Network Working Group   Chandra Ramachandran
Internet Draft   Juniper Networks
Intended status: Standards Track   Ina Minei
                                                        Google, Inc
                                                     Dante Pacella
                                                           Verizon
                                                        Tarek Saad
                                               Cisco Systems Inc.

Expires: August 12, 2017   February 12, 2017

Refresh Interval Independent FRR Facility Protection
draft-ietf-mpls-ri-rsvp-frr-01

Status of this Memo

Copyright Notice

Abstract

   RSVP-TE relies on periodic refresh of RSVP messages to synchronize
   and maintain the LSP related states along the reserved path. In the
   absence of refresh messages, the LSP related states are
   automatically deleted. Reliance on periodic refreshes and refresh
   timeouts are problematic from the scalability point of view. The
   number of RSVP-TE LSPs that a router needs to maintain has been
   growing in service provider networks and the implementations should
   be capable of handling increase in LSP scale.

   RFC 2961 specifies mechanisms to eliminate the reliance on periodic
   refresh and refresh timeout of RSVP messages, and enables a router
   to increase the message refresh interval to values much larger than
   the default 30 seconds defined in RFC 2205. However, the protocol
   extensions defined in RFC 4090 for supporting fast reroute (FRR)
   using bypass tunnels implicitly rely on short refresh timeouts to
   cleanup stale states.

   In order to eliminate the reliance on refresh timeouts, the routers
   should unambiguously determine when a particular LSP state should be
   deleted. Coupling LSP state with the corresponding RSVP-TE signaling
   adjacencies as recommended in RSVP-TE Scaling Recommendations
   (draft-ietf-teas-rsvp-te-scaling-rec) will apply in scenarios other
   than RFC 4090 FRR using bypass tunnels. In scenarios involving RFC
   4090 FRR using bypass tunnels, additional explicit tear down
   messages are necessary. Refresh-interval Independent RSVP FRR (RI-
   RSVP-FRR) extensions specified in this document consists of
   procedures to enable LSP state cleanup that are essential in
   scenarios not covered by procedures defined in RSVP-TE Scaling
   Recommendations.

Requirements Language

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
   document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

1. Introduction

   RSVP-TE Fast Reroute [RFC4090] defines two local repair techniques
   to reroute label switched path (LSP) traffic over pre-established
   backup tunnel. Facility backup method allows one or more LSPs
   traversing a connected link or node to be protected using a bypass
   tunnel. The many-to-one nature of local repair technique is
   attractive from scalability point of view. This document enumerates
   facility backup procedures in RFC 4090 that rely on refresh timeout

and hence make facility backup method refresh-interval dependent.
The RSVP-TE extensions defined in this document will enhance the
facility backup protection mechanism by making the corresponding
procedures refresh-interval independent.

1.1. Motivation

Standard RSVP [RFC2205] maintains state via the generation of RSVP
Path/Resv refresh messages. Refresh messages are used to both
synchronize state between RSVP neighbors and to recover from lost
RSVP messages. The use of Refresh messages to cover many possible
failures has resulted in a number of operational problems.

  - One problem relates to RSVP control plane scaling due to periodic
    refreshes of Path and Resv messages, another relates to the
    reliability and latency of RSVP signaling.

  - An additional problem is the time to clean up the stale state
    after a tear message is lost. For more on these problems see
    Section 1 of RSVP Refresh Overhead Reduction Extensions
    [RFC2961].

The problems listed above adversely affect RSVP control plane
scalability and RSVP-TE [RFC3209] inherited these problems from
standard RSVP. Procedures specified in [RFC2961] address the above
mentioned problems by eliminating dependency on refreshes for state
synchronization and for recovering from lost RSVP messages, and by
eliminating dependency on refresh timeout for stale state cleanup.
Implementing these procedures allows to improve RSVP-TE control
plane scalability. For more details on eliminating dependency on
refresh timeout for stale state cleanup, refer to "Refresh Interval
Independent RSVP" section in [TE-SCALE-REC].

However, the procedures specified in [RFC2961] do not fully address
stale state cleanup for facility backup protection [RFC4090], as
facility backup protection still depends on refresh timeouts for
stale state cleanup. Thus [RFC2961] is insufficient to address the
problem of stale state cleanup when facility backup protection is
used.

The procedures specified in this document, in combination with
[RFC2961], eliminate facility backup protection dependency on
refresh timeouts for stale state cleanup. These procedures, in
combination with [RFC2961], fully address the above mentioned
problem of RSVP-TE stale state cleanup, including the cleanup for
facility backup protection.

The procedures specified in this document assume reliable delivery
of RSVP messages, as specified in [RFC2961]. Therefore this document
makes support for [RFC2961] a pre-requisite.

2. Terminology

The reader is assumed to be familiar with the terminology in
[RFC2205], [RFC3209], [RFC4090] and [RFC4558].

