Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 15, 2013

J. Jeganathan
H. Gredler
Y. Shen
Juniper Networks

Oct 12, 2012

RSVP-TE LSP egress fast-protection-00
draft-minto-rsvp-lsp-egress-fast-protection-00

Abstract

RFC4090 defines an RSVP fast reroute mechanism for local repairing LSP tunnel in the order of 10s milliseconds, in the event of a downstream link or node failure. However, the mechanism does not provide node protection for LSP egress nodes. This document describes two methods to establish a bypass LSP from the penultimate-hop node of an LSP to a backup egress node, which could be used to protect the LSP against egress node failure. The methods enable local repair in the order of 10s of millisecond, in the event of the egress node failure. These methods are only applicable if traffic carried by the LSP could be rerouted to ultimate destination by the backup egress node.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

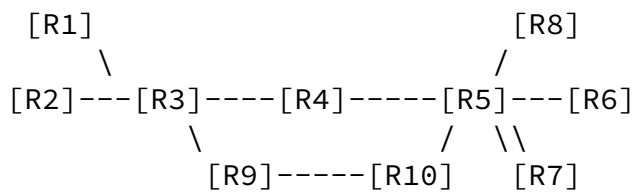
This Internet-Draft will expire on April 15, 2013.

Copyright Notice

Table of Contents

1. Introduction

This document define two methods that could enable fast protection for egress node failure for RSVP-TE signaled LSP tunnels. Both methods have a common concept of primary egress node and backup egress node for a tunnel endpoint address. The methods differ by how tunnel endpoints are modeled in the network. The primary egress node of an LSP (called protected LSP) terminates the LSP in steady state, while a bypass LSP is established from the penultimate-hop node to the backup egress node. The penultimate-hop node, serving as a PLR (point of local repair), redirects traffic to the backup egress node of the LSP via the bypass LSP in the event of primary egress node failure, and the backup egress node forwards the traffic to the ultimate destination. How the backup egress node forwards traffic is beyond the scope this document. For one example, the backup egress node could mirror from the primary egress node the inner labels (e.g. layer-2/3 VPN service labels) carried by the traffic, and forward the traffic based on those labels by using the mechanisms specified in [pwe3-endpoint-fast-protection] and [l3vpn-egress PE-fast-protection].



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]
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]
x, y, z: Tunnel destination addresses.
R6 has x,y destination addresses.

Figure 1

In Figure 1, 4 LSPs are required egress protection. R6 and R8 are the primary egresses for 4 LSPs, R7 is backup egress and R5 is penultimate hop node for all LSPs. R5 establish bypass LSP to R7 for fast protection to handle the R6 or R8 failure. Below table shows the protected LSP and bypass LSP in R5.

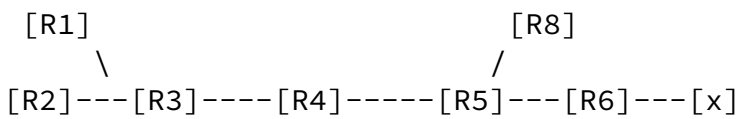
Protected LSP	Egress Bypass LSP
to-R6.x	by-R7.x
to-R6.y	by-R7.y
to-sec-R6.x	by-R7.x
to-R8.z	by-R7.z

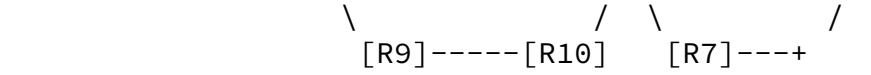
Two methods defined in the documents that enable the backup LSP to establish to backup egress.

a. Proxy node method

b. Alias method

In the proxy method, an LSP endpoint address is represent 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. With this representation, the penultimate-hop node of the LSP could use the normal procedure of RSVP fast-reroute PLR to set up a bypass LSP to the backup egress node, by avoiding the primary egress node. This methed has the advantage of not requiring software upgrade on the penultimate-hop node, and thus can ease the deployment this technology.





