Network Working Group — Ping Pan, Ed. (Juniper Networks)
Internet Draft — Der-Hwa Gan (Juniper Networks)
Expiration Date: May 2002 — George Swallow (Cisco Systems)
Network Working Group — Jean Philippe Vasseur (Cisco Systems)
Dave Cooper (Global Crossing)
Alia Atlas (Avici Systems)
Markus Jork (Avici Systems)

Fast Reroute Techniques in RSVP-TE

draft-ping-rsvp-fastreroute-00.txt

Status of this Memo

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 bypass tunnel to backup a set of protected LSPs (making use of label stacking) to reroute both data and control traffic. Both methods can protect both link and node during network failure, and make use of local protection techniques.

0. Background

This draft is based on [FR-GAN] [FR-SWALLOW] and [FR-ATLAS]. It represents the first step toward having a single MPLS fast reroute solution. Additional details about merging procedure and interop issues to be added in a further release.

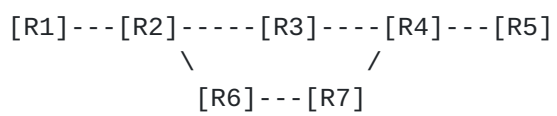
1. Introduction

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 LSP backup and facility backup. LSP 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.



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.

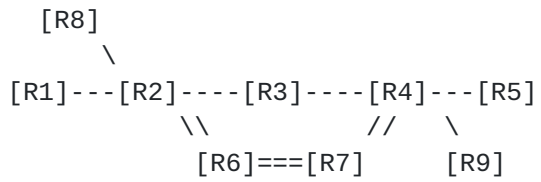
To fully protect a LSP that traverses through N nodes, there could be as many as (N - 1) detours. To minimize processing overhead, it is desirable to merge detours back to a main LSP wherever possible.

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.



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

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 - An MPLS LSP 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.

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

MP - Merge Point. The LSR where detour or backup tunnels meet the protected LSP. In case of one-to-one backup, this is where multiple detours converge. A MP may also be a PLR.

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.

Reroutable LSP - Any LSP for with the "Local protection desired" bit is set in the Flag field of the SESSION_ATTRIBUTE object of its Path messages.

CSPF - Constraint-based Shortest Path First.

3. 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.

Initially, the users from head-end LSRs specify the backup service requirements of a particular LSP. Each LSR can interface with CSPF to compute the most suitable detour path for the LSP automatically (see Section 5.2). The PLR needs to setup the detour LSP immediately. During network failure, the PLR redirects the data packets into the detour LSP.

3.1. RSVP Extensions

Two new RSVP objects are defined here, FAST_ROUTE and DETOUR. Both objects can only be carried in RSVP Path messages. To support this detour method, an implementation MUST support both objects.

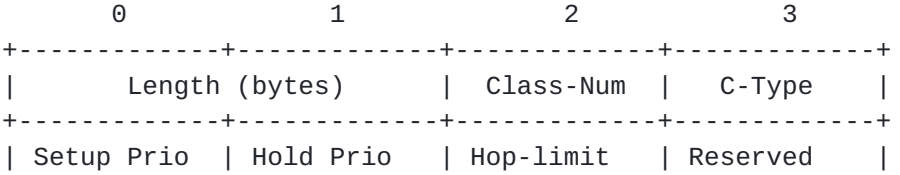
Both objects are defined to be backward compatible for LSRs that do not recognize them (see Section 3.10 in [RSVP]). For the LSRs that do not support the FAST_REROUTE objects, they MUST forward the objects downstream unchanged. For the LSRs that do not support the DETOUR objects, the LSRs MUST reject the message and send a PathErr to notify the PLR at head-end.

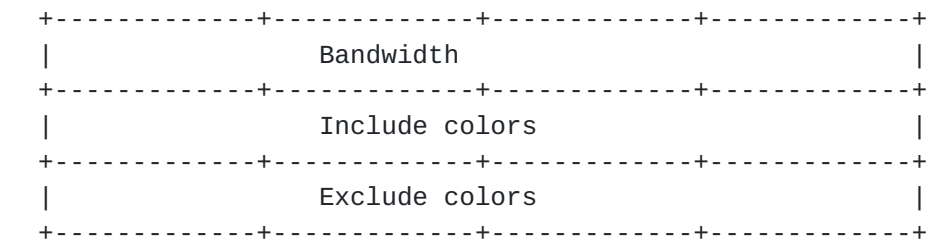
Thus, even if some LSRs along a protected LSP do not recognize or support the new objects, it is still possible to establish detour LSPs between the LSRs that can support the new objects. At worst, the detour LSPs will not be established to protect the links between the non-supporting nodes. This feature is useful and important for deployment.

