Fast Reroute Extensions to RSVP-TE for LSP Tunnels

draft-ietf-mpls-rsvp-lsp-fastreroute-00.txt

Abstract

This document describes the use of RSVP [RSVP, RSVP-TE] to establish backup LSP tunnels for local repair of LSP tunnels.

Two methods are presented here. One is to setup one-to-one detour LSPs according to the requirements defined by the head-end users. The other is to setup many-to-one bypass LSP using a single tunnel to backup a set of protected LSPs (making use of label stacking). Both methods can be used to protect links and nodes during network failure. The described use of RSVP allows both one-to-one and many-to-one backups to interoperate.
1. Introduction

This document describes the use of RSVP [RSVP] to establish backup LSP tunnels for local repair of LSP tunnels. By the term LSP tunnel we mean an explicitly routed LSP. In this document, we often refer to LSPs. In all cases we mean explicitly routed LSPs. Applicability of the techniques discussed herein to LSPs which dynamically change their routes such as those used in unicast IGP routing is beyond the scope of this document.

In order to meet the needs of real-time applications such as voice over IP, it is highly desirable to be able to re-direct user traffic onto backup LSP tunnels in 10s of milliseconds. The backup LSPs have to be placed as close to the failure point as possible, since reporting failure between nodes may cost significant delay. We use the term local repair when referring to techniques which accomplish this, and refer the LSP that is associated to one or more backup tunnels as a protected LSP. There are two basic strategies for setting up backup tunnels. These are one-to-one backup and facility backup. One-to-one backup operates on the basis of a backup LSP for each protected LSP. The facility backup aims at using a single LSP to back up a set of protected LSPs.

1.1. One-to-one backup

In the one to one case, a label switched path is established which intersects the original tunnel somewhere downstream of the point of link or node failure. For each LSP which is backed up, another backup LSP is established.

```
[R1]----[R2]-----[R3]----[R4]----[R5]
   \           /       \\
    [R6]----[R7]
```

For example, suppose that in the simple topology above, R1 creates a tunnel to R5 via the path [R1->R2->R3->R4->R5]. R2 can provide user traffic protection by creating a partial backup tunnel [R2->R6->R7->R4] which merges with the original tunnel [R1->R2->R3->R4->R5] at R4. We refer a partial one-to-one backup tunnel [R2->R6->R7->R4] as a detour.

To fully protect a LSP that traverses through N nodes, there could be as many as (N - 1) detours. To minimize processing overhead, it is desirable to merge detours back to a main LSP wherever possible.
1.2. Facility backup

A second means of backing up LSPs is to take advantage of the label stack. Instead of creating a separate LSP for every backed-up LSP, a single LSP is created which serves to backup up a set of LSPs. We call such a LSP tunnel a bypass tunnel.

The bypass tunnel must intersect the path of the original LSP(s) somewhere downstream of the point of local repair. This of course implies that the set of LSPs being backed up all pass through some common downstream node. All LSPs which pass through the point of local repair and through this common node which do not also use the facilities involved in the bypass tunnel are candidates for this set of LSPs.

To effect the repair of the protected LSPs, packets belonging to a LSP are redirected onto the bypass tunnel. An additional label representing the bypass tunnel is stacked onto the redirected packets. At the penultimate hop of the bypass tunnel, the label for the bypass tunnel is popped off the stack, revealing the label which represents the LSP being backed up.

```
[R8]  \\
[R1]-->[R2]-->[R3]-->[R4]-->[R5]  \\
     //  \\
[R6]-->[R7]  [R9]
```

In the above example, R2 in this case would build a bypass tunnel [R2->R6->R7->R4]. The doubled lines represent this tunnel. The backup path for [R1->R2->R3->R4->R5] again rejoins the original path at R4, but its path is now [R1->R2->R4->R5] with the bypass tunnel as the connection between R2 and R4.

In this example, the backup tunnel is a Next-Next-Hop (NNHOP) bypass tunnel. That is, it bypasses a single node (R3) of the protected path. NNHOP bypass tunnels may protect against Link (R2-R3) failure and/or Node (R3) failure as NHOP bypass tunnel only protects against link failure.

The scalability improvement comes in that this bypass tunnel can also be used to backup LSPs from any of R1, or R2, R8 to any of R4, R5, or R9 which traverse the link R2->R3.
2. 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] and [RSVP-TE].

LSR - Label Switch Router

LSP - An MPLS Label Switched Path

Local Repair - Techniques used to repair LSP tunnels quickly when a node or link along the LSPs path fails.

Protected LSP - An LSP is said to be protected at a given hop if it has one or multiple associated backup tunnels originating at that hop.

Detour LSP - The LSP that is used to re-route traffic around a failure in one-to-one backup.

Bypass Tunnel - An LSP that is used to protect a set of LSPs passing over a common facility.

Backup Tunnel - The LSP that is used to backup up one of the many LSPs in many-to-one backup.

NHOP Bypass Tunnel - Next-Hop Bypass Tunnel. A backup tunnel which bypasses a single link of the protected LSP.

NNHOP Bypass Tunnel - Next-Next-Hop Bypass Tunnel. A backup tunnel which bypasses a single node of the protected LSP.

Backup Path - The LSP that is responsible for backing up one protection LSP. A backup path refers to either a detour LSP or a backup tunnel.

PLR - Point of Local Repair. The head-end of a backup tunnel or a detour LSP.

