Jean-Philippe Vasseur (Editor)
Cisco Systems, Inc.
Raymond Zhang
Infonet Services Corporation

IETF Internet Draft
Expires: December, 2003

June, 2003

draft-vasseur-inter-as-te-01.txt

Inter-AS MPLS Traffic Engineering

Status of this Memo

This document is an Internet-Draft and is in full conformance with all
provisions of Section 10 of RFC2026. Internet-Drafts are Working documents of
the Internet Engineering Task Force (IETF), its areas, and its working
groups. Note that other groups may also distribute working documents as
Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at
http://www.ietf.org/ietf/1id-abstracts.txt.
The list of Internet-Draft Shadow Directories can be accessed at
http://www.ietf.org/shadow.html.

Abstract

This draft proposes a set of mechanisms to establish and maintain MPLS
Traffic Engineering Label Switched Path that span multiple Autonomous
Systems, considering the set of requirements for inter-AS Traffic Engineering
listed in [INTERnter-AS-TE-REQS]. Each mechanism is described along with its
applicability to the set of requirements.

Content

[1](1). Terminology

LSR - Label Switch Router

LSP - An MPLS Label Switched Path

PCS - Path Computation Server (may be any kind of LSR (ABR, ...) or a
      centralized path computation server

PCC - Path Computation Client (any head-end LSR) requesting a path
      computation of the Path Computation Server.

Local Repair - Techniques used to repair LSP tunnels quickly when a node or
               link along the LSPs path fails.

Protected LSP - An LSP is said to be protected at a given hop if it has one
                or multiple associated backup tunnels originating at that
                hop.

Bypass Tunnel - An LSP that is used to protect a set of LSPs passing over a
                common facility.

PLR - Point of Local Repair. The head-end of a bypass tunnel.

MP - Merge Point. The LSR where bypass tunnels meet the protected LSP.

NHOP Bypass Tunnel - Next-Hop Bypass Tunnel. A backup tunnel which bypasses
                    a single link of the protected LSP.

NNHOP Bypass Tunnel - Next-Next-Hop Bypass Tunnel. A backup tunnel which
                      bypasses a single node of the protected LSP.

Reroutable LSP - Any LSP for with the "Local protection desired" bit is set
                 in the Flag field of the SESSION_ATTRIBUTE object of its
                 Path messages.

CSPF - Constraint-based Shortest Path First.

Inter-AS MPLS TE - TE LSP whose Head-end LSR and Tail-end LSR do not reside
                   within the same Autonomous System (AS) or both Head-end
                   LSR and Tail-end LSR are in the same AS but the TE tunnel
                   transiting path may be across different ASes. Note that
                   this definition also applies to TE LSP whose Head-end and
                   Tail-end LSRs reside in different sub-ASes
                   (BGP confederations).

Interconnect or ASBR Routers - Routers used to connect to another AS of a
                               different or the same Service Provider via one
                               or more Inter-AS links.

Region - a "region" is either a single IGP area, or a single Autonomous
         System (AS).

Boundary LSR - a boundary LSR is either an ABR if the region is an IGP area
               /level or an ASBR is the region is an AS.

2. Introduction

Considering the set of requirements for inter-AS Traffic Engineering listed
in [INTER-AS-TE-REQS], this draft proposes a set of mechanisms to establish
and maintain MPLS Traffic Engineering Label Switched Path that span:

    (1) multiple Autonomous Systems, or
    (2) sub-ASes in the context of BGP confederation. In this later case, the

draft applies to the case of multiple sub-ASes having their own
        routing domain.

Two scenarios are proposed in this draft that differ in terms of complexity
and optimality. For each scenario, the set of mechanisms is described along
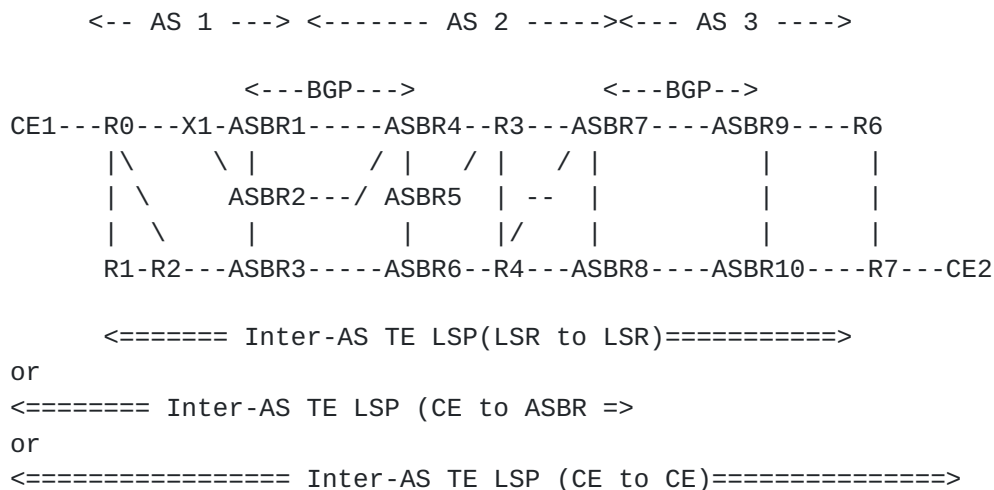with its applicability to the inter-AS TE requirements.

For the sake of reference, [INTER-REGION] focuses on signaling a TE LSP with
LSP segments or FA-LSPs in different regions (ASes) and exercise more
localized (per region) control on the inter-region TE LSP. There are also
other differences between this draft and [INTER-REGION] that will not be
detailed here.

Definition of an optimal Path

Qualifying a path as optimal requires clarification. Indeed, a globally
optimal TE LSP placement usually refers to a set of TE LSP whose placements
optimize the network resources (i.e a placement that reduces the maximum or
average network load for instance). By contrast, a optimal path for a TE LSP,
is the shortest path that obeys the set of required constraints (bandwidth,
affinities,...), minimizing either the IGP or TE metric cost (See [SECOND
-METRIC] and [MULTIPLE-METRICS]). In this draft, an optimal inter-AS TE path
is defined as the optimal path that would be obtained in the absence of AS
/Areas, in a totally flat network between the source and destination of the
TE LSP. Note that, the path calculated for a TE LSP in a distributed
environment, depends on the TE LSP set up order, as in a single area
environment.

3. General assumptions

In the rest of this draft, we will consider the following general case, built
on a superset of the various scenarios defined in [INTER-AS-TE-REQS]:

      <-- AS 1 ---> <------- AS 2 -----><--- AS 3 ---->

              <---BGP--->   <---BGP-->
CE1---R0---X1-ASBR1-----ASBR4--R3---ASBR7----ASBR9----R6
      |\   \ | / | / | / | | |
      | \   ASBR2---/ ASBR5 | -- | | |
      | \ | | |/ | | |
      R1-R2---ASBR3-----ASBR6--R4---ASBR8----ASBR10----R7---CE2

      <======= Inter-AS TE LSP(LSR to LSR)===========>
or
<======== Inter-AS TE LSP (CE to ASBR =>
or
<================= Inter-AS TE LSP (CE to CE)===============>

The diagram depicted above allows covering all the inter-AS TE deployment
cases described in [INTER-AS-TE-REQS].

Assumptions:

- Three interconnected ASes, respectively AS1, AS2, and AS3. Note that AS3
might be AS1 in some scenarios described in [INTER-AS-TE-REQS],

- The various ASBRs are BGP peers, without any IGP running on the single
hop link interconnecting the ASBRs,

- Each AS runs an IGP (IS-IS or OSPF) with the required IGP TE extensions
(see [OSPF-TE] and [IS-IS-TE]). In other words, the ASes are TE enabled,

- Each AS can be made of several areas. In this case, the TE LSP will rely on
the inter-area TE techniques to compute and set up a TE LSP traversing
multiple IGP areas (as described in [MULTI-AREA-TE]. For the sake of
simplicity, Eeach routing domain will be considered as single area in this
draft, but the solutions described in this draft does not prevent from making
use of multi-area techniques.

