MPLS Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 13, 2013

Katherine. Zhao
Renwei. Li
Huawei Technologies
Christian. Jacquenet
France Telecom Orange
January 9, 2013

Fast Reroute Extensions to Receiver-Driven RSVP-TE for Multicast Tunnels
draft-zlj-mpls-mrsvp-te-frr-01.txt

Abstract

This document specifies fast reroute procedures to protect multicast LSP tunnels built by mRSVP-TE, a receiver-driven extension to RSVP-TE specified by [I-D.draft-lzj-mpls-receiver-driven-multicast-rsvp-te]. This document is motivated by the observation that the existing FRR solution specified by [RFC4090] and [RFC4875] for the sender-driven RSVP-TE is no longer applicable to the receiver-driven RSVP-TE.

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 July 13, 2013.

Copyright Notice

Table of Contents

1. Terminology

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 RFC2119 [RFC-WORDS]. The reader is assumed to be familiar with the terminology in [RSVP], [RSVP-TE] and [mRSVP-TE].

This document uses same terminologies stated in [I-D.draft-lzj-mpls-receiver-driven-multicast-rsvp-te], [RFC4090] and [RFC4875]. In addition, some key notions and terminologies for this document are explained as follows:

- mLSP, Multicast Label Switched Path, is either a P2MP or MP2MP LSP consisting of one or more sub-LSPs.

- mRSVP-TE, Multicast Resource Reservation Protocol-Traffic Engineering, is used to distinguish from the regular sender-driven RSVP-TE. One major difference between RSVP-TE and mRSVP-TE is that the tunnel setup is initiated by the data receiver instead of the data sender.

- PLR: Point of Local Repair, an LSR that detects a local failure event and redirects traffic from protected mLSP to a backup mLSP tunnel which is designed to take over traffic forwarding until the protected tunnel is repaired.

- MP: Merge Point, an LSR that merges the traffic from backup tunnels with primary LSP at the level of forwarding plane. In the receiver-driven RSVP-TE for approach, the MP is the LSR that initiates backup mLSP setup taking PLR as the root of the backup LSP.

- N: The node to be protected.

- Pn: The node(s) on the backup path for protecting node N.

- Root: A router where an mLSP is rooted at. Multicast contents enter the root and then are distributed to leaf routers along the P2MP/MP2MP LSP.

- FRR Domain: A set of links and LSRs that compose a protected sub-LSP and backup LSP, and which is located between PLR and MP(s).

2. Introduction

Fast Reroute technology has been well accepted and deployed to provide millisecond-level protection in case of node/link failures. FRR employs some local repair mechanisms to meet the fast reroute requirements by computing and provisioning backup tunnels in advance of failure and by redirecting traffic to such backup tunnels as close to the failure point as possible.

The fast reroute extensions to RSVP-TE are specified in [RFC4090] and [RFC4875]. Such extensions work well with the sender-driven RSVP-TE, but they are no longer applicable to the receiver-driven RSVP-TE for multicast tunnels described in the draft [I-D.draft-lzj-mpls-receiver-driven-multicast-rsvp-te].

In the receiver-driven paradigm of mRSVP-TE, the procedure to set up an LSP tunnel is inverted from that in the sender-driven RSVP-TE, and thus the backup mLSP setup and failover handling mechanism will have to be different from what has been specified for the sender-driven RSVP-TE. From the signaling point of view, the behavior of PLR and MR is inverted from the sender-driven paradigm of RSVP-TE: the setup for a backup mLSP is initiated by MP with PLR being taken as the root of a P2MP/MP2MP tree. The RSVP PATH message is sent from MP towards PLR with the FAST_REROUT, DETOUR as well as other FRR related objects conveyed in the PATH message. RSVP RESV message is sent from PLR towards MP carrying FRR information such as the inner label used to represent a protected mLSP tunnel, etc.

On the other hand, from the packet forwarding point of view, the behavior of PLR and MP is similar to the sender-driven RSVP-TE. The traffic switchover and redirecting are still initiated by PLR, and the data traffic is merged at MP in the same way as what is specified for the sender-driven RSVP-TE.

This document describes various FRR protection methods and behavior changes for the receiver-driven mRSVP-TE, and specify fast-reroute extensions to the RSVP-TE messages, mechanisms and procedures specified in the mRSVP-TE draft [I-D.draft-lzj-mpls-receiver-driven-multicast-rsvp-te].

2.1. Link Protection and Node Protection with mRSVP-TE

FRR link protection aims to protect a direct link between two LSRs (Label Switch Routers). An LSR at one end of the link is called PLR (Point of Local Repair), and the other LSR located at the other end of the link is called MP (Merge Point). A backup LSP whose setup is originated at MP and terminated at PLR will be established to protect the primary LSP crossing over the link. The LSRs over the backup

path are called Pn. These connected LSRs and links are called an FRR domain in this document. An example of an FRR domain supporting link protection is shown in Figure 1.

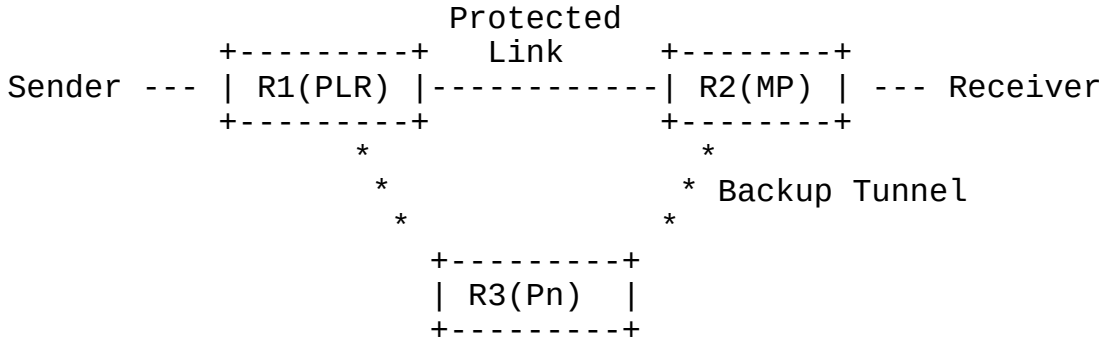

Figure 1: Basic FRR Link Protection

In an FRR domain constructed by mRSVP-TE, the MP initiates both the primary and the backup LSP setup at the signaling control plane, and merges the traffic from the backup LSP into the primary LSP at the data forwarding plane. The PLR works with the MP to set up LSP at the signaling control plane accordingly, and detects link failure and initiates local repair at the data forwarding plane. In Figure 1, we use hyphens (-)to denote a primary tunnel between LSRs; and asterisks (*) to denote a backup tunnel. The same symbols will be applied to all figures throughout the document.