MP - Merge Point. The LSR where one or more backup tunnels rejoin the path of the protected LSP, downstream of the potential failure. In the case of one-to-one backup, a Merge Point may also be an LSR where multiple detours converge and only one detour is signaled beyond that LSR; this type of merge point may be referred to as a Detour Merge Point. A MP may also
3. RSVP Extensions

We propose two additional objects, FAST_REROUTE and DETOUR, that extend RSVP-TE for fast-reroute signaling. The new objects are defined to be backward compatible for LSRs that do not recognize them (Section 3.10 in [RSVP]). Both objects can only be carried in RSVP Path messages.

The SESSION_ATTRIBUTE and RECORD_ROUTE objects are also extended to support bandwidth and node protection features:

In many circumstances, it may be desirable for the head-end LSR not only to signal an LSP as fast reroutable but also to specify to every PLR along its path that the LSP must be rerouted onto a backup path offering an equivalent bandwidth.

It may be desirable to signal the need for the fast reroutable LSP to be node protected along its path. By node protected we mean that each PLR along the path must protect the reroutable LSP with a detour LSP or a NNHOP backup tunnel (except for the penultimate hop LSR that will just require a NHOP backup tunnel). This way the reroutable LSP is being protected against any link or node failure.

3.1. FAST_REROUTE Object

The FAST_REROUTE object carries the control information, such as setup and hold priorities and bandwidth. A protected LSP uses the FAST_REROUTE object to specify the level of protection that is required during local repair. The FAST_REROUTE object can be used for both one-to-one and facility backup, and has the following format:
Class = TBD  (use form 11bbbbb for compatibility)
C-Type = 1

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Length (bytes)</td>
<td>Class-Num</td>
<td>C-Type</td>
</tr>
<tr>
<td>Setup Prio</td>
<td>Hold Prio</td>
<td>Hop-limit</td>
<td>Flags</td>
</tr>
<tr>
<td>Bandwidth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Included-Any</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excluded-Any</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include-all</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Setup Priority

The priority of the backup path with respect to taking resources, in the range of 0 to 7. The value 0 is the highest priority. Setup Priority is used in deciding whether this session can preempt another session. See [RSVP-TE] for the usage on priority.

Holding Priority

The priority of the backup path with respect to holding resources, in the range of 0 to 7. The value 0 is the highest priority. Holding Priority is used in deciding whether this session can be preempted by another session. See [RSVP-TE] for the usage on priority.

Hop-limit

The maximum number of extra hops the backup path is allowed to take, from current node (a PLR) to a MP, with PLR and MP excluded in counting. For example, hop-limit of 0 means only direct links between PLR and MP can be considered.

Flags

0x01  One-to-one Backup Desired

Indicates that the LSP should be protected via the one-to-one backup mechanism described in Section 5. This flag can only be set by the head-end LSRs.
0x02 Facility Backup Desired

Indicates that the LSP should be protected via the facility backup mechanism described in Section 6. This flag can only be set by the head-end LSRs.

Bandwidth

Bandwidth estimate (32-bit IEEE floating point integer) in bytes-per-second.

Exclude-any

A 32-bit vector representing a set of attribute filters associated with a backup path any of which renders a link unacceptable.

Include-any

A 32-bit vector representing a set of attribute filters associated with a backup path any of which renders a link acceptable (with respect to this test). A null set (all bits set to zero) automatically passes.

Include-all

A 32-bit vector representing a set of attribute filters associated with a backup path all of which must be present for a link to be acceptable (with respect to this test). A null set (all bits set to zero) automatically passes.

The C-Class must be assigned in such a way that, for the LSRs that do not support the FAST_REROUTE objects, they MUST forward the objects downstream unchanged.

Some of the existing implementations use the FAST_REROUTE object with a different C-type value, and slightly different object format (shown below). For backward compatible purposes, it is documented here for information purpose.
C-Type = 7

3.2. DETOUR Object

The DETOUR object is used in one-to-one backup to setup and identify detour LSPs. It has the following format:

Class = TBD (to conform 0bbbbbbb format for compatibility)
C-Type = 7

PLR ID (1 - n)

IPv4 address identifying the beginning point of detour which is a PLR. Any local address on the PLR can be used.
Avoid Node ID  (1 - n)

IP address identifying the immediate downstream node that the PLR is trying to avoid. Router ID of downstream node is preferred. This field is mandatory, and is used by the MP for merging rules discussed below.

There could be more than one pair of (PLR_ID, Avoid_Node_ID) entry in a DETOUR object. If detour merging is desired, after each merging operation (Section 5.3), the MP should combine all the merged detours in the subsequent Path messages.

The C-Class must be assigned in such a way that, for the LSRs that do not support the DETOUR objects, the LSRs MUST reject the message and send a PathErr to notify the PLR.

### 3.3. SESSION_ATTRIBUTE Modification

To explicitly require bandwidth and node protection, two new flags are defined in the SESSION_ATTRIBUTE object:

<table>
<thead>
<tr>
<th>Current Flags:</th>
</tr>
</thead>
</table>

Local protection desired: 0x01

This flag permits transit routers to use a local repair mechanism which may result in violation of the explicit route object. When a fault is detected on an adjacent downstream link or node, a transit node can reroute traffic for fast service restoration.
Label recording desired: 0x02

This flag indicates that label information should be included when doing a route record.