- An inter-AS TE LSP T1 originated at R0 in AS1 and terminating at R6 in AS3,

- An inter-AS TE LSP T2 originated at CE1 (connected to R0) and terminating
  at CE2 (connected to R7),

- A set of backup tunnels:

   o B1 from X1 to ASBR4 following the X1-ASBR2-ASBR4 path and protecting
     against a failure of the ASBR1 node,

   o B2 from ASBR1 to ASBR4 following the ASBR1-ASBR2-ASBR4 path and
     protecting against a failure of the ASBR1-ASBR4 link,

   o B3 from ASBR1 to R3 following the ASBR1-ASBR2-ASBR3-ASBR6-ASBR5-R3 path
     and protecting against a failure of the ASBR4 node.

- ASBR1, ASBR8 and ASBR9 have the function of PCS for the respective ASes
  1, 2 and 3 (the notion of PCS applies to the scenario 2 of this draft)..

4. Flooding of inter-AS non TE enabled links

As mentioned earlier, the links interconnecting ASBRs are usually not TE
enabled and no IGP is running at the AS boundaries.

This has two following implications:

   (1) The links interconnecting the ASBRs are not TE enabled,

(2) No IGP with TE extensions is running between ASBRs

An implementation supporting inter-AS MPLS TE MUST obviously allow the set
up of inter-AS TE LSP over the region interconnecting multiple ASBRs. In
other words, an ASBR compliant with this draft MUST support the set up of TE
LSP over ASBR to ASBR link, performing all the usual operations related to
MPLS Traffic Engineering (Call Admission Control,...) as defined in
[RSVP-TE]. So the limitation (1) must MUST be removed.

Regarding the second limitation (2), in the very vast majority of the cases,
two SPs usually do not run an IGP between ASBRs. Although this imposes for
the two ASBR to be interconnected via single hop link, this does not
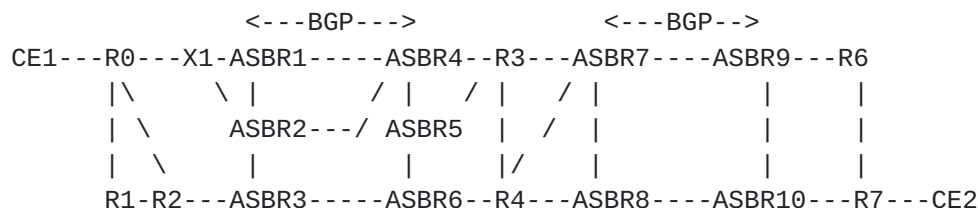constitute a severe limitation.

That being said, an interesting solution MUST allow the ASBRs to flood the TE
information related to the ASBR-ASBR link(s) although no IGP TE is enabled
over those links (and so there is no IGP adjacencies over the ASBR to ASBR
links). This allows a Head-end LSR to make a more appropriate route selection
up to the first ASBR in the next hop AS in the case of scenario 1 described
bellow and will significantly reduce the number of signaling steps in route
computation described in scenario 2. Note that the TE information is only
related to the ASBR-ASBR link. In other words, the TE LSA/LSP flooded by the
ASBR includes not only the links contained in the AS but also the ASBR-ASBR
links. No summarized TE information is leaked between ASes in any of the
proposed scenarios in this draft.

Example:

```
              <---BGP--->              <---BGP-->
CE1---R0---X1-ASBR1-----ASBR4--R3---ASBR7----ASBR9---R6
     |\     \ |        / |  / |  / |            |      |
     | \      ASBR2---/ ASBR5  | / |            |      |
     | \     |        |    |/  |            |      |
      R1-R2---ASBR3-----ASBR6--R4---ASBR8----ASBR10---R7---CE2
```

For example, in the diagram depicted above, when ASBR1 floods an its IGP TE
LSA (opaque LSA for OSPF) /LSP (TLV 22 for IS-IS) in its routing domain, it
reflects reflecting the reservation states and TE properties of the following
links: X1-ASBR1, ASBR1-ASBR4, ASBR1-ASBR2.

5. Definition of the LSP ATTRIBUTE Object

The object is optional and only required in environments where mid-point
LSRs may make use of different behaviors (contiguous LSP, stitching, nesting,
...). It allows a Head-end LSR to have a tighter control on mid-point

behaviors.

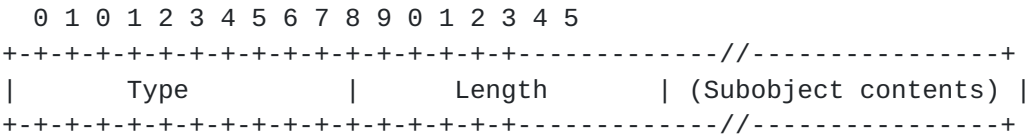The LSP-ATTRIBUTE object class-num is TBD (with a form 10bbbbbb) and C-type=1

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Reserved         |              Flags              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                             |
//                       subobjects                          //
|                                                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
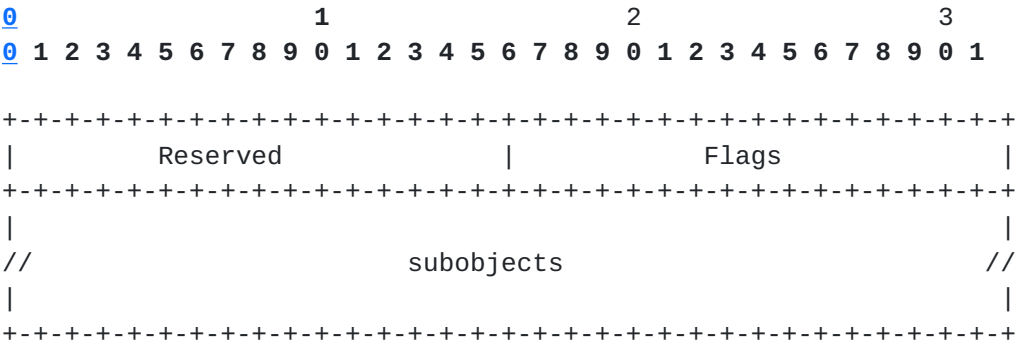
Each sub-object has the form:

```
  0 1 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-------------//---------------+
|      Type       |      Length      | (Subobject contents) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-------------//---------------+
```

No sub-object is currently defined.

Flags:

0x01: Crankback allowed bit: When set, the boundary LSP can either decide to forward the Path Error message upstream to the Head-end LSR or try to select another egress boundary LSR (which is also referred to as crankback). When cleared, this bit indicates that a boundary region LSR (an ABR or an ASBR LSR), when receiving a Path Error message from a downstream boundary LSR MUST propagate the Path Error message up to the inter-region head-end LSR.

0x02: Contiguous LSP required bit: when set, this indicates that a boundary LSR MUST not perform any stitching or nesting on the TE LSP and the TE LSP MUST be routed as any other TE LSP (it must be contiguous end to end). When this bit is cleared, a boundary LSR can decide to perform stitching or nesting. This bit MUST not be modified by any downstream node. See [INTER-REGION] for a definition of TE LSP stitching.

0x04: LSP stitching required bit: this flag is set for the segment LSP. This

flag SHOULD not be modified by any LSRs in between. If the egress LSR for the segment LSP does not understand this flag then it will simply forward the object unmodified and will send a Path Error message upstream with error sub-code=16. See [INTER-REGION] for a detailed description of the mode of operation.

