Jean-Philippe Vasseur(Editor)
Cisco Systems, Inc.
Yuichi Ikejiri
NTT Communications Corporation

IETF Internet Draft
Expires: December, 2003
June, 2003

draft-vasseur-mpls-loose-path-reopt-02.txt

Reoptimization of MPLS Traffic Engineering loosely routed explicit LSP paths

Status of this Memo

Abstract

The aim of this document is to propose a mechanism for the reoptimization of MPLS Traffic Engineering loosely routed explicit LSP paths. A loosely routed explicit LSP path is a path specified as a combination of strict and loose hop(s) that contains at least one loose hop and zero or more strict hop(s). The path calculation (which implies ERO expansion) to reach a loose hop is made on the previous hop defined in the TE LSP path. This draft proposes a mechanism that allows:

- The TE LSP Head-end LSR to trigger a new ERO expansion on every hop having a next hop defined as a loose hop,

- An LSR to signal to the TE LSP head-end that a better path exists to reach a loose hop (than the current path in use). A better path is defined as a path with a lower cost, where the cost is determined by the metric used to compute the path.

This primarily applies to inter-area TE LSPs and inter-AS TE LSPs when the path is defined as a list of loose hops (generally the loose hops are the ABRs/ASBRs) but the following mechanism is also applicable to any loosely routed explicit path within a single routing domain.

1. Establishment of a loosely routed explicit TE LSP path

A loosely routed explicit path is as a path specified as a combination of strict and loose hop(s) that contains at least one loose hop and zero or more strict hop(s). Loose hops are listed in the ERO object of the RSVP Path message with the L flag of the Ipv4 prefix sub-object set, as defined in [RSVP-TE]. In this case, each LSR along path can perform a partial route computation to reach the next loose hop and then performs an ERO expansion, before forwarding the RSVP Path message downstream.

Note that the examples in the rest of this draft will be provided in the context of MPLS inter-area TE but the proposed mechanism also

applies to loosely routed path within a single routing domain. Furthermore, this mechanism could also be used in the context of loosely routed paths in the context of TE LSPs spanning several autonomous systems and as such abides by the requirements for inter-AS TE define in [INTER-AS-TE-REQS]
The examples below will be provided with OSPF as the IGP but the described mechanisms similarly apply to IS-IS.

An example of an explicit loosely routed TE LSP signalling (see also [MULTI-AREA-TE scenario 1]

```
<---area 1--><-area 0--><-area 2->

 R1---R2----R3---R6    R8-----R10
```



```
  |          |    |   / |\    |
  |          |    | --  | --\ |
  |          |    |/    |    \|
  |---R4----R5---R7----R9-----R11
```

Assumptions
- R3, R5, R8 and R9 are ABRs
- A TE LSP1 from R1 (Head-End LSR) to R11 (Tail-end LSR) is defined with the following loosely routed path: R1-R3(loose)-R8(loose)-R11(loose):R3, R8 and R11 are defined as loose hops.

Step 1: LSP 1's Head-end (R1) builds the following ERO object: R1(S)-R2(S)-R3(S)-R8(L)-R11(L)
where:
S: Strict hop (L=0)
L: Loose hop (L=1)
The R1-R2-R3 path obeys the TE LSP1's set of constraints

Step 2: the RSVP Path message is then forwarded by R1 following the ERO path 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 expand operation to reach the next loose hops R8. The new ERO becomes: R6(S)-R7(S)-R8(S)-R11(L).

Step 4: the same procedure applies at R8.
...

2. Reoptimization of a loosely routed explicit TE LSP path

Once the TE LSP is set up, it is maintained through normal RSVP procedures. Then a more optimal path might appear between an LSR and its next loose hop (suppose in the example above that a link between R6 and R8 is added that provides a shorter path between R3 and R8 (R3-R6-R8) than the existing R3-R6-R7-R8 path). Currently, if the better path is not visible from the Head-end LSR, it cannot make use of this better path (and perform a make before break) when appropriate. This is for instance the case in the example above as the better path does not appear in the Head-end area.

This draft proposes a mechanism that allows:

- The TE LSP Head-end LSR to trigger on every LSR whose next hop is a loose hop the re evaluation of the current path in order to detect a potential more optimal path,

- An LSR whose next hop is a loose-hop to signal (using a new ERROR-SPEC sub code carried in a Path Error Notify message) to

the TE LSP head-end that a better path exists (a path with a lower cost, where the cost is defined by the metric used to compute the path - see [SEC-METRIC], [METRIC]).

Then once the existence of a better path is notified to the Head-end LSR, it can perform a make before break. This allows the Head-end LSR to reoptimize a TE LSP making use of the non disruptive make before break procedure if and only if a better path exists.

3. Signalling extensions

3.1. ERO expansion signaling request

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

ERO Expansion request: 0x20

This flag indicates that a new ERO expansion is requested.

