MPLS Working Group

Zafar Ali
Rakesh Gandhi
Tarek Saad
Cisco Systems

Internet Draft
Intended status: Standard Track
Expires: January 15, 2013

July 16, 2012

Signaling RSVP-TE P2MP LSPs in an Inter-domain Environment
draft-ali-mpls-inter-domain-p2mp-rsvp-te-lsp-08.txt

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 January 15, 2013.

^L

Abstract

Point-to-MultiPoint (P2MP) Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs) may be established using signaling techniques described in [RFC4875]. However, [RFC4875] does not address many issues that comes when a P2MP-TE LSP is signaled in an inter-domain networks. Specifically, one of the issues in inter-domain networks is how to allow computation of a loosely routed P2MP-TE LSP such that it is re-merge free. Another issue is reoptimization of a P2MP-TE tree vs. reoptimization of an individual destination, as loosely routing domain border node is not aware of the reoptimization scope. This document provides a framework and required protocol extensions needed for establishing, controlling and reoptimizing P2MP MPLS and GMPLS TE LSPs in inter-domain networks.

This document borrows inter-domain TE terminology from [RFC 4726], e.g., for the purposes of this document, a domain is considered to be any collection of network elements within a common sphere of address management or path computational responsibility. Examples of such domains include Interior Gateway Protocol (IGP) areas and Autonomous Systems (ASes).

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 [RFC2119].

Table of Contents

^L

1. Introduction

[RFC4875] describes how to set up point-to-multipoint (P2MP) Traffic Engineering Label Switched Paths (TE LSPs) for use in MultiProtocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks.

As with all other RSVP controlled LSPs, P2MP LSP state is managed using RSVP messages. While the use of RSVP messages is mostly similar to their P2P counterpart, P2MP LSP state differs from P2P LSP in a number of ways. In particular, the P2MP LSP must also handle the "re-merge" problem described in [RFC4875] section 18.

The term "re-merge" refers to the situation when two S2L sub-LSPs branch at some point in the P2MP tree, and then intersect again at a another node further down the tree. This may occur due to discrepancies in the routing algorithms used by different nodes, errors in path calculation or manual configuration, or network topology changes during the establishment of the P2MP LSP. Such re-merges are inefficient due to the unnecessary duplication of data. Consequently one of the requirements for signaling P2MP LSPs is to choose a P2MP path that is re-merge free. In some

deployments, it may also be required to signal P2MP-TE LSPs that are both re-merge and crossover free [RFC4875].

This requirement becomes more acute to address when P2MP LSP spans multiple domains. For the purposes of this document, a domain is considered to be any collection of network elements

^L

within a common sphere of address management or path computational responsibility. Examples of such domains include Interior Gateway Protocol (IGP) areas and Autonomous Systems (ASes). This is because in an inter-domain environment, the ingress node may not have topological visibility into other domains to be able to compute and signal a re-merge free P2MP LSP. In that case, the border node for a new domain will be given loose next hops for one or more destinations in a P2MP LSP. A border node computes paths in its domain by individually expanding the loose next hops for the destinations when signaled to it. A border node can ensure that it computes the remerge free paths while performing loose hop ERO expansions by individually grafting destinations.
Note that computed P2MP tree by a border node in this fashion may not be optimal. When processing a path message, the border node may not have knowledge of all of the destinations of the P2MP LSP, for example in the case when not all S2L sub-LSPs pass through this border node. In that case, existing protocol mechanisms do not provide sufficient information for it to be able to expand the loose hop(s) in such a way that the overall P2MP LSP path is guaranteed to be re-merge free.

RFC 4875 specifies two approaches to handle re-merge conditions. In the first method that is based on control plane handling, the re-merge node initiates the removal of the re-merge branch(es) by sending a Path Error message. In the second method that is based on data plane handling, the node detecting the re-merge case, i.e., the re-merge node, allows the re-merge to persist, but data from all but one incoming interface is dropped at the re-merge node. This ensures that duplicate data is not sent on any outgoing interface.

This document proposes RSVP-TE signaling procedures for P2MP LSP to handle re-merge for both using control plane approach and data plane approach.

