Network Working Group
Internet Draft
Category: Standard Track
Expires: September 2007

J.L. Le Roux
France Telecom

R. Aggarwal
Juniper Networks

J.P. Vasseur
Cisco Systems, Inc.

M. Vigoureux
Alcatel-Lucent

March 2007

P2MP MPLS-TE Fast Reroute with P2MP Bypass Tunnels

draft-leroux-mpls-p2mp-te-bypass-01.txt

Status of this Memo

Internet Draft draft-leroux-mpls-p2mp-te-bypass-01.txt March 2007

Abstract

This document defines procedures for fast reroute protection of Point-To-MultiPoint (P2MP) Traffic Engineering Label Switched Paths (TE-LSP) in MultiProtocol Label Switching (MPLS) networks, based upon Point-To-MultiPoint bypass tunnels. The motivation for using P2MP bypass tunnels is to avoid potentially expensive data duplication along the backup path that could occur if point-to-point bypass tunnels where used, i.e. to optimize the bandwidth usage, during fast reroute protection of a link or a node. During link or node failure the traffic carried onto a protected P2MP TE-LSP is tunnelled within one or several P2MP bypass tunnels towards a set of Merge Points. To avoid data duplication backup labels (i.e. inner labels) are assigned by the Point of Local Repair (PLR) following the RSVP-TE upstream label assignment procedure.

Conventions used in this document

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.

Table of Contents

1. Terminology....3
2. Introduction....3
3. Solution overview....4
4. PLR procedures....6
4.1. Before failure....6
4.1.1. P2MP Bypass Tunnel(s) Selection....6
4.1.2. P2MP Backup LSP Signalling....7
4.2. During failure....8
5. MP Procedures....8
6. To be included in future revisions....9
7. Security Considerations....9
8. Acknowledgments....9
9. References....9
9.1. Normative references....9
10. Authors' Addresses:....10
11. Intellectual Property Statement....11

1. Terminology

This document uses terminologies defined in [RFC3031], [RFC3209], [RFC4090] and [RFC4461]. It defines the following new terms:

P2MP Bypass tunnel: Point-to-Multipoint Bypass Tunnel. A P2MP TE-LSP that is used to protect a set of P2MP TE-LSPs traversing a common facility (link or node).

P2MP Facility Backup: A local repair method in which a P2MP bypass tunnel is used to protect one or more P2MP TE-LSPs that traverse the Point of Local Repair (P2MP Bypass Ingress) and the resource being protected.

Backup P2MP LSP: The LSP that is used to backup up one of the many protected P2MP LSPs in P2MP Facility Backup.

Backup S2L sub-LSP: A S2L sub-LSP of a backup P2MP LSP.

PLR: Point of Local Repair: Head-end LSR of the bypass tunnel

MP: Merge Point: LSR where a primary LSP and its backup LSP merge.

.

2. Introduction

[RFC4090] defines fast reroute extensions to RSVP-TE [RFC3209] for local protection of Point-To-Point (P2P) Traffic Engineered Label Switched Paths (TE LSP) in MultiProtocol Label Switching (MPLS) networks. Two techniques are defined: the one-to-one backup method, which creates a detour LSP for each protected LSP at each point of local repair (PLR), and the facility backup method, which creates a

bypass tunnel that can be used to protect a set of TE LSPs by taking advantage of MPLS label stacking.

[RSVP-P2MP] defines extensions to RSVP-TE for setting up Point-To-Multipoint (P2MP) TE LSPs. It specifies extensions to one-to-one and facility backup Fast Reroute procedures defined in [RFC4090] so as to support fast reroute protection of P2MP TE LSPs.
The facility backup solution defined in [RSVP-P2MP] only relies on P2P bypass tunnels for link and node protection. This faces the following limitations:

- The protection of a downstream link of a P2MP TE LSP on a branch LSR may require a P2P Bypass LSP that uses another downstream link of the P2MP LSP, and this leads to twice the traffic on that link during failure, which is inefficient. Finding a bypass path that avoids all downstream links on the P2MP LSP would be a solution but this is often not achievable in lowly meshed topologies.

- The protection of a P2MP TE LSP against node failures requires, when the protected node is a Branch LSR, a set of P2P Next-Next-Hop (NNHOP) Bypass tunnels toward all LSRs downstream to the protected node. During failure the PLR has to replicate traffic on each P2P NNHOP bypass tunnel. If there are K next-next-hops, this may lead to K times the traffic on some links, which is not acceptable (as K is of the order of magnitude of the squared node degree).

