Networking Working Group   JP. Vasseur, Ed.
Internet-Draft   Cisco Systems, Inc
Intended status: Standards Track   A. Ayyangar, Ed.
Expires: May 19, 2008   Juniper Networks, Inc
                                                           R. Zhang
                                                                 BT
                                                  November 16, 2007

A Per-domain path computation method for establishing Inter-domain Traffic Engineering (TE) Label Switched Paths (LSPs)
draft-ietf-ccamp-inter-domain-pd-path-comp-06

Status of this Memo

Copyright Notice

Abstract

   This document specifies a per-domain path computation technique for
   establishing inter-domain Traffic Engineering (TE) Multiprotocol
   Label Switching (MPLS) and Generalized MPLS (GMPLS) Label Switched
   Paths (LSPs). In this document a domain refers to a collection of

network elements within a common sphere of address management or path
computational responsibility such as IGP areas and Autonomous
Systems.

Per-domain computation applies where the full path of an inter-domain
TE LSP cannot be or is not determined at the ingress node of the TE
LSP, and is not signaled across domain boundaries. This is most
likely to arise owing to TE visibility limitations. The signaling
message indicates the destination and nodes up to the next domain
boundary. It may also indicate further domain boundaries or domain
identifiers. The path through each domain, possibly including the
choice of exit point from the domain, must be determined within the
domain.

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].

Table of Contents

1. Terminology

Terminology used in this document

AS: Autonomous System.

ABR: Area Border Router: routers used to connect two IGP areas (areas
in OSPF or levels in IS-IS).

ASBR: Autonomous System Border Router: routers used to connect
together ASes of a different or the same Service Provider via one or
more Inter-AS links.

Boundary LSR: a boundary LSR is either an ABR in the context of
inter-area TE or an ASBR in the context of inter-AS TE.

ERO: Explicit Route Object

IGP: Interior Gateway Protocol

Inter-AS TE LSP: A TE LSP that crosses an AS boundary.

Inter-area TE LSP: A TE LSP that crosses an IGP area.

LSR: Label Switching Router.

LSP: Label Switched Path.

TE LSP: Traffic Engineering Label Switched Path.

PCE: Path Computation Element: an entity (component, application or
network node) that is capable of computing a network path or route
based on a network graph and applying computational constraints.

TED: Traffic Engineering Database.

The notion of contiguous, stitched and nested TE LSPs is defined in
[RFC4726] and will not be repeated here.

2. Introduction

The requirements for inter-domain Traffic Engineering (inter-area and
inter-AS TE) have been developed by the Traffic Engineering Working
Group and have been stated in [RFC4105] and [RFC4216]. The framework
for inter-domain MPLS Traffic Engineering has been provided in
[RFC4726].

Some of the mechanisms used to establish and maintain inter-domain TE LSPs are specified in [I-D.ietf-ccamp-inter-domain-rsvp-te] and [I-D.ietf-ccamp-lsp-stitching].

This document exclusively focuses on the path computation aspects and defines a method for establishing inter-domain TE LSP where each node in charge of computing a section of an inter-domain TE LSP path is always along the path of such TE LSP.

When the visibility of an end to end complete path spanning multiple domains is not available at the Head-end LSR (LSR that inititiated the TE LSP), one approach described in this document consists of using a per-domain path computation technique during LSP setup to determine the inter-domain TE LSP as it traverses multiple domains.

The mechanisms proposed in this document are also applicable to MPLS TE domains other than IGP areas and ASs.

The solution described in this document does not attempt to address all the requirements specified in [RFC4105] and [RFC4216]. This is acceptable according to [RFC4216] which indicates that a solution may be developed to address a particular deployment scenario and might, therefore, not meet all requirements for other deployment scenarios.

It must be pointed out that the inter-domain path computation technique proposed in this document is one among many others and the choice of the appropriate technique must be driven by the set of requirements for the paths attributes and the applicability to a particular technique with respect to the deployment scenario. For example, if the requirement is to get an end-to-end constraint-based shortest path across multiple domains, then a mechanism using one or more distributed PCEs could be used to compute the shortest path across different domains (see [RFC4655]). Other offline mechanisms for path computation are not precluded either. Note also that a Service Provider may elect to use different inter-domain path computation techniques for different TE LSP types.