SE Style desired: 0x04

This flag indicates that the tunnel ingress node may choose to reroute this tunnel without tearing it down. A tunnel egress node SHOULD use the SE Style when responding with a Resv message. When requesting fast reroute, the head-end LSR MUST set this flag.

New Flags:

Bandwidth protection desired: 0x08

This flag indicates to the PLRs along the protected LSP path that a backup path with a bandwidth guarantee is desired. The bandwidth which must be guaranteed is that of the protected LSP, if no FAST_REROUTE object is included in the PATH message; if a FAST_REROUTE object is in the PATH message, then the bandwidth specified in there is that which must be guaranteed.

Node protection desired: 0x10

This flag indicates to the PLRs along a protected LSP path that they must select a backup path that bypasses at least the next node of the protected LSP.

3.4. RRO Modification

To record bandwidth and node protection, we define two new flags in the RRO IPv4 sub-object.

RRO IPv4 sub-object address:
Type: 0x01 IPv4 address

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Length</td>
<td>IPv4 address (4 bytes)</td>
<td></td>
</tr>
<tr>
<td>IPv4 address (continued)</td>
<td>Prefix Len</td>
<td>Flags</td>
<td></td>
</tr>
</tbody>
</table>

Current Flags:

Local protection available: 0x01

Indicates that the link downstream of this node is protected via a local repair mechanism, which can be either one-to-one or facility backup.

Local protection in use: 0x02

Indicates that a local repair mechanism is in use to maintain this tunnel (usually in the face of an outage of the link it was previously routed over, or an outage of the neighboring node).

New Flags:

Bandwidth protection: 0x04

The PLR will set this when the protected LSP has a backup path which provides the desired bandwidth, which is that in the FAST_REROUTE object or the bandwidth of the protected LSP, if no FAST_REROUTE object was included. The PLR may set this whenever the desired bandwidth is guaranteed; the PLR MUST set this flag when the desired bandwidth is guaranteed and the "bandwidth protection desired" flag was set in the SESSION_ATTRIBUTE object.

Node protection: 0x08

When set, this indicates that the PLR has a backup path providing protection against link and node failure on the corresponding path section. In case the PLR could only setup a link-protection backup path, the "Local protection
available" bit will be set but the "Node protection" bit will be cleared.

3.5. New RRO sub-object: MAX_PROTECTED_BANDWIDTH

This sub-object is carried in the RRO object and is optional. An implementation MAY support it. An LSR MUST ignore and silently propagate this sub-object, if it is not understood.

RRO MAX_PROTECTED_BANDWIDTH sub-object:

<table>
<thead>
<tr>
<th>Type</th>
<th>Length</th>
<th>Flags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bandwidth protection ratio

Let’s call T the bypass tunnel selected for the protected LSP. The bandwidth protection ratio is the sum of the bandwidths of all the protected LSPs having selected T as their bypass tunnel / bandwidth of the bypass tunnel T. The bandwidth protection ratio is a 32-bit IEEE floating point integer in bytes-per-second.

The minimum value for the protected ratio is 1, which means "the TE LSP is bandwidth protected".

Note that the PLR must select a backup tunnel in such a way that the bandwidth protected ratio is 1 for the TE LSP having required bandwidth protection in the SESSION_ATTRIBUTE object of their Path message.

The bandwidth protected ratio may be used for troubleshooting purpose.
or to trigger appropriate decision the head-end LSR (outside the scope of this document).

4. Signaling for Backup Path

A number of objectives must be met to obtain a satisfactory signaling solution. These are summarized as follows:

1. Unambiguously and uniquely identify backup paths
2. Unambiguously associate protected LSPs with their backup paths
3. Work with both global and non-global label spaces
4. Allow for merging of backup paths
5. Maintain RSVP state during and after fail-over.

LSP tunnels are identified by a combination of the SESSION and SENDER_TEMPLATE objects. The relevant fields are as follows.

IPv4 tunnel end point address

IPv4 address of the egress node for the tunnel.

Tunnel ID

A 16-bit identifier used in the SESSION that remains constant over the life of the tunnel.

Extended Tunnel ID

A 32-bit identifier used in the SESSION that remains constant over the life of the tunnel. Normally set to all zeros. Ingress nodes that wish to narrow the scope of a SESSION to the ingress-egress pair may place their IPv4 address here as a globally unique identifier.

IPv4 tunnel sender address

IPv4 address for a sender node

LSP ID

A 16-bit identifier used in the SENDER_TEMPLATE and the FILTER_SPEC that can be changed to allow a sender to share resources with itself.

The first three of these are in the SESSION object and are the basic identification of the tunnel. The last two are in the
In particular, setting the "Extended Tunnel ID" to the original IPv4 sender address allows the PLR to identify to which protected LSP a message (from MP) corresponds. For example, when a Resv message arrives at the PLR, the Extended Tunnel ID identifies the original sender, allowing the PLR to identify the state to be refreshed.

4.1. Identification and association of backup paths

We propose two different approaches to identify backup paths:

- Path Message Specific:

The backup paths use the same SESSION and SENDER_TEMPLATE objects as the ones used in the protected LSP. However, the Path messages need to provide enough information that allow the LSRs to differentiate the backup paths from the protected LSPs.

In case of one-to-one backup, the presence of DETOUR object in Path messages signifies a backup path, while the presence of FAST_REROUTE object indicates a protected LSP.

- Sender-Template Specific:

In this approach, the SESSION object and the LSP_ID are copied from the protected LSP. The IPv4 tunnel sender address is set to an address of the PLR node. If the head-end of a tunnel is also acting as the PLR, it must choose an IP address different from the one used in the SENDER_TEMPLATE of the original LSP tunnel.