Both objects are defined as the following:

3.1.1. FAST_REROUTE Object

```
Class = 205  (use form 11bbbbbb for compatibility)
C-Type = 7
```





Setup Priority

The priority of the detour 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 usage of priority.

Holding Priority

The priority of the detour 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 usage of priority.

Hop-limit

The maximum number of extra hops the detour 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.

Reserved

This field is reserved. It MUST be set to zero on transmission and MUST be ignored on receipt.

Bandwidth

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

Include colors

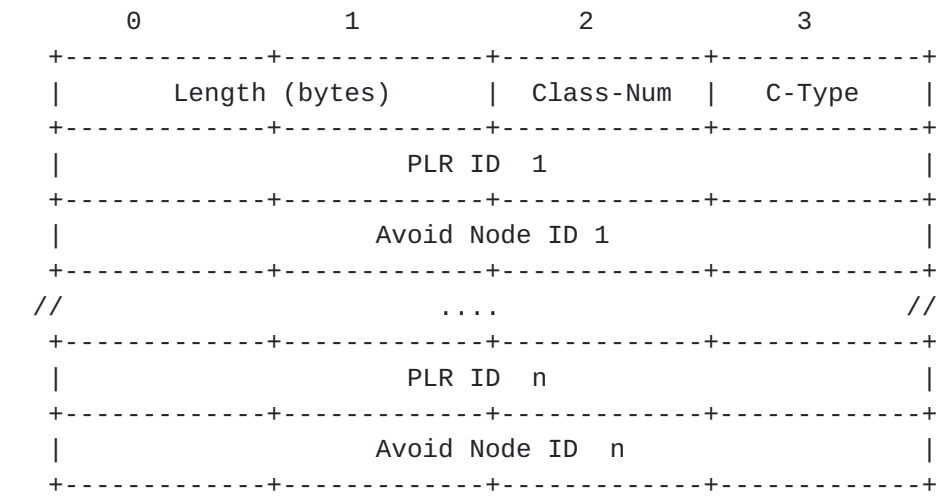
A 32-bit vector representing a set of attribute filters associated with a detour any of which renders a link acceptable (with respect to this test). A null set (all bits set to zero) automatically passes.

Exclude colors

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

3.1.2. DETOUR Object

Class = 63 (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 optional, and may be used by the MP for debugging purposes.

There could be more than one pair of (PLR_ID, Avoid_Node_ID) entry in a DETOUR object. After each detour merging operation (Section 3.2.2), the MP should list all the merged detours in the subsequent Path messages.

3.2. Operations

To initiate the setup of a detour for a protected LSP, the head-end LSR MUST insert a FAST_REROUTE object in the Path messages. 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 support the detour LSPs MUST store all received FAST_REROUTE and/or DETOUR objects for Path refreshes.

The detour LSPs must be processed and maintained separately from the protected LSPs. Given that LSPs will work over nodes that cannot support MPLS label switching, detour LSPs MUST NOT go over non-MPLS cloud.

It is possible to have the detour LSPs traversing the same LSRs as the protected LSPs. This is because during detour path computation, CSPF will only exclude the protected links and nodes, and provide a result that may include the links and nodes that belong to the protected LSP. The LSRs must process the detour LSPs independent of the protected LSPs to avoid triggering the LSP loop detection procedure described in [RSVP-TE].

3.2.1. 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 RECORD_ROUTE (RRO) from the protected LSP's Resv messages, to compute a detour route. More details on CSPF computation are described in Section 5.2.

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 detour LSPs MUST use the same reservation style as the protected LSP. This must be correctly reflected in the SESSION_ATTRIBUTE object.

- 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.

- 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.

3.2.2. Procedures for the Merge Point

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. If one LSP is originated from this node, this must be the final LSP. Quit.