Control plane solution is using crankback signaling in RSVP. [RFC5151] describes mechanisms for applying crankback to inter-domain P2P LSPs, but does not cover P2MP LSPs. Also, crankback mechanisms for P2MP LSPs are not addressed by [RFC4875]. This document describes how crankback signaling extensions for MPLS and GMPLS RSVP-TE defined in [RFC4920] can be used for setting up

P2MP TE LSPs to resolve re-merges.

Date plane solution described in [RFC4875] is extended by using a new flag in RRO Attributes Sub-object in RSVP. The proposed solution makes use of RRO Attributes Sub-object as defined in [RFC5420] for this purpose. This document describes how new RRO Attributes Flag can be used to handle P2MP re-merge conditions efficiently.

[RFC4736] defines procedures and signaling extensions for reoptimizing an inter-domain LSP. Specifically a head-end node sends a "path re-evaluation request" to a border node by setting a flag (0x20) in SESSION_ATTRIBUTES object in a path message. A border node sends a path error code 25 (notify) with sub-code 6 to indicate "preferable path exists" to the head-end node.
This path error can be sent by the border node unsolicited or upon receiving a "path re-evaluation request". The head-end node upon receiving this path error may initiate reoptimization of the LSP.

For P2MP LSP, head-end may reoptimize the entire P2MP LSP by resignaling all destinations or may reoptimize only one or some of the destinations in the P2MP LSP.
For P2MP LSP, a border node may have loosely routed entire or part of the P2MP LSP by expanding EROs in path messages of the destinations. Border node does not know with the signaling procedure defined in [RFC4736] if head-end is requesting a reoptimization for a single destination or reoptimization of the P2MP tree. Signaling extension and procedure are defined in this document to support reoptimization of inter-domain P2MP LSP and a single destination.

The solutions presented in this document do not guarantee optimization of the overall P2MP tree across all domains. PCE can

^L

be used, instead, to address optimization of the overall P2MP tree.

2. Framework

2.1. Signaling Options

The four signaling options defined for P2P inter-domain LSPs in [RFC4726] are also applicable to P2MP LSPs.

- LSP nesting, using hierarchical LSPs [RFC4206].
- A single contiguous LSP, using the same SESSION and LSP ID along its whole path.
- LSP stitching [RFC5150].
- A combination of the above.

In the case of LSP nesting using hierarchical LSPs, the tunneled LSP MUST use upstream-assigned labels to ensure that the same label is used at every leaf of the H-LSP ([RFC5331], [I-D.ietf-mpls-rsvp-upstream]). The H-LSP SHOULD request non-PHP behavior and out-of-band mapping as defined in [I-D.ietf-mpls-rsvp-te-no-php-oob-mapping].

2.2. Path Computation Techniques

This document focuses on the case where the ingress node does not have full visibility of the topology of all domains, and is therefore not able to compute the complete P2MP tree. Rather, it has to include loose hops to traverse domains for which it does not have full visibility, and the border node(s) on entry to each domain are responsible for expanding those loose hops.

3. Control Plane Solution

It is RECOMMENDED that boundary re-routing or segment-based re-routing is requested for P2MP LSPs traversing multiple domains. This is because border nodes that are expanding loose hops are typically best placed to correct any re-merge errors that occur

within their domain, not the ingress node.

^L

3.1. Single Border Node

The ingress node is RECOMMENDED to select the same border node as an ERO loose hop for all sibling S2L sub-LSPs that transit a given domain. This reduces the chances of the sibling S2L sub-LSPs in remerging states, because a single border node has the necessary state to ensure that the path that they take through the domain is re-merge free.

3.2. Crankback and Path Error Signaling Procedure

As mentioned in [RFC4875], in order to avoid duplicate traffic, the re-merge node MAY initiate the removal of the re-merge S2L sub-LSPs by sending a Path Error message to the ingress node of the S2L sub-LSP.

Crankback procedures for rerouting around failures for P2P RSVP-TE LSPs are defined in [RFC4920]. These techniques can also be applied to P2MP LSPs to handle re-merge conditions, as described in this section.