- Similarly the protection of a P2MP TE LSP against the failure of a LAN interface that connects a branch LSR and a set of K downstream LSRs requires one P2P bypass tunnel per downstream LSR, which may lead to K times the traffic on some links during failure.

To overcome these limitations it is highly desirable to define extensions to the fast reroute facility backup solution, so as to support P2MP bypass tunnels. This retains the scalability advantages of MPLS label stacking and avoids sending multiple copies of a packet on some links during failure.

This draft specifies extensions to the Fast ReRoute (FRR) procedures defined in [RFC4090] and [RSVP-P2MP] to support local repair of P2MP

TE LSP with P2MP bypass tunnels.

Procedures defined in [RFC3209], [RFC4090] and [RSVP-P2MP] MUST be followed unless specified below.

3. Solution overview

The P2MP Facility Backup method defined in this document relies on the use of P2MP bypass tunnels. Similarly to the P2P case, the same P2MP bypass tunnel can be used to protect a set of P2MP TE LSPs, by taking advantage of MPLS label stacking.

A P2MP Bypass tunnel can be used to protect a P2MP TE-LSP against downstream link or node failures. There are various options for the protection of a downstream link or node:

- Rely on a single P2MP bypass tunnel whose leaf LSRs exactly matches the set of Merge Points (MP). Merge points are transit or egress LSRs on the protected P2MP LSP downstream to the PLR or downstream to the protected element (link or node).
- Rely on a single P2MP Bypass tunnel whose set of leaf LSRs is a superset of the set of MPs. Leaf LSRs which are not MP have to drop the traffic.
- Rely on a combination of P2MP bypass LSPs whose leaf LSRs are a subset of the set of MP but there combination encompass all MPs.

These three options differ in terms of bandwidth optimization and control plane state minimization. Option 1 increases the number of states compared to option 2 (it implies more P2MP bypass LSPs), but is less expensive in terms of bandwidth (traffic only sent to MPs). With point-to-multipoint hierarchy there is always a tension between minimizing the amount of control plane state and minimizing bandwidth consumption. Choosing one of these options is a decision local to the PLR. The choice depends on the desired trade-off between control plane and data plane optimization, and the operational complexity associated with the different options.

When the P2MP facility backup method is used, during failure the PLR MUST send data for each protected P2MP LSP into the set of one or more P2MP bypass tunnel. Label stacking is used: the inner label is the backup label for the backup P2MP LSP, that will be used on the MP

to forward traffic to the corresponding protected P2MP LSP, and the outer label is the P2MP bypass tunnel label.

To avoid data replication on the PLR, the same backup label MUST be used for all S2L sub-LSPs of a given backup P2MP LSP, tunneled within the same P2MP bypass tunnel. This backup label will indicate to the Merge Points that packets received with that label should be switched along the protected P2MP LSP.

For that purpose upstream label assignment procedures defined in [MPLS-UPSTREAM] and RSVP-TE extensions for upstream label assignment defined in [RSVP-UP] MUST be used. To signal a backup P2MP LSP, the same backup label, is distributed by the PLR to all MPs belonging to a same P2MP Bypass tunnel, in the context of this P2MP bypass tunnel. This requires the backup P2MP LSP to be signalled prior to the failure.

On the MP, backup S2L sub-LSPs (i.e. S2L sub-LSPs of the backup P2MP LSP) are merged with protected S2L sub-LSPs. A MP (i.e. the bypass tunnel leaf LSRs), maintains a context specific ILM for the P2MP Bypass tunnel. This can be implemented by maintaining a different context specific ILM for each LSR that is the root of a P2MP Bypass tunnel, or by maintaining a different context specific ILM for each P2MP Bypass tunnel. The context of an inner label (i.e a backup label) is determined by the underlying P2MP bypass tunnel on which it is received. This requires deactivating PHP on the P2MP bypass tunnel. A label, in a given Bypass tunnel specific ILM, is mapped to the outgoing interface(s) and label(s) of the corresponding protected P2MP LSP.

4. PLR procedures

4.1. Before failure

4.1.1. P2MP Bypass Tunnel(s) Selection

To protect a P2MP TE LSP against a downstream link or node failure, a PLR MUST select a set of one or more P2MP bypass tunnel(s), denoted {B1.Bm}, as follows:

- The bypass tunnel(s) MUST NOT traverse the protected link/node/SRLG.

