Network Working Group J. Jeganathan
Internet-Draft H. Gredler
Intended status: Standards Track Y. Shen
Expires: October 28, 2013 Juniper Networks
April 26, 2013

# RSVP-TE LSP egress fast-protection
# draft-minto-rsvp-lsp-egress-fast-protection-02

## Abstract

RFC4090 defines a fast reroute mechanism for locally repairing an RSVP-TE LSP in the order of 10s of milliseconds, in the event of a downstream link or node failure. However, this mechanism does not provide node protection for LSP egress nodes, even when an alternate egress node (a backup egress) is available that could carry the traffic to its ultimate destination. This document addresses this scenario and describes how to provide egress protection by establishing a bypass LSP from the penultimate-hop node of a LSP to the backup egress node. The methods described in this document enable local repair in the order of 10s of milliseconds, in the event of the egress node failure. These methods are only applicable if traffic carried by the LSP can be rerouted to its ultimate destination by the backup egress node.

## Status of This Memo

Copyright Notice

Table of Contents

## 1. Introduction

This document describes procedures for providing fast protection for RSVP-TE LSPs in case of the egress node failure. Such protection can only be provided when an alternate egress node exists that can carry the traffic destined for the protected egress to its ultimate destination. The primary egress node of an LSP (the protected egress) terminates the LSP in steady state, while the alternate egress node (the backup egress) does so when the primary fails. A bypass LSP is established from the penultimate-hop node to the backup egress. The penultimate-hop node, serving as a PLR (point of local repair), redirects traffic to the backup egress node of the LSP using this bypass LSP in the event of primary egress node failure.

The backup egress node forwards the traffic to its ultimate destination using methods that are beyond the scope this document. For example, backup egress node could use the service specific mechanism specified in [pwe3-endpoint-fast-protection] and [l3vpn-egress PE-fast-protection] and mirror the inner labels (e.g. layer-2 /3 VPN service labels) from the primary on the backup. The backup would then repair the traffic to its destination based on the mirrored labels. This document focuses on the methods for setting up the bypass LSP to the backup egress so that service specific mechanism could build top on this.

```
                 [R1]                        [R8]
                     \                      /
                [R2]---[R3]----[R4]-----[R5]---[R6]
                          \             /  \\
                           [R9]-----[R10]   [R7]

                 Protected LSP to-R6.x:   [R1->R3->R4->R5->R6.x]
                 Protected LSP to-R6.y:   [R1->R3->R4->R5->R6.y]
                 Protected LSP to-sec-R6.x:   [R1->R3->R9->R10->R5->R6.x]
                 Protected LSP to-R8.z:   [R2->R3->R4->R5->R8.z]
                 x, y, z: Tunnel destination addresses.
                 R6 has x,y destination addresses.
                 Egress-Bypass LSP Tunnel by-R7.x: [R5->R7.x]
                 Egress-Bypass LSP Tunnel by-R7.y: [R5->R7.y]
                 Egress-Bypass LSP Tunnel by-R7.z: [R5->R7.z]
```

Figure 1

In Figure 1, four LSPs require egress protection. R6 and R8 are the primary egresses. R7 is backup egress for both R6 and R8. R5 is the penultimate hop node. R5 establishes a bypass LSP to R7 to provide fast protection in case R6 or R8 fail. Table 1 shows the bypass LSPs for each of the protected LSPs at R5.

| Protected LSP | Egress Bypass LSP |
|---|---|
| to-R6.x | by-R7.x |
| to-R6.y | by-R7.y |
| to-sec-R6.x | by-R7.x |
| to-R8.z | by-R7.z |

Table 1

This draft describes two methods for setting up the bypass LSP to the backup egress node, the proxy node method and the alias method.

In the proxy method, an LSP endpoint address is represented as a virtual node in the TE domain, attached to the primary egress node and the backup egress node via bidirectional point-to-point TE links.

```
 [R1]                        [R8]
     \                      /
[R2]---[R3]----[R4]-----[R5]---[R6]---[x]
          \             /  \         /
           [R9]-----[R10]   [R7]---+
```

x: Tunnel destination addresses in the proxy method.

Figure 2

With the proxy method, when providing egress protection to the LSPs with destination address x, terminating on primary R6, with backup egress R7, from Figure 1, the topology is modeled as shown in Figure 2.

With this representation, penultimate-hop node R5 could use RFC 4090 RSVP fast-reroute PLR procedures to set up a bypass LSP to the backup egress node R7, by avoiding the primary egress node R6.