The LSP-ATTRIBUTE object is carried in RSVP Path message. If a Path message contains multiple LSP-ATTRIBUTE objects, only the first object is meaningful. Subsequent LSP-ATTRIBUTES object MUST be silently discarded.

6. Scenario 1: Per-AS Traffic Engineering Path computation

In scenario 1, two modes are supported:

(1) Static loose hops configuration

Every inter-AS TE LSP path is defined as a set of loose and strict hops but at least the ASBRs traversed by the inter-AS TE LSP MUST be specified as loose hops on the Head-End LSR.

Examples of configuration of T1 on R0:

- Example 1 (set of strict hops end to end):
          R0-X1-ASBR1-ASBR4-ASBR5-R3-ASBR7-ASBR9-R6

- Example 2 (set of loose hops):
          R0-ASBR4(loose)-ASBR9(loose)-R6(loose)

- Example 3: (set of loose hops):
           R0-ASBR1(loose)-ASBR4(loose)-ASBR7(loose)-ASBR9(loose)-R6(loose)

- Example 4 (mix of strict and loose hops):
          R0-R2-ASBR3-ASBR2-ASBR1-ASBR4(loose)-ASBR10(loose)-ASBR9-R6

- Example 5 (mix of strict and loose hops):
          R0-ASBR2(loose)-ASBR4(loose)-ASBR7(loose)-ASBR9(loose)-R6

When a next hop is defined as a loose hop, a dynamic path calculation (also called ERO expand) is required taking into account the topology and TE information of its own AS and the set of TE LSP constraints. In the example 1 above, the inter-AS TE LSP path is statically configured as a set of strict hops, so in this case, no dynamic computation is required. In the example 2, a per-AS path computation is performed, respectively on the R0 for AS1, ASBR4 for AS2 and ASBR98 for AS3.

Vasseur and Zhang   7

draft-vasseur-inter-AS-TE-01.txt   June 2003

Note that when an LSR has to perform an ERO expansion, the next hop must either belong to the same AS, or must be the ASBR directly connected to the next hops AS. In this later case, the ASBR reachability must be announced in

the IGP TE LSA/LSP originated by its neighboring ASBR. In the example 2
above, the TE LSP path is defined as: R0-ASBR4(loose)-ASBR9(loose)-R6(loose).
This implies that the ERO expansion performed by R0 must compute the path
from R0 to ASBR4. As stated in [section 4](#), the TE reservation state related to
the ASBR1-ASBR4 link is flooded in ASBR1Æs IGP TE LSA/LSP. In addition, ASBR
[1](#) MUST also announce the IP address of ASBR4 used specified in the T1 path
configuration. If not possible/desired, then the set of loose hops must
include every traversed ASBRs as in the example 1 above.

(2) Dynamic mode

Every LSR receiving an RSVP Path message with an ERO object such that the
next hop is loose, will have to determine automatically the next hop ASBR,
based on the IGP/BGP reachability of the TE LSP destination.

Example of configuration of T1 on R0 in dynamic mode:
T1 Path: R0-R6(loose)

[6.1](#). Static loose hops configuration

At least, the set of ASBRs from the TE LSP Head-end to the Tail-End MUST be
specified as a set of loose hops. Optionally, a set of paths can be
configured on the HE LSR, ordered by priority.

Example: the T1 path can be defined as an ordered set of paths specified as
loose hops:

Priority 1: R0-ASBR1(loose)-ASBR4(loose)-ASBR9(loose)-R6(loose)
Priority 2: R0-ASBR2(loose)-ASBR7(loose)-ASBR9(loose)-R6(loose)
Priority 3: R0-ASBR3(loose)-ASBR6(loose)-ASBR7(loose)-R6 (loose)
Priority 4: R0-ASBR3(loose)-ASBR6(loose)-ASBR8(loose)-ASBR10(loose)-R6(loose)

Each priority path can be associated with a different set of constraint.
Typically, it might be desirable to systematically have a last resort option
with no constraint to ensure that the inter-AS TE could always be set up if
at least a path exist between the inter-AS TE LSP source and destination.

Note that in case of set up failure or when an RSVP Path Error is received
indicating the TE LSP has suffered a failure, an implementation might support
the possibility to retry a particular path option a specific amount of time
(optionally with dynamic intervals between each trial) before trying a lower
priority path option.

Although this mode requires some manual configuration, it allows exerting
control over which ASBRs are traversed, with an order of priority. Moreover,
the required extra configuration can be acceptable in environment where both
the number of ASBRs and traversed ASes is not too large. In the case of three
ASes AS1, AS2 and AS3 having respectively X1, X2 and X3 ASBRs the maximum
number of path-options can be as large as X1*X2*X3 combinations.

Note:

- Any path can be defined as a set of loose and strict hops. In other words, in some case, it might be desirable to rely on the dynamic path computation in some area, and exert a strict control on the path in other areas (defining strict hops).

- It is also worth mentioning that a SP may desire to have a strict control within its AS on the path followed by inter-AS TE LSP signaled with loose hops. When receiving the RSVP Path message with an ERO specifying a loose hop, the user has the ability to specify by local configuration a list of strict hops for the TE LSP instead of triggering a CSPF to dynamically compute the path.

6.2. Dynamic loose next hop hops computation

With this scheme, the HE LSR and every downstream ASBR loose hop (except the last loose hop that computes the path to the final destination) automatically computes the path up to the next ASBR,the next loose hop based on the IGP/BGP reachability of the TE LSP destination. If a particular destination is reachable via multiple loose hops (ASBRs), local heuristics may be implemented by the HE LSR/ASBRs to select the next hop an ASBR among a list of possible choices (closest exit point, metric advertised for the IP destination (ex: OSPF LSA External - Type 2), local policy,...). Once the next ASBR has been determined, an ERO expansion is performed as in the previous case.

   (1) IGP reachability

   In the case of IGP, this implies for every ASBR to redistribute in their
   IGP the loopback address of every potential TE LSP destination LSR. Note
   that this may introduce some undesirable effect on the IGP both in term of
   IGP scalability as the number of redistributed loopback addresses might
   not be negligible and stability (any change of a loopback reachability in
   AS x will trigger an IGP LSA flooding and SPF computation in AS y). Note
   that the ASBR configuration should make sure that loopback reachability
   information are redistributed in IGP by every ASBR for redundancy
   purposes. In some cases (for instance with MPLS VPN), the information
   is available anyway.

   (2) BGP reachability

   In the case of BGP, this implies for every ASBR to announce to their BGP
   peers the loopback address of every potential TE LSP destination LSR. Some
   filtering may be applied at each ASBR.

6.3. Inter-AS TE LSP path set up

Once the next loose hop has been determined (via static configuration or dynamic computation), the HE LSR initiates the TE LSP path set up according to the procedures defined in [RSVP-TE]. Loose hops are specified in the ERO object of the Path message with the L flag of the Ipv4 prefix sub-object set, as per [RSVP-TE]. Then each loose hopsASBR along the path reiterates the same operation as described above until the TE LSP set up is completed.

In case of failure of an LSR to find a path up to the next loose hop obeying the set of required constraint:

- With the static exit loose point configuration, an RSVP Path Error is sent to the HE LSR,

- With the dynamic exit next hop point computation: if an LSP-ATTRIBUTE object is present in the RSVP Path message and the crankback allowed bit is set,, the node might either decide to try to find another loose hop next hop ASBR to reach the next hop AS or to send an RSVP Path Error up to the HE LSR. This is a matter of local policy. If the bit is cleared, then the node MUST send a Path Error to the Head-end LSR.

Vasseur and Zhang   9

Example (with static loose hops configuration):

Assumption: an inter-AS TE LSP T1 originated at R0 in AS1 and terminating at R6 in AS3 must be set up.