In the path-message-specific approach, when an LSR receives multiple Path message which have the same Session and Sender Template objects and which have the same next-hop, that LSR must merge the Path messages; without this behavior, the multiple RESV messages received back would not be distinguishable as to which backup path each belongs to. This merging behavior does reduce the total number of RSVP states inside the network. One merging example is given in Section 5.3.

When using the sender-template-specific approach, the protected LSPs and the backup paths SHOULD use the Shared Explicit (SE) style. This allows bandwidth sharing between multiple backup paths. The backup paths and the protected LSP can be merged by the Merge Points, when the ERO from the MP to the egress is the same on each LSP to be merged, as specified in [RSVP].
5. One-to-one backup protection

In this section, we describe an one-to-one backup method that has the feature to protect both network links and nodes.

To support the one-to-one backup, the users at head-end LSRs must specify the backup service requirements for the protected LSPs. The LSRs must be able to interface with CSPF to compute the most suitable detour route for the protected LSPs. Upon receiving the local protection requests for a protected LSP, the PLRs must try to establish the detour LSPs immediately. During network failure, the PLR must redirect the data packets into the detour LSPs in a timely fashion.

5.1. Operation Overview

If a one-to-one backup for a protected LSP is explicitly desired, the head-end LSR SHOULD insert into the Path message a FAST_REROUTE object, with the "One-to-one Backup desired" flag set. A one-to-one backup for a protected LSP may also be created based upon a PLR’s local policy if either the "local protection desired" flag is set in the SESSION_ATTRIBUTE object or a FAST_REROUTE object is included or both.

When processed at a PLR, the PLR initiates a detour LSP by sending a new Path message that contains a DETOUR object. Since an LSP cannot be a protected and a detour LSP at the same time, any Path message MUST NOT contain both FAST_REROUTE and DETOUR objects.

The LSRs that initiate the detour LSPs SHOULD support both FAST_REROUTE and DETOUR objects. It is possible that some LSRs along a protected LSP do not support this standard. If that is the case, those LSRs will not establish protection for their immediate links or nodes. Any LSR which does support this standard SHOULD provide protection.

The LSRs that support the detour LSPs MUST store all received FAST_REROUTE and/or DETOUR objects for Path refreshes. The LSRs must process the detour LSPs independent of the protected LSPs to avoid triggering the LSP loop detection procedure described in [RSVP-TE].

The one-to-one backup can use either path-message-specific or sender-template-specific to identify the detour LSPs.

When using the sender-template-specific approach, the protected and detour LSPs should have the "SE Style desired" bit set in the SESSION_ATTRIBUTE objects. At the MP, the detour LSPs merge into the
protected LSPs according to the merging rules defined for SE style reservations in [RSVP].

In the case of one-to-one backup, there is no need for the PLRs to learn about the backup labels used at the merging points.

5.2. Procedures for the PLR

Upon receiving a Path message that contains a FAST_REROUTE object, a PLR needs to run CPSF based on the information provided in the FAST_REROUTE, as well as the downstream interface and nexthop router information, to compute a detour route. More details on CSPF computation are described in Section 7.

Once a detour is successfully computed and established, the PLR needs not to compute the detour routes again, unless (1) the contents of FAST_REROUTE have changed, or (2) the downstream interface and/or the nexthop router for a protected LSP have changed.

After a successful detour computation, the PLR generates a Path message to setup a detour path. The Path consists of the following:

- A DETOUR object that specifies the current PLR ID and Avoid Node ID. Only one pair of (PLR_ID, Avoid_Node_ID) permitted.
- An EXPLICIT_ROUTE object toward the egress. The ERO information comes from the CSPF computation.
- The SENDER_TSPEC object contains the bandwidth information from the previously received FAST_REROUTE objects.
- The RSVP_HOP object contains the PLR’s IP address.
- The detour LSP may generate and process its own RRO object.
- The FAST_REROUTE object MUST NOT be included.
- When using the sender-template-specific approach, the "IPv4 tunnel sender address" in the SENDER_TEMPLATE must be set to an address belonging to the PLR.
- The detour LSPs MUST use the same reservation style as the protected LSP. This must be correctly reflected in the SESSION_ATTRIBUTE object.
- All other objects SHOULD be identical to those of the protected LSP.

The PLR MUST not mix the messages for the protected and the detour LSPs. When a PLR receives Resv, ResvTear and PathErr messages from the downstream detour destination, the messages MUST not be forwarded upstream. Similarly, when a PLR receives ResvErr and ResvConf messages from a protected LSP, it MUST not propagate them onto the associated detour LSP.

A session tear-down request is normally originated by the sender via PathTear messages. When a PLR node receives a PathTear message from upstream, it MUST delete both protected and detour LSPs. The PathTear messages MUST propagate to both protected and detour LSPs.

During error conditions, the LSRs may send ResvTear messages to fix problems on the failing path. When a PLR node receives the ResvTear messages from downstream for a protected LSP, as long as a detour is up, the ResvTear messages MUST not sent further upstream.

5.3. Procedures for the MP using the Path-Specific Approach

An LSR (that is, a MP) may receive multiple Path messages from different interfaces with identical SESSION and SENDER TEMPLATE objects. Path state merging is REQUIRED.