If a node on the path of the P2MP LSP is unable to find a route that can supply the required resources or that is re-merge free, it SHOULD generate a Path Error message for the subset of the S2L sub-LSPs which it is not able to route. For this purpose the node SHOULD try to find a minimum subset of S2L sub-LSPs for which the Path Error needs to be generated. This rule applies equally to the case where multiple S2L sub-LSPs are signaled using one Path message, as to the case where a single S2L sub-LSP is signaled in each Path message. RSVP-TE Notify messages do not include S2L_SUB_LSP objects and cannot be used to send errors for a subset of the S2L sub-LSPs in a Path message. For that reason, the node SHOULD use a Path Error message rather than a Notify message to communicate the error. In the case of a re-merge error, the node SHOULD use the error code "Routing Problem" and the error value "ERO resulted in re-merge" as specified in [RFC4875].

A border node receiving a Path Error message for a set of S2L sub-LSPs MAY hold the message and attempt to signal an alternate path that can avoid re-merge through its domain for those S2L sub-LSPs that pass through it. However, in the case of a re-merge

error for which some of the re-merging S2L sub-LSPs do not pass through the border node, it SHOULD propagate the Path Error upstream to the ingress node. If the subsequent attempt by the border node is successful, the border node discards the held Path Error and follows the crank back roles of [RFC4920] and

^L

[RFC5151]. If all subsequent attempts by the border node are unsuccessful, the border node MUST send the held Path Error upstream to the ingress node.

If the ingress node receives a Path Error message with error code "Routing Problem" and error value "ERO resulted in re-merge", then it SHOULD attempt to signal an alternate path through a different domain or through a different border node for the affected S2L sub-LSPs.

However, it may be that the ingress node or a border node does not have sufficient topology information to compute an Explicit Route that is guaranteed to avoid the re-merge link or node. In this case, Route Exclusions [RFC4874] may be particularly helpful. To achieve this, [RFC4874] allows the re-merge information to be presented as route exclusions to force avoidance of the re-merge link or node.

As discussed in [RFC4090] section 3.3, border node MAY keep the history of Path Errors. In case of P2MP LSPs, ingress node and border nodes may keep re-merge Path Errors in history table until S2L sub-LSPs have been successfully established or until local timer expires.

4. Data Plane Solution

As mentioned in [RFC4875], node may accept the remerging S2Ls but only send the data from one of these interfaces to its outgoing interfaces. That is, the node MUST drop data from all but one incoming interface. This ensures that duplicate data is not sent on any outgoing interface.

It is desirable to avoid the persistent re-merge condition associated with data plane based solution in the network in order to optimize bandwidth resources in the network.

RSVP-TE signaling extensions are defined in the following to request P2MP-TE Re-merge Recording and indicate P2MP-TE Re-merge Presence.

4.1. P2MP-TE Re-merge Recording Request Flag

In order to indicate nodes that P2MP-TE Re-merge Recording is desired, a new flag in the Attribute Flags TLV of the LSP_ATTRIBUTES object defined in [RFC5420] is defined as follows:

^L

Bit Number (to be assigned by IANA): P2MP-TE Re-merge Recording Request flag

The P2MP-TE Re-merge Recording Request flag is meaningful in a Path message and can be inserted by the ingress node or a border node.

If the P2MP-TE Re-merge Recording Flag is set to 1, it means that "P2MP-TE Re-merge Presence" defined in the next section should be used to indicate to the ingress and border nodes along the setup of the LSP that a remerge is present but accepted and that incoming traffic is being dropped for the given S2L.

The rules of the processing of the Attribute Flags TLV of the LSP_ATTRIBUTES object follow [RFC5420].

4.2. P2MP-TE Re-merge Present Flag

The P2MP-TE Re-merge Present Flag is the counter part of the P2MP-TE Re-merge Recording Request flag defined above. Specifically, RSVP signaling extension is defined to indicate to the upstream node of the re-merge condition and that incoming traffic is being dropped for the given S2L.

When a node decides to accept remerge and drop traffic from an incoming interface for an S2L due to the re-merge condition, and understands the "P2MP-TE Re-merge Recording Request in the Attribute Flags TLV of the LSP_ATTRIBUTES object of the Path message, the node SHOULD set the newly defined "P2MP-TE Re-merge Present" flag in the RRO Attributes sub-object defined in [RFC 5420] in RRO.