In alias method, an LSP endpoint address is associated with a primary egress and a explicit backup egress. The penultimate-hop node of the protected LSP may learn the backup for the LSP from backup egress IGP advertisement or by a local configuration. With this method, the penultimate-hop node can set up a bypass LSP to the backup egress node, by avoiding the primary egress node.

```
            [R1]                      [R8]
                \                     /
            [R2]---[R3]----[R4]-----[R5]---[R6]x
                     \             /  \
                       [R9]-----[R10]   [R7](x)
           x: Tunnel destination addresses.
           R6 x: R6 primary egress for x.
           R7(x): R7 Backup egress for x.
```

Figure 3

In Figure 3, let say x is tunnel destination address and R6 advertise x as secondary loopback address. With this alias representation R5 see the x as x{R6,R7} where R6 is primary and R7 is backup for x. This primary to backup mapping is either learn through R7's IGP backup availability advertisement or by a local configuration in R5.

## 2. Specification of Requirements

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

## 3. Terminology

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

PHN: Penultimate Hop Node for an LSP.

Primary egress node: Node terminates a LSP in steady state.

Primary: Primary egress node.

Egress Protected LSP: A Protected LSP that also required protection from primary egress node failure

Backup egress node: Node could rerouted/repaired data carried in a protected LSP

Backup node: Backup egress node.

Protector: Backup egress node.

Protector and Backup node are used interchangeably but convey the same meaning.

## 4. Proxy method

In this method, an LSP endpoint address is represented as a virtual TE node connected to a primary egress node and a backup egress node with bidirectional TE links, as shown in Figure 2. With this model, node protection establishment and bypass LSP path computation on the penultimate hop of an LSP can follow the procedure described in RFC4090.

### 4.1. Tunnel destination Advertisement in IGP

Advertising the tunnel destination as a stub proxy TE node requires two steps: 1) a node representation (proxy-node) and 2)links to and from primary egress and backup egress.

The primary advertises a proxy node with two links, to the primary egress and the backup egress, respectively. The router ID of the proxy node is LSP end point address. The system-ID of the proxy is derived from the LSP end point address with BCD encoding. The resulting system-ID and router-ID MUST be unique within the IGP routing domain.

Both stub links are advertised with maximum routable metric and TE metric, and zero bandwidth, as shown in Figure 4. This avoids the proxy node serving as a transit node for any path. The router-ID or system-ID of the backup egress can be dynamically learned from the IGP link state database or can be configured on the primary egress.

```
     primary egress -
                      \ metric 1, TE metric 1, bandwidth max
                       \
                        \
                         \
                          \ metric max, TE metric max, bandwidth 0
                           |
                       proxy node [stub node]
                           |
                          / metric max, TE metric max, bandwidth 0
                         /
                        /
                       /
                      / metric max, TE metric max, bandwidth 0
       backup egress-
```

Figure 4

The primary egress advertises an unnumbered transit link to the proxy node, with metric 1, TE metric 1, and maximum bandwidth. It may be necessary for the primary node to have the capabilities to advertise multiple TE unnumbered transit links between primary node and proxy-node. The upper bound on the number of such links is the number of the links the primary node advertises into TE.

The backup egress advertises an unnumbered transit link to the proxy node, with MAX metric, MAX TE metric, and zero bandwidth. Other TE characteristic of the links can be configured and advertised as well.

### 4.1.1. IS-IS proxy-node (Non-Normative)

When IS-IS is used as IGP to advertise the proxy node, only zeroth fragment of the proxy-node advertisement is valid. All other fragments SHOULD be ignored. The zeroth fragment MUST include the area address TLV and MAY include the hostname TLV.

The set of area addresses advertised in proxy node zeroth fragment link-state PDU MUST be a subset of Area Addresses advertised by the primary egress in the zeroth fragment of the link-state PDU of the corresponding IS-IS level. The advertisement SHOULD be syntactically identical to the primary egress zeroth fragment at corresponding IS-IS level. The hostname SHOULD be set as <tunnel-destination + primary egress hostname>.

The Overload (OL) MUST be set to 1. The Attached (ATT), and Partition Repair (P) bits MUST be set to 0.

### 4.1.2. OSPF proxy-node (Non-Normative)

The advertising router and Link State ID of router LSA MUST be LSP end point address. All options bits in router LSA MUST be set to zero.