3. General assumptions

3.1. Common assumptions

- Each domain in all the examples below is assumed to be capable of doing Traffic Engineering (i.e. running OSPF-TE or ISIS-TE and RSVP-TE). A domain may itself comprise multiple other domains. E.g. An AS may itself be composed of several other sub-AS(es) (BGP confederations) or areas/levels. In this case, the path computation technique described for inter-area and inter-AS MPLS Traffic

Engineering just recursively applies.

- The inter-domain TE LSPs are signaled using RSVP-TE ([RFC3209] and
[RFC3473]).

- The path (specified by an ERO (Explicit Route Object) in an RSVP-TE
Path message)) for an inter-domain TE LSP may be signaled as a set of
(loose and/or strict) hops.

- The hops may identify:

* The complete strict path end-to-end across different domains

* The complete strict path in the source domain followed by boundary
LSRs (or domain identifiers, e.g. AS numbers)

* The complete list of boundary LSRs along the path

* The current boundary LSR and the LSP destination.

The set of (loose or strict) hops can either be statically configured
on the Head-end LSR or dynamically computed. A per-domain path
computation method is defined in this document with an optional Auto-
discovery mechanism based on IGP, BGP, policy routing information
yielding the next-hop boundary node (domain exit point, such as Area
Border Router (ABR) or an Autonomous System Border Router (ASBR)
along the path as the TE LSP is being signaled, along with potential
crankback mechanisms. Alternatively the domain exit points may be
statically configured on the Head-end LSR in which case next-hop
boundary node auto-discovery would not be required.

- Boundary LSRs are assumed to be capable of performing local path
computation for expansion of a loose next-hop in the signaled ERO if
the path is not signaled by the Head-end LSR as a set of strict hops
or if the strict hop is an abstract node (e.g. an AS). In any case,
no topology or resource information needs to be distributed between
domains (as mandated per [RFC4105] and [RFC4216]), which is critical
to preserve IGP/BGP scalability and confidentiality in the case of TE
LSPs spanning multiple routing domains.

- The paths for the intra-domain Hierarchical LSPs (H-LSP) or
Stitched LSPs (S-LSP) or for a contiguous TE LSP within the domain
may be pre-configured or computed dynamically based on the arriving
inter-domain LSP setup request (depending on the requirements of the
transit domain). Note that this capability is explicitly specified
as a requirement in [RFC4216]. When the paths for the H-LSPs/S-LSP
are pre-configured, the constraints as well as other parameters like
local protection scheme for the intra-domain H-LSP/S-LSP are also

pre-configured.

- While certain constraints like bandwidth can be used across
different domains, certain other TE constraints like resource
affinity, color, metric, etc. as listed in [RFC2702] may need to be
translated at domain boundaries. If required, it is assumed that, at
the domain boundary LSRs, there will exist some sort of local mapping
based on policy agreement in order to translate such constraints
across domain boundaries. It is expected that such an assumption
particularly applies to inter-AS TE: for example, the local mapping
would be similar to the Inter-AS TE Agreement Enforcement Polices
stated in [RFC4216].

- The procedures defined in this document are applicable to any node
(not just boundary node) that receives a Path message with an ERO
that constains a loose hop or an abstract node that is not a simple
abstract node (that is, an abstract node that identifies more than
one LSR).

3.2. Example of topology for the inter-area TE case

The following example will be used for the inter-area TE case in this
document.

```
 <-area 1-><-- area 0 --><--- area 2 --->
  ------ABR1------------ABR3-------
 |   /   |              |  \     |
R0--X1    |              |   X2---X3--R1
 |        |              |  /     |
  ------ABR2-----------ABR4--------
<=========== Inter-area TE LSP =======>
```

   Figure 1 - Example of topology for the inter-area TE case

Description of Figure 1:

- ABR1, ABR2, ABR3 and ABR4 are ABRs,
- X1: an LSR in area 1,
- X2, X3: LSRs in area 2,
- An inter-area TE LSP T0 originated at R0 in area 1 and terminating
  at R1 in area 2.