Node protection is a technique used to protect a node N that resides between PLR and MP over a primary LSP. An example of node protection is shown in Figure 2.

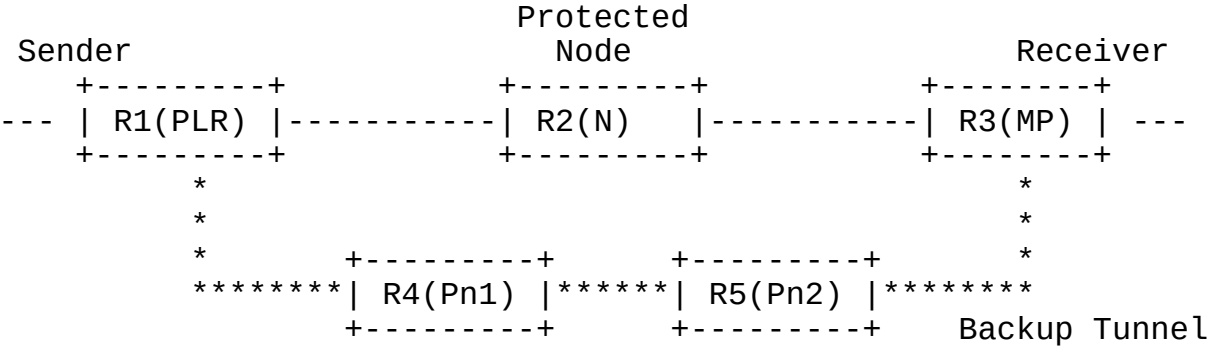

Figure 2: Basic FRR Node Protection

N (R2) denotes a node being protected over a primary LSP, its upstream node plays the role of PLR while the downstream node plays the role of MP. Pn denotes a transit node over its backup LSP. Note that there can be multiple Pn's over a backup tunnel. Pn does not play a significant role for FRR but works as a regular LSR to receive and transmit multicast data and signaling messages over backup LSPs.

Besides the basic P2P node protection, mRSVP-TE suggests P2MP and MP2MP node protection as shown in Figures 3 and 4. Because the same protection mechanism can be commonly used for both P2MP and MP2MP, this document uses P2MP as example for the discussion, and mention MP2MP only if there is a difference from P2MP.

There are two typical methods to protect a P2MP multicast tree, one that uses a P2MP tree as a backup LSP to protect a primary mLSP (see Figure 3), and the other that uses multiple P2P LSPs to protect a P2MP mLSP(see Figure 4).

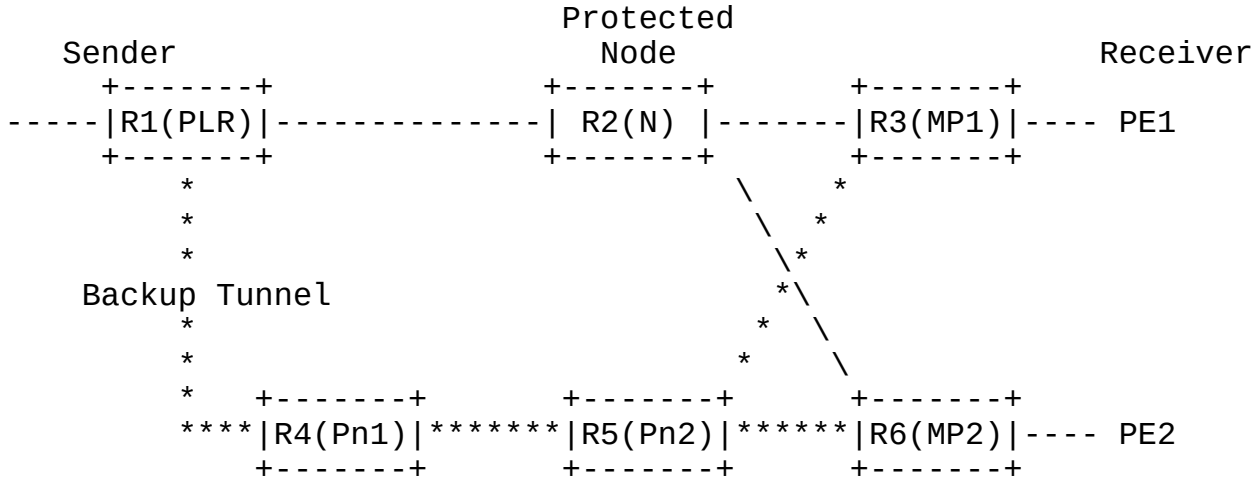

Figure 3: P2MP Node Protection in Facility Mode

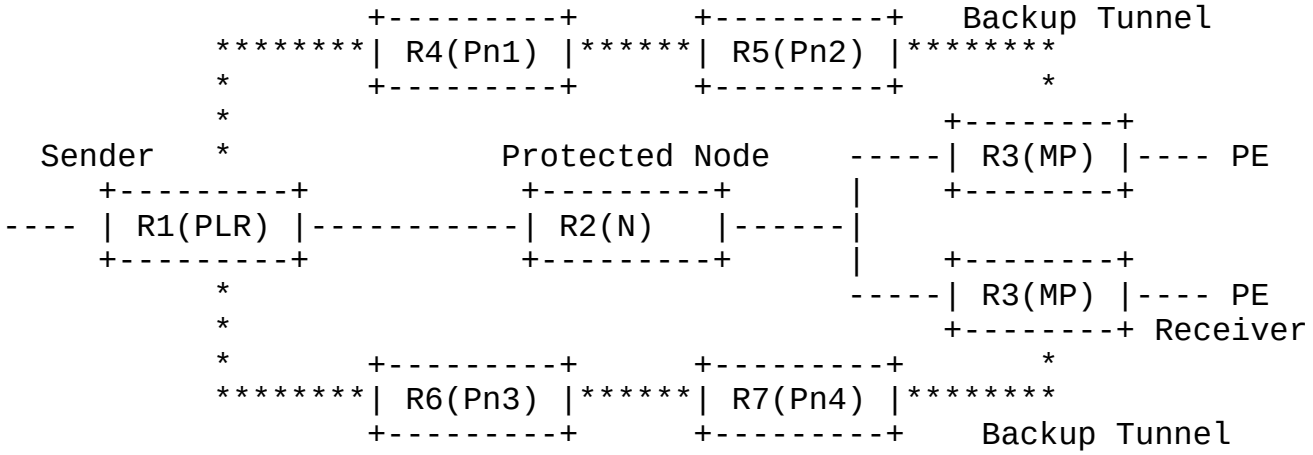

Figure 4: Multiple P2Ps Protecting a P2MP LSP

2.2. Primary and Backup LSP

A router that detects a node/link failure must have pre-determined which alternate reroute path it should use to forward traffic while the failure is being fixed. The alternate backup path should be established before a protected LSP is broken. Anything such as backup route computation and configuration required for local repair purposes should be done prior to failure occurrence so that the failover time can be reduced to minimum.