The following new flag for RRO Attributes Sub-object is defined as follows:

Bit Number (same as bit number assigned for P2MP-TE Re-merge Recording Request flag): P2MP-TE Re-merge Present flag

The presence of P2MP-TE Re-merge Present flag indicates that the S2L is causing a re-merge. The re-merge has been accepted but the incoming traffic on this S2L is dropped by the

reporting node.

^L

4.3. Signaling Procedure

When a node receives an S2L sub-LSP Path message with LSP Attributes sub-object that has "P2MP-TE Re-merge Recording Request" Flag set, and the node does not support data plane based re-merge handling, and the S2L is causing a re-merge, the node SHOULD reject the S2L sub-LSP path message and send the Path Error with the error code "Routing Problem" and the error value "ERO resulted in re-merge" as specified in [RFC4875].

When a path message is received at a transit node and "P2MP-TE Re-merge Recording Request" Flag is set in the LSP Attributes sub-object, the node MAY decide to accept the re-merge S2L sub-LSP. In this case, before the Resv message is sent to the upstream node, the node adds the RRO Attributes sub-object to the RRO and sets the "P2MP-TE Re-merge Recording Request" Flag. .

When a transit node receives a reservation message for an S2L that is causing a re-merge, the node SHOULD set the "P2MP-TE Re-merge Present" flag in the RRO Attributes sub-object in the reservation message if it decides to drop the incoming traffic of this S2L. "P2MP-TE Re-merge Present" flag in RRO Attribute sub-object is not set for the S2Ls if the node has selected the incoming interface of the S2Ls to forward the traffic.

An ingress node MAY immediately start sending traffic on all S2Ls in up state even when re-merges are present on some S2Ls of the P2MP LSP.

Proposed signaling extensions allow an ingress node and a border node to have a complete view of the re-merges on entire S2L path and on all S2Ls of the P2MP tree and can take appropriate actions to resolve re-merges and optimize network bandwidth resources. The proposed signaling extensions are equally applicable to single domain scenarios.

A node may need to select a different incoming interface to forward traffic in future. In that case, a reservation change message is sent upstream indicating the change by marking or clearing the "P2MP-TE Re-merge Present" flag appropriately for all effected S2Ls.

The re-merge node SHOULD NOT dynamically change incoming interface to forward traffic to avoid unnecessary race conditions.

^L

A border node due to local policy MAY remove the record route object from the reservation message of the S2L sub-LSP and propagate reservation message towards the ingress node. When such a policy is provisioned, the border node may attempt to correct the re-merge condition in its domain. If the border node is not able to resolve the re-merge condition, the border node SHOULD send the Path Error with the error code "Routing Problem" and the error value "ERO resulted in re-merge" as specified in [RFC4875].

5. Reoptimization Signaling Procedure

Using signaling procedure defined in [RFC4736], a head-end node MAY initiate "path re-evaluation request" query to reoptimize a destination in a P2MP LSP. Note that this message SHOULD be used to reoptimize a single or a sub-set of the destinations in a P2MP LSP. Head-end sends query in path message for each destination it is reoptimizing.

When a path message for a destination in a P2MP LSP with "path re-evaluation request" flag [RFC4736] is received at the border node, it SHOULD recompute the loose-hop ERO to see if a preferred path exists for that destination. A border node MAY send path error 25 with "preferred path exists" sub-code to indicate that a preferred path exists for the requested destination AND border node is capable of per destination reoptimization. A border node MAY terminate the path query OR respond with new "Preferred P2MP-TE Tree Exists" path error (defined below) by checking for a preferred P2MP tree instead of a single destination if it is not capable of per destination reoptimization.

It is often desired to reoptimize the entire P2MP LSP. In order to query border nodes to check if a preferred P2MP tree exists, head-end node MAY send path message with newly defined flag in Attributes Flags TLV of the LSP_ATTRIBUTES object [RFC5420] as follows:

Bit Number (to be assigned by IANA): P2MP-TE tree re-evaluation Request flag

The P2MP-TE tree Re-evaluation Request flag is meaningful in a Path message and can be inserted by the ingress node. A head-end node

MAY send this message for all destinations in a P2MP LSP or a sub-set of the destinations.

