Network Working Group
Request for Comments: 4736
Category: Informational

JP. Vasseur, Ed.
Cisco Systems, Inc.
Y. Ikejiri
NTT Communications Corporation
R. Zhang
BT Infonet

November 2006

Reoptimization of Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Loosely Routed Label Switched Path (LSP)

Status of This Memo

Copyright Notice

Abstract

This document defines a mechanism for the reoptimization of loosely routed MPLS and GMPLS (Generalized Multiprotocol Label Switching) Traffic Engineering (TE) Label Switched Paths (LSPs) signaled with Resource Reservation Protocol Traffic Engineering (RSVP-TE). This document proposes a mechanism that allows a TE LSP head-end Label Switching Router (LSR) to trigger a new path re-evaluation on every hop that has a next hop defined as a loose or abstract hop and a mid-point LSR to signal to the head-end LSR that a better path exists (compared to the current path) or that the TE LSP must be reoptimized (because of maintenance required on the TE LSP path). The proposed mechanism applies to the cases of intra- and inter-domain (Interior Gateway Protocol area (IGP area) or Autonomous System) packet and non-packet TE LSPs following a loosely routed path.

Table of Contents

1. Introduction

This document defines a mechanism for the reoptimization of loosely routed MPLS and GMPLS (Generalized Multiprotocol Label Switching) Traffic Engineering LSPs signaled with RSVP-TE (see [RFC3209] and [RFC3473]). A loosely routed LSP is defined as one that does not contain a full, explicit route identifying each LSR along the path of the LSP at the time it is signaled by the ingress LSR. Such an LSP is signaled with no Explicit Route Object (ERO), with an ERO that contains at least one loose hop, or with an ERO that contains an abstract node that is not a simple abstract node (that is, an abstract node that identifies more than one LSR).

The Traffic Engineering Working Group (TE WG) has specified a set of requirements for inter-area and inter-AS MPLS Traffic Engineering (see [RFC4105] and [RFC4216]). Both requirements documents specify the need for some mechanism providing an option for the head-end LSR to control the reoptimization process should a more optimal path exist in a downstream domain (IGP area or Autonomous System). This document defines a solution to meet this requirement and proposes two mechanisms:

(1) The first mechanism allows a head-end LSR to trigger a new path re-evaluation on every hop that has a next hop defined as a loose hop or abstract node and get a notification from the mid-point as to whether a better path exists.

(2) The second mechanism allows a mid-point LSR to explicitly signal to the head-end LSR either that a better path exists to reach a loose/abstract hop (compared to the current path) or that the TE LSP must be reoptimized because of some maintenance required along the TE LSP path. In this case, the notification is sent by the mid-point LSR without being polled by the head-end LSR.

A better path is defined as a lower cost path, where the cost is determined by the metric used to compute the path.

2. Terminology

ABR: Area Border Router.

ERO: Explicit Route Object.

LSR: Label Switching Router.

TE LSP: Traffic Engineering Label Switched Path.

TE LSP head-end: head/source of the TE LSP.

TE LSP tail-end: tail/destination of the TE LSP.

Interior Gateway Protocol Area (IGP Area): OSPF Area or IS-IS level.

Intra-area TE LSP: A TE LSP whose path does not transit across areas.

Inter-area TE LSP: A TE LSP whose path transits across at least two different IGP areas.

Inter-AS MPLS TE LSP: A TE LSP whose path transits across at least two different Autonomous Systems (ASes) or sub-ASes (BGP confederations).

2.1. Requirements Language

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].

3. Establishment of a Loosely Routed TE LSP

The aim of this section is purely to summarize the mechanisms involved in the establishment of a loosely routed TE LSP, as specified in [RFC3209]. The reader should see RFC 3209 for a more detailed description of these mechanisms.

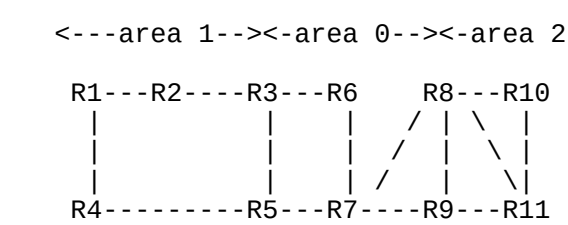
In the context of this document, a loosely routed LSP is defined as one that does not contain a full, explicit route identifying each LSR along the path of the LSP at the time it is signaled by the ingress LSR. Such an LSP is signaled with no ERO, with an ERO that contains at least one loose hop, or with an ERO that contains an abstract node that is not a simple abstract node (that is, an abstract node that identifies more than one LSR). As specified in [RFC3209], loose hops are listed in the ERO object of the RSVP Path message with the L flag of the IPv4 or the IPv6 prefix sub-object set.