On the control plane, the backup LSP will be set up along with its primary LSP setup. The PATH/RESV refresh messages are transmitted over both protected and backup LSPs before failover. However on the data plane, there are two implementation options for traffic forwarding. One option is that traffic is not forwarded on backup LSP tunnel until a failure is detected and the local repair takes place. The second option is to forward traffic on both protected and backup mLSPs before failover, and the LSR at Merge Point will then drop packets coming from the backup path before switchover. The second option can further reduce traffic switchover time at the cost of extra overhead and bendwidth sub-optimization. This document leaves the flexibility for implementation to decide which option to choose, but will use the first option for the discussion, i.e. we assume that the traffic is forwarded on the primary LSP only before switchover.

2.3. Detour Backup and Facility Backup

Due to historical reasons and implementation preferences, two independent methods for doing fast reroute have been developed. One backup method is called detour backup and is especially designed for

1:1 protection. The other one is called facility backup and is especially designed for 1: N protection, where N can be equal to or greater than 1. From the point of view of applications, the facility backup method can support both 1:N and 1:1, but from the technical point of view, these are two different methods requiring different implementations with respect to their label stacks when forwarding packets.

The detour backup creates a dedicated LSP to protect an LSP and uses a single MPLS label for packet encapsulation; its implementation is simpler but consumes more label resources. The facility backup creates a common LSP to protect a set of LSPs that have similar backup constraints. This method takes advantage of MPLS label stacking and uses dual-label encapsulation, thus it can save some label resources compared to the detour backup method.

These two solutions have co-existed as options for vendors and service providers to choose. This document will specify both the methods applied to mRSVP-TE. Throughout the document, the detour method is used to represent 1:1 protection while the facility method is used to represent 1:N protection. The term "detour LSP" is especially used for 1:1 protection while "backup LSP" is used for 1: N protection. Sometimes the latter can be used for both kinds of protection schemes when no ambiguity arises.

3. Detour Backup for mRSVP-TE

This section specifies mechanisms and procedures for mRSVP-TE fast reroute by using the detour backup method. The term "detour LSP" will be used to denote the LSP in the detour mode and the 1:1 protection scheme.

3.1. Link Protection in Detour Backup Mode

3.1.1. Detour LSP Setup Scenario for Link Protection

A detour LSP setup is initiated by MP along with the setup of the protected LSP (Figure 1), which is one of the major differences from the procedure stated in [RFC4090] and [RFC4875]. Following the LSP setup procedure specified by the draft [I-D.draft-lzj-mpls-receiver-driven-multicast-rsvp-te], MP sends RSVP PATH messages towards the sender over a primary path. For link protection purpose, both the MP and PLR are directly connected by the link being protected, hence the PATH message is sent from the MP to the PLR directly upstream.

The MP is not necessarily the originator of the primary LSP, but is

the first LSR entering an FRR domain along the primary route. Once the PATH message is sent out by the MP, the MP will check whether there is a detour route available for link protection. The detour route calculation can be done by running CSPF on the link state database produced by IGP protocols with TE extensions. There is no change required for backup route computation, and the detour LSP computation will be based on this assumption.

If the CSPF stack returns 'no detour route found' after the detour calculation, MP will not perform the detour LSP setup. If at least one detour route is found by CSPF stack, MP selects the shortest route and initiates the detour LSP setup. MP considers PLR as the end point of the detour LSP and sends a PATH message towards PLR hop-by-hop. In the example of Figure 1, the PATH message will be sent to Pn (R3) and then relayed to PLR (R1).

Upon receipt of the PATH message, the PLR sends back a RESV message towards the MP through the Pn(s). The transit Pn(s) nodes relay the PATH/RESV messages without any special process required for the link protection. The detour LSP setup is completed once the RESV message is received and processed by the MP.

3.1.2. Label Allocation for Link Protection

Because the detour method uses a dedicated backup LSP to protect a primary LSP, one-to-one binding can be made for a pair of primary and backup LSPs, a single MPLS label encapsulation will be sufficient for packet forwarding and local failure repair purpose. DLA (downstream label allocation) can be used as the label assignment method over the detour tunnel for the link protection. With mRSVP-TE, a downstream label is assigned by an LSR that is sending a PATH message to its upstream router, and an upstream label is assigned by an LSR that is sending the RESV message to its downstream router. The label allocation, however, is more complicated when the primary LSP is a P2MP or MP2MP tree. A specific upstream label allocation and resource preemption method is defined in this document to handle the protection of P2MP and MP2MP tree structures.

An example of the label allocation for link protection in the detour mode is provided in Figure 5. For the sake of readability, we use label Lp to represent the label assigned to the primary tunnel, and label Lb for the labels assigned to the backup tunnel. For example, Lp2 represent a downstream label assigned for LSR R2 to receive incoming data over the primary tunnel. Lb2 represents a downstream label assigned for R2 to receive data over a detour LSP.