A border node receiving the new "P2MP-TE tree re-evaluation Request" SHOULD check for a better P2MP LSP for the destinations it is loosely routing by loose-hop ERO expansions and if a preferred P2MP-TE tree is found, it SHOULD reply with "Preferred P2MP-TE Tree Exists" path error and terminate the path query. If no preferred path is found it SHOULD propagate query downstream.

Following new sub-code for path error code 25 is defined:

Sub-code (to be assigned by IANA): Preferred P2MP-TE Tree Exists flag

When a preferred P2MP tree is found, the border node MAY send a newly defined sub-code "Preferred P2MP-TE tree exists" with path error code 25 to indicate head-end node to reoptimize the entire P2MP LSP. A border node may send this path error message unsolicited or in a response to "path re-evaluation query" for a destination or newly defined "P2MP-TE tree re-evaluation request" query.

If a head-end node initiated a "path re-evaluation request" query for a single destination for per S2L sub LSP reoptimization and receives "Preferred P2MP-TE Tree Exists" path error, head-end MAY cancel the per S2L reoptimization and initiate P2MP-TE tree reoptimization. This may happen in cases when a border node is not capable of per destination reoptimization.

6. Security Considerations

This document does not introduce any additional security issues above those identified in [RFC3209], [RFC4875], [RFC5151], [RFC4920] and [RFC5920].

7. IANA Considerations

The following new flag is defined for the Attributes Flags TLV in the LSP_ATTRIBUTES object. The numeric values are to be assigned by IANA.

o P2MP-TE Re-merge Recording Request Flag:

- Bit Number: To be assigned by IANA.

- Attribute flag carried in Path message: Yes

- Attribute flag carried in Resv message: No

The following new flag is defined for the RRO Attributes sub-object in the RECORD_ROUTE object. The numeric values are to be assigned by IANA.

o P2MP-TE Re-merge Recording Present Flag:

- Bit Number: To be assigned by IANA.

- Attribute flag carried in Path message: No

- Attribute flag carried in RRO Attributes sub-object in RRO of the Resv message: Yes

The following new flag is defined for the Attributes Flags TLV in the LSP_ATTRIBUTES object. The numeric values are to be assigned by IANA.

o P2MP-TE tree Re-evaluation Request Flag:

- Bit Number: To be assigned by IANA.

- Attribute flag carried in Path message: Yes

- Attribute flag carried in Resv message: No

As defined in [RFC3209], the Error Code 25 in the ERROR SPEC object corresponds to a Notify Error. This document adds a new flag as follows:

o Sub-code for Notify Path Error code 25:

- Sub-code To be assigned by IANA: Preferred P2MP-TE Tree exists.

^L

8. Acknowledgments

The authors would like to thank N. Neate for his contributions on the draft.

9. References

9.1. Normative References

[RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.

[RFC5151] Farrel, A., Ayyangar, A., and JP. Vasseur, "Inter-Domain MPLS and GMPLS Traffic Engineering -- Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 5151, February 2008.

[RFC4920] Farrel, A., Satyanarayana, A., Iwata, A., Fujita, N., and G. Ash, "Crankback Signaling Extensions for MPLS and GMPLS RSVP-TE", RFC 4920, July 2007.

[RFC5920] L. Fang, Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

[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.

[RFC4736] Vasseur, JP., Ikejiri, Y. and Zhang, R, "Reoptimization of Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Loosely Routed Label Switched Path (LSP)", RFC 4736, November 2006.

9.2. Informative References

[RFC4726] Farrel, A., Vasseur, J., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.

[RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.

[RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, February 2008.

^L

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

[I-D.ietf-mpls-rsvp-upstream] Aggarwal, R. and J. Roux, "MPLS Upstream Label Assignment for RSVP-TE", draft-ietf-mpls-rsvp-upstream-05 (work in progress), March 2010.

[I-D.ietf-mpls-rsvp-te-no-php-oob-mapping] Ali, Z. and G. Swallow, "Non PHP Behavior and out-of-band mapping for RSVP-TE LSPs", draft-ietf-mpls-rsvp-te-no-php-oob-mapping-04 (work in progress), March 2010.

Author's Addresses

Zafar Ali
Cisco Systems
Email: zali@cisco.com

Rakesh Gandhi
Cisco Systems
Email: rgandhi@cisco.com

Tarek Saad
Cisco Systems
Email: tsaad@cisco.com

^L