Each LSR along the path whose next hop is specified as a loose hop or a non-specific abstract node triggers a path computation (also referred to as an ERO expansion), before forwarding the RSVP Path message downstream. The computed path may be either partial (up to the next loose hop) or complete (set of strict hops up to the TE LSP destination).

Note that although the examples in the rest of this document are provided in the context of MPLS inter-area TE, the proposed mechanism applies equally to loosely routed paths within a single routing

domain and across multiple Autonomous Systems. The examples below are provided with OSPF as the IGP, but the described set of mechanisms similarly apply to IS-IS.

An example of an explicit loosely routed TE LSP signaling follows.



Assumptions

- R3, R5, R8, and R9 are ABRs.

- The path of an inter-area TE LSP T1 from R1 (head-end LSR) to R11 (tail-end LSR) is defined on R1 as the following loosely routed path: R1-R3(loose)-R8(loose)-R11(loose). R3, R8, and R11 are defined as loose hops.

  Step 1: R1 determines that the next hop (R3) is a loose hop (not directly connected to R1) and then performs an ERO expansion operation to reach the next loose hops R3. The new ERO becomes: R2(S)-R3(S)-R8(L)-R11(L), where S is a strict hop (L=0) and L is a loose hop (L=1).

  The R1-R2-R3 path satisfies T1’s set of constraints.

  Step 2: The RSVP Path message is then forwarded by R1 following the path specified in the ERO object and reaches R3 with the following content: R8(L)-R11(L).

  Step 3: R3 determines that the next hop (R8) is a loose hop (not directly connected to R3) and then performs an ERO expansion operation to reach the next loose hops R8. The new ERO becomes: R6(S)-R7(S)-R8(S)-R11(L).

  Note: In this example, the assumption is made that the path is computed on a per-loose-hop basis, also referred to as a partial route computation. Note that other path computation techniques may result in complete paths (set of strict hops up to the final destination).

  Step 4: The same procedure is repeated by R8 to reach T1’s destination (R11).

4. Reoptimization of a Loosely Routed TE LSP Path

Once a loosely routed, explicit TE LSP is set up, it is maintained through normal RSVP procedures. During the TE LSP lifetime, a more optimal path might appear between an LSR and its next loose hop (for the sake of illustration, suppose that in the example above a link between R6 and R8 is added or restored that provides a preferable path between R3 and R8 (R3-R6-R8) than the existing R3-R6-R7-R8 path). Since a preferable (e.g., shorter) path might not be visible from the head-end LSR by means of the IGP if the head-end LSR does not belong to the same IGP area where the associated topology change occurred, the head-end cannot make use of this shorter path (and reroute the LSP using a make-before-break technique as described in [RFC3209]) when appropriate. Thus, a new mechanism specified in this document is required to detect the existence of such a preferable path and to notify the head-end LSR accordingly.

This document defines a mechanism that allows

- a head-end LSR to trigger on every LSR whose next hop is a loose hop or an abstract node the re-evaluation of the current path in order to detect a potentially more optimal path; and

- a mid-point LSR whose next hop is a loose-hop or an abstract node to signal (using a new Error Value sub-code carried in a RSVP PathErr message) to the head-end LSR that a preferable path exists (a path with a lower cost, where the cost definition is determined by some metric).

Once the head-end LSR has been notified of the existence of such a preferable path, it can decide (depending on the TE LSP characteristics) whether to perform a TE LSP graceful reoptimization such as the "make-before-break" procedure.

There is another scenario whereby notifying the head-end LSR of the existence of a better path is desirable: if the current path is about to fail due to some (link or node) required maintenance.

This mechanism allows the head-end LSR to reoptimize a TE LSP by making use of the non-disruptive make-before-break procedure if and only if a preferable path exists and if such a reoptimization is desired.

5. Signaling Extensions

A new flag in the SESSION ATTRIBUTE object and new Error Value sub-codes in the ERROR SPEC object are proposed in this document.

5.1. Path Re-Evaluation Request

The following new flag of the SESSION_ATTRIBUTE object (C-Type 1 and 7) is defined:

Path re-evaluation request: 0x20

This flag indicates that a path re-evaluation (of the current path in use) is requested. Note that this does not trigger any LSP Reroute but instead just signals a request to evaluate whether a preferable path exists.

Note: In case of link bundling, for instance, although the resulting ERO might be identical, this might give the opportunity for a mid-point LSR to locally select another link within a bundle. However, strictly speaking, the ERO has not changed.

5.2. New Error Value Sub-Codes

As defined in [RFC3209], the Error Code 25 in the ERROR SPEC object corresponds to a Notify Error.