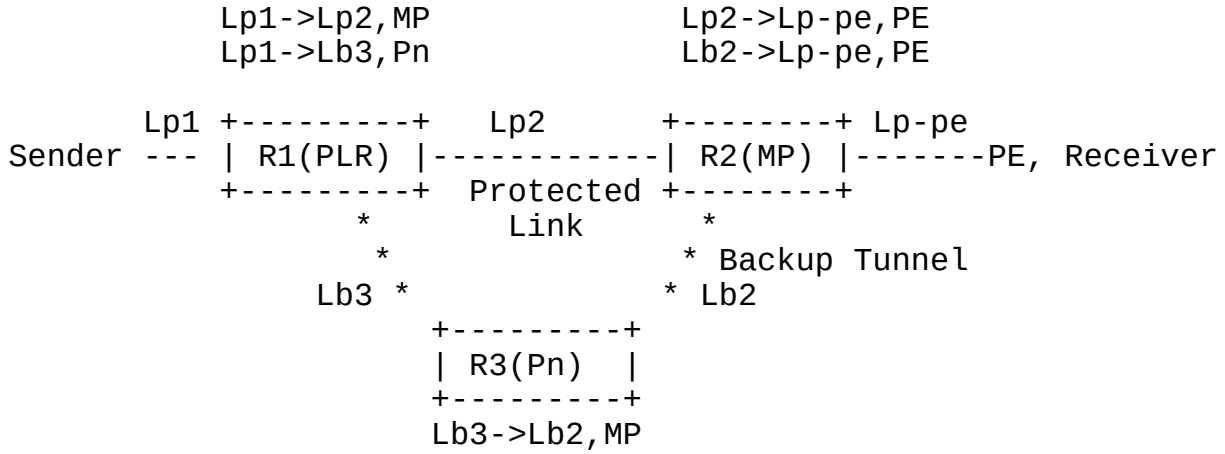

Figure 5: Label Allocation for Link Protection in Detour Mode

In the example of Figure 5, MP assigns label Lp2 and sends it to PLR via the PATH message over the link {MP-PLR} to set up the primary LSP. For the detour route {MP-Pn-PLR}, MP assigns a label Lb2 and sends it to Pn via the PATH message. MP binds label Lp2 with label Lb2 for this pair of primary and detour LSPs. An entry 'Lp2->Lp-pe, PE' will be added into MP's FIB to forward packets over the protected LSP. Another entry 'Lb2-> Lp-pe, PE' will be added and used when traffic is received from the detour tunnel upon switchover.

Pn (transit node) on the detour tunnel receives Lb2 from MP. Pn assigns a downstream label Lb3 and sends it to the PLR via a PATH message. Pn will add an entry 'Lb3->Lb2, MP' to its FIB for packet forwarding. Note that Pn is not aware of the primary LSP, so there is only one forwarding entry needed in its FIB.

PLR receives two PATH messages from MP and Pn respectively. Then it binds label Lp2 from the primary LSP with label Lb3 from the detour LSP. The detour LSP ends at PLR while the primary LSP may not end at PLR if the PLR is not the root of the P2MP tree. PLR will allocate a downstream label Lp1 and sends it to its upstream router, which is outside of the FRR domain in this example, hance not shown in Figure 5. There will be two entries added into PLR's FIB: one entry 'Lp1->Lp2, MP' for the primary traffic forwarding, and another entry 'Lp1->Lb3, Pn' for the detour traffic forwarding upon failover.

PLR processes PATH messages from MP and sends RESV messages towards MP. If the primary sub-LSP is part of a RD P2MP tree, PLR will not allocate upstream labels for receiving traffic from the downstream node (MP or Pn in this example) because traffic is uni-directionally forwarded. If the primary sub-LSP is part of a RD MP2MP tree, PLR will allocate an upstream label for receiving traffic from the opposite direction, and Pn(s) do the same and allocate upstream label

for the detour sub-LSP accordingly. Detour LSP setup is completed once MP has received and processed the RESV message originated by PLR. Figures 5 shows the summary of labels allocated and FIB entries created on each node in the FRR domain.

3.1.3. Link Failure Repair in Detour Mode

Link failure can be detected by, for example, BFD (Bidirectional Forwarding Detection, [RFC5880],[RFC5884])along the protected LSP, The failure detection algorithm is the same as what is used for the sender-driven RSVP-TE.

Once a link failure is detected by PLR and all switchover criteria are met, PLR will redirect the traffic to the detour LSP based on the forwarding entry 'Lp1->Lb3, Pn'. The entry 'Lp1->Lp2, MP' for the primary path will be withdrawn.

Pn works as a normal label switch router and forward MPLS packets to MP. MP receives the packet and figures out that such packets come from the detour path, so they will be forwarded to PE based on the entry 'Lb2->Lp-pe, PE', in the example of Figure 5. The detour traffic is therefore merged back to the primary LSP towards PE, which completes the link failure repairing by detouring and merging the traffic.

3.1.4. Re-convergence after Local Repair

Routers that do not belong to the FRR domain are not impacted by the link failure and local repair. Traffic is transmitted over a detour LSP after a link failure and local repair. Usually, the detour path is not the shortest path so the network will eventually re-converge and a new shortest path will be calculated by the MPLS control plane. Once a new primary path is determined, the traffic is no longer transmitted through the detour LSP and PLR will be notified to tear down the detour LSP and clean up its internal LIB. PLR will send a PathTear message to Pn and MP for tearing down the detour LSP and release backup labels. Re-convergence procedure is the same as the procedure used for sender-driven RSVP-TE FRR.

3.2. Node Protection in Detour Backup Mode

3.2.1. Detour LSP Setup for Node Protection

The detour LSP setup for the node protection is similar to the link protection. Take Figure 2 as an example, where protected node N resides between MP and PLR. In this case the two sub-links {MP-N} and {N-PLR} are also to be protected in addition to the node N protection. It is assumed that the link protection mechanism

discribed in the previous sub-section is applicable to the sub-link protection in this situation. Hence this section will focus on the procedure to handle node protection. A combined solution for providing node protection with link protection can be derived from the discussions of section 3.1 and this section.

For the node protection shown in Figure 2, MP(R3) sends a PATH message to N for the primary LSP setup, the primary LSP in the FRR domain goes through the route {MP-N-PLR}. Once the PATH message is sent out to N, MP checks whether there is a detour path available for node N by using CSPF computation, which would indicate N as a node to be avoided on the detour path. If no detour route is found, MP skips the detour LSP setup. If a detour route is found, MP initiates the detour LSP setup and considers PLR as the end-point of the detour LSP. MP sends a PATH message towards PLR over the detour route hop-by-hop. In the example of Figure 2, the detour route is in the order of {MP-Pn2-Pn1-PLR}. Similar to the link protection, PLR sends back a RESV message towards MP through Pn(s). Transit node Pn(s) just relay the PATH and RESV messages without any specific node protection procedure. The detour LSP setup is completed once the RESV message is received and processed by MP.

Figure 2 shows a typical example of node protection where N is not a branch node; it will be more complicated when N is a branch node that is part of a RD P2MP/MP2MP tree structure. The corresponding mechanism is described in section 5.2.2.

3.2.2. Label Allocation and Binding for Node Protection

Similar to link protection, node protection uses the single label encapsulation and downstream label allocation method in the detour backup mode. An example of the label allocation for node protection is provided in Figure 6.