2. If only one LSP contains FAST_REROUTE object, this must be the final LSP. Quit.

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

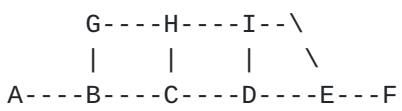
4. If several candidates remain (that is, there are both detour and protected LSPs), prefer the ones with FAST_REROUTE object.

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

6. If several candidate LSPs still remain, chose the one with the shortest ERO path length. If there are more than one LSP having the same path length, choose one randomly.

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.

Consider the following example:



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 rely the message 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.

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.

3.2.3. 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.

If DETOUR object becomes too big, MP is free to truncate some of the (PLR_ID, Avoid_Node_ID) pairs.

3.2.4. Local reroute of the traffic onto the detour LSP

Detour LSPs are regular LSPs in operation. To perform 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.

4. 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 flags in the SESSION_ATTRIBUTE object 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 rerouted LSP onto the bypass tunnel. The control messages of the backed-up LSPs are also sent over the bypass tunnel (see details in the sections bellow).

4.1. RSVP Extensions

- Bandwidth protection desired

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 tunnel offering an equivalent bandwidth.

- Node protection desired

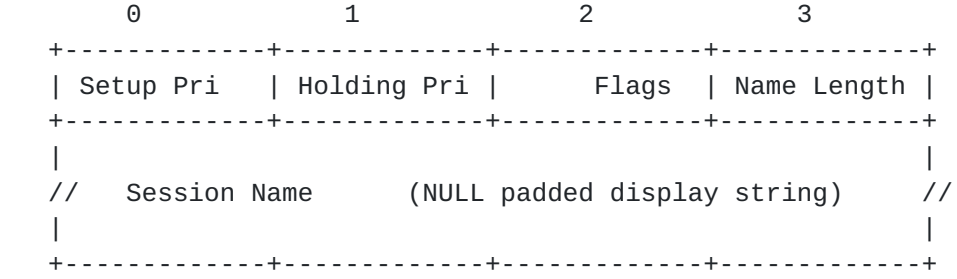
It may be desirable to signal the need for the fast reroutable LSP to be node protected along its path. By node protected we mean that each PLR along the path must protect the fast reroutable LSP with a NNHOP backup tunnel (except for the penultimate hop LSR that will just require a NHOP backup tunnel).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

4.1.1. New SESSION_ATTRIBUTE Flags

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

SESSION_ATTRIBUTE

Class = 207
C-Type = 7 (LSP_TUNNEL)



Current Flags:

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 primary LSP path that they must select a backup tunnel providing the same bandwidth guarantee as the protected LSP.

Node protection desired: 0x10

This flag indicates to the PLRs along the primary LSP path that they must select a NNHOP backup tunnel.

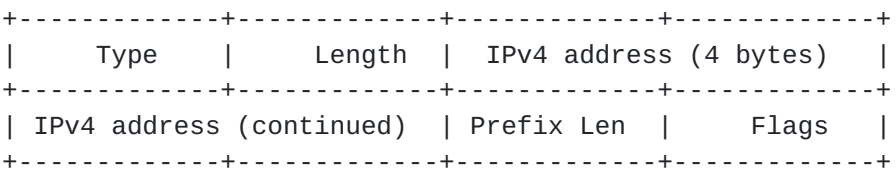
4.1.2. RRO Modification

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

RRO IPv4 sub-object address:

Type: 0x01 IPv4 address

0 1 2 3



```
+-------------+-------------+-------------+-------------+
|     Type    |     Length  |  IPv4 address (4 bytes)   |
+-------------+-------------+-------------+-------------+
| IPv4 address (continued)  | Prefix Len  |     Flags   |
+-------------+-------------+-------------+-------------+
```

Current Flags:

Local protection available: 0x01

Indicates that the link downstream of this node is protected via a local repair mechanism. This flag can only be set if the Local protection flag was set in the SESSION_ATTRIBUTE object of the corresponding Path message.

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).

New Flags:

Bandwidth protection: 0x04

When set, this indicates that the PLR could select a bypass tunnel providing the same bandwidth guaranty as the protected LSP for the protected section.