- Step 1: discovery (via configuration) of the next loose hop to reach the destination AS by the HE LSR.

- Step 2: the HE LSR computes the TE LSP path up to the first loose hop and builds the ERO as a set of strict hops up (to the next loose hop,) followed by a list of loose hops.

  Ex: ERO generated at the HE LSR (R0): R0(strict)-X1(strict)-ASBR1(strict)
      -ASBR4(strict)-ASBR8(loose hop)-ASBR9(loose hop)-R6

- Step 3: the inter-AS TE LSP is signaled along the computed path up to the next loose hop that performs a similar operation up to the next loose hop until the TE LSP destination node is reached.

  Ex: in the example above, the next loose hop is ASBR48. ASBR4 computes the
      TE LSP path up to ASBR8 (ERO= ASBR4-R3-R4-ASBR6-ASBR8) and signals the
      TE LSP. Then, ASBR8 computes the TE LSP Path up to ASBR9 and expands
      the ERO which becomes: ASBR10-ASBR9. Finally, ASBR9 computes the last
      segment of the inter-AS TE LSP path: ASBR9-R6.

Notes: in case of TE LSP set up failure:

- With the static loose hop configuration: an RSVP Path Error messages in
  sent to the HE LSR that will in turn select another sequence of loose
  hops, if configured. Alternatively, the HE LSR may decide to retry the
  same path; this can be useful in case of set up failure due an outdated
  IGP TE database in some downstream AS. An alternative could also be for
  the HE LSR to retry to same sequence of loose hops after having relaxed
  some constraint(s).

- With the dynamic loose next hop computation, as explained above, there are
  two possible modes: an RSVP Path Error message is sent to the HE LSR (as in
  the previous case) or the previous loose node tries another to find another
  loose next hop.

  Ex: ASBR4 selects ASBR75 as its next loose hop and receives an RSVP Path
      Error from R3 (not enough bandwidth on the R3-ASBR75 link - IGP TE
      database not up to date). In this case, instead of relaying the RSVP
      Path Error up to the HE LSR RO,R0, ASBR4 can try to find an alternative
      loose next hop hopASBR: ASBR8.

  This mode is usually referred as "crankback". Note that depending on the
  network topology crankback may or may not result in more optimal solutions
  than Head-end rerouting.

6.4. Support of diversely routed paths

There are several circumstances where the ability to set up a set of
diversely routed TE LSP paths between two LSRs might be desirable:

(1) Load balancing

When a single TE LSP path satisfying the required constraints cannot be found
between two LSRs, an alternative may consist in setting up N TE LSPs such
that the sum of their bandwidth is equal to the total required bandwidth. In
addition, having diverse paths allows in case of network element failure to
limiting the traffic disruption between the two nodes to the set of affected
TE LSPs.

(2) Path Protection

In some networks, Path protection is used to protect TE LSP from link/SRLG
/node failure. This requires setting up, for each TE LSP, a diversely routed
TE LSP. In case of failure along the primary TE LSP path, the node directly
attached to the failed resources signals to the HE LSR that the TE LSP has
failed, sending an RSVP Path Error. The HE LSR can also detect that the TE

LSP has suffered a failure when receiving an IGP update reflecting the
failed resource. Note that the HE LSR cannot rely on the IGP topology
database to detect the failure if the failure does not occur in the Head-End
area/AS. Once the HE LSR learns the failure, the traffic is switched onto the
pre-established backup TE LSP. Note that a set of TE LSP can potentially
share a single backup TE LSP (1:N protection).

In the case of scenario 1, the set up of N diversely routed TE LSP paths can
be done using the following scenario:

- Set up the first TE LSP among the set of N TE LSPs and include an RSVP RRO
  object in the RSVP Resv message to record the Path,

- For i=2 to N, Set up the TE LSPi, excluding the elements traversed by the
  already set up TE LSP1,..., TE LSP i-1. The exclusion of a set of resources
  from a TE LSP path can be performed on the HE LSR by the CSPF and in other
  ASes by the loose hops along the path, each of them performing the
  computation of a part of the TE LSP. This requires from the HE to pass the
  "exclude" constraints (see [EXCLUDE-ROUTE]);

Important note: such an algorithm does not guarantee that a diverse path can
be found for the successive TE LSPs since the TE LSP path are not
simultaneously computed, even if such a possible solution exists.

6.5. Optimality

In this scenario, the resulting TE LSP path is the first path obeying the
required set of constraints. As such, the path might not be optimal
(the shortest path). This gets particularly true as TE LSP gets rerouted due
the network element failures..

7. Scenario 2: Path Computation Server

In this scenario, the inter-AS TE LSP path computation is off-loaded on a
PCS. Note that a PCS may be an LSR (like an ASBR) or a centralized path
computation tool (in this case the PCS must get the IGP topology and
resources information by some means, outside of the scope of this draft).

Note that the PCS may or may not be along the TE LSP Path. This implies that
the PCS is just responsible for the TE LSP path computation only, not for its
maintenance. Moreover, the PCS may compute just a path segment, not the whole
path end to end; in this case, the returned computed path will contain loose
hops.

Vasseur and Zhang   11

7.1. Path Computation Server discovery

Any HE LSR requiring an inter-AS TE path computation first needs to learn the PCS address in order to send an RSVP path computation request. This can be done via static configuration or dynamic PCS discovery.

(1) Static configuration

A set of PCS(es) addresses is statically configured on every HE LSR having potentially some inter-AS TE LSP to set up. A dedicated PCS per AS can be configured or a default PCS serving the requests for all the destination ASes.

Optionally, more than one PCS address can be configured in case of PCS failure of a PCS. In this case, the HE LSR must implement an algorithm to select a backup PCS in case of non-response of the preferred PCS after a certain amount of time.

(2) Dynamic PCS discovery

Dynamic PCS discovery can be performed via IGP advertisement. [IGPOSPF-CAP] and [ISIS-CAP] defines IGP extensions for IP/LSR to advertise optional router capabilities. A Router capability TLV is defined for IS-IS while an router information opaque LSA is defined for OSPF. [OSPF-TE-CAP] defines a set of Traffic Engineering TLVs carried in the router capability LSA. Two TE TLVs are currently defined:
- The Path Computation Server Discovery (PCSD) TLV that allows a router or a centralized path computation tool to advertise its PCS capabilities within a routing domain,
- The Mesh-group TLV used by an LSR to indicate its desire to participate to a mesh of TE LSPs.

In particular, the PCSD TLV allows PCS to advertise the following capabilities:
- IPv4 to IPV6 address to be used by any PCC to send RSVP path computation request to the PCS,
- Path Computation server capabilities (ability to manage Path computation request priorities, support of diverse routes computation, support of network element exclusion in the path computation, multiple metric path computation,...)
- Set of ASes for which the PCS can compute inter-AS TE paths for.

Note that if the AS is made of multiple areas/levels, [IGPOSPF-CAP] and [ISIS-CAP] supports the capabilities announcements across the entire routing domain (making use of TLV leaking procedure for IS-IS and OSPF opaque LSA type 11 for OSPF)..

7.2. PCC-PCS Signaling protocol

Any PCC (i.e LSR) can send an RSVP path computation request to a PCS that will in turn compute a set of TE LSP(s) path and return the corresponding path parameters via an RSVP path computation reply message. The format of the RSVP path computation requests and reply messages are defined in [PATH-COMP]

as well as the set of optional objects characterizing the constraints:

REQUEST-ID object: must be present in any RSVP Path computation request and reply message and specifies:

   - The request-ID-number,

   - Several requests characteristics. Whether:
      o the Path cost is requested by the PCC,
      o the Path must be complete (set of strict hops) or can be partial,
      o the request concerns a reoptimization. In this case, the current path
        of the TE LSP to be reoptimized is provided in the path computation
        request to avoid double bandwidth booking,
      o the TE LSP is a bi-directional TE LSP,
      o if a path fully satisfying the required constraint cannot be found,
        the PCC has the possibility to indicate that a less-constrained TE
        LSP path is of interest,
      o the PCC can also indicate the urgency of the request (two priorities
        are currently defined: normal and high).