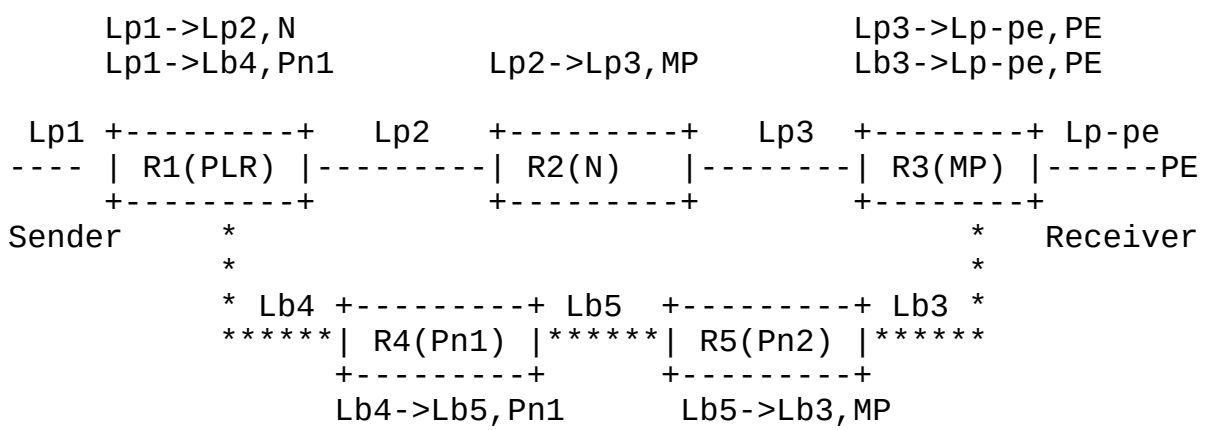

Figure 6: Node Protection in Detour Mode

MP (R3) assigns a label Lp3 for the primary LSP and sends it to node N via a PATH message over the protected route {MP-N-PLR}. N will allocate a downstream label Lp2 and sends it to PLR via a PATH message. MP also assigns a label Lb3 for the detour LSP and sends it to Pn2 via a PATH message over the detour route {MP-Pn2-Pn1-PLR}. MP binds label Lp3 with label Lb3 for this pair of primary and backup LSPs. An entry 'Lp3->Lp-pe, PE' will be added to MP's FIB for packet forwarding over the primary LSP. Another entry 'Lb3->Lp-pe, PE' will be kept in the FIB and used when a failover takes place and traffic is redirected to the detour LSP.

There could be multiple transit nodes Pn(s) along the detour LSP, each of which will allocate a downstream label and sends it to its upstream router. Eventually PLR receives the PATH message from the protected node N and the transit node Pn1 in this example. PLR binds primary label Lp2 with the detour label Lb4, and adds two entries into its FIB: One entry 'Lp1->Lp3, N' for the traffic forwarding over the primary LSP, and another entry 'Lp1->Lb4, Pn1' for the traffic forwarding over the detour LSP. An example of the allocated labels and FIB entries in the FRR domain are mentioned in Figure 6.

3.2.3. Node Failure Repair in Detour Mode

Once the node N failure is detected by PLR, it will redirect the traffic from the primary LSP to its detour LSP based on the binding and forwarding entry 'Lp1->Lb4, Pn1'. The traffic is forwarded through LSR->Pn1-Pn2->MP. Eventually, MP will receive packets from the detour path. Consulting its FIB forwarding entry 'Lb3->Lp-pe, PE', traffic will then be forwarded to PE in the example of Figure 6, so that the detoured traffic gets merged into the primary path.

The local repair mechanism for the node protection is the same as the link protection in the detour mode except that there are two links {MP-N} and {N-PLR} to be protected in conjunction with the node N protection. The FRR domain must be configured so that both the link and node failure detection methods are specified. For example, BFD needs to be activated between MP abd N, N and PLR, and PLR and MP. PLR and MP can be used for either link repair, node repair or both depending on the results of BFD detection.

3.2.4. Re-Convergence after Local Repair

After a node failure takes place, the network topology will change. As a consequence, the network will eventually re-converge and a new best path will be computed to establish the primary LSP. PLR will be notified as soon as the new primary LSP is signaled and set up. PLR

will send notification messages to Pn1 and MP for tearing down the detour LSP and withdraw backup labels.

4. Facility Backup for mRSVP-TE

This section specifies mechanisms and procedures for mRSVP-TE fast reroute by using the facility backup method. The term backup LSP will be used to denote the LSP in the facility mode for 1: N protection. Note that the term 'detour LSP' is no longer used in this section for the Facility backup.

The backup LSP differs from the detour LSP in that one single backup LSP is used to protect multiple primary LSPs. General speaking, two labels will be used for the backup LSP with the inner label being used to indicate which primary LSP is being protected.

4.1. Link Protection in Facility Backup Mode

4.1.1. Backup LSP Setup for Link Protection

Similar to the detour LSP setup, MP sends a RSVP PATH message towards PLR over the primary route. Once the PATH message is sent out, MP will execute the backup LSP procedures as per the following steps:

- Check whether there has been a backup LSP created to protect the link between PLR and MP. If a backup LSP is found, skip the further process at MP, e.g., do not send a PATH message over the backup route for LSP setup. However, this does not mean that no process is needed for link protection. Later on, the PLR will allocate an inner label for each newly created primary LSP and send it to Pn(s) and MP via RESV messages. Details for label allocation and packet encapsulation are discussed in section 4.1.2.

- If there is no backup LSP available, MP initiates the backup LSP setup: MP calculates a backup route by using CSPF taking PLR as the endpoint of the backup LSP and sends a PATH message towards PLR hop-by-hop over the backup route. In the example of Figure 1, PATH messages will be sent from MP to Pn (R3) and relayed to PLR (R1). PLR will then send a RESV message to MP, so as to complete the backup LSP setup. Section 4.1.2 specifies the details about the label allocation and binding.

4.1.2. Label Allocation for Link Protection

As a backup LSP protects one or more primary LSPs, the facility protection scheme uses two labels for packet forwarding. The outer label is used for regular packet forwarding hop-by-hop over the backup LSP, while the inner label is used to represent a primary LSP and used by MP to merge traffic forwarded over the backup LSP to its corresponding primary LSP. Multiple primary LSPs will share the common outer label while the inner label is unique for each protected LSP. Figure 7 below shows how the two labels are assigned and used for the facility backup. There are two primary LSPs to be protected by a common backup LSP in this example.