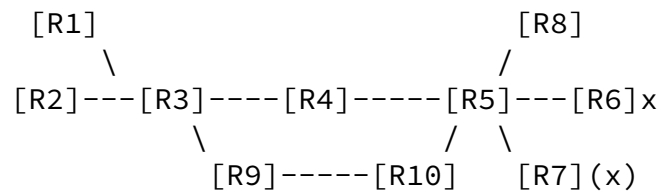
x: Tunnel destination addresses in proxy method.

Figure 2

With proxy method, topology is modeled as figure 2 in the rest of the network for LSP destination address x which required egress protection and R6 is primary R7 is backup.

In alias method, an LSP endpoint address is associated with an dedicated IP address on the backup egress node. This IP address is called an alias. The penultimate-hop node of the LSP may learn the alias via IGP or configuration, and use it as the destination when computing a path for the bypass LSP. With this method, the penultimate-hop node can set up a bypass LSP to the backup egress node, by avoiding the primary egress node. This method requires software upgrade penultimate-hop node, but is flexile to support all

traffic engineering constraints.



x: Tunnel destination addresses in alias method.

Figure 3

In figure 3, let say x is tunnel destination address and R6 is primary and R7 is backup then with alias method, R6 advertises x as secondary loopback address and R5 knows x has backup either by configuration or R7 advertisement in IGP.

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.

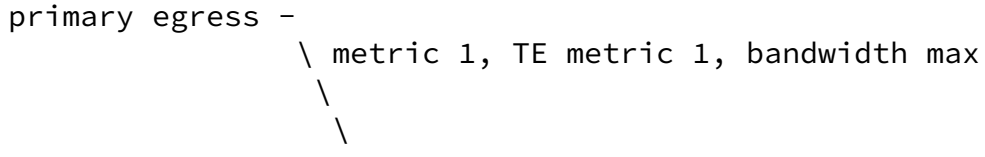
Protector: Backup egress node.

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 figure 1. 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.



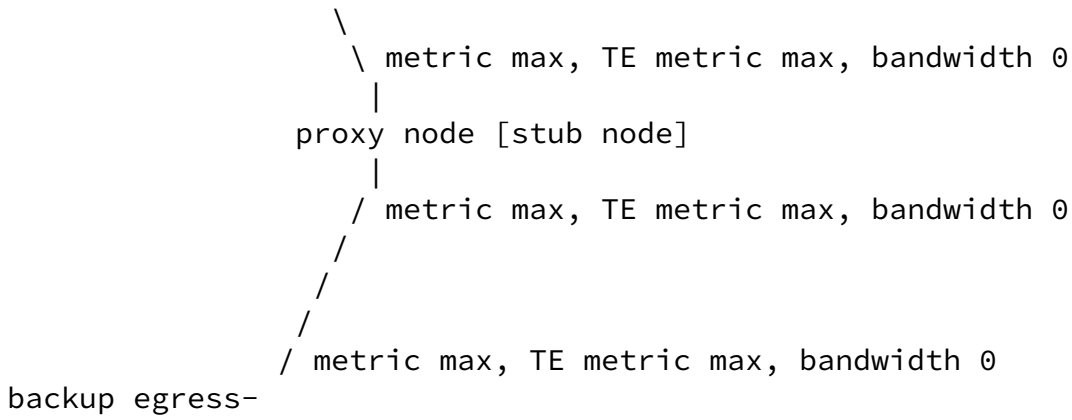


Figure 4

4.1. Tunnel destination Advertisement in IGP

Tunnel destination advertised as stub proxy TE node required two parts. A node representation (proxy-node) and links to and from primary egress and backup egress..

The primary advertises proxy node with two links to primary egress and backup egress, respectively. The router ID of the proxy node is LSP end point address. The system-ID is derived from the LSP end point address with BCD encoding. The resulting system-ID and router-ID MUST be unique with in IGP routing domain. Both stub links are advertised with maximum routable metric and TE metric, and zero bandwidth. This avoids the proxy node serve as a transit node for any paths. The router-ID or system-ID of the protector could be dynamically learned from IGP link state database or could be configured in primary.