Phop node: Previous-hop router along the label switched path

PPhop node: Previous-Previous-hop router along the LSP

LP-MP node: Merge Point router at the tail of Link-protecting bypass
tunnel

NP-MP node: Merger Point router at the tail of Node-protecting
bypass tunnel

TED: Traffic Engineering Database

Conditional PathTear: PathTear message containing a suggestion to a
receiving downstream router to retain Path state if the receiving
router is NP-MP

Remote PathTear: PathTear message sent from Point of Local Repair
(PLR) to MP to delete state on MP if PLR had not reliably sent
backup Path state before

3. Problem Description

```
                          E
                        /    \
                       /       \
                      /          \
                     /             \
                    /                \
                   /                   \
                  A ----- B ----- C ----- D
                           \             /
```
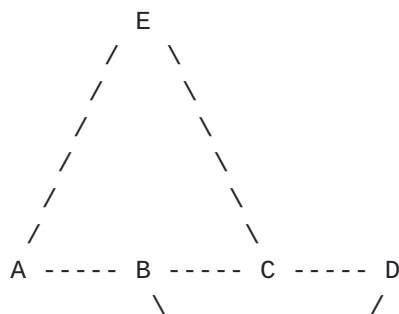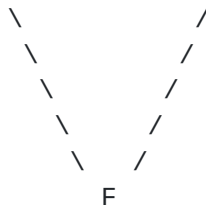
```
                           \           /
                            \         /
                             \       /
                              \     /
                               \   /
                                F
```
                      Figure 1: Example Topology

   In the topology in Figure 1, consider a large number of LSPs from A
   to D transiting B and C. Assume that refresh interval has been
   configured to be large of the order of minutes and refresh reduction
   extensions are enabled on all routers.

   Also assume that node protection has been configured for the LSPs
   and the LSPs are protected by each router in the following way

   - A has made node protection available using bypass LSP A -> E ->
      C; A is the Point of Local Repair (PLR) and C is Node Protecting
      Merge Point (NP-MP)

   - B has made node protection available using bypass LSP B -> F ->
      D; B is the PLR and D is the NP-MP

   - C has made link protection available using bypass LSP C -> B -> F
      -> D; C is the PLR and D is the Link Protecting Merge Point (LP-
      MP)

   In the above condition, assume that B-C link fails. The following is
   the sequence of events that is expected to occur for all protected
   LSPs under normal conditions.

   1. B performs local repair and re-directs LSP traffic over the bypass
      LSP B -> F -> D.
   2. B also creates backup state for the LSP and triggers sending of
      backup LSP state to D over the bypass LSP B -> F -> D.
   3. D receives backup LSP states and merges the backups with the
      protected LSPs.
   4. As the link on C, over which the LSP states are refreshed has
      failed, C will no longer receive state refreshes. Consequently the
      protected LSP states on C will time out and C will send tear down
      message for all LSPs. As each router should consider itself as a

   Merge Point, C will time out the state only after waiting for an
   additional duration equal to refresh timeout.
   While the above sequence of events has been described in [RFC4090],
   there are a few problems for which no mechanism has been specified
   explicitly.

   - If the protected LSP on C times out before D receives signaling
      for the backup LSP, then D would receive PathTear from C prior to
      receiving signaling for the backup LSP, thus resulting in deleting
      the LSP state. This would be possible at scale even with default
      refresh time.

   - If upon the link failure C is to keep state until its timeout,
      then with long refresh interval this may result in a large amount
      of stale state on C. Alternatively, if upon the link failure C is
      to delete the state and send PathTear to D, this would result in
      deleting the state on D, thus deleting the LSP. D needs a reliable
      mechanism to determine whether it is MP or not to overcome this
      problem.

   - If head-end A attempts to tear down LSP after step 1 but before
      step 2 of the above sequence, then B may receive the tear down
      message before step 2 and delete the LSP state from its state
      database. If B deletes its state without informing D, with long
      refresh interval this could cause (large) buildup of stale state
      on D.

   - If B fails to perform local repair in step 1, then B will delete
      the LSP state from its state database without informing D. As B
      deletes its state without informing D, with long refresh interval
      this could cause (large) buildup of stale state on D.

   The purpose of this document is to provide solutions to the above
   problems which will then make it practical to scale up to a large
   number of protected LSPs in the network.

4. Solution Aspects

   The solution consists of five parts.

   - Utilize MP determination mechanism specified in [SUMMARY-FRR]
      that enables the PLR to signal availability of local protection to
      MP. In addition, introduce PLR and MP procedures to establish
      Node-ID hello session between the PLR and the MP to detect router
      failures and to determine capability. See section 4.1 for more

   details. This part of the solution re-uses some of the extensions
   defined in [SUMMARY-FRR] and [TE-SCALE-REC], and the subsequent
   sub-sections will list the extensions in these drafts that are
   utilized in this document.