Notes:

- The terminology used in the example above corresponds to OSPF but
the path computation technique proposed in this document equally
applies to the case of an IS-IS multi-level network.

- Just a few routers in each area are depicted in the diagram above
for the sake of simplicity.

- The example depicted in Figure 1 shows the case where the Head-end
and Tail-end areas are connected by means of area 0. The case of an
inter-area TE LSP between two IGP areas that does not transit through
area 0 is not precluded.

3.3. Example of topology for the inter-AS TE case

We consider the following general case, built on a superset of the
various scenarios defined in [RFC4216]:


```
      <-- 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)===============>
```

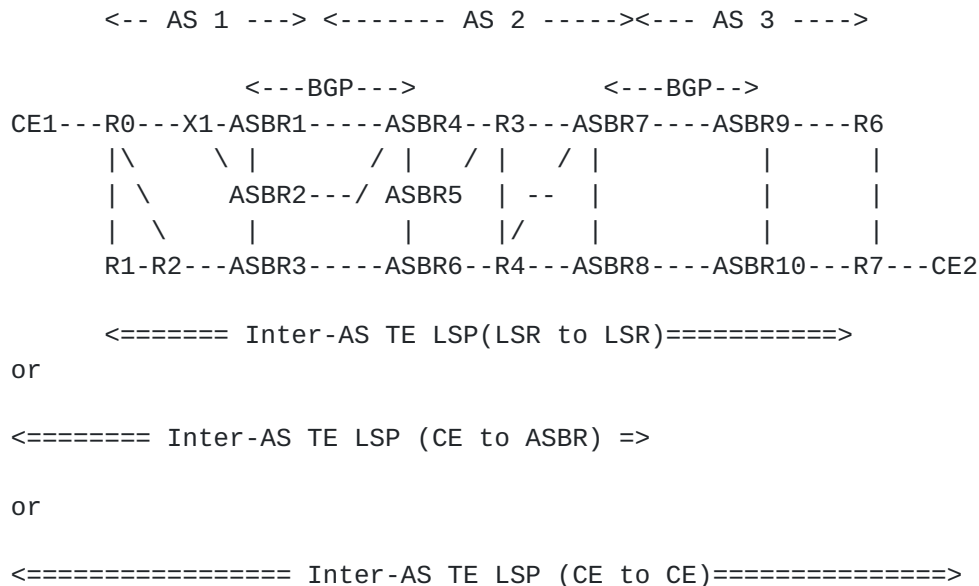   Figure 2 - Example of topology for the inter-AS TE case

The diagram depicted in Figure 2 covers all the inter-AS TE
deployment cases described in [RFC4216].

Description of Figure 2:

- Three interconnected ASs, respectively AS1, AS2, and AS3. Note
that in some scenarios described in [RFC4216] AS1=AS3.

- The ASBRs in different ASs are BGP peers. There is usually no IGP
running on the single hop links interconnecting the ASBRs and also
referred to as inter-ASBR links.

- Each AS runs an IGP (IS-IS or OSPF) with the required IGP TE
extensions (see [RFC3630], [RFC3784], [RFC4203] and [RFC4205]). In

other words, the ASes are TE enabled,

- Each AS can be made of several IGP areas. The path computation
technique described in this document applies to the case of a single
AS made of multiple IGP areas, multiples ASs made of a single IGP
areas or any combination of the above. For the sake of simplicity,
each routing domain will be considered as single area in this
document. The case of an Inter-AS TE LSP spanning multiple ASs where
some of those ASs are themselves made of multiple IGP areas can be
easily derived from the examples above: the per-domain path
computation technique described in this document is applied
recursively in this case.

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

4. Per-domain path computation procedures

The mechanisms for inter-domain TE LSP computation as described in
this document can be used regardless of the nature of the inter-
domain TE LSP (contiguous, stitched or nested).