## 4.2. Ingress Node Behavior

The ingress node of an LSP requesting egress protection SHOULD follow the procedures described in RFC 2205 and RFC 4090 to signal the LSP. In particular, it SHOULD set the destination to the endpoint address (i.e. the proxy node), and the "link protection desired" flag and the "node protection desired" flag in the SESSION_ATTRIBUTE object of the Path message. In path computation, it MAY optionally exclude MAX metric links to avoid the link between the backup egress and the proxy node.

### 4.3. Primary Egress Node Behavior

When the primary egress node receives a Path message for the LSP with destination matching the proxy node address, it MUST append two entities in the RRO object of Resv message: 1) the proxy node as a virtual downstream node, and 2) itself as a virtual transit node. The entity for the proxy node is encoded as {proxy node address, proxy link ID, implicit NULL}.

### 4.4. Penultimate Hop Node

When the penultimate hop node receives a Resv message from the primary egress, it sees itself as two hops away from LSP's destination rather than one hop, based on the RRO. Thus, it can set up node protection for the LSP by following the procedure described in RFC 4090. It SHOULD set up a bypass LSP to the backup egress node. When computing a path for the bypass LSP, it SHOULD avoid the primary egress node and choose a path via the backup egress node to reach the proxy node.

#### 4.4.1. Backup LSP Signaling during Local Repair

The penultimate hop node SHOULD uses the same procedure as defined RFC4090 to signal the backup Path, in the event of failure of the primary egress node.

### 4.5. Backup Egress Node Behavior

When the backup egress node receives Path message of the bypass LSP, it MUST terminate the Path message based on match between the LSP destination and the proxy node address. It SHOULD assign a non-reserved label to the bypass LSP. This non-reserved label provide forwarding context during repair.

#### 4.5.1. Backup LSP Signaling during Local Repair

During local repair, the backup egress node will receive Path message of egress-protected LSP from the penultimate hop node. The backup egress node SHOULD terminate the Path message, and respond with a Resv message.

### 4.6. Proxy method solution characteristics

The biggest advantage of the proxy method is that it does not require protocol extensions and can be implemented locally at the tunnel egress node. Thus, no software upgrades are required in the core of the network.

The proxy method has the following caveats:

1. To support TE constrains like colors and SRLG for a protected LSP the primary needs to have the capability to advertise multiple links to between proxy and primary.

2. Bypass LSP with constrains cannot be supported.

3. If IS-IS is used as the IGP then the Primary node should not be configured with overload bit.

4. Backup egress could be used as primary end point in the forwarding plane if the protected LSP established to backup instead of primary in transient condition.

Due to its characteristics, the proxy method is suitable for mixed environments, where an upgrade of the entire network is not feasible.

## 5. Alias model

In this model Penultimate hop node (PHN) of a protected LSP understands that LSP end point has a backup egress and it could repair traffic carried in the protected LSP in the event of primary egress failure. After the primary egress failure, the PHN reroutes traffic using a bypass tunnel to backup egress. The tunnel endpoint address and backup egress mapping could be configured in penultimate hop node or signaled through IGP from the backup. Table 2 illustrates the PHN mapping primary to backup mapping for the topology in Figure 1.

| Primary Egress | Backup egress | Backup LSP destination |
|---|---|---|
| Router ID | router ID | address. |
| 10.1.2.6 | 10.1.1.6 | 10.1.1.7 |
| 10.1.2.6 | 10.1.3.6 | 10.1.1.6 |
| 10.1.1.7 | 10.1.3.6 | 10.1.2.8 |
| 10.1.1.8 | 10.1.1.7 | 10.1.2.8 |

Table 2: Table mapping

### 5.1. Ingress Behavior

The ingress should follow the procedure in RFC 3209 with tunnel endpoint address. The path computation could use the tunnel endpoint address advertised using the procedures of RFC 5786.

### 5.2. Primary Egress node

Primary egress node advertises tunnel end points that require protection using RFC 5786 in OSPF and/or IP interface addresses TLV(132) in IS-IS.  These TLVs are defined as Local address advertisement in TE.  The rest of the behavior is same RFC 4090.

### 5.3. Backup egress node

When backup receives a Path message not through a bypass tunnel for a destination address it protects with ERO constrains only one self sub objects then it MUST accept and respond with RRO objects in Resv message.  The RRO object {node ID, Ip address, label} for tunnel end address set with {Node ID, tunnel endpoint address, non-reserved label}. This non-reserved label provide forwarding context during local repair.

#### 5.3.1. Procedures for the Backup egress during Local Repair

The Backup egress sends Resv, ResvTear, and PathErr messages by sending them directly to the address in the RSVP_HOP object, as specified in [RSVP-TE].