   - Handle upstream link or node failures by cleaning up LSP states
      if the node has not found itself as MP through the MP
      determination mechanism. See section 4.2 for more details.

      The combination of "path state" maintained as Path State Block
      (PSB) and "reservation state" maintained as Reservation State
      Block (RSB) forms an individual LSP state on an RSVP-TE speaker.

   - Introduce extensions to enable a router to send tear down message
      to downstream router that enables the receiving router to
      conditionally delete its local state. See section 4.3 for more
      details.

   - Enhance facility protection by allowing a PLR to directly send
      tear down message to MP without requiring the PLR to either have a
      working bypass LSP or have already signaled backup LSP state. See
      section 4.4 for more details.

   - Introduce extensions to enable the above procedures to be
      backward compatible with routers along the LSP path running
      implementation that do not support these procedures. See section
      4.5 for more details.

4.1. **Signaling Handshake between PLR and MP**

4.1.1. **PLR Behavior**

   As per the procedures specified in RFC 4090, when a protected LSP
   comes up and if the "local protection desired" flag is set in the
   SESSION_ATTRIBUTE object, each node along the LSP path attempts to
   make local protection available for the LSP.

      - If the "node protection desired" flag is set, then the node
      tries to become a PLR by attempting to create a NP-bypass LSP to
      the NNhop node avoiding the Nhop node on protected LSP path. In
      case node protection could not be made available after some time
      out, the node attempts to create a LP-bypass LSP to Nhop node
      avoiding only the link that protected LSP takes to reach Nhop

- If the "node protection desired" flag is not set, then the PLR
  attempts to create a LP-bypass LSP to Nhop node avoiding the link
  that the protected LSP takes to reach Nhop

With regard to the PLR procedures described above and that are
specified in RFC 4090, this document specifies the following
additional procedures.

- While selecting the destination address of the bypass LSP, the
   PLR SHOULD attempt to select the router ID of the NNhop or Nhop
   node. If the PLR and the MP are in same area, then the PLR may
   utilize the TED to determine the router ID from the interface
   address in RRO (if NodeID is not included in RRO). If the PLR and
   the MP are in different IGP areas, then the PLR SHOULD use the
   NodeID address of NNhop MP if included in the RRO of RESV. If the
   NP-MP in a different area has not included NodeID in RRO, then the
   PLR SHOULD use NP-MP's interface address present in the RRO. The
   PLR SHOULD use its router ID as the source address of the bypass
   LSP. The PLR SHOULD also include its router ID in a NodeID sub-
   object in PATH RRO unless configured explicitly not to include
   NodeID. While including its router ID in NodeID sub-object, the
   PLR SHOULD include the NodeID sub-object after including its
   IPv4/IPv6 address or unnumbered interface ID sub-object.

- In parallel to the attempt made to create NP-bypass or LP-bypass,
   the PLR SHOULD initiate a Node-ID based Hello session to the NNhop
   or Nhop node respectively to establish the RSVP-TE signaling
   adjacency. This Hello session is used to detect MP node failure as
   well as determine the capability of the MP node. If the MP sets I-
   bit in CAPABILITY object [TE-SCALE-REC] carried in Hello message
   corresponding to NodeID based Hello session, then the PLR SHOULD
   conclude that the MP supports refresh-interval independent FRR
   procedures defined in this document.

- If the bypass LSP comes up, then the PLR SHOULD include Bypass
   Summary FRR Extended Association object and triggers PATH to be
   sent. If Bypass Summary FRR Extended Association object is
   included in PATH message, then the encoding rules specified in
   [SUMMARY-FRR] MUST be followed.

4.1.2. Remote Signaling Adjacency

A NodeID based RSVP-TE Hello session is one in which NodeID is used
in source and destination address fields in RSVP Hello. [RFC4558]
formalizes NodeID based Hello messages between two routers. This
document extends NodeID based RSVP Hello session to track the state
of RSVP-TE neighbor that is not directly connected by at least one
interface. In order to apply NodeID based RSVP-TE Hello session
between any two routers that are not immediate neighbors, the router
that supports the extensions defined in the document SHOULD set TTL
to 255 in the NodeID based Hello messages exchanged between PLR and
MP. The default hello interval for this NodeID hello session SHOULD
be set to the default specified in [TE-SCALE-REC].

In the rest of the document the term "signaling adjacency", or
"remote signaling adjacency" refers specifically to the RSVP-TE
signaling adjacency.

4.1.3. MP Behavior

When the NNhop or Nhop node receives the triggered PATH with a
"matching" Bypass Summary FRR Extended Association object, the node
should consider itself as the MP for the PLR IP address
"corresponding" to the Bypass Summary FRR Extended Association
object. The matching and ordering rules of Bypass Summary FRR
Extended Association specified in [SUMMARY-FRR] SHOULD be followed
by implementations supporting this document.