Note that any path can be defined as a set of loose and strict hops.
In other words, in some cases, it might be desirable to rely on the
dynamic path computation in some domain, and exert a strict control
on the path in other domains (defining strict hops).

When an LSR that is a boundary node such as an ABR/ASBR receives a
Path message with an ERO that contains a strict node, the procedures
specified in [RFC3209] apply and no further action is needed.

When an LSR that is a boundary node such as an ABR/ASBR receives a
Path message with an ERO that contains a loose hop or an abstract
node that is not a simple abstract node (that is, an abstract node
that identifies more than one LSR), then it MUST follow the
procedures as described in [I-D.ietf-ccamp-inter-domain-rsvp-te].

In addition, the following procedures describe the path computation
procedures that SHOULD be carried out on the LSR:

1) If the next hop is not present in the TED, the two following
conditions MUST be checked:

o If the IP address of the next hop boundary LSR is outside of the
   current domain,

o If the next hop domain is PSC (Packet Switch Capable) and uses in-
   band control channel

If the two conditions above are satisfied then the boundary LSR
SHOULD check if the next-hop has IP reachability (via IGP or BGP).
If the next-hop is not reachable, then a signaling failure occurs and
the LSR SHOULD send back an RSVP PathErr message upstream with error
code=24 ("Routing Problem") and error subcode as described in section
4.3.4 of [RFC3209]. If the available routing information indicates
that next-hop is reachable, the selected route will be expected to
pass through a domain boundary via a domain boundary LSR. The
determination of domain boundary point based on routing information
is what we term as "auto-discovery" in this document. In the absence
of such an auto-discovery mechanism, a) the ABR in the case of inter-
area TE or the ASBR in the next-hop AS in the case of inter-AS TE
should be the signaled loose next-hop in the ERO and hence should be
accessible via the TED or b) there needs to be an alternate scheme
that provides the domain exit points. Otherwise the path computation
for the inter-domain TE LSP will fail.

An implementation MAY support the ability to disable such IP
reachability fall-back option should the next hop boundary LSR not be
present in the TED. In other words, an implementation MAY support
the possibility to trigger a signaling failure whenever the next-hop
is not present in the TED.

2) Once the next-hop boundary LSR has been determined (according to
the procedure described in 1)) or if the next-hop boundary is present
in the TED

o Case of a contiguous TE LSP. Unless not allowed by policy, the
   boundary LSR that processes the ERO SHOULD perform an ERO
   expansion (process consisting of computing the constrained path up
   to the next loose hop and add the list of hops as strcit nodes in
   the ERO). If no path satisfying the set of constraints can be
   found, then this is treated as a path computation and signaling
   failure and an RSVP PathErr message SHOULD be sent for the inter-
   domain TE LSP based on section 4.3.4 of [RFC3209].

o Case of stitched or nested LSP

   * If the boundary LSR is a candidate LSR for intra-area H-LSP/
      S-LSP setup (the boundary has local policy for nesting or
      stitching), the TE LSP is a candidate for hierarchy/nesting
      (the "Contiguous LSP" bit defined in
      [I-D.ietf-ccamp-inter-domain-rsvp-te] is not set) and if there
      is no H-LSP/S-LSP from this LSR to the next-hop boundary LSR
      that satisfies the constraints, it SHOULD signal a H-LSP/S-LSP

to the next-hop boundary LSR. If pre-configured H-LSP(s) or
S-LSP(s) already exist, then it will try to select from among
those intra-domain LSPs. Depending on local policy, it MAY
signal a new H-LSP/S-LSP if this selection fails. If the
H-LSP/S-LSP is successfully signaled or selected, it propagates
the inter-domain Path message to the next-hop following the
procedures described in [I-D.ietf-ccamp-inter-domain-rsvp-te].
If, for some reason the dynamic H-LSP/S-LSP setup to the next-
hop boundary LSR fails, then this SHOULD be treated as a path
computation and signaling failure and an RSVP PathErr message
SHOULD be sent upstream for the inter-domain LSP. Similarly,
if selection of a preconfigured H-LSP/S-LSP fails and local
policy prevents dynamic H-LSP/S this SHOULD be treated as a
path computation and signaling failure and an RSVP PathErr
SHOULD be sent upstream for the inter-domain TE LSP. In both
these cases procedures described in section 4.3.4 of [RFC3209]
SHOULD be followed to handle the failure.