METRIC-TYPE object: allows the PCC to indicate to the PCS the metric to be used to compute the shortest path (currently two metrics are defined: the IGP or TE metric).

PATH-COST object: object inserted in the RSVP path computation reply message to indicate the cost of a computed TE LSP in addition to the path. This object is mandatory if the cost has been explicitly requested in the RSVP path computation request and optional in any other case.

NO-PATH-AVAILABLE object: when no path satisfying the set of required constraints for a TE LSP can be found, some PCS might have the capability to specify the constraint responsible of the path computation failure. Optionally, a less-constrained path can be provided by the PCS (if an interest in less-constrained TE LSP has been requested in the RSVP path computation request). A PCS might also suggest new constraints for which a path could be found.

NB-PATH, PATH-CORRELATED, MIN-BW and LSP-BANDWIDTH objects: a PCC may desire to request the computation of more than one TE LSP. This object allows the PCC to request the computation of a set of TE LSPs such that the sum of their bandwidth is equal to X, with potentially a constraint on the number and minimum bandwidth size of each of the TE LSPs in the set. The N TE LSP may or not share the same set of constraints and can be correlated (diversely routed).

EXCLUDE-ELEMENT object: allows the PCC to request a TE LSP path such that the

path does not include a set of network elements (link, node or AS).

The protocol state machine is also defined in [PATH-COMP].

7.3. Inter-AS TE LSP Path set up

The end-to-end inter-AS TE LSP set up will be made of the following steps (at each step, an example is provided, based on the assumptions made in section 3):

Hypothesis: an inter-AS TE LSP must be set up between AS1 and ASn (in the example provided bellow n=3 - an inter-AS TE LSP T1 originated at R0 in AS1 and terminating at R6 in AS3 must be set up).

Step 1: discovery by the HE LSR of the PCS serving the source AS that can compute inter-AS TE LSP path terminating in ASn. As mentioned above, the PCS can either be statically configured or dynamically discovered via IGP extensions.

Ex: R1 selects ASBR1 as the PCS serving its request for the T1 path computation.

Step 2: an RSVP Path computation request is sent to the selected PCS. Note that the RSVP path computation request can be sent either:

(1) to a PCS in the same AS as the PCC that will relay the request to a PCS in the next hop AS.

   Ex: R0 sends a RSVP path computation request to ASBR1

or
(2) can be directly sent to the PCS in the next hop AS if the HE LSR has a complete topology and TE view up to the next hop PCS (depending on the configuration).

   Ex: R0 sends an RSVP path computation request to ASBR4

Note on the PCC-PCS and PCS-PCS communication (when a PCC sends a request to a PCS in another AS) and PCS-PCS (when the PCSes belong to different ASes)
   - some policy may be put in place between SPs,
   - the communication MIGHT be secured making use of RSVP authentication

It is expected that (1) will be the most common inter-AS TE deployment scenario.

Step i:

The PCS of ASi relays the path computation request to the targeted
PCS peer in AS(i+1) located in the next hop AS. Note that the address of the
list of PCS(es) in the next hop AS must be statically configured on the PCS.
This implies some static configuration on the PCS only as opposed to the PCS
address configuration required on every potential PCC within an AS.

Ex: ASBR1 sends an RSVP path computation request to ASBR4

Step i+1:

If the TE LSP destination is in ASi, then the PCS provides the list of
shortest path (with their corresponding ERO (potentially partial ERO) +
Path-cost) from every ASBR to the inter-AS TE LSP destination. See a detailed
example bellow.

If the TE LSP destination is not in ASi, the PCS relays the RSVP path
computation request to the targeted PCS peer in AS(i+2) in the next hop AS.

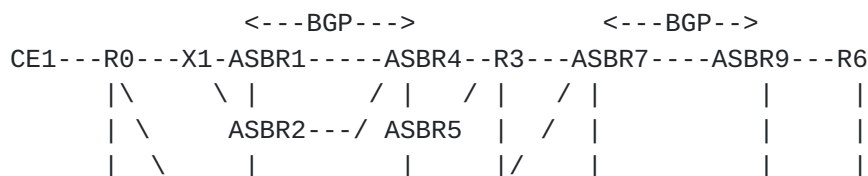The process is iterated until the destination AS is reached.

Ex: ASBR4 relays the RSVP path computation request to ASBR9 which determines
that the T1Æs destination address belongs to its own AS. ASBR9 will in turn
return a path computation reply to the requesting PCS (which is considered as
a PCC in this case), e.g ASBR4. The RSVP path computation reply contains two
routes:

    - ERO 1: ASBR9-R6, Cost=c1,
    - ERO 2: ASBR10-R7-R6, Cost=c2

Note that the ERO object might be made of loose hop to preserve
confidentiality.

Step i+2:

Once a requesting PCSi receives an RSVP Path computation reply, the
shortest path is computed from ASi to ASi+1 by route concatenation. This is
done by running a virtual SPT (Shortest Path Tree) computation using SPF
where the nodes are the ABSRs connected by virtual link whose cost are equal
to the shortest path connecting the ASBRs.

Vasseur and Zhang   14

```
            <---BGP--->            <---BGP-->
CE1---R0---X1-ASBR1-----ASBR4--R3---ASBR7----ASBR9---R6
     |\     \ |      / |  / |  / |            |     |
     | \     ASBR2---/ ASBR5 |  / |           |     |
     | \     |        |    |/    |            |     |
```

```
      R1-R2---ASBR3-----ASBR6--R4---ASBR8----ASBR10---R7---CE2
```

Resulting Virtual SPT computed by ASBR 4

```
ASBR4---ASBR7---ASBR9---R6
| |  \ /  /| \  / |    /
| |   \ / |  \/  |   /
| |  / \/  |  /\  |  /
| | /  /\  | /  \ | /
|ASBR5--ASBR8---ASBR10
| |   / /
| |  / /
| | / /
| |/ /
ASBR6
```

Within ASi, the cost of each ASBR-ASBR virtual link is equal to the shortest
path cost. This information is known by PCSi.

The cost of the ASBR-ASBR link between ASBR of different ASes is also known
by the PCSi (see [section 4](#)).

Within ASi+1, the cost of the ASBR-ASBR virtual link is provided in the RSVP
path computation reply of the PCSi+1.

Ex: ASBR4 will then compute the shortest path for the TE LSP traversing AS2
and AS3.

Then the process is reiterated recursively until the end-to-end Path
computation is computed. The whole path may not be seen by each PCS for
confidentiality reason but this process will ensure that the shortest path is
selected.

Example: the resulting computed virtual SPT computed by ASBR1 will finally
be:

```
R0-----ASBR1-----ASBR4----R6
| \       |  |\ \ /  /||   / /
|  \      |  | \ \  / ||  / /
|   \     |  |  / \/  || / /
|    \    |  | / \/\  ||/  |
|     \-|-ASBR2-ASBR5     |
|       |  |\  /\/  ||   |
|       |  | \/ /\  ||   /
|       |  | /\/  \ ||  /
|       |  |/ /\   \|| /
--------|=ASBR3--ASBR6/
```

Then once the optimal end to end path is computed, the HE LSR sets up the
inter-AS TE using a complete list of ERO if the various PCS have provided the
list of ERO or some loose hops in the contrary, similarly to the scenario 1.