In addition to the above procedures, the node SHOULD check the
presence of remote signaling adjacency with PLR (this check is
needed to detect network being partitioned). If a matching Bypass
Summary FRR Extended Association object is found in PATH and the
RSVP-TE signaling adjacency is present, the node concludes that the
PLR will undertake refresh-interval independent FRR procedures
specified in this document. If the PLR has included NodeID in PATH
RRO, then that NodeID is the remote neighbor address. Otherwise, the
PLR's interface address in RRO will be the remote neighbor address.
If a matching Bypass Summary FRR Extended Association object is
included by PPhop node, then it is NP-MP. If a matching Bypass
Summary FRR Extended Association object is included by Phop node, it
concludes it is LP-MP.

4.1.4. "Remote" state on MP

Once a router concludes it is MP for a PLR running refresh-interval
independent FRR procedures, it SHOULD create a remote path state for

the LSP. The "remote" state is identical to the protected LSP path
state except for the difference in RSVP_HOP object. The RSVP_HOP
object in "remote" Path state contains the address that the PLR uses
to send NodeID hello messages to MP.

The MP SHOULD consider the "remote" path state automatically deleted
if:

- MP later receives a PATH with no matching Bypass Summary FRR
  Extended Association object corresponding to the PLR RRO, or

- Node signaling adjacency with PLR goes down, or

- MP receives backup LSP signaling from PLR or

- MP receives PathTear, or

- MP deletes the LSP state on local policy or exception event

Unlike the normal path state that is either locally generated on
Ingress or created from PATH message from Phop node, the "remote"
path state is not signaled explicitly from PLR. The purpose of
"remote" path state is to enable the PLR to explicitly tear down
path and reservation states corresponding to the LSP by sending tear
message for the "remote" path state. Such message tearing down
"remote" path state is called "Remote PathTear.

The scenarios in which "Remote" PathTear is applied are described in
Section 4.4 - Remote State Teardown.

4.2. Impact of Failures on LSP State

This section describes the procedures for routers on the LSP path
for different kinds of failures. The procedures described on
detecting RSVP control plane adjacency failures do not impact the
RSVP-TE graceful restart mechanisms ([RFC3473], [RFC5063]). If the
router executing these procedures act as helper for neighboring
router, then the control plane adjacency will be declared as having
failed after taking into account the grace period extended for
neighbor by the helper.

Immediate node failures are detected from the state of NodeID hello
sessions established with immediate neighbors. [TE-SCALE-REC]
recommends each router to establish NodeID hello sessions with all
its immediate neighbors. PLR or MP node failure is detected from the

state of remote signaling adjacency established according to [Section 4.1.2](#) of this document.

[4.2.1](#). Non-MP Behavior

When a router detects Phop link or Phop node failure and the router is not an MP for the LSP, then it SHOULD send Conditional PathTear (refer to Section "Conditional PathTear" below) and delete PSB and RSB states corresponding to the LSP.

[4.2.2](#). LP-MP Behavior

When the Phop link for an LSP fails on a router that is LP-MP for the LSP, the LP-MP SHOULD retain PSB and RSB states corresponding to the LSP till the occurrence of any of the following events.

- Node-ID signaling adjacency with Phop PLR goes down, or

- MP receives normal or "Remote" PathTear for PSB, or

- MP receives ResvTear RSB.

When a router that is LP-MP for an LSP detects Phop node failure from Node-ID signaling adjacency state, the LP-MP SHOULD send normal PathTear and delete PSB and RSB states corresponding to the LSP.

[4.2.3](#). NP-MP Behavior

When a router that is NP-MP for an LSP detects Phop link failure, or Phop node failure from Node-ID signaling adjacency, the router SHOULD retain PSB and RSB states corresponding to the LSP till the occurrence of any of the following events.

- Remote Node-ID signaling adjacency with PPhop PLR goes down, or

- MP receives normal or "Remote" PathTear for PSB, or

- MP receives ResvTear for RSB.

When a router that is NP-MP does not detect Phop link or node failure, but receives Conditional PathTear from the Phop node, then the router SHOULD retain PSB and RSB states corresponding to the LSP till the occurrence of any of the following events.

- Remote Node-ID signaling adjacency with PPhop PLR goes down, or

- MP receives normal or "Remote" PathTear for PSB, or

- MP receives ResvTear for RSB.

Receiving Conditional PathTear from the Phop node will not impact
the "remote" state from the PLR. Note that Phop node would send
Conditional PathTear if it was not an MP.

In the example topology in Figure 1, assume C & D are NP-MP for PLRs
A & B respectively. Now when A-B link fails, as B is not MP and its
Phop link signaling adjacency has failed, B will delete LSP state
(this behavior is required for unprotected LSPs - Section 4.2.1). In
the data plane, that would require B to delete the label forwarding
entry corresponding to the LSP. So if B's downstream nodes C and D
continue to retain state, it would not be correct for D to continue
to assume itself as NP-MP for PLR B.