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 could be configured and

advertised in to TE.

4.1.1. ISIS proxy-node (Non-Normative)

Only zeroth fragment of the proxy-node is only valid. All Other fragments SHOULD be ignored. Zeroth fragment MUST include area address TLV and MAY include hostname TLV.

The set of area addresses advertised MUST be a subset of the set of Area Addresses advertised in the protected LSP number zero at the corresponding level. Preferably, the advertisement SHOULD be syntactically identical to that included in the normal LSP number zero at the corresponding level. The hostname could be set as <tunnel-destination + protected 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 be LSP end point address. All options bits in router LSA MUST be set to zero. The number of links MUST be 2

4.2. Ingress Node Behavior

The ingress node of an LSP should follow same procedure 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 SESSION_ATTRIBUTE of Path message. In path computation, it MAY optionally set not to use MAX metric link, as another constraint, to avoid the link between the backup egress and the proxy node.

4.3. Primary Egress Node Behavior

When the primary egress node receives Path message for the LSP with destination matching the proxy node address, it MUST append two entities in the RRO object of Resv message, first for the proxy node as a virtual downstream node, and second for itself as 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 Resv message from 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 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 the Path message of the bypass LSP, it MUST terminate the Path message based on the match bewteen the LSP destination and the proxy node address. It SHOULD assign a non-reserved label to the bypass LSP, and point the label to a specific label table where the labels learned from the primary egress node are installed. This can facilitate forwarding of traffic when the backup egress node receives traffic over the bypass LSP during local repair. In this case, the traffic will be carrying inner labels assigned by the primary egress node, and a further label lookup in the specific label table SHOULD enable the backup egress node to forward traffic to the ultimate destination.

4.5.1. Backup LSP Signaling during Local Repair

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

4.6. Pros/Cons

Pros

1. Protocol extension not required. Changes required only in tunnel egress nodes. Core router software upgrade required is not required.

Cons

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

2. Bypass LSP with constrains cant be supported.

3. If ISIS used as IGP then Primary node should not configured with overload bit.

4. if OSPF as IGP then a Proxy node could be used in transit even if primary is down.

5. Protector could be used as primary end point in the forwarding plane if the protected LSP established to protector instead of primary in transient condition

5. Alias model

In this model Penultimate hop node understand tunnel end point has a backup egress which is may not protected LSP path and backup egress could repair traffic carried protected LSP in the event of primary egress failure. After primary egress failure PHN reroute using 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. Following table illustrate the PNH mapping primary to backup mapping for the figure 1.

Primary Egress Router ID	Backup egress router ID	Backup LSP destination 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 1: Table mapping

5.1. Head-End Behavior

Ingress should follow same procedure in RFC 3209 with tunnel endpoint address and path computation could use RFC 5786 advertised tunnel endpoint address.

5.2. Primary Egress node

Primary egress node advertises tunnel end points that required protection using RFC 5786 in OSPF and/or IP interface addresses TLV(132) in ISIS. These TLVs are defines as Local address advertisement in TE. And rest of 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 constains 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-NULL}. This non-NULL will be used for identify LSP it protects in forwarding. Backup could also signals protection availability for tunnel end point addresses through IGP.

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. PNH 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 PNH for the protected LSP needs node protection then destination for backup path MUST be backup egress router ID with constrain that 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 explicitly signaling backup path before failure or setup the UHP bypass tunnel to 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 same procedure as defined RFC4090 during the local repair.

5.5. Pros/Cons

Pro

1. Will work with any TE constrains

Cons

1. Protocol changes required in RSVP. Also IGP extension required to avoid PLR static protector configuration.

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 Nischal Sheth, Nitin Bahadur, Yimin shen, 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. 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. and 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