The merging rule is the following:

For all Path messages that do not have either a FAST_REROUTE or a DETOUR object, or the MP is the egress of the LSP, no merging is required. The messages are processed according to [RSVP-TE].

Otherwise, the MP MUST record the Path state as well as their incoming interface. If the Path messages do not share outgoing interface and next-hop LSR, the MP must consider them as independent LSPs, and must not merge them.

For all the Path messages that share the same outgoing interface and next-hop LSR, the MP runs the following procedure to select one of them as the final LSP.

1. Eliminate from consideration those that traverse nodes that other LSPs want to avoid.

2. If one LSP is originated from this node, this must be the final LSP. Quit.
3. If only one LSP contains FAST_REROUTE object, this must be the final LSP. Quit.

4. If there are several LSPs, and not all of them have a DETOUR object, then eliminate those with DETOUR from final LSP considerations.

5. If several candidates remain (that is, there are both detour and protected LSPs), prefer the ones with FAST_REROUTE object.

6. If none found, prefer the ones without DETOUR object. If none found, prefer the ones with DETOUR object.

7. If several candidate LSPs still remain, it is a local decision to choose which one will be the final LSP. The decision can be based on the number of IP hops in ERO, bandwidth requirements, or others.

Once the final LSP has been identified, the MP MUST only transmit the Path messages that are corresponding to the final LSP. Other LSPs are considered merged at this node.

The MP may receive PathTear messages for some of the merging LSPs. No PathTear message should be propagated downstream until the MP has received tear-down from all merging LSPs.

When an LSR receives a ResvTear for an LSP and it is not a PLR for that LSP, then the LSR SHOULD propagate the ResvTear towards the LSP’s ingress. For each backup LSP where the LSR is the merge node, the ResvTear should also be propagated along the backup LSP towards the backup LSP’s ingress, a PLR.

5.3.1. An Example on Path Message Merging

Consider the following example:

```
G----H----I--\n |    |    |   \nA----B----C----D----E---F
```

The protected LSP is A-B-C-D-E-F. After running CSPF, let the detour ERO from B be B-G-H-I-D-E-F, and the detour ERO from C be C-H-I-E-F.

H will receive Path messages that have the same SESSION and
SENDER TEMPLATE from detours for B and C. During merging at H, since detour C has a shorter ERO path length (that is, ERO is I-E-F, and path length is 3), H will select it as the final LSP, and only propagate its Path messages downstream. Upon receiving a Resv (or a ResvTear) message, H must relay on the messages toward both B and C.

E needs to merge as well, and will select the main LSP, since it has the FAST_REROUTE object. Thus, the detour LSP terminates at E.

5.3.2. Creating new DETOUR object at MP

If several LSPs are merged, the MP uses the following algorithm to format its outgoing DETOUR object for the final LSP:

- If final LSP is protected LSP itself (that is, it contains FAST_REROUTE object), no DETOUR object needed.

- Otherwise, combine all the (PLR_ID, Avoid_Node_ID) pairs from all the DETOUR objects of all merged LSPs, and create a new object with all listed. Ordering is insignificant.

5.4. Local reroute of the traffic onto the detour LSP

Detour LSPs are regular LSPs in operation. They are established as soon as the protected LSPs are up. During local repair, packets belonging to a protected LSP are simply switched (for example, label swapping) onto the corresponding detour LSP. At the Merge Point, the packets arrived from the detour LSP are merged to the final LSP.

In the example above, if there is a node failure at D, C will switch traffic onto the pre-established detour LSP (C-H-I-E-F). At E, the traffic switches onto the protected LSP again.
6. Facility protection using label stacked bypass tunnel

In this section, we describe a method where a single backup tunnel can be used to protect many LSPs. The LSPs can be protected against both link and node failures.

Each PLR makes use of one or more NHOP or NNHOP bypass tunnels. Each bypass tunnel will be used to backup a set of protected LSP. Those bypass tunnels may be setup initially or may also be dynamically setup. The users at head-end initiate the fast reroute process by setting the appropriated fields in the SESSION_ATTRIBUTE and/or FAST_REROUTE objects in an LSP’s Path messages. At each PLR, one bypass tunnel is selected to reroute an LSP’s data packets in case of network failure. The process of selecting a bypass tunnel for a protected LSP is performed by the PLR when the LSP is first setup.

During failure, the PLR reroutes the data packets of each protected LSP onto the bypass tunnel. The control messages of the backed-up LSPs are also sent over the bypass tunnel. The facility backup uses the sender-template-specific approach to identify the backup tunnels.

6.1. Discovering downstream labels

When global labels are in use at MPs, the PLR may learn backup labels in a very efficient manner. The labels are learned during normal signaling of the protected LSP by observing the contents of the RRO object in the Resv message.

When a protected LSP is first signaled through a PLR, the PLR can learn about the incoming labels that are used by all downstream nodes for this LSP. In particular, it can learn incoming labels used by downstream MPs, whether they are one hop or multiple hops away from the PLR. The labels are learned during normal signaling of the protected LSP by observing the contents of the RRO object in the Resv message.

Two methods are available for discovering/obtaining the label used at the merge node. One relies on explicit signaling over the bypass tunnel prior to any failure of the primary path. If the nodes in the network use a global-to-the-node label space, then the label can be discovered by using the RRO object without additional signaling.