The mechanism that enables D to stop considering itself as NP-MP and
delete "remote" path state is given below.

    1. When C receives Conditional PathTear from B, it decides to
    retain LSP state as it is NP-MP of PLR A. C also SHOULD check
    whether Phop B had previously signaled availability of node
    protection. As B had previously signaled NP availability in its
    PATH RRO, C SHOULD remove Bypass Summary FRR Extended
    Association object containing Association Source set to B from
    the PATH message and trigger PATH to D.
    2. When D receives triggered PATH, it realizes that it is no
    longer NP-MP and so deletes the "remote" path state. D does not
    propagate PATH further down because the only change is that the
    Bypass Summary FRR Extended Association object corresponding to
    Association Source B is no longer present in the PATH message.

4.2.4. Behavior of a Router that is both LP-MP and NP-MP

A router may be both LP-MP as well as NP-MP at the same time for
Phop and PPhop nodes respectively of an LSP. If Phop link fails on
such node, the node SHOULD retain PSB and RSB states corresponding
to the LSP till the occurrence of any of the following events.

- Both Node-ID signaling adjacencies with Phop and PPhop nodes go
down, or

- MP receives normal or "Remote" PathTear for PSB, or

- MP receives ResvTear for RSB.

If a router that is both LP-MP and NP-MP detects Phop node failure, then the node SHOULD retain PSB and RSB states corresponding to the LSP till the occurrence of any of the following events.

- Remote Node-ID signaling adjacency with PPhop PLR goes down, or

- MP receives normal or "Remote" PathTear for PSB, or

- MP receives ResvTear for RSB.

4.3. Conditional Path Tear

In the example provided in the Section 4.2.5 "NP-MP Behavior on PLR link failure", B deletes PSB and RSB states corresponding to the LSP once B detects its link to Phop went down as B is not MP. If B were to send PathTear normally, then C would delete LSP state immediately. In order to avoid this, there should be some mechanism by which B can indicate to C that B does not require the receiving node to unconditionally delete the LSP state immediately. For this, B SHOULD add a new optional object called CONDITIONS object in PathTear. The new optional object is defined in Section 4.3.3. If node C also understands the new object, then C SHOULD delete LSP state only if it is not an NP-MP - in other words C SHOULD delete LSP state if there is no "remote" PLR state on C.

4.3.1. Sending Conditional Path Tear

A router that is not an MP for an LSP SHOULD delete PSB and RSB states corresponding to the LSP if Phop link or Phop Node-ID signaling adjacency goes down (Section 4.2.1). The router SHOULD send Conditional PathTear if the following are also true.

- Ingress has requested node protection for the LSP, and

- PathTear is not received from upstream node

4.3.2. Processing Conditional Path Tear

When a router that is not an NP-MP receives Conditional PathTear, the node SHOULD delete PSB and RSB states corresponding to the LSP, and process Conditional PathTear by considering it as normal PathTear. Specifically, the node SHOULD NOT propagate Conditional PathTear downstream but remove the optional object and send normal PathTear downstream.

When a node that is an NP-MP receives Conditional PathTear, it
SHOULD NOT delete LSP state. The node SHOULD check whether the Phop
node had previously included Bypass Summary FRR Extended Association
object in PATH. If the object had been included previously by Phop,
then the node processing Conditional PathTear from Phop SHOULD
remove the corresponding object and trigger PATH downstream.

If Conditional PathTear is received from a neighbor that has not
advertised support (refer to Section 4.5) for the new procedures
defined in this document, then the node SHOULD consider the message
as normal PathTear. The node SHOULD propagate normal PathTear
downstream and delete LSP state.

4.3.3. CONDITIONS object

As any implementation that does not support Conditional PathTear
SHOULD ignore the new object but process the message as normal
PathTear without generating any error, the Class-Num of the new
object SHOULD be 10bbbbbb where 'b' represents a bit (from Section
3.10 of [RFC2205]).

The new object is called as "CONDITIONS" object that will specify
the conditions under which default processing rules of the RSVP-TE
message SHOULD be invoked.