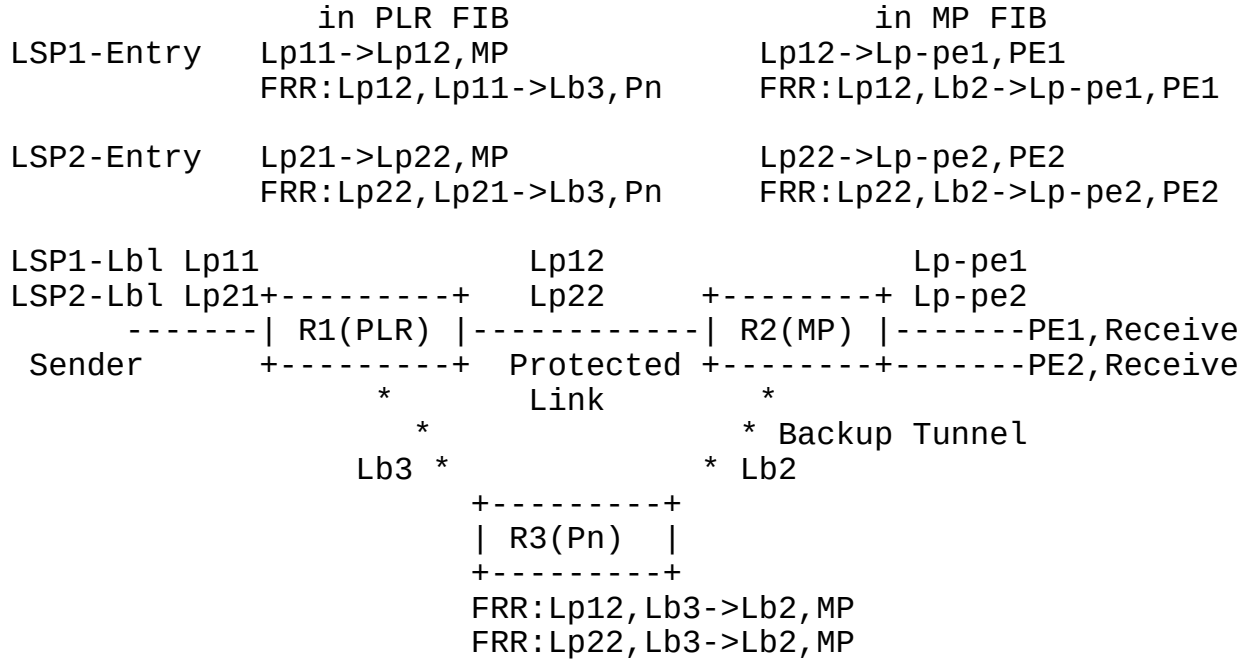

Figure 7: Label Allocation for Link Protection in Facility Mode

Assume that primary LSP1 is created first, MP assigns a downstream label Lp12 for LSP1 being protected and sends the label to PLR via a PATH message over route {MP-PLR}. Because the primary LSP1 is the first LSP created over this route, MP also assigns a downstream label Lb2 for the backup LSP and sends it to Pn via a PATH message over the backup route {MP-Pn-PLR}. Pn allocates a downstream label Lb3 and sends it to PLR via a PATH message.

Once PATH messages are received from MP and Pn respectively, PLR will allocate an inner label to represent the primary LSP1 for the backup LSP. The method to allocate the inner label is implementation-specific. In this example, label Lp12 is used as the inner label to

represent primary LSP1 over the backup LSP. LSR at merge point uses the inner label to locate the corresponding primary LSP. The inner label is propagated from PLR to MP by a RESV message. Note that PLR and MP are the only LSRs that actually see, use or process the inner label, while other transit node Pns do not process the inner label.

The process for the second or additional primary LSPs protected by the same backup LSP is different from that for the first one. MP does not allocate any new downstream label for the backup LSP since the backup LSP for the first primary LSP is shared between all the primary LSPs protected by the same backup LSP. But the PLR is required to allocate an inner label for each newly created primary LSP and sends it to MP hop-by-hop via a RESV message.

We use Figure 7 as an example to show the packet forwarding FIB entry by using the following format:

FRR:(inner label),(incoming outer label)->(outgoing outer label),NHOP

When MP allocates the downstream label Lp12 for the primary LSP1, an entry ’Lp12->Lp-pe1, PE1’ is added into MP’s FIB. Another FRR entry ’FRR: Lg12, Lb2->Lp-pe1, PE1’ is added when MP receives a RESV message that carries an inner label Lg12 and binding information with LSP1. So the MP will have two forwarding entries for each protected LSP. In this example MP will maintain four entries in its FIB for the two protected paths LSP1 and LSP2:

Lp12->Lp-pe1, PE1

Lp22->Lp-pe2, PE2

FRR: Lp12, Lb2 -> Lp-pe1, PE1

FRR: Lp22, Lb2 -> Lp-pe2, PE2

PLR creates a forwarding entry for a primary LSP whenever it receives a PATH message for the setup of a new primary LSP. For each primary path LSP1, once PLR receives the PATH message from the backup route, PLR allocates an inner label for the primary LSP and creates an FRR entry in its FIB. The PLR FIB will have these entries for the two protected LSP LSP1 and LSP2:

Lp11 ->Lp12, MP

Lp21->Lp22, MP

FRR: Lp12, Lp11 -> Lb3, MP

FRR: Lp22, Lp21 -> Lb3, MP

Note that the transit routers Pn use the outer label for packet forwarding and keep the inner label untouched.

4.1.3. Link Failure Repair in Facility Mode

Before a link failure is detected, PLR encapsulates user packets with a single label Lp1 and forwards the packet to MP. MP also uses a single label encapsulation and forwards the packet to PE (as per Figure 7).

After a link failure is detected, the PLR (for example, R1 in Figure 7) will encapsulate traffic with two labels: the outer label Lb2 is used for packet forwarding over the backup path, while the inner label Lp2 is used to map traffic to the corresponding primary LSP. MP will pop out outer label Lb2 if needed, swap inner label Lp12 with Lp-pe1, and then forward packets to PE1, as per the example of Figure 7.

4.1.4. Re-Convergence after Local Repair

After a link failure occurs, the network will reconverge. PLR will be notified as soon as a new best path for the primary LSP will be found and activated. Then PLR will tear down the backup LSP, release backup labels and clean up entries in its FIB.

4.2. Node Protection in Facility Backup Mode

4.2.1. Backup LSP setup in Facility Mode

Two methods for node protection in the facility protections scheme have been illustrated in Figures 3 and 4. The method shown in Figure 3 uses a P2MP or MP2MP backup LSP to protect a branch node N; the method shown in Figure 4 uses multiple LSPs to protect the node N. The first method is likely to reduce traffic replication on the backup LSP; the second method suffers from traffic overhead because multiple backup sub-LSPs are used. Which method to use is design option. In this document, we will use the method shown in Figure 3 to describe the node protection mechanism in the facility protection scheme.