7.4. **Support of diversely routed paths**

As mentioned above in section 5.2, the RSVP signaling protocol defined in
[PATH-COMP] allows an LSR to request the computation of a set of N diversely
routed TE LSPs. Then in this scenario, a set of diversely routed TE LSP
between two LSRs can be computed.

7.5. **Optimality**

As opposed to scenario 1, and end-to-end shortest path obeying the set of
required constraint is computed. In that sense, the path is optimal.

Note that in order for this path computation to be meaningful, a metric
normalization between ASes is required. One solution to avoid IGP metric
modification would be for the SPs to agree on a TE metric normalization and
use the TE metric for TE LSP path computation (in that case, this must be
requested in the RSVP Path computation request via the METRIC-COST object).

8. **Routing traffic onto inter-AS TE LSPs**

Once an inter-AS TE LSP has been set up, the Head-end LSR has to determine
the set of traffic to be routed onto the inter-AS TE LSP.

In the case of intra-area TE LSP, various options are available:

(1) modify the IGP SPF such that shortest path calculation can be performed
taking into account existing TE LSP, with some path preference,

(2) make use of static routing. Note that the recursive route resolution of
BGP allows routing any traffic to a particular (MP)BGP peer making use of a
unique static route pointing the BGP peer address to the TE LSP. So any
routes advertised by the BGP peer (IPv4/VPNv4 routes) will be reached using
the TE LSP.

With an inter-AS TE LSP, just the mode (2) is available, as the TE LSP
Head-end does not have any topology information related to the destination
AS.

9. **Applicability of MPLS TE Fast Reroute to Inter-AS TE LSP**

MPLS Traffic Engineering Fast Reroute ([FAST-REROUTE]) defines a local
protection scheme that provides fast recovery (in 10s of msecs) of protected
TE LSPs upon link/SRLG/Node failure. A backup TE LSP is configured and
signaled at each hop and upon detecting a network element failure (via link
failure detection mechanisms providing via layer 2 protocol, or IGP/RSVP fast

hellos), the node immediately upstream to the failure (called the PLR (Point
of Local Repair)) reroutes the set of protected TE LSPs onto the appropriate
backup tunnel(s) around the failed resources. In the context of Inter-AS TE
LSP, one must consider the following failure scenarios and analyze for each
of them the potential required extensions for MPLS TE FRR. [FAST-REROUTE]
specifies two modes referred to as the one to one and facility back up mode.
This section specifies the use of MPLS TE Fast Reroute for the facility backup
mode.

9.1. Within an Autonomous System (link or node failure)

The current set of mechanisms defined in [FAST-REROUTE] applies without any
restriction to any link/SRLG/Node failure within an AS. In other words, a
protected inter-AS TE LSP (an LSP signaled with the "local protection
desired" bit set in the SESSION-ATTRIBUTE object or with a FAST-REROUTE
object) can be protected via the MPLS TE Fast Reroute mechanism regardless
of whether the TE LSP is an intra-AS or inter-AS TE LSP in case of link/SRLG
/node failure within the AS.

9.2. At AS boundaries

9.2.1. Failure of an ASBR-ASBR link

To protect a link connecting two ASBRs with MPLS TE Fast Reroute, the following
actions are required:

- A set of backup tunnels must be configured between the two ASBRs diversely
routed from the protected ASBR to ASBR link. Typically, the region connecting
two ASes is not TE enabled. So an implementation will have to support the set
up of TE LSP over a non-TE enabled region. The backup tunnel path will be
configured on each ASBR as a set of strict hops and then signaled via the
RSVP-TE procedure defined in RFC3209.

The reason why a set of NHOP backup tunnels might be required is in case of
requirement for bandwidth protection if a single backup tunnel satisfying the
bandwidth requirement cannot be found (see [BANDWIDTH-PROTECTION]).

- For each protected inter-AS TE LSP traversing the protected link, a NHOP
backup must be selected by a PLR (i.e ASBR), when the TE LSP is first set up.
This requires for the PLR to select a backup tunnel terminating at the NHOP.

Finding the NHOP backup tunnel of an inter-AS LSP can be achieved by
analyzing the content of the RRO object received in the RSVP Resv message of
both the backup tunnel and the protected TE LSP(s). As defined in [RSVP-TE],
the addresses specified in the RRO IPv4 subobjects can be node-ids and/or
interface addresses (with specific recommendation to use the interface

address of the outgoing Path messages). Within a single area, the PLR can easily find whether the backup tunnel intersects the protected TE LSP regardless of whether the node-id or the interfaces are specified in the RRO object as it has the complete topology knowledge in its IGP database. This is not the case when the MP resides in a different AS. [NODEID] proposes a solution to this issue, defining an additional RRO IPv4 suboject that specifies a node-id address.

Example: The ASBR1-ASBR4 link is protected by the backup tunnel B1 that follows the ASBR1-ASBR2-ASBR4 path.

9.2.2. Failure of an ASBR LSR node

To protect inter-AS TE LSP from an ASBR node failure using MPLS TE Fast Reroute, the following actions are required:

(1) a set of backup tunnel(s) must be configured from the penultimate hop in AS1 (penultimate node directly connected to the last ASBR in AS1) to the first ASBR in AS2 to protect against the failure of the last ASBR in AS1.

   Ex: B1 from X1 to ASBR4 follows the X1-ASBR2-ASBR4 path and protects against the failure of the ASBR1 node.

(2) a set of backup tunnel(s) must be configured from the last ASBR in AS1 to the next hop of the first ASBR in AS2 to protect against the failure of the first ASBR in AS2.

   Ex: B3 from ASBR1 to R3 follows the ASBR1-ASBR2-ASBR3-ASBR6-ASBR5-R3 path and protects against the failure of the ASBR4 node.

- for each protected inter-AS TE LSP traversing the protected link/node, a NNHOP backup must be selected by a PLR (i.e LSR/ASBR). This requires for the PLR to select a backup tunnel terminating at the NNHOP.

Finding the NNHOP backup tunnel of an inter-AS LSP can be achieved by analyzing the content of the RRO object received in the RSVP Resv message of both the backup tunnel and the protected TE LSP(s).

9.3. Procedures of PLR during Fast Reroute

For the sake of illustration, procedures related to MPLS TE Fast Reroute that indifferently apply to intra-AS or inter-AS TE LSP will be first described followed by the required changes specific to inter-AS TE LSP.

Upon link/SRLG/Node failure detection, a PLR must immediately start rerouting the protected TE LSP onto their respective backup tunnel (which has been

selected when the protected inter-AS TE LSP has first been set up). In
addition, the RSVP control traffic must be sent over the backup tunnel (this
includes RSVP Path, PathTear, and ResvConf messages - the MP (Merge Point)
sends Resv, ResvTear, and PathErr messages to the address specified in the
RSVP_HOP object).

The following changes are performed:

   - The SESSION object is unchanged,

   - The SESSION-ATTRIBUTE object is unchanged except:
     The "Local protection desired", "Bandwidth protection desired", and
     "Node protection desired" flags SHOULD be cleared.
     The "Label recording desired" MAY be modified.

   - The IPv4 (or IPv6) tunnel sender address of the SENDER_TEMPLATE is set
     to an address belonging to the PLR (if the PLR is also the Head-End LSR
     an IP address different from the IP address used in the original
     SESSION-TEMPLATE must be chosen by the PLR).

   - The RSVP_HOP object is set to an IP source address belonging to the PLR.
     Consequently, the MP will send messages back to the PLR using as a
     destination that IP address.

   - The PLR must update the EXPLICIT_ROUTE object toward the egress. As
     mentioned in [FAST-REROUTE], the ERO object must be updated by the PLR
     before sending the RSVP Path message over the backup tunnel; otherwise
     the MP would receive an incorrect ERO object and would likely discard
     the Path message with the cause "Bad initial subobject" (in case of NHOP
     backup tunnel, the ERO would contain the IP address of a down interface
     - in case of NNHOP backup tunnel, the MP would receive the PLRÆs NHOP
     address).