When this second method is intended, the head-end router includes an RRO object and sets the label-recording-requested flag in the Session_Attribute object. This will cause (as specified in [RSVP-TE]) all nodes to record their INBOUND labels and to note via a flag

```
draft-ietf-mpls-rsvp-lsp-fastreroute-00.txt
```

^L[Page 20]
if the label is global to the node.

Note that when global labels are used, no Path message need be sent via the bypass tunnel prior to failure.

When MPs use per-interface-label spaces, the PLR must send Path messages (for each Reroutable LSP) via the bypass tunnel prior to the failure in order to discover the appropriate MF label. The signaling procedures for this are identical to those in section 6.3 below.

6.2. Procedures for the PLR before fast-reroute

When a protected LSP in first signaled, all the PLRs along the path which determine to create a backup tunnel via a bypass tunnel should perform the following:

- If the "Local protection desired" bit is set in the SESSION_ATTRIBUTE and there is no Fast_Reroute object, or there is a Fast_Reroute object with the Facility-Backup-Desired flag set, the PLR should select or create a bypass tunnel for the reroutable LSP.

- If the PLR can find a NNHOP bypass tunnel, the PLR MUST set the "Node protection" bit and the "Local protection available" flags of its IPv4 or IPv6 RRO subobject if an RRO object is included in the Resv message.

- If the PLR cannot find a NNHOP bypass tunnel, but can find a NHOP bypass tunnel, the PLR must clear the "Node protection" bit and must set the "local protection available" flags in the RRO object of the Resv message.

- If the PLR can find a bypass tunnel with bandwidth guarantee, the PLR must set the "Bandwidth protection" flag in the above mentioned RRO subobject.

- If the PLR cannot find a bypass tunnel with the requested bandwidth guarantee, the PLR must clear the "Bandwidth protection" flag in the above mentioned RRO subobject.

Based on this additional information the head-end may take appropriate actions.

Note that when global labels are used, no Path message need be sent via the bypass tunnel prior to failure.
6.3. Procedures for the PLR during fast-reroute

When the PLR detects a link or/and node failure condition, it needs to reroute the data traffic onto the bypass tunnel and to start sending the control traffic for the protected LSP onto the bypass tunnel.

The backup tunnel is identified as follows:

- The SESSION and SESSION_ATTRIBUTE are unchanged.

- The IPv4 tunnel sender address of the SENDER_TEMPLATE is changed (set to an address belonging to the PLR).

- The RSVP_HOP object must contain the IPv4 source address (and LIH) of the bypass tunnel. Consequently, the MP will send messages back to the PLR with HOP objects containing this same IPv4 address.

- The PLR must generate an EXPLICIT_ROUTE object toward the egress. Detailed ERO processing is described below.

- The RRO object may need to be updated, as described below.

Messages sent by PLR via the backup tunnel include Path, PathTear, and ResvConf. Messages sent by MP via the same RSVP_HOP object contents include Resv, and ResvTear.

6.3.1. Processing backup tunnel’s ERO

Procedures for ERO processing are described in [RSVP-TE]. If normal ERO processing rules are followed by the Merge Point, and the PLR sends a Path message via the backup tunnel, the Merge Point would examine the first sub-object and likely reject it (Bad initial sub-object).

This is because the ERO may contain the IP address of a bypassed node (in the case of a NNHOP Backup Tunnel), or of an interface which is currently down (in the case of a NHOP Backup Tunnel). For this reason, the PLR must update the ERO before sending Path messages onto Backup Tunnels.

This is done by operating on the original ERO:

Sub-objects belonging to abstract nodes which precede the Merge Point are removed, along with the first Sub-object belonging to the MP. A
Sub-object identifying the Backup Tunnel destination is then added.

More specifically, the PLR must:

- remove all the sub-objects proceeding the first address belonging to the MP.
- replace this first MP address with the IP destination address of the backup tunnel.

The procedure described above ensures successful ERO processing at the Merge Point.

6.3.2. Processing backup tunnel’s RRO

During fast reroute, for each protected LSP containing an RRO object, the PLR must update the RRO by inserting an IPv4 sub-object with the IPv4 address of the backup tunnel source address in the Path messages.

For each rerouted LSP in the backup tunnel, the PLR must update the RRO object in Resv messages sent upstream in the following manner:

- In the IPv4 or IPv6 sub-object inserted by this node, set the "Local protection available" and "Local protection in use" flags according to the current state of the local repair mechanism.
- Update the label sub-object recording the INBOUND label (same label value as the one sent the Resv message).

6.4. Procedures for state maintenance during fast-reroute

We will describe how state is maintained using an example:

```
[R8]
 \   
[R1]---[R2]-X--[R3]----[R4]---[R5]
   \   /   \  
   [R6]===[R7]     [R9]
```

We assume that:
- a bypass tunnel is set up and follows the R2-R6-R7-R4 path;
- PLR (R2) performs 1:N protection;
- various protected LSPs exist and follow the R2-R3-R4 segment;
- link R2-R3 fails, and all protected LSPs are rerouted via the bypass tunnel.

Note that the same procedure as the one described below would apply in case of a node (R3) failure.

6.4.1. Path state

Path state for every locally repaired LSPs is refreshed downstream by the PLR. These Path messages use a new SENDER_TEMPLATE value (the IPv4 tunnel sender address is set to a PLR address), and are sent onto the bypass tunnel with changed PHOP, ERO and RRO.

When a local link fails, there could be some protected LSPs using this link. At this point, the LSR MUST NOT remove the state (Path and Resv) and send PathTear and ResvErr messages that are corresponding to these LSPs immediately. We always assume that these LSPs may have been repaired upstream, and new Path messages will soon arrive via the bypass tunnels.