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

3.2. New Path Error sub-code

The format of a Path Error is the following:

```
<PathErr message> ::= <Common Header> [ <INTEGRITY> ]

                                      <SESSION> <ERROR_SPEC>

                                      [ <POLICY_DATA> ...]

                                     [ <sender descriptor> ]

           <sender descriptor> ::= (see earlier definition)
```

IPv4 ERROR_SPEC object: Class = 6, C-Type = 1

```
           +-------------+-------------+-------------+-------------+
           |            IPv4 Error Node Address (4 bytes)          |
           +-------------+-------------+-------------+-------------+
           |    Flags    |  Error Code |        Error Value        |
           +-------------+-------------+-------------+-------------+
```

Various Error Codes and Error values have been defined in RFC2205 and RFC3209.

The ERROR-CODE 25 corresponds to a Path Error - Notify Error. We propose to add three new sub-codes:

6 Better path exists
7 Local link maintenance required
8 Local node maintenance required

See details about Local maintenance required modes in section 4.3.2

4. Mode of operation

4.1. TE LSP reroute

The notification process of a better path is by nature de-correlated from the reoptimization operation. In other words, the location where a potentially more optimal path is discovered does not have to be where the TE LSP is actually reoptimized. In particular, when a better path is discovered one could conceivably envisage reoptimizing the TE LSP on a mid-point LSR or on the Head-end LSR of the TE LSP. In the former case, this would require some RSVP extensions but more importantly this may not be desirable in some circumstances: indeed, the reoptimization process inevitably generates some jitter, potentially packet reordering. By the way, the only LSR having the complete view of the end to end path and TE LSP set of attributes/constraints is the Head-end LSR. For those reasons, this draft applies to the context of a head-end LSR reoptimization. It is just worth mentioning that in some other contexts, mid-point reoptimization may also be desirable.

4.2. Reoptimization triggers

There are two 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).

4.3. Head-end reoptimization request versus mid-point reoptimization indication

This draft defines two modes:

- The request for a new path evaluation of an explicit loosely routed TE LSP is requested by the Head-end LSR.

- A mid-point LSR having determined that a better path (than the current path is use) exists or having the desire to perform a link/node local maintenance explicitly notifies the head-end LSR which will in turn perform a make before break.

4.3.1. Head-end reoptimization request

In this mode, 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 a Path message with the "ERO Expansion request" bit of the SESSION_ATTRIBUTE object set. This bit is then cleared in subsequent RSVP path messages sent downstream.

Upon receiving a Path message with the "ERO expansion request" bit of the SESSION_ATTRIBUTE object set, every LSR for which the next abstract node contained in the ERO is defined as a loose hop, must perform the following set of actions:

- A new ERO expansion is triggered and the newly computed path is compared to the existing path:

    - If a better path can be found, the LSR MUST immediately send a Path Error to the head-end (Error code 25 (Notify), sub-code=6 (better path exists)). At this point, the LSR MAY decide to clear the ERO expansion request bit of the SESSION-ATTRIBUTE object in subsequent RSVP Path messages sent downstream: this mode is the recommended mode. Indeed, the sending of a Path Error Notify message "Better path exists" to the Head-end LSR will trigger a Head-end reoptimization so triggering ERO expansions on downstream nodes is unnecessary. The only motivation to forward subsequent RSVP Path messages with the "Expansion 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 delay of the
reoptimized TE LSP.

- No better path can be found: as previously stated,
the recommended mode is for an LSR to relay the request
(by setting the ERO expansion bit of the SESSION-
ATTRIBUTE object in RSVP path message sent downstream)
only if no better path has been found on this LSR.

Note: by better path, we mean a path having a lower cost. By default, an LSR uses the IGP metric in their CSPF to detect the shortest path that obeys a set of constraints. Note that the head-end might use the METRIC-TYPE object (defined in [PATH-COMP]) in its path message to

request the LSR having a next hop defined as a loose hop in the ERO to use the TE metric to determine the best path.

Example 2:

Let call Ln the list of LSRs defined as loose hops in the ERO sent in the Path message by the Head-end LSR: Ln=<l1, l2, ..., ln>. Let's now call Pn=<p1, p2, ..., pn> the list of LSRs pi such that li is a next (loose) hop of pi for i=1...n

```
<---area 1--><-area 0--><-area 2->

 R1---R2----R3---R6    R8-----R10
  |          |    |   / |\    |
  |          |    | --  | --\ |
  |          |    |/    |    \|
  |---R4----R5---R7----R9-----R11
```

A TE LSP1 from R1 (Head-End LSR) to R11 (Tail-end LSR) is defined with the following loosely routed path: R1-R3(loose)-R8(loose)-R11(loose). R3, R8 and R11 are defined as loose hops.