The object has the following format:

```
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           Length              |    Class     |    C-type     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                     Reserved                               |M|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Length

This contains the size of the object in bytes and should be set to
eight.

Class

To be assigned

C-type

1

M bit

If M-bit is set to 1, then the PathTear message SHOULD be processed
based on the condition if the receiver router is a Merge Point or
not.

If M-bit is set to 0, then the PathTear message SHOULD be processed
as normal PathTear message.

4.4. Remote State Teardown

If the Ingress wants to tear down the LSP because of a management
event while the LSP is being locally repaired at a transit PLR, it
would not be desirable to wait till backup LSP signaling to perform
state cleanup. To enable LSP state cleanup when the LSP is being
locally repaired, the PLR SHOULD send "remote" PathTear message
instructing the MP to delete PSB and RSB states corresponding to the
LSP. The TTL in "remote" PathTear message SHOULD be set to 255.

Consider node C in example topology (Figure 1) has gone down and B
locally repairs the LSP.

1. Ingress A receives a management event to tear down the LSP.
2. A sends normal PathTear to B.
3. To enable LSP state cleanup, B SHOULD send "remote" PathTear with
   destination IP address set to that of D used in Node-ID signaling
   adjacency with D, and RSVP_HOP object containing local address
   used in Node-ID signaling adjacency.
4. B then deletes PSB and RSB states corresponding to the LSP.
5. On D there would be a remote signaling adjacency with B and so D
   SHOULD accept the remote PathTear and delete PSB and RSB states
   corresponding to the LSP.

4.4.1. PLR Behavior on Local Repair Failure

If local repair fails on the PLR after a failure, then this should
be considered as a case for cleaning up LSP state from PLR to the
Egress. PLR would achieve this using "remote" PathTear to clean up
state from MP. If MP has retained state, then it would propagate
PathTear downstream thereby achieving state cleanup. Note that in

the case of link protection, the PathTear would be directed to LP-MP
node IP address rather than the Nhop interface address.

4.4.2. PLR Behavior on Resv RRO Change

When a router that has already made NP available detects a change in
the RRO carried in RESV message, and if the RRO change indicates
that the router's former NP-MP is no longer present in the LSP path,
then the router SHOULD send "Remote" PathTear directly to its former
NP-MP.

In the example topology in Figure 1, assume A has made node
protection available and C has concluded it is NP-MP. When the B-C
link fails then implementing the procedure specified in Section
4.2.4 of this document, C will retain state till: remote NodeID
control plane adjacency with A goes down, or PathTear or ResvTear is
received for PSB or RSB respectively. If B also has made node
protection available, B will eventually complete backup LSP
signaling with its NP-MP D and trigger RESV to A with RRO changed.
The new RRO of the LSP carried in RESV will not contain C. When A
processes the RESV with a new RRO not containing C - its former NP-
MP, A SHOULD send "Remote" PathTear to C. When C receives a "Remote"
PathTear for its PSB state, C will send normal PathTear downstream
to D and delete both PSB and RSB states corresponding to the LSP. As
D has already received backup LSP signaling from B, D will retain
control plane and forwarding states corresponding to the LSP.

4.4.3. LSP Preemption during Local Repair

If an LSP is preempted when there is no failure along the path of
the LSP, the node on which preemption occurs would send PathErr and
ResvTear upstream and only delete the forwarding state and RSB state
corresponding to the LSP. But if the LSP is being locally repaired
upstream of the node on which the LSP is preempted, then the node
SHOULD delete both PSB and RSB states corresponding to the LSP and
send normal PathTear downstream.

4.4.3.1. Preemption on LP-MP after Phop Link failure

If an LSP is preempted on LP-MP after its Phop or incoming link has
already failed but the backup LSP has not been signaled yet, then
the node SHOULD send normal PathTear and delete both PSB and RSB
states corresponding to the LSP. As the LP-MP has retained LSP state
because the PLR would signal the LSP through backup LSP signaling,
preemption would bring down the LSP and the node would not be LP-MP
any more requiring the node to clean up LSP state.

4.4.3.2. **Preemption on NP-MP after Phop Link failure**

If an LSP is preempted on NP-MP after its Phop link has already
failed but the backup LSP has not been signaled yet, then the node
SHOULD send normal PathTear and delete PSB and RSB states
corresponding to the LSP. As the NP-MP has retained LSP state
because the PLR would signal the LSP through backup LSP signaling,
preemption would bring down the LSP and the node would not be NP-MP
any more requiring the node to clean up LSP state.

Consider B-C link goes down on the same example topology (Figure 1).
As C is NP-MP for PLR A, C will retain LSP state.

1. The LSP is preempted on C.
2. C will delete RSB state corresponding to the LSP. But C cannot
   send PathErr or ResvTear to PLR A because backup LSP has not
   been signaled yet.
3. As the only reason for C having retained state after Phop node
