd.
Mount Laurel, NJ 08054
USA
Email: mehmet_toy@cable.comcast.com

Lei Liu
UC Davis
USA

Email: liulei.kddi@gmail.com