- The set of leaf LSRs of bypass tunnels {B1.Bm}, denoted {LSR1. LSRn} must include a set of Merge Points (MP), on the protected P2MP LSP. These Merge Points are transit or egress LSRs on the protected P2MP LSP downstream to the PLR or downstream to the protected element (link or node). We will denote this set of Merge Points as {MP1.MPq}. Note that the case where some MPs are LSRs downstream to the PLR but not downstream to the failed element allows avoiding sending twice the traffic on downstream links during failure.

- In the event of failure of the protected link or node, traffic received on the protected P2MP LSP by the PLR, can be delivered to all the leaves of the protected P2MP LSP downstream to the PLR, if it is tunnelled to {MP1.MPq} over the set of one or more P2MP bypass tunnel(s) {B1.Bm}.

The PLR will assign upstream labels to Merge Points {MP1.MPq} for the backup P2MP LSP. The same backup label will be assigned to all Merge Points belonging to the same P2MP Bypass tunnel.

A MP may actually be leaf LSR of multiple bypass tunnels, but will be associated to only one bypass tunnel. That is a PLR will signal the P2MP backup LSP to that MP, for a single P2MP bypass tunnel context.

{LSR1.LSRn} may be a superset of {MP1.MPq}, that is some leaf LSRs of a given P2MP bypass tunnel, noted {LSRx.LSRy}, may not belong to {MP1.MPq}. The PLR will not assign upstream labels for the backup P2MP LSP to these LSRs {LSRx.LSRy}. During failure packets with a backup label will also be delivered onto the P2MP bypass tunnel to {LSRx.LSRy} which will discard these packets based on no entry for this label in the context specific ILM for that bypass tunnel. This requires that {LSRx.LSRy} create a context specific ILM for that bypass tunnel.

PHP MUST be deactivated on the P2MP Bypass tunnel, in order to allow MPs to determine the context for the backup labels assigned by the PLR.

Note that P2MP bypass LSPs may be signalled in advance either automatically or via configuration, or may be dynamically setup upon protected P2MP LSP signalling. Such procedures rely on local implementation issues and are beyond the scope of this document.

4.1.2. P2MP Backup LSP Signalling

The same backup label (i.e. the inner label) MUST be used for all backup S2L sub-LSPs which are tunneled within the same P2MP Bypass tunnel, so as to avoid traffic replication on the PLR. This label MUST be assigned by the PLR using upstream label assignment procedures.

Backup P2MP LSPs MUST be signaled prior to the failure. To signal the backup P2MP LSP, the PLR will send one or more Path messages, referred to as a backup LSP's Path message, to each MP, as specified in [RSVP-P2MP]. A backup LSP's Path message to a given MP comprises one or more backup S2L sub-LSPs that transit through this MP. A backup Path message MUST be sent to the MP using directed signaling; i.e., it is addressed to the MP, without Router Alert option.

As specified in [RSVP-P2MP] it is RECOMMENDED that the PLR use the sender template specific method to identify a backup LSP's Path message, that is, the PLR will set the source address in the sender template to a local PLR address.

The backup label MUST be assigned by the PLR, in the context of the underlying P2MP Bypass tunnel, following upstream label assignment and P2MP RSVP-TE context identification procedures defined in [RSVP-UP]. Hence, a backup LSP's Path message sent to a given MP MUST include an Upstream Assigned Label object carrying the value of the backup label. It MUST also include an RSVP-TE P2MP LSP TLV within an IF_ID RSVP object, that carries the session object of the underlying P2MP Bypass tunnel. This allows the MP to identify the label space of the backup label assigned by the PLR. The same backup label MUST be sent to all MPs belonging to a given P2MP Bypass tunnel.

Note that the PLR MUST continue to refresh Path messages for the protected P2MP TE LSP along the nominal route.

The processing of backup S2L sub-LSP SEROs/SRROs MUST follow

backup LSP ERO/RRO processing described in [RFC4090].

4.2. During failure

When the PLR detects a link or/and node failure condition, it has to reroute a protected P2MP LSP onto a set of one or more P2MP bypass tunnels using as inner label(s) the backup label(s) assigned for this P2MP LSP.