Vasseur and Zhang   18

Within a single area, the PLR must:

   - Remove all the sub-objects preceding the first address belonging to the
     MP.

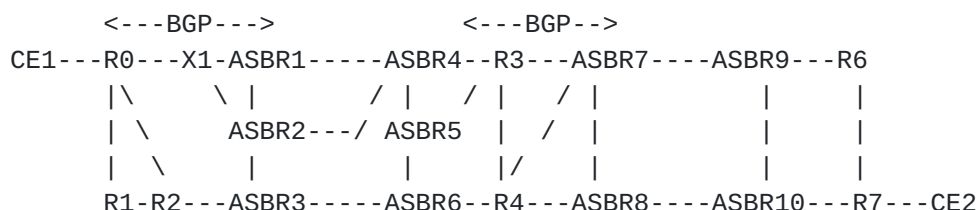   - Replace this first MP address with an IP address of the MP.

In the context of inter-AS TE LSP, there is a specific action that must be
performed when protecting the first ASBR of the next AS via a NNHOP backup
tunnel (see 5.6.3 (1)). The PLR must:

   - Identify the MP address in the RRO received in the corresponding Resv
     message,

- Remove all the sub-objects preceding the first address belonging to the
        MP,

      - Replace this first MP address with the IP address of the MP (its node-id
        address).

Note that in order to support ASBR Node protection for inter-AS TE LSP in
case of failure of the first ASBR in the next hop AS, some specific action
must be performed.

Example:

```
            <---BGP--->              <---BGP-->
        CE1---R0---X1-ASBR1-----ASBR4--R3---ASBR7----ASBR9---R6
             |\      \ |        / |   / |   / |            |      |
             | \     ASBR2---/ ASBR5  |  /   |            |      |
             |  \      |         |    |/     |            |      |
             R1-R2---ASBR3-----ASBR6--R4---ASBR8----ASBR10---R7---CE2
```

- T1: a protected inter-AS TE LSP from R0 to R6, whose path is defined on R0
      as a set of loose hops: R0-ASBR1(loose)-ASBR4(loose)-ASBR9(loose)-R6

- B3: a backup tunnel from ASBR1 to R3 following the ASBR1-ASBR2-ASBR3-ASBR6
      -ASBR5-R3 path and protecting against a failure of the ASBR4 node.

- The ERO subobject content signaled in the rerouted RSVP Path message of T1
  over B3 by ASBR1 (PLR) must content the MPÆs as the next hop address (R3).
  Otherwise, R3 will receive an incorrect ERO.

A similar mechanism is required when rerouting an inter-AS TE LSP from the
failure of the last ASBR of an AS.

- The RRO object may need to be updated by inserting an IPv4 or IPv6
subobject corresponding to the outbound interface address the rerouted
traffic is forwarded onto (both the "Local protection in use" and "Local
Protection Available" flags must be set).

Notification of Local repair

For each rerouted TE LSP, the PLR MUST send a notification of the local
repair by sending an RSVP Path Error message with error code of "Notify"
(Error code =25) and an error value field of ss00 cccc cccc cccc where ss=00
and the sub-code = 3 ("Tunnel locally repaired") (see [RSVP-TE]). In the
context of inter-AS TE LSP, if the failure occurs in an area/AS different

from the HE LSR, the HE LSR exclusively relies on the Path Error message to get informed that a local repair has been performed in order to potentially perform a reoptimization. The RSVP Path Error message SHOULD be sent in reliable mode ([REFRESH-REDUCTION]).

10. Failure handling of inter-AS TE LSP

In the context of MPLS Inter-AS Traffic Engineering, if a link/SRLG/Node failure occurs in an AS different from the HE LSR, the HE LSR exclusively relies on the receipt of an RSVP Path Error message to get informed that the TE LSP has suffered a failure in a downstream AS (a "Notify" Path Error "Notify" message if the inter-AS TE has been locally repaired via MPLS TE Fast Reroute. For those reasons, the Path Error message SHOULD be sent in reliable mode ([REFRESH-REDUCTION]). Note that this requires to configure the reliable messaging mechanisms proposed in [REFRESH-REDUCTION] between every pair of LSRs in the network (more precisely between every PLR and any potential HE LSRs).

Upon receiving an RSVP Path Error message, a HE LSR must perform a TE reroute (new route computation) in a make before break fashion.

It is worth highlighting that the set up of inter-AS TE LSP might be significantly slower than in the case of intra-area TE LSP:

- In scenario 1, the process involves several ASBRs performing policy control, partial route computation (ERO expansions), ... In case of set up failure, the number of trials can be significant, which even more increases the set up time.

Furthermore, in case of dynamic loose hop computation, both the IGP and BGP reachability solutions have drawbacks in term of convergence upon failure. This is due to the slow convergence property of BGP. With BGP redistribution within ASes, the convergence might be even slower especially when BGP Route Reflector are in use with no multi-paths load balancing.

- In scenario 2, some signaling exchange between potentially quite a few PCC and PCS are performed prior to setting up the TE LSP. Note that in scenario 2, the probability of TE LSP set up failure is limited to some lack of synchronization of the TE databases and as such is significantly lower than in the case of scenario 1.

Moreover, in case of a large amount of inter-AS TE LSP set up, some non negligible extra signaling and routing computation load will be required on the loose hops (scenario 1) and loose hops/PCS (scenario 2). Some implementation may implement some pacing of inter-AS TE LSP set up rate.

Typically a link/node/SRLG failure may impact a large number of TE LSPs. Relying on a local repair mechanism like MPLS TE Fast Reroute allows to relax the load on ASBR/PCS and reduces the need for urgent inter-AS TE LSP reroute. This is the recommended approach.

11. Reoptimization of Inter-AS TE LSP

After an inter-AS TE LSP has been set up, a more optimal route might appear
in the various traversed ASes. Then in this case, it is desirable to get the
ability to reroute an inter-AS TE LSP in a non-disruptive fashion (making use
of the so called Make Before Break procedure) to follow this more optimal
path. This is a known as a TE LSP reoptimization procedure.

[LOOSE-REOPT] proposes a mechanisms allowing:

- The 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 potentially more
  optimal path. This is done, via explicit signaling request: the HE LSR sets
  the "ERO Expansion request" bit of the SESSION-ATTRIBUTE object carried in
  the RSVP Path message.

- An LSR whose next hop is a loose-hop to signal to the Head-end LSR that a
  better path exists. This is performed by sending an RSVP Path Error Notify
  message (ERROR-CODE = 25):
.
  Two new sub-codes are defined:

     4 - Better path exists: sent by any LSR upon detecting that a better path
         exists.
     5 - No better path: sent by any LSR polled by the HE LSR to indicate that
         no better path exists.

This indication may be sent either:

  - In response to a query sent by the HE LSR,

  - Spontaneously by any LSR having detected a more optimal path

The reoptimization event can be either be:

  - Timer-based: a timer expires,

  - Event-driven: a link UP event for instance.

Note that the reoptimization MUST always be performed in a non disruptive
fashion.

Once the HE LSR is informed of the existence of a more optimal path either in
its head-end area/AS or in another AS, the inter-AS TE Path computation is
triggered using the same set of mechanisms as when the TE LSP is first set up
(per-AS path computation as in scenario 1 or Path Computation Server in

scenario 2). Then the inter-AS TE LSP is set up following the more optimal path, making use of the make before break procedure.

12. Inter-AS MPLS TE Management

[LSPPING] proposes a solution which can be adopted for inter-AS TE LSPs whereby a HE LSR sends MPLS echo requests over the LSP being tested.á When the destination LSR receives the message, it needs to acknowledge the source LSR by sending an LSP_ECHO object in a RSVP Resv message.