However, the state will be removed if they have not been refreshed by a PLR after the soft-state lifetime has expired.

6.4.2. Resv state

Resv state is refreshed by the MP by sending Resv messages to the IP destination contained in the PHOP object of the Path message received via the bypass tunnel.

The PLR receives these Resv messages, refreshes the original state (corresponding to the protected LSP), and hence continues refreshing the state upstream of the PLR to the head-end.
6.5. Local reroute of the traffic onto the bypass tunnel

To perform Local Repair, packets belonging to a protected LSP are sent on the corresponding backup tunnel in case of local failure.

An additional label (representing the bypass tunnel) is pushed onto the stack. At the penultimate hop of the bypass tunnel, the additional label is popped off the stack. The packet thus arrives at the Merge Point with the same top-level label it would have carried when arriving prior to failure (although it would have arrived on a different interface prior to failure).

7. Procedures for detour and bypass tunnel computation

To setup the detours described in Section 5 and the bypass tunnels in Section 6, CSPF may be used to find the optimal route. Before CSPF computation, the following information should be collected at a PLR:

- The list of downstream nodes that the protected LSP passes through. This information is readily available from the RECORD_ROUTE objects during LSP setup. Note, a protected LSP’s ERO may not provide adequate information since the LSP could be a loose routed path.

- The downstream links/nodes that we want to protect against. Once again, this information is learnt from the RECORD_ROUTE objects.

- The upstream uni-directional links that the protected LSP passes through, this information is learnt from the RECORD_ROUTE objects. This information is only needed for setting up one-to-one protection in path-message-specific approach.

- The LSP resource information, such as bandwidth. Such information can be found in the FAST_REROUTE objects.

When applying a CSPF algorithm to compute the backup route, the following constraints should be satisfied:

- The source address of the backup LSP is the current PLR, For setting detours (Section 5), the destination MUST be the tail-end of the protected LSP, whereas for setting up bypass tunnels (Section 6), the destination MUST be the address of the MP.

- When setting up one-to-one protection using the path-specific approach, a detour MUST not traverse the upstream links of
the protected LSP in the same direction. This prevents the possibility of early merging of the detour into the protected LSP.

- The backup LSP cannot traverse the downstream nodes and links that we are trying to protect against. However, if the PLR is the penultimate hop, avoid traversing downstream link only. The detour LSP/bypass tunnel may also be SRLG disjoint from the protected section (see the note at the end of this section).

- The backup path must satisfy the resource requirements of the protected LSP.

If such computation succeeds, the PLR should trigger RSVP to establish a backup path. The PLR may schedule a re-computation at a later time. The backup path should be as short as possible, and must merge back into the protected LSP at its MP. If for any reason, the PLR is unable to bring up a backup path, it must schedule a retry at a later time.

The PLR has the option to apply other constraints during the CSPF computation. For example, a simple method can be to terminate the computation as soon as a backup path is found. On the other hand, an implementation may wish to continue exhaustive search to discover an optimal path with lowest cost (or highest available bandwidth).

The PLR also has the option to re-compute the backup path periodically even after the backup is up and running to ensure continuous adaptation to the latest network conditions. However, during the replacement of a functional backup path with a more optimal one, the protected LSP may not have any backup path available for a short interval. Except, if the PLR supports both one-to-one and facility backup schemes, the protected LSP could be protected by multiple backup LSPs. In this case, the LSP is fully protected at all time.

Nevertheless, the exact CSPF algorithms to be used to compute back-up tunnels or detour LSPs are beyond the scope of this document. Both [OSPF-TE] and [ISIS-TE] may provide more insight on this subject.

Note also that the backup tunnel path computation may be performed by a centralized path computation server or may use some distributed backup path computation algorithms.
7.1. Notion of diverse routing

Two TE LSPs are said link diverse if and only if their paths do not have any link in common. Two TE LSPs are said node diverse if and only if their paths do not have any node in common. It is straightforward to demonstrate that two node diverse paths are also link diverse.

To be effective a backup tunnel must imperatively be diversely routed from the protected LSP path section it is protecting. That is, a one-hop NHOP backup tunnel path must not contain the protected link. In the example provided in Section 6, the backup LSP path must not contain the R2-R3 link. A NNHOP backup tunnel must not contain the protected link nor the PLR's next hop. In the first example provided in Section 1, the backup tunnel must not traverse the R2-R3 link nor the R3 node.

The notion of SRLG diverse path also exists. A set of links constitute a SRLG ("Shared Risk Link Group") if they share a resource whose failure may affect all the links in the set. So the backup tunnel may be SRLG disjoint from the protected LSP path section it is protecting.

Note that in the case of Path protection, the whole paths of the protected LSP and the backup tunnel must be entirely link/node diverse.

Well-known algorithms can be used to compute link/node/SRLG diversely routed paths.

8. Network Failure Detection, Notification and Troubleshooting

8.1. Notification of local repair

In many situations, the route used during a Local Repair will be less than optimal. The point of the Local Repair is to keep high priority and loss sensitive traffic flowing while a more optimal re-routing of the tunnel can be effected by the head-end of the tunnel. Thus the head-end needs to know of the failure so it may re-signal an LSP which is optimal.