Node protection: 0x08

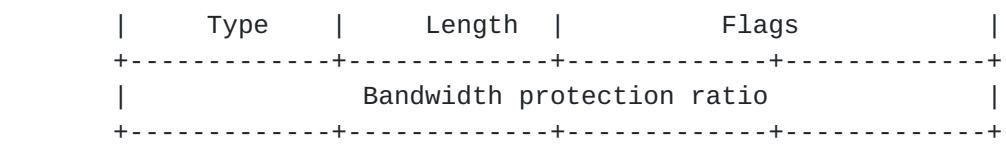
When set, this indicates that the PLR could find a NNHOP backup tunnel providing protection against link and node failure on the corresponding path section. In case the PLR could just find a NHOP backup tunnel, the "Local protection available" bit will be set but the "Node protection" bit will be cleared.

4.1.3. New RRO sub-object: MAX_PROTECTED_BANDWIDTH

This sub-object is carried in the RRO object and is optional.

RRO MAX_PROTECTED_BANDWIDTH sub-object:

```
     0             1             2             3
+-------------+-------------+-------------+-------------+
```



Type: 0x04

Length: 32

Flags:

No Flags are currently defined

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.

4.2. 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.

When originating a Resv message for a fast reroutable LSP, each LSR along the path should include an RRO object. As specified in [RSVP-TE], the RRO object should contain IPv4 sub-objects recording the OUTBOUND interface addresses (with the "Local protection available" and "Local protection in use" bits set accordingly). The only exception is the LSP tail-end that will record the INBOUD IP address. The RRO object should also contain Label Record sub-objects recording the INBOUND labels (same label value as the one sent the Resv message).

Also, the PLR needs to update the RRO to set the "Local protection available" and "Local protection in use" bits. In addition, the inbound IP address of the LSP tail-end should be recorded as well.

When per interface label space is used by the Merge Point, the PLR cannot learn the backup labels using this scheme. This is because the incoming labels learned via the RRO are interface-specific, and do not correspond to the interface on which the MP would receive a rerouted LSP's data packets. When MPs use per interface label space, the PLR must send Path messages (for each Reroutable LSP) via the bypass tunnel prior to the failure in order to discover the appropriate MP label.

4.3. Procedures for the PLR before fast-reroute

When a protected LSP in first signaled, all the PLRs along the path should perform the following:

- If the "Local protection desired" bit is set in the SESSION_ATTRIBUTE, the PLR should select 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, 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. The PLR can optionally add the MAX_PROTECTED_BANDWIDTH subobject in the RRO object of the Resv message.

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

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

4.4. 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 also starts sending the control traffic for the rerouted LSP onto the bypass tunnel.

4.4.1. 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).

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

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

4.4.2. Identification and association of backup tunnels

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

In the SESSION object:

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.

In the SENDER_TEMPLATE object:

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 LSP_ID is used to differentiate multiple LSPs during a tunnel reroute procedure. During this procedure, multiple LSPs each with their own LSP_ID may be active simultaneously. It is quite possible that a node which is downstream of the PLR on the LSP being backed up is also upstream of the PLR for some other LSP associated with the tunnel. It is thus necessary to properly associate the LSP_ID of a detour LSP with the LSP_ID of the LSP being protected.

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.

4.4.3. Backup tunnel message format

When an LSP is being backed up by means of a bypass tunnel, the backup LSP is identified as follows. The SESSION object and the LSP_ID are copied from the protected LSP. The IPv4 tunnel sender address of the SENDER_TEMPLATE is set to an address belonging to the PLR. If the PLR is also the head-end of the protected LSP, it must choose an IP address different from the one used in the SENDER_TEMPLATE originally used to signal the protected LSP.

The Path message for the backup tunnel must obey the following:

SESSION_ATTRIBUTE, SESSION objects are unchanged. Just the IPv4 tunnel sender address of the SENDER_TEMPLATE is changed (set to an address belonging to the PLR).

When multiple protected LSPs use a common backup tunnel, the PLR should use the some IP sender address in the SENDER_TEMPLATE object in all Path messages sent via this backup tunnel.

4.4.4. Procedures for PHOP processing

When the PLR sends RSVP messages via the bypass tunnel, the PHOP object must contain the IPv4 source address (and LIH) of the bypass tunnel (normal procedure as defined in [RSVP]).

Consequently, the MP will send messages back to the PLR with HOP objects containing this same IPv4 address.

Messages sent by PLR via the Backup Tunnel include Path, PathTear, and ResvConf.