Specific procedures are needed for the P2MP or MP2MP tree setup and label allocation. Assume that LSR PE1 joins a primary P2MP tree structure in the example of Figure 3. PE1 sends a RSVP PATH message to MP1 for LSP setup, this PATH message will be relayed to PLR through node N being protected. MP1 calculates the backup route with a constraint to avoid node N; it initiates the backup LSP setup by

sending a PATH message over the backup path {MP1-Pn2-Pn1-PLR}. RSVP RESV messages will then be sent in return by PLR to MP1 through the primary {PLR-N-MP1} and the backup {PLR-Pn1-Pn2-MP1} routes respectively.

Later on, another LSR PE2 joins the P2MP tree by sending a PATH message to MP2. MP2 will relay the PATH message to node N being protected. Then N becomes a branch node and it is therefor not necessary to send PATH messages to the PLR anymore. MP2 performs the same procedure as MP1 did for the first branch {PE1-MP1-N}, a backup route {MP2-Pn2-Pn1-PLR} will be computed by CSPF, and the node Pn2 now becomes a branch node that belongs to the backup P2MP tree. The PATH message that used to be sent by Pn2 towards the PLR is not necessary anymore. RSVP RESV messages will be sent back by the PLR to MP2 through the primary route {PLR-N-MP2} and the backup route {PLR-Pn1-Pn2-MP2} respectively.

Whenever additional primary LSP(s) are set up as far as the same node N and PLR are connected, all these primary LSPs can be protected by the single backup LSP. The procedure to setup the primary LSP is the same as what is used for the first primary LSP setup, the key technique is to allocate a unique identifier to a primary LSP and bind it with the backup LSP, as per the mechanism discribed in section 4.2.2.

4.2.2. Label Allocation for Node Protection

In order to achieve 1:n protection in Facility mode, a unique identifier must be assigned to represent each primary LSP being protected. This identifier should be advertized to all the LSRs in a FRR domain and used for traffic switchover in case of node N failure. There are many ways to assign and use the identifier, and this document gives a sample mechanism based upon ULA (Upstream Label Allocation) to assign a MPLS label and use it as the identifier of a primary LSP. Figure 8 provides an example of label allocation and FIB entry creation for the node protection in Facility mode.

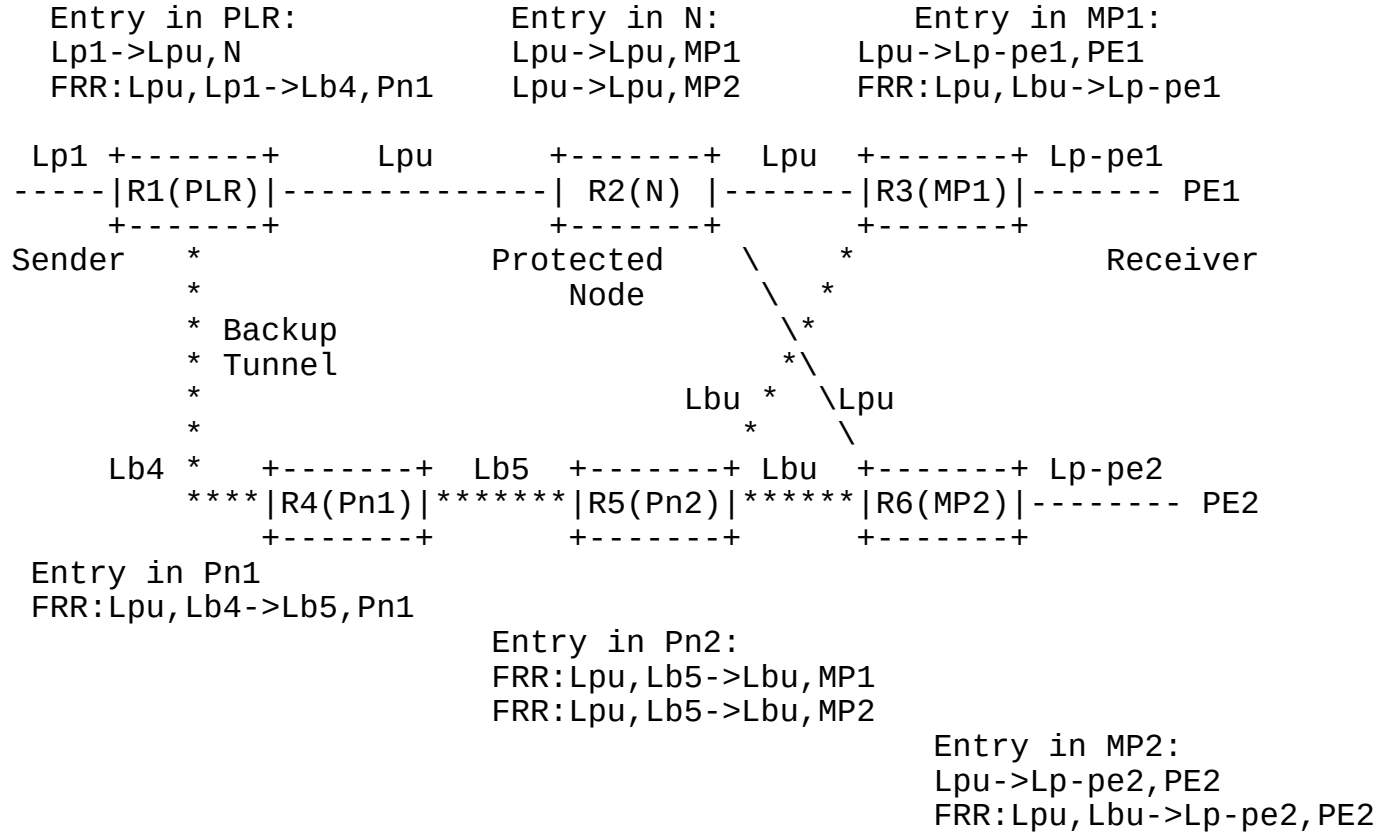

Figure 8: Label Allocation for P2MP Node Protection in Facility Mode

In the FRR domain of Figure 8, an identical label Lpu is assigned to these sub-LSPs over the primary LSP: {PLR-N}, {N-MP1} and {N-MP2}. Lpu can be allocated by the branch node N for the primary LSP and used as the identifier of the primary LSP. If there are multiple primary LSPs that cross the same node N and need to be protected by the single backup LSP, there will be multiple Lpu labels assigned for each of the primary LSPs accordingly. In order to guarantee the uniqueness of Lpu in node N and MPs, the LSRs are required to have ULA capability in FRR domain. In addition, an algorithm for ULA assignment and negotiation among the LSRs needs to be further specified by a yet-to-be-published internet draft.