To provide this notification, the PLR SHOULD send a Path Error message with error code of "Notify" (Error code = 25) and an error value field of ss00 cccc cccc cccc where ss=00 and the sub-code = 3 ("Tunnel locally repaired") (see [RSVP-TE])
Note also that in the case of inter-area TE LSP (TE LSP spanning areas), the head-end LSR will exclusively rely on the Path Error message to be informed that the LSP has suffered a failure if the failure occurs in another area than the area it belongs to. In the case of a failure occurring in the head-end area or in the case of intra-area TE LSP, the head-end could also detect the TE LSP failure through the IGP notification.

8.2. Failure detection mechanisms

Link failure detection can be performed through layer-2 failure detection mechanism. Node failure detection can be done through IGP loss of adjacency or RSVP hellos messages extensions as per defined in [RSVP-TE]. However, it is beyond the scope of this document to define and describe the exact mechanisms on failure detection.

When a network failure is detected, the PLR MUST immediately switch traffic from the protected LSP to the backup path. At the same time, the PLR MAY send a PathErr messages toward the head-end LSR to notify the failure condition. The PLR MUST send a RESV with an updated RRO which indicates that local protection is in use.

8.3. Troubleshooting of local repair

For troubleshooting purposes, an RRO object may be inserted in the Path message sent by the head-end. The previously described mechanisms do not require the Path message to carry an RRO object. On the other hand, the RRO object MUST be inserted in the Resv message for the protected LSP if the "Local protection desired" bit of the SESSION_ATTRIBUTE has been set in the corresponding Path message, or if FAST_REROUTE object is present in Path messages.
9. Interoperability considerations

   The following guidelines are useful when running one-to-one and/or facility backups.

9.1. Requesting local-protection and recognizing those requests

   The head-end LSR of a protected LSP MUST either set the "Local protection desired" flag in the SESSION_ATTRIBUTE object, or include the FAST_REROUTE object, or both. A PLR MUST consider that a PATH message with either a set "Local protection desired" flag in the SESSION_ATTRIBUTE object, or the presence of the FAST_REROUTE object, or both to be a request for local protection.

   A PLR SHOULD consider the constraints signaled via a received FAST_REROUTE object, or a received SESSION_ATTRIBUTE object (Bandwidth and Node protection constraints on the bypass tunnel can also be specified by setting the "Bandwidth protection desired" and "Node protection desired" bits in the SESSION_ATTRIBUTE object), when determining the backup path to use. If signaled backup constraints and bandwidth are desired, the PATH message SHOULD contain the FAST_REROUTE object.

   A head-end LSR MUST set the "Label recording desired" flag in the SESSION_ATTRIBUTE object if a backup tunnel through a bypass tunnel is desired.

   If local protection was not requested for the current LSP of a tunnel and it is then desired for that tunnel, the head-end LSR MUST send a new Path message reflecting the change ("Local protection desired" flag set in the SESSION_ATTRIBUTE object or include a FAST_REROUTE object). When a node detects a change in the SESSION_ATTRIBUTE object it SHOULD forward the Path message immediately.

9.2. Backups for local protection

   A PLR that recognizes that local protection is required on a protected LSP MUST try to protect the LSP’s data path immediately, by either setting up an one-to-one detour LSP or a bypass tunnel.

   When a network has a mix of PLRs that support either one-to-one backup, or facility backup, or both, it is up to the network operators to decide which backup mechanism to use.

   When using both schemes, the PLR has the option to backup data
traffic on an one-to-one detour LSP, as well as on a bypass tunnel. In case of a network failure, the PLR can re-reroute traffic using one of the two backup path initially. If the backup path failed also, the other backup path can be used to re-reroute user traffic.

If no established detour LSP or backup tunnel exists, or the detour LSP and the backup tunnel is in "DOWN" state, the PLR MUST clear the "local protection available" flag in its IPv4 (or IPv6) address subobject of the RRO and SHOULD send the updated RESV. When a detour LSP or backup tunnel is established, the PLR MUST set the "local protection available" flag and the appropriated "bandwidth protection" and "node protection" bits, and SHOULD send the updated Resv.

10. Security Considerations

This document does not introduce new security issues. The security considerations pertaining to the original RSVP protocol [RSVP] remain relevant.

11. IANA Guidelines

IANA [RFC-IANA] will assign RSVP C-class numbers for FAST_ROUTE and DETOUR objects. Currently, in production networks, FAST_REROUTE uses C-class 205, and DETOUR uses C-class 63.

12. Intellectual Property Considerations

Cisco Systems and Juniper Networks may have intellectual property rights claimed in regard to some of the specification contained in this document.
13. Full Copyright Statement

Copyright (C) The Internet Society (2002). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

14. Acknowledgments

We acknowledge the helpful comments from Arthi Ayyangar, Riza Cetin, Rob Goguen, Carol Iturralde, Kireeti Kompella, Manoj Leelanivas, Yakov Rekhter and Nischal Sheth.
15. References


16. Author Information

Ping Pan
Juniper Networks
1194 N.Mathilda Ave
Sunnyvale, CA 94089
e-mail: pingpan@juniper.net
phone: +1 408 745 3704

Der-Hwa Gan
Juniper Networks
1194 N.Mathilda Ave
Sunnyvale, CA 94089
e-mail: dhg@juniper.net
phone: +1 408 745 2074

George Swallow
Cisco Systems, Inc.
250 Apollo Drive
Chelmsford, MA 01824
e-mail: swallow@cisco.com
phone: +1 978 244 8143