This document adds three new Error Value sub-codes:

6 Preferable path exists

7 Local link maintenance required

8 Local node maintenance required

The details about the local maintenance required modes are in Section 6.3.2.

6. Mode of Operation

6.1. Head-End Reoptimization Control

The notification process of a preferable path (shorter path or new path due to some maintenance required on the current path) is by nature de-correlated from the reoptimization operation. In other words, the location where a potentially preferable path is discovered does not have to be where the TE LSP is actually reoptimized. This document applies to the context of a head-end LSR reoptimization.

6.2. Reoptimization Triggers

There are several possible reoptimization triggers:

- Timer-based: A reoptimization is triggered (process evaluating whether a more optimal path can be found) when a configurable timer expires.

- Event-driven: A reoptimization is triggered when a particular network event occurs (such as a "Link-UP" event).

- Operator-driven: A reoptimization is manually triggered by the Operator.

It is RECOMMENDED that an implementation supporting the extensions proposed in this document support the aforementioned modes as path re-evaluation triggers.

6.3. Head-End Request versus Mid-Point Explicit Notification Functions

This document defines two functions:

1) "Head-end requesting function": The request for a new path evaluation of a loosely routed TE LSP is requested by the head-end LSR.

2) "Mid-point explicit notification function": Having determined that a preferable path (other than the current path) exists or having the need to perform a link/node local maintenance, a mid-point LSR explicitly notifies the head-end LSR, which will in turn decide whether to perform a reoptimization.

6.3.1. Head-End Request Function

When a timer-based reoptimization is triggered on the head-end LSR or the operator manually requests a reoptimization, the head-end LSR immediately sends an RSVP Path message with the "Path re-evaluation request" bit of the SESSION-ATTRIBUTE object set. This bit is then cleared in subsequent RSVP path messages sent downstream. In order to handle the case of a lost Path message, the solution consists of relying on the reliable messaging mechanism described in [RFC2961].

Upon receiving a Path message with the "Path re-evaluation request" bit set, every LSR for which the next abstract node contained in the ERO is defined as a loose hop/abstract node performs the following set of actions:

A path re-evaluation is triggered, and the newly computed path is compared to the existing path:

- If a preferable path can be found, the LSR performing the path re-evaluation MUST immediately send an RSVP PathErr to the head-end LSR (Error code 25 (Notify), Error sub-code=6 (better path exists)). At this point, the LSR MAY decide not to propagate this bit in subsequent RSVP Path messages sent downstream for the re-evaluated TE LSP; this mode is the RECOMMENDED mode for the reasons described below.

  The sending of an RSVP PathErr Notify message "Preferable path exists" to the head-end LSR will notify the head-end LSR of the existence of a preferable path (e.g., in a downstream area/AS or in another location within a single domain). Therefore, triggering additional path re-evaluations on downstream nodes is unnecessary. The only motivation to forward subsequent RSVP Path messages with the "Path re-evaluation request" bit of the SESSION-ATTRIBUTE object set would be to trigger path re-evaluation on downstream nodes that could in turn cache some potentially better paths downstream, with the objective to reduce the signaling setup delay, should a reoptimization be performed by the head-end LSR.

- If no preferable path can be found, the recommended mode is for an LSR to relay the request (by setting the "Path re-evaluation" bit of the SESSION-ATTRIBUTE object in RSVP path message sent downstream).

  Note that, by preferable path, we mean a path with a lower cost.

  If the RSVP Path message with the "Path re-evaluation request" bit set is lost, then the next request will be sent when the next reoptimization trigger will occur on the head-end LSR. The solution to handle RSVP reliable messaging has been defined in [RFC2961].

  The network administrator may decide to establish some local policy specifying to ignore such request or not to consider those requests more frequently than at a certain rate.

  The proposed mechanism does not make any assumption of the path computation method performed by the ERO expansion process.

6.3.2. Mid-Point Explicit Notification

By contrast with the head-end request function, in this case, a mid-point LSR whose next hop is a loose hop or an abstract node can locally trigger a path re-evaluation when a configurable timer expires, some specific events occur (e.g., link-up event), or the user explicitly requests it. If a preferable path is found, the LSR sends an RSVP PathErr to the head-end LSR (Error code 25 (Notify), Error sub-code=6 ("preferable path exists").

There is another circumstance whereby any mid-point LSR MAY send an RSVP PathErr message with the objective for the TE LSP to be rerouted by its head-end LSR: when a link or a node will go down for local maintenance reasons. In this case, the LSR where a local maintenance must be performed is responsible for sending an RSVP PathErr message with Error code 25 and Error sub-code=7 or 8, depending on the affected network element (link or node). Then the first upstream node that has performed the ERO expansion MUST perform the following set of actions:

- The link (sub-code=7) or the node (sub-code=8) MUST be locally registered for further reference (the TE database must be updated).