Messages sent by MP with this same HOP object contents include Resv and ResvTear.

4.4.5. Procedures for ERO processing

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.

It does this 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.

4.4.6. Procedures for RRO processing

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

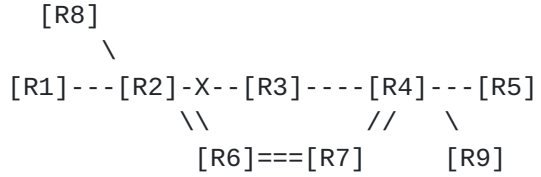
For each rerouted LSP in the backup tunnel, the PLR must update the RRO object in Resv messages sent upstream in the following manner:

- update the IPv4 sub-objects recording the OUTBOUND interface address (with the "Local protection available" and "Local protection in use" bits set accordingly). The only exception is the LSP tail-end that will record the INBOUD IP address.

- update the label sub-object recording the INBOUND label (same label value as the one sent the Resv message)

4.5. Procedures for state maintenance during fast-reroute

We will describe how state is maintained using an example:



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.

4.5.1. 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 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.

4.5.2. 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.

5. Procedures for detour and bypass tunnel computation

To setup the detours described in Section 3 and the bypass tunnels in Section 4, 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 LSP resource information, such as bandwidth. Depending on backup mechanism, such information can be found in the FAST_REROUTE or the SENDER_TSPEC objects. The resource information for a specific bypass tunnel (section 4) may be a function of the protected LSP used for this bypass tunnel but could also be based on some configurable percentage of the link bandwidth to the downstream node. In such a case the amount of bandwidth requested for the bypass tunnel may not be a function of the actual amount of reserved capacity on the protected link.

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 3), the destination MUST be the tail-end of the protected LSP, whereas for setting up bypass tunnels (Section 4), the destination MUST be the address of the MP.

- 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. We also say that the detour LSP/bypass tunnel are diversely routed 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, and 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, the exact CSPF algorithms to be used to compute back-up tunnels paths are beyond the scope of this document. Both [OSPF-TE] and [ISIS-TE] may provide more information 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.

5.1. 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 above 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 this section, the backup tunnel must not traverse the R2-R3 link nor the R3 node.

The notion of SRLG diverse 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.

6. Failure detection mechanisms

Link failure detection can be performed through layer2 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.

7. 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. 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.

The insertion of an RRO object in the Path message may help for troubleshooting purposes.

8. 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.

9. Security Considerations

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

10. 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

11. Acknowledgments

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

12. References

[RSVP] R. Braden, Ed., et al, "Resource ReSerVation protocol (RSVP) -- version 1 functional specification," RFC2205.

[RSVP-TE] D. Awduche, et al, "RSVP-TE: Extensions to RSVP for LSP tunnels" Internet Draft.

[FR-GAN] D. Gan, et al, "A Method for MPLS LSP Fast-Reroute Using RSVP Detours", Internet Draft.

[FR-SWALLOW] R. Goguen and G. Swallow, "RSVP Label Allocation for Backup Tunnels", Internet Draft.

[FR-ATLAS] A. Atlas, et al, "MPLS RSVP-TE Local Protection to Support Fast-Reroute", Internet Draft.

[OSPF-TE] Katz, Yeung, Traffic Engineering Extensions to OSPF, draft-katz-yeung-ospf-traffic-05.txt, June 2001.

[ISIS-TE] Smit, Li, IS-IS extensions for Traffic Engineering, draft-ietf-isis-tr affic-03.txt, June 2001.

13. 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
email: swallow@cisco.com
phone: +1 978 244 8143

Jean Philippe Vasseur
Cisco Systems, Inc.
11, rue Camille Desmoulins
92782 Issy les Moulineaux Cedex 9,
France
email: jvasseur@cisco.com
phone: +33 689108267

Dave Cooper
Global Crossing
960 Hamlin Court
Sunnyvale, CA 94089
email: dcooper@gblx.net
phone: +1 916 415 0437

Alia Atlas
Avici Systems
101 Billerica Avenue
N. Billerica, MA 01862
email: aatlas@avici.com
phone: +1 978 964 2070

Markus Jork
Avici Systems
101 Billerica Avenue
N. Billerica, MA 01862
email: mjork@avici.com
phone: +1 978 964 2142