Note that when some MPs are LSRs downstream to the PLR but not downstream to the failed element, the PLR MUST stop sending traffic directly within the protected P2MP TE LSP towards these MPs. This allows avoiding sending twice the traffic on downstream links during failure.

The PLR MUST continue to send Path messages for the backup P2MP LSP. The RRO/ERO flags MUST be updated as per defined in [RFC4090]

5. MP Procedures

A MP receives one or more Path messages for the protected P2MP TE LSP and one or more Path messages for the backup P2MP LSP.

Note that, as specified in [RFC4090], the reception of a backup LSP's Path message does not indicate that a failure has occurred or that the incoming protected LSP will no longer be used.

A S2L sub-LSP is received within a Path message for the protected P2MP LSP and within a Path message for the backup P2MP LSP. These two Path messages are distinguished thanks to the sender-template specific method. As specified in [RFC4090], each of these Path messages will have a different sender address. The protected LSP can be recognized because it will include the FAST_REROUTE object or have the "local protection desired" flag set in the SESSION_ATTRIBUTE object, or both.

A MP MUST maintain one context specific ILM table per PLR or per P2MP bypass tunnel for which it is a leaf.

A MP MUST install the upstream assigned label received in a backup

LSP's Path message, within an ILM specific to the underlying bypass tunnel, which is identified by its session object, carried within the IF_ID RSVP_HOP object of the backup LSP's Path message. An upstream assigned label for a backup P2MP LSP MUST be mapped to the outgoing interface(s) and label(s) of the corresponding protected P2MP LSP.

As specified in [RSVP-UP], the Resv message sent by a MP to the PLR, does not carry any Label Object.

The processing of backup S2L sub-LSP SEROs/SRROs MUST follow backup tunnel ERO/RRO processing described in [RFC4090].

6. To be included in future revisions

The following items will be included in further revisions of this document:

- Combination of P2P and P2MP bypass tunnels to protect a given link/node. This will allow backward compatibility with LSRs that do not support upstream label assignment.

- Cases where the PLR is not directly upstream to the protected element.

- Partial protection: that is the case where only a subset of Merge Points can be covered.

- New RSPV-TE Attribute flags:

  - o A flag in the ATTRIBUTE FLAGS TLV to indicate that protection with P2MP bypass tunnels is desired, and to record such protection.
  - o A flag in the ATTRIBUTE FLAGS TLV to indicate whether partial protection is allowed or not and to record partial protection.
  - o A flag in the ATTRIBUTE FLAGS TLV to indicate that PHP must be deactivated, and to record PHP status (this has a broader scope so this may belong to a dedicated draft).

7. Security Considerations

No new security issues are raised in this document.

8. Acknowledgments

We would like to thank Kireeti Kompella and Venu Hemige, for the useful comments and discussions.

9. References

9.1. Normative references

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

[RFC3031] E. Rosen, A. Viswanathan, R. Callon, "MPLS Architecture", RFC 3031.

[RFC3209] D. Awduche et al., "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209.

[RFC4461] S. Yasukawa et al., "Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)", RFC4461.

[RFC4090] Pan, Swallow, Atlas, et al., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC4090.

[RSVP-P2MP] Aggarwal, Papadimitriou, Yasukawa et al. "Extensions to RSVP-TE for Point to Multipoint TE LSPs", draft-ietf-mpls-rsvp-te-p2mp, work in progress.

[MPLS-UPSTREAM] Aggarwal, Rekhter, Rosen, "MPLS Upstream Label Assignment and Context Specific Label Space", draft-ietf-mpls-upstream-label, work in progress.

[RSVP-UP] Aggarwal, Le Roux, " MPLS Upstream Label Assignment for RSVP-TE", draft-ietf-mpls-rsvp-upstream, work in progress.

10. Authors' Addresses:

Jean-Louis Le Roux
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex
FRANCE
Email: jeanlouis.leroux@orange-ftgroup.com

Rahul Aggarwal
Juniper Networks
1194 North Mathilda Ave.
Sunnyvale, CA 94089
USA
Email: rahul@juniper.net

Jean-Philippe Vasseur
Cisco Systems, Inc.
1414 Massachusetts avenue
Boxborough , MA - 01719
USA
Email: jpv@cisco.com

M. Vigoureux
Alcatel-Lucent France
Route de Villejust
91620 Nozay
FRANCE
Email: martin.vigoureux@alcatel-lucent.fr

11. Intellectual Property Statement

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

Copyright Statement