The TTL processing over inter-AS TE primary or local backup LSPs will be supported as per [MPLS-TTL].

13. Scalability and extensibility

All the features related to intra-area TE LSP are also applicable to inter-AS TE LSP, without any restriction.

Vasseur and Zhang   21

14. Policy Control at the AS boundaries

As stated in section 5.11 of [INTER-AS-TE-REQS], a set of configurable options may be made available upon which ingress control policies can be implemented governing or honoring inter-AS TE agreements made by two interconnect SPs.

During the path-computation process, the inter-AS RSVP path message sent to its downstream loose hop (ASBR also) in a different AS can be firstly passed through an inter-AS TE policy control process on the downstream ASBR prior to its ERO expansion.á The inter-AS RSVP path setup request will get rejected resulting in an path-error message will be sent to the HE LSP should it fail the control policy,áfor example, requesting bandwidth reservation more than agreed upon or wrong preemption priorities.

15. Confidentiality

As mentioned in [INTER-AS-REQS], since an inter-AS TE LSP may span multiple ASes belonging to different SPs, the solution should allow to preserve AS confidentiality, by hiding the set of hops followed by the inter-AS TE LSP within an AS.

In scenario 1, this requirement can be met via the proper RRO filtering at the AS boundaries (this applies to the RRO object carried in both the Path and Resv message). Note that, if MPLS TE Fast Reroute is required to protect inter-AS TE LSP against the failure of an ASBR, the RRO object carried in the Resv message of an inter-AS TE LSP must not be completely filtered, as mentioned in section 7. As least, the information (label, IPv4 or IPv6

subobject (node-id subobject)) pertaining to the ASBRÆs next hop must be preserved.

In scenario 2, the RSVP Path computation reply can be filtered to provide a partial ERO (an ERO containing loose hops). If the agreement between SPs at AS boundary is such that confidentiality must be guaranteed, then the PCC in AS x MUST send a Path computation request to the PCS in AS y, with the "Partial" flag of the REQUEST-ID object set. The PCS SHOULD control that this flag is appropriately set; if not, the PCS might decide either to provide a partial ERO or to drop the request.

Even when the returned ERO is partial, the PCS might provide the cost of the computed path. This can be done by including a PATH-COST object in the RSVP Path computation reply message. If the agreement between SPs at AS boundaries is such that path cost might be provided, then the PCC in AS x might send a Path computation request to the PCS in AS y, with the "cost" flag of the REQUEST-ID object set. The PCS controls that this flag is appropriately set; if set, the PCS should include a PATH-COST object in its RSVP Path Computation reply message. Note that this is required to compute end to end shortest path.

16. Security Considerations

When signaling an inter-AS TE, one may make use of the already defined security features related to RSVP (authentication). This may require some coordination between SPs to share the keys (see RFC 2747 and RFC 3097).

17. Intellectual Property Considerations

Cisco Systems may have intellectual property rights claimed in regard to some of the specification contained in this document.

18. Acknowledgments

We would like to acknowledge input and helpful comments from Carol Iturralde, Francois le Faucheur and Rob Goguen.

References

[RSVP] Braden, et al, " Resource ReSerVation Protocol (RSVP) - Version 1, Functional Specification", RFC 2205, September 1997.

[RSVP-TE] Awduche, et al, "Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

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

Extensions", [RFC2961](), April 2001.

[FAST-REROUTE] Ping Pan, et al, "Fast Reroute Extensions to RSVP-TE for LSP
Tunnels", [draft-ietf-mpls-rsvp-lsp-fastreroute-01.txt](), May 2003.

[BANDWIDTH-PROTECTION] Vasseur et all, "MPLS Traffic Engineering Fast
reroute: bypass tunnel path computation for bandwidth protection",
[draft-vasseur-mpls-backup-computation-01.txt](), October 2002,
Work in progress.

[OSPF-TE] Katz, D., Yeung, D., Kompella, K., "Traffic Engineering Extensions
to OSPF Version 2", [draft-katz-yeung-ospf-traffic-09.txt]()(work in progress).

[ISIS-TE] Li, T., Smit, H., "IS-IS extensions for Traffic Engineering",
[draft-ietf-isis-traffic-04.txt]() (work in progress)

[SECOND-METRIC] Le faucheur, "Use of IGP Metric as a second TE Metric",
Internet draft, [draft-lefaucheur-te-metric-igp-02.txt](). (work in progress)

[MULTIPLE-METRICS] Fedyk D., Ghanwani A., Ash J., Vedrenne A. "Multiple
Metrics for Traffic Engineering with IS-IS and OSPF", Internet draft,
[draft-fedyk-isis-ospf-te-metrics-01.txt]() (work in progress)

[PATH-COMP] Vasseur et al, "RSVP Path computation request and reply
messages", [draft-vasseur-mpls-computation-rsvp-03.txt](), November 2001.
(work in progress)

[OSPF-TE-CAP] Vasseur, Psenak. "OSPF TE TLV capabilities",
[draft-vasseur-mpls-ospf-te-cap-00.txt](), October 2002 (Work in Progress).

[IGPOSPF-CAP] Lindem et all "Extensions to OSPF for Advertising Optional
Router Capabilities", [draft-lindem-ospf-cap-01.txt](), work in progress.

[ISIS-CAP] Aggarwal et all, "Extensions to IS-IS for Advertising Optional
Router Capabilities", work in progress

[EXCLUDE-ROUTE] Lee et all, "Exclude Routes - Extension to RSVP-TE"
, [draft-ietf-ccamp-rsvp-te-exclude-route-00.txt]()>, work in progress.

[MULTI-AREA-TE] Kompella eat all,"Multi-area MPLS Traffic Engineering",
[draft-kompella-mpls-multiarea-te-053.txt](), June 20032.

[LSPPING] Kompella, K., Pan, P., Sheth, N., Cooper, D.,Swallow, G., Wadhwa,
S., Bonica, R., "Detecting Data Plane Liveliness in MPLS", Internet Draft
<[draft-ietf-mpls-lsp-ping-02.txt]()>, October 2002. (Work in Progress)

[MPLS-TTL], Agarwal, et al, "Time to Live (TTL) Processing in MPLS Networks",
RFC 3443 Updates RFC 3032) ", January 2003

[INTER-AS-TE-REQS] Zhang et al, "MPLS Inter-AS Traffic Engineering
requirements", draft-zhangietf-tewg-interas-mpls-te-req-001.txt
(work in progress).

[LOOSE-PATH-REOPT] Vasseur and Ikejiri, "Reoptimization of an explicit
loosely routed MPLS TE paths", draft-vasseur-mpls-loose-path-reopt-021.txt,
February June 2003, Work in Progress.

[NODE-ID] Vasseur, Ali and Sivabalan, "Definition of an RRO node-id
subobject", draft-vasseurietf-mpls-nodeid-subobject-010.txt, June 2003,
Work in progress.

[LSP-HIER] Kompella K., Rekhter Y., "LSP Hierarchy with Generalized
MPLS TE", draft-ietf-mpls-lsp-hierarchy-08.txt, March 2002.

[INTER-REGION] Ayyangar et Vasseur, "Inter-region MPLS Traffic Engineering",
draft-ayyangar-inter-region-te-00.txt, June 2003.

Authors' Address:

Jean Philippe Vasseur
Cisco Systems, Inc.
300 Apollo Drive
Chelmsford, MA 01824
300 Beaver Brook Road
Boxborough , MA - 01719
USA
Email: jpv@cisco.com

Raymond Zhang
Infonet Services Corporation
2160 E. Grand Ave.
El Segundo, CA 90025
USA
Email: raymond_zhang@infonet.com