   failure was that it was NP-MP, C SHOULD send normal PathTear to
   D and delete PSB state also. D would also delete PSB and RSB
   states on receiving PathTear from C.
4. B starts backup LSP signaling to D. But as D does not have the
   LSP state, it will reject backup LSP PATH and send PathErr to B.
5. B will delete its reservation and send ResvTear to A.

4.5. **Backward Compatibility Procedures**

The "Refresh interval Independent FRR" or RI-RSVP-FRR referred below
in this section refers to the changes that have been proposed in
previous sections. Any implementation that does not support them has
been termed as "non-RI-RSVP-FRR implementation". The extensions
proposed in [SUMMARY-FRR] are applicable to implementations that do
not support RI-RSVP-FRR. On the other hand, changes proposed
relating to LSP state cleanup namely Conditional and remote PathTear
require support from one-hop and two-hop neighboring nodes along the
LSP path. So procedures that fall under LSP state cleanup category
SHOULD be turned on only if all nodes involved in the node
protection FRR i.e. PLR, MP and intermediate node in the case of NP,
support the extensions. Note that for LSPs requesting only link
protection, the PLR and the LP-MP should support the extensions.

4.5.1. **Detecting Support for Refresh interval Independent FRR**

   An implementation supporting the extensions specified in previous
   sections (called RI-RSVP-FRR here after) SHOULD set the flag
   "Refresh interval Independent RSVP" or RI-RSVP in CAPABILITY object
   in Hello messages. The RI-RSVP flag is specified in [TE-SCALE-REC].

   - As nodes supporting the extensions SHOULD initiate Node Hellos
      with adjacent nodes, a node on the path of protected LSP can
      determine whether its Phop or Nhop neighbor supports RI-RSVP-FRR
      enhancements from the Hello messages sent by the neighbor.

   - If a node attempts to make node protection available, then the
      PLR SHOULD initiate remote Node-ID signaling adjacency with NNhop.
      If the NNhop (a) does not reply to remote node Hello message or
      (b) does not set "Enhanced facility protection" flag in CAPABILITY
      object in the reply, then the PLR can conclude that NNhop does not
      support RI-RSVP-FRR extensions.

   - If node protection is requested for an LSP and if (a) PPhop node
      has not included a matching Bypass Summary FRR Extended
      Association object in PATH or (b) PPhop node has not initiated
      remote node Hello messages, then the node SHOULD conclude that PLR
      does not support RI-RSVP-FRR extensions. The details are described
      in the "Procedures for backward compatibility" section below.

   Any node that sets the I-bit is set in its CAPABILITY object MUST
   also set Refresh-Reduction-Capable bit in common header of all RSVP-
   TE messages.

4.5.2. **Procedures for backward compatibility**

   The procedures defined hereafter are performed on a subset of LSPs
   that traverse a node, rather than on all LSPs that traverse a node.
   This behavior is required to support backward compatibility for a
   subset of LSPs traversing nodes running non-RI-RSVP-FRR
   implementations.

4.5.2.1. **Lack of support on Downstream Node**

   - If the Nhop does not support the RI-RSVP-FRR extensions, then the
      node SHOULD reduce the "refresh period" in TIME_VALUES object
      carried in PATH to default small refresh default value.

   - If node protection is requested and the NNhop node does not
      support the enhancements, then the node SHOULD reduce the "refresh

   period" in TIME_VALUES object carried in PATH to a small refresh
   default value.

   If the node reduces the refresh time from the above procedures, it
   SHOULD also not send remote PathTear or Conditional PathTear
   messages.

   Consider the example topology in Figure 1. If C does not support the
   RI-RSVP-FRR extensions, then:

   - A and B SHOULD reduce the refresh time to default value of 30
      seconds and trigger PATH

   - If B is not an MP and if Phop link of B fails, B cannot send
      Conditional PathTear to C but SHOULD time out PSB state from A
      normally. This would be accomplished if A would also reduce the
      refresh time to default value. So if C does not support the RI-
      RSVP-FRR extensions, then Phop B and PPhop A SHOULD reduce refresh
      time to a small default value.

4.5.2.2. Lack of support on Upstream Node

   - If Phop node does not support the RI-RSVP-FRR extensions, then
      the node SHOULD reduce the "refresh period" in TIME_VALUES object
      carried in RESV to default small refresh time value.

   - If node protection is requested and the Phop node does not
      support the RI-RSVP-FRR extensions, then the node SHOULD reduce
      the "refresh period" in TIME_VALUES object carried in PATH to
      default value.

   - If node protection is requested and PPhop node does not support
      the RI-RSVP-FRR extensions, then the node SHOULD reduce the
      "refresh period" in TIME_VALUES object carried in RESV to default
      value.

   - If the node reduces the refresh time from the above procedures,
      it SHOULD also not execute MP procedures specified in Section 4.2
      of this document.

4.5.2.3. Incremental Deployment

   The backward compatibility procedures described in the previous sub-
   sections imply that a router supporting the RI-RSVP-FRR extensions
   specified in this document can apply the procedures specified in the
   document either in the downstream or upstream direction of an LSP,