#### 5.3.2. Processing Backup Tunnel's ERO

When backup receive Path message through a bypass tunnel with one sub-object for destination address it protects then it should accept ERO.

### 5.4. Penultimate hop node

PLR learns/configured backup egress for tunnel a end point address advertised by primary egress.  When PLR setup bypass for node protection LSP it will also lookup for the backup egress if PLR is penultimate hop of the LSP.  If backup egress is available for LSP tunnel end point address then it setup bypass-LSP to backup egress if it is not setup already.  The constrains will be exclude egress node. PHN could setup bypass-LSP with destination as backup egress node or tunnel endpoint address.  If the bypass tunnel endpoint address is not the protected LSP tunnel endpoint then it also initiates backup LSP for tunnel end point address through bypass tunnel to learn the label to use in failure.

#### 5.4.1. Signaling a Backup Path

PHP SHALL uses the same procedure as defined RFC4090 to signal the backup Path.

#### 5.4.2. Procedures for Backup Path Computation

PLR has to find the desired explicit route for the backup path. This can be done using a CSPF computation. If PLR is PHN for the protected LSP needs node protection then destination for the backup path MUST be backup egress router ID with the constraint that the LSP cannot traverse the primary egress node and/or link whose failure is being protected against. For other constrains SHOULD follow RFC4090.

#### 5.4.3. Signaling for Facility Protection

A PHN use one or more bypass tunnels to protect against the failure of a egress primary node. This bypass tunnels set up in advance or dynamically created as new protected LSPs are signaled.

##### 5.4.3.1. Discovering Downstream Labels

To support facility backup, the PHN must determine the label that will indicate to the backup egress that packets received with that label should be processed by primary egress context. This can be done by setting up the UHP bypass tunnel to the backup egress with tunnel endpoint address as destination.

##### 5.4.3.2. Processing Backup Tunnel's ERO

Sub-objects belonging to abstract nodes that precede the tunnel endpoint Point are removed. A sub-object identifying the Backup Tunnel destination is then added.

##### 5.4.3.3. PLR Procedures during Local Repair

PHN SHALL uses the procedures defined in RFC4090 during the local repair.

### 5.5. Alias method solution characterization

The alias method will work with arbitrary TE constraints and suitable for networks that required LSP with those TE constraints. To avoid PLR static backup egress configuration, IGP extension is required.

## 6. Security Considerations

The security considerations discussed in RFC 5036, RFC 5331, RFC 3209, and RFC 4090 apply to this document.

## 7. Acknowledgements

This document leverages work done by Hannes Gredler, Yakov Rekhter and several others on LSP tail-end protection.  Thanks to Ina Minei, Nischal Sheth, Nitin Bahadur, Ashwin Sampath and Kaliraj Vairavakkalai for their contribution.

## 8. References

### 8.1. Normative References

[RFC5331]  Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, August 2008.

[RFC4364]  Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC5036]  Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

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

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

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

[RFC3471]  Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.

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

[LDP-UPSTREAM]
Aggarwal, R. and J. L. Le. Roux, "MPLS Upstream Label Assignment for LDP", draft-ietf-mpls-ldp-upstream (work in progress), 2011.

[RSVP-NON-PHP-OOB]
Ali, A., Swallow, Z., and R. Aggarwal, "Non PHP Behavior and out-of-band mapping for RSVP-TE LSPs", draft-ietf-mpls-rsvp-te-no-php-oob-mapping (work in progress), 2011.

## 8.2. Informative References

[RFC5286] Atlas, A. and A. Zinin, "Basic Specification for IP Fast Reroute: Loop-Free Alternates", RFC 5286, September 2008.

[RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, January 2010.

[pwe3-endpoint-fast-protection]
Shen, Y., Ed., Aggarwal, R., , "PW Endpoint Fast Failure Protection", 2011, <pwe3-endpoint-fast-protection>.

[l3vpn-egress-PE-fast-protection]
Jeganathan, J. and G. Gredler, "2547 egress PE Fast Failure Protection", 2011, <2547-egress-PE-fast-protection>.

Authors' Addresses

Jeyananth Minto Jeganathan
Juniper Networks
1194 N Mathilda Avenue
Sunnyvale, CA  94089
USA

Email: minto@juniper.net

Hannes Gredler
Juniper Networks
1194 N Mathilda Avenue
Sunnyvale, CA  94089
USA

Email: hannes@juniper.net

Yimin Shen
Juniper Networks
10 Technology Park Drive
Westford, MA  01886
USA

Email: yshen@juniper.net