Ln=<R3,R8,R11>
Pn=<R1,R3,R8>

As soon as a positive response is received from an LSR pi (sub-code=6, "Better path exists"), the Head-end LSR MUST perform a make before break.

Note that if the Path message with the ERO expansion request bit set of the SESSION-ATTRIBUTE object is lost, then the next request will be sent when the reoptimization event will trigger on the Head-end LSR. The solution to handle RSVP reliable messaging has been defined in [REFRESH-REDUCTION].

4.3.2. Mid-point reoptimization indication

In this mode, an LSR whose next abstract node is a loose hop can locally trigger an ERO expansion (when a configurable timer expires or on event-driven basis (link-up event for example) or the user explicitly requests it). If a better path is found compared to the existing one, the LSR sends a Path Error to the head-end (Error code 25 (Notify), sub-code=6 (better path exists)). The Head-end LSR MUST then immediately perform a make before break.

There are other circumstance by which a mid-point may send an RSVP Path Error Notify 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 mid-point on which the local

maintenance must be performed is responsible for sending an RSVP Path Error Notify message with the sub-code=7 or 8 depending on the affected network element (link or node). Then the first upstream node having 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 Path Error message MUST be immediately forwarded unchanged upstream to the Head-end LSR.

Upon, receiving a Path Error Notify message with sub-code 7 or 8, the Head-end LSR MUST perform a make before break.

Note that those modes are not exclusive: both the timer and even-driven reoptimization triggers can be implemented on the Head-end and/or any mid-point LSR with potentially different timer values for the timer driven reoptimization case.

4.3.3. ERO caching

Once a mid-point LSR has determined that a better path exists (after a reoptimization request has been received by the Head-end LSR or the reoptimization timer on the mid-point has fired), the more optimal path MAY be cached on the mid-point for a limited amount of time to avoid having to recompute a route once the Head-LSR performs a make before break. This mode is optional.

5. Interoperability

An LSR non supporting the "ERO expansion request" bit of the SESSION-ATTRIBUTE object SHOULD just ignore it.

Any Head-end LSR non supporting this draft receiving a Path Error Notify message with sub-code = 6, 7 or 8MUST just silently ignore the Path message.

6. Security Considerations

The practice described in this draft does not raise specific security issues beyond those of existing TE.

7. Acknowledgment

The authors would like to thank Carol Iturralde, Miya Kohno, Francois Le Faucheur, Philip Matthews and Jim Gibson for their useful and valuable comments.

8. Intellectual Property

The contributor represents that he has disclosed the existence of any proprietary or intellectual property rights in the contribution that are reasonably and personally known to the contributor. The contributor does not represent that he personally knows of all potentially pertinent proprietary and intellectual property rights owned or claimed by the organization he represents (if any) or third parties.

References

[TE-REQ] Awduche et al, Requirements for Traffic Engineering over MPLS, RFC2702, September 1999.

[OSPF-TE] Katz, Yeung, Traffic Engineering Extensions to OSPF, draft-katz-yeung-ospf-traffic-05.txt, June 2001.

[ISIS-TE] Smit, Li, IS-IS extensions for Traffic Engineering, draft-ietf-isis-traffic-03.txt, June 2001.

[RSVP-TE] Awduche et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC3209, December 2001.

[METRICS] Fedyk et al, "Multiple Metrics for Traffic Engineering with IS-IS and OSPF", draft-fedyk-isis-ospf-te-metrics-01.txt, November 2000.

[DS-TE] Le Faucheur et al, "Requirements for support of Diff-Serv-aware MPLS Traffic Engineering", draft-ietf-tewg-diff-te-reqts-01.txt, June 2001.

[MULTI-AREA-TE] Kompella at all, "Multi-area MPLS Traffic Engineering", draft-kompella-mpls-multiarea-te-03.txt, June 2002.

[PATH-COMP] Vasseur et al, "RSVP Path computation request and reply messages", draft-vasseur-mpls-computation-rsvp-03.txt, June 2002.

[SEC-METRIC] Le Faucheur et all," Use of Interior Gateway Protocol (IGP) Metric as a second MPLS Traffic Engineering Metric", draft-ietf-tewg-te-metric-igp-02.txt, September, 2002.

[INTER-AS-TE-REQS] Zhang et al, "MPLS Inter-AS Traffic Engineering

requirements", draft-ietf-tewg-interas-mpls-te-req-00.txt, June 2003,
Work in progress.

[INTER-AS-TE] Vasseur and Zhang, "MPLS Inter-AS Traffic Engineering",
draft-vasseur-inter-as-te-00.txt, February 2003, work in progress.

[REFRESH-REDUCTION] Berger et al, "RSVP Refresh Overhead Reduction
Extensions", April 2001

Authors' addresses:

Jean Philippe Vasseur
Cisco Systems, Inc.
300 Beaver Brook Road
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