During normal operation, PLR encapsulates packets with the label Lpu and forwards them to node N over the primary LSP. The node N as a branch node will replicate traffic to MP1 and MP2 using label Lpu in the example of Figure 8. When a node failure is detected, PLR will redirect traffic to the backup LSP, and the two labels will be used for packet encapsulation over the backup LSP. The inner label is Lpu and uniquely identifies a primary LSP; the outer label is allocated by MP and Pn(s) using DLA (Downstream Label Allocation), which is used for packet forwarding over the backup LSP by means of RSVP-TE

mechanism.

Detailed label allocation on each LSR is described below.

1. Label Allocation and FRR Entry on MP1 and MP2:

For the first primary LSP setup, MP1 assigns a downstream label Lpdla for the primary LSP and sends it to the protected node N via a PATH message. Node N discards Lpdla and uses ULA to assign a new label Lpu that will be used as a downstream label for N to send packets to MP1.

Node N sends the label Lpu to MP1 via a RESV message; MP1 replaces its downstream assigned label Lpdla with Lpu. If Lpu has been used by another LSP on the LSR, MP1 will request node N to assign another Lpu by a RSVP notify message. In case of conflict, an ULA negotiation procedure has to be executed (this procedure is TBD).

MP1 also assigns a downstream label Lbdla for the backup LSP and sends it to Pn2 via a PATH message over the backup route {MP1-Pn2-Pn1-PLR in Figure 8}. Pn2 is a branch node and will therefore execute the same procedure as the branch node N on the primary LSP. Pn2 discards label Lbdla received from the PATH message, assigns a new label Lbu and sends it to MP1 via a RESV message.

Once a RESV message is originated by PLR and sent through the backup route, MP1 will get an inner label Lpu that represents the primary LSP in this example. MP1 adds a FRR entry with both inner and outer label in its FIB. MP1 FIB will have two forwarding entries for the LSP being protected in Facility mode:

Lpu->Lp-pe1, PE1

FRR: Lpu, Lbu->Lp-pe2, PE2

With the same process, MP2 will have two forwarding entries for the LSP being protected:

Lpu->Lp-pe2, PE2

FRR: Lpu, Lbu->Lp-pe2, PE2

2. Label Allocation and FRR Entry on Pn2 and Pn1:

As mentioned in the last paragraph, when Pn2 (transit branch node) receives PATH message from MP1 and MP2 respectively, it will allocate label Lbu and sends it to each MP. Pn2 will have two forwarding entries for the LSP being protected:

FRR: Lpu, Lb5->Lbu, MP1

FRR: Lpu, Lb5->Lbu, MP2

Pn1 is a transit node and has only one FRR entry for the LSP being protected:

FRR: Lpu, Lb4->Lb5, Pn2

3. Label Allocation and FRR Entry on PLR:

PLR receives a PATH message from node N that carries a downstream label Lpu and a PATH message from Pn1 that carries a downstream label Lb5. PLR uses Lpu as an inner label for the primary LSP and sends it towards MPs through Pn1 by means of RESV message. PLR will maintain two entries in its FIB for a goiven protected LSP:

Lp1->Lpu, N

FRR: Lpu, Lp1->Lb1, Pn1

For every add-in primary LSP being protected by the same backup LSP, PLR will assign an inner label and send it to LSRs across the backup LSP so that each LSR can add the corresponding FRR entry in its FIB and use this entry to forward traffic over the backup LSP.

4.2.3. Node Failure Repair and Packet Encapsulation

Once protected node N fails and the failure is detected by PLR, it will initiate a switchover by redirecting traffic to the backup LSP. Packet encapsulation in each LSR over the backup LSP will be done based on the FRR entries of its FIB. For example (Figure 8), a packet that arrives at PLR and which is supposed to be forwarded to node N by using entry 'Lp1->Lpu, N', will be redirected to Pn1 based on entry 'FRR: Lpu,Lp1->Lb4, Pn1'. PLR encapsulates the packet with Lpu as inner label, Lb4 as outer label and forwards it to Pn1. Pn1 will swap outer label for packet forwarding and keep inner label unchanged.

Once the packet reaches MP1, MP1 will pop out the outer label, swap the inner label with outgoing label Lp-pe1 and forward the packet to NHOP PE1 with a single label Lp-pe1, the packet de-capsulation/ encapsulation is based on the 'FRR: Lpu, Lbu->Lp-pe1, PE1' entry. Once traffic reaches MP1, it is then merged with the primary path. The same procedure is applicable to receiver LSR MP2.

4.2.4. Re-convergence after Local Repair

Routers that do not belong to the FRR domain are not impacted by the link failure and local repair procedures. However, the network will eventually re-converge and a new best path to reach the root of the RD P2MP tree structure will be computed by PE1 and PE2 (Figure 8). PLR will be notified as soon as the new primary path is determined. PLR will send notification message to Pn and MP sp that they tear down the detour LSP and withdraw backup labels. There is no difference between facility and detour methods in terms of re-convergence process.

5. IANA Considerations

TBD.

6. Manageability Considerations

TBD.

7. Security Considerations

TBD.

8. Acknowledgements

We would like to thank Quintin Zhao, Lin Han, Emily Chen, and Robert Tao for discussions and comments.

9. References

9.1. Normative References

[I-D.lzj-mpls-receiver-driven-multicast-rsvp-te] Li, R., Zhao, Q., and C. Jacquenet, "Receiver-Driven Multicast Traffic Engineered Label Switched Paths", draft-lzj-mpls-receiver-driven-multicast-rsvp-te-00 (work in progress), March 2012.

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

[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.

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

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

9.2. Informative References

[RFC3468] Andersson, L. and G. Swallow, "The Multiprotocol Label Switching (MPLS) Working Group decision on MPLS signaling protocols", RFC 3468, February 2003.

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

[RFC3564] Le Faucheur, F. and W. Lai, "Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering", RFC 3564, July 2003.

Authors' Addresses

Katherine Zhao
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: katherine.zhao@huawei.com

Renwei Li
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA

Email: renwei.li@huawei.com

Christian Jacquenet
France Telecom Orange
4 rue du Clos Courtel
35512 Cession Sevigne,
France

Email: christian.jacquenet@orange.com