- The RSVP PathErr message MUST be immediately forwarded upstream to the head-end LSR. Note that in the case of TE LSP spanning multiple administrative domains, it may be desirable for the boundary LSR to modify the RSVP PathErr message and insert its own address for confidentiality.

Upon receiving an RSVP PathErr message with Error code 25 and Error sub-code 7 or 8, the head-end LSR SHOULD perform a TE LSP reoptimization.

Note that the two functions (head-end and mid-point driven) are not exclusive of each other: both the timer and event-driven reoptimization triggers can be implemented on the head-end or on any mid-point LSR with a potentially different timer value for the timer-driven reoptimization case.

A head-end LSR MAY decide upon receiving an explicit mid-point notification to delay its next path re-evaluation request.

6.3.3. ERO Caching

Once a mid-point LSR has determined that a preferable path exists (after a reoptimization request has been received by the head-end LSR or the reoptimization timer on the mid-point has expired), the more optimal path MAY be cached on the mid-point LSR for a limited amount

of time to avoid having to recompute a path once the head-LSR performs a make-before-break. This mode is optional. A default value of 5 seconds for the caching timer is suggested.

7. Applicability and Interoperability

The procedures described in this document are entirely optional within an MPLS or GMPLS network. Implementations that do not support the procedures described in this document will interoperate seamlessly with those that do. Further, an implementation that does not support the procedures described in this document will not be impacted or implicated by a neighboring implementation that does implement the procedures.

An ingress implementation that chooses not to support the procedures described in this document may still achieve re-optimization by periodically issuing a speculative make-before-break replacement of an LSP without trying to discovery whether a more optimal path is available in a downstream domain. Such a procedure would not be in conflict with any mechanisms already documented in [RFC3209] and [RFC3473].

An LSR not supporting the "Path re-evaluation request" bit of the SESSION-ATTRIBUTE object SHALL forward it unmodified.

A head-end LSR not supporting an RSVP PathErr with Error code 25 message and Error sub-code = 6, 7, or 8 MUST just silently ignore such an RSVP PathErr message.

8. IANA Considerations

IANA assigned three new error sub-code values for the RSVP PathErr Notify message (Error code=25):

6 Preferable path exists

7 Local link maintenance required

8 Local node maintenance required

9. Security Considerations

This document defines a mechanism for a mid-point LSR to notify the head-end LSR of the existence of a preferable path or the need to reroute the TE LSP for maintenance purposes. Hence, in the case of a TE LSP spanning multiple administrative domains, it may be desirable for a boundary LSR to modify the RSVP PathErr message (Code 25, Error sub-code = 6, 7, or 8) so as to preserve confidentiality across

domains. Furthermore, a head-end LSR may decide to ignore explicit notification coming from a mid-point residing in another domain. Similarly, an LSR may decide to ignore (or to accept up to a pre-defined rate) path re-evaluation requests originated by a head-end LSR of another domain.

10. Acknowledgements

The authors would like to thank Carol Iturralde, Miya Kohno, Francois Le Faucheur, Philip Matthews, Jim Gibson, Jean-Louis Le Roux, Kenji Kumaki, Anca Zafir, and Dimitri Papadimitriou for their useful comments. A special thanks to Adrian Farrel for his very valuable inputs.

11. References

11.1. Normative References

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

[RFC2961] Berger, L., Gan, D., Swallow, G., Pan, P., Tommasi, F., and S. Molendini, "RSVP Refresh Overhead Reduction Extensions", RFC 2961, April 2001.

[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.

[RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

11.2. Informative References

[RFC4105] Le Roux, J.-L., Vasseur, J.-P., and J. Boyle, "Requirements for Inter-Area MPLS Traffic Engineering", RFC 4105, June 2005.

[RFC4216] Zhang, R. and J.-P. Vasseur, "MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements", RFC 4216, November 2005.

Authors' Addresses

JP Vasseur (Editor)
Cisco Systems, Inc
1414 Massachusetts Avenue
Boxborough, MA 01719
USA

EMail: jpv@cisco.com

Yuichi Ikejiri
NTT Communications Corporation
1-1-6, Uchisaiwai-cho, Chiyoda-ku
Tokyo, 100-8019
Japan

EMail: y.ikejiri@ntt.com

Raymond Zhang
BT Infonet
2160 E. Grand Ave.
El Segundo, CA 90025
USA

EMail: raymond_zhang@bt.infonet.com

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.