depending on the capability of the routers downstream or upstream in
the LSP path.

- RI-RSVP-FRR extensions and procedures are enabled for downstream
  Path, PathTear and ResvErr messages corresponding to an LSP if
  link protection is requested for the LSP and the Nhop node
  supports the extensions

- RI-RSVP-FRR extensions and procedures are enabled for downstream
  Path, PathTear and ResvErr messages corresponding to an LSP if
  node protection is requested for the LSP and both Nhop & NNhop
  nodes support the extensions

- RI-RSVP-FRR extensions and procedures are enabled for upstream
  PathErr, Resv and ResvTear messages corresponding to an LSP if
  link protection is requested for the LSP and the Phop node
  supports the extensions

- RI-RSVP-FRR extensions and procedures are enabled for upstream
  PathErr, Resv and ResvTear messages corresponding to an LSP if
  node protection is requested for the LSP and both Phop and PPhop
  nodes support the extensions

For example, if an implementation supporting the RI-RSVP-FRR
extensions specified in this document is deployed on all routers in
particular region of the network and if all the LSPs in the network
request node protection, then the FRR extensions will only be
applied for the LSP segments that traverse the particular region.
This will aid incremental deployment of these extensions and also
allow reaping the benefits of the extensions in portions of the
network where it is supported.

5. **Security Considerations**

This security considerations pertaining to [RFC2205], [RFC3209] and
[RFC5920] remain relevant.

This document extends the applicability of Node-ID based Hello
session between immediate neighbors. The Node-ID based Hello session
between PLR and NP-MP may require the two routers to exchange Hello
messages with non-immediate neighbor. So, the implementations SHOULD
provide the option to configure Node-ID neighbor specific or global
authentication key to authentication messages received from Node-ID
neighbors. The network administrator MAY utilize this option to
enable RSVP-TE routers to authenticate Node-ID Hello messages
received with TTL greater than 1. Implementations SHOULD also

   provide the option to specify a limit on the number of Node-ID based
   Hello sessions that can be established on a router supporting the
   extensions defined in this document.

6. IANA Considerations

6.1. New Object - CONDITIONS

   RSVP Change Guidelines [RFC3936] defines the Class-Number name space
   for RSVP objects. The name space is managed by IANA.

   IANA registry: RSVP Parameters
   Subsection: Class Names, Class Numbers, and Class Types

   A new RSVP object using a Class-Number from 128-183 range called the
   "CONDITIONS" object is defined in Section 4.3 of this document. The
   Class-Number from 128-183 range will be allocated by IANA.

7. Normative References

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

   [RFC3209] Awduche, D., "RSVP-TE: Extensions to RSVP for LSP
             Tunnels", RFC 3209, December 2001.

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

   [RFC2961] Berger, L., "RSVP Refresh Overhead Reduction Extensions",
             RFC 2961, April 2001.

   [RFC2205] Braden, R., "Resource Reservation Protocol (RSVP)", RFC
             2205, September 1997.

   [RFC4558] Ali, Z., "Node-ID Based Resource Reservation (RSVP) Hello:
             A Clarification Statement", RFC 4558, June 2006.

   [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching
             Signaling Resource Reservation Protocol-Traffic Engineering
             Extensions", RFC 3473, January 2003.

   [RFC5063] Satyanarayana, A., "Extensions to GMPLS Resource
             Reservation Protocol Graceful Restart", RFC5063, October
             2007.

   [RFC3936] Kompella, K. and J. Lang, "Procedures for Modifying the
             Resource reSerVation Protocol (RSVP)", [BCP 96](#), [RFC 3936](#),
             October 2004.

   [TE-SCALE-REC] Vishnu Pavan Beeram et. al, "Implementation
             Recommendations to improve scalability of RSVP-TE
             Deployments", [draft-ietf-teas-rsvp-te-scaling-rec](#) (work in
             progress)

   [SUMMARY-FRR]   Mike Tallion et. al, "RSVP-TE Summary Fast Reroute
             Extensions for LSP Tunnels", [draft-mtaillon-mpls-summary-
             frr-rsvpte](#) (work in progress)

[8](#). Informative References

   [RFC5439] Yasukawa, S., "An Analysis of Scaling Issues in MPLS-TE
             Core Networks", [RFC 5439](#), February 2009.

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

[9](#). Acknowledgments

We are very grateful to Yakov Rekhter for his contributions to the
development of the idea and thorough review of content of the draft.
Thanks to Raveendra Torvi and Yimin Shen for their comments and
inputs.

[10](#). Contributors

Markus Jork
Juniper Networks
Email: mjork@juniper.net

Harish Sitaraman
Juniper Networks
Email: hsitaraman@juniper.net

Vishnu Pavan Beeram
Juniper Networks
Email: vbeeram@juniper.net

Ebben Aries
Juniper Networks

      Email: exa@juniper.net

      Mike Tallion
      Cisco Systems Inc.
      Email: mtallion@cisco.com

[11]. Authors' Addresses

      Chandra Ramachandran
      Juniper Networks
      Email: csekar@juniper.net

      Ina Minei
      Google, Inc
      inaminei@google.com

      Dante Pacella
      Verizon
      Email: dante.j.pacella@verizon.com

      Tarek Saad
      Cisco Systems Inc.
      Email: tsaad@cisco.com