* If, however, the boundary LSR is not a candidate for intra-
   domain H-LSP/S-LSP (the boundary LSR does not have local policy
   for nesting or stitching) or the TE LSP is a not candidate for
   hierarchy/nesting (the "Contiguous LSP" bit defined in
   [I-D.ietf-ccamp-inter-domain-rsvp-te] is set), then it SHOULD
   apply the same procedure as for the contiguous case.

The ERO of an inter-domain TE LSP may comprise abstract nodes such as
ASes. In such a case, upon receiving the ERO whose next hop is an
AS, the boundary LSR has to determine the next-hop boundary LSR which
may be determined based on the "auto-discovery" process mentioned
above. If multiple ASBRs candidates exist the boundary LSR may apply
some policies based on peering contracts that may have been pre-
negotiated. Once the next-hop boundary LSR has been determined a
similar procedure as the one described above is followed.

Note related to the inter-AS TE case:

In terms of computation of an inter-AS TE LSP path, an interesting
optimization technique consists of allowing the ASBRs to flood the TE
information related to the inter-ASBR link(s) although no IGP TE is
enabled over those links (and so there is no IGP adjacency over the
inter-ASBR links). This of course implies for the inter-ASBR links
to be TE-enabled although no IGP is running on those links.

```
     <-- 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
```
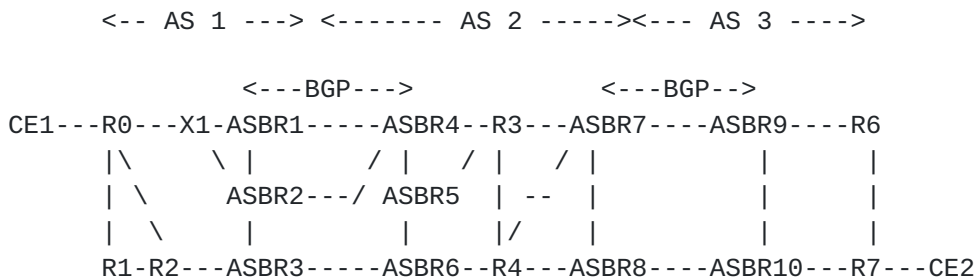
  Figure 3 - Flooding of the TE-related information for the Inter-ASBR links

   Referring to figure 3, ASBR1 for example would advertise in its OSPF
   LSA/IS-IS LSP the Traffic Engineering TLVs related to the link ASBR1-
   ASBR4.

   This allows an LSR (could be entry ASBR) in the previous AS to make a
   more appropriate route selection up to the entry ASBR in the
   immediately downstream AS taking into account the constraints
   associated with the inter-ASBR links. This reduces the risk of call
   set up failure due to inter-ASBR links not satisfying the inter-AS TE
   LSP set of constraints. Note that the TE information is only related
   to the inter-ASBR links: the TE LSA/LSP flooded by the ASBR includes
   not only the TE-enabled links contained in the AS but also the inter-
   ASBR links.

   Note that no summarized TE information is leaked between ASes which
   is compliant with the requirements listed in [RFC4105] and [RFC4216].

   For example, consider the diagram depicted in Figure 2: when ASBR1
   floods its IGP TE LSA ((opaque LSA for OSPF)/LSP (TLV 22 for IS-IS))
   in its routing domain, it reflects the reservation states and TE
   properties of the following links: X1-ASBR1, ASBR1-ASBR2 and ASBR1-
   ASBR4.

   Thanks to such an optimization, the inter-ASBRs TE link information
   corresponding to the links originated by the ASBR is made available
   in the TED of other LSRs in the same domain that the ASBR belongs to.
   Consequently, the path computation for an inter-AS TE LSP path can
   also take into account the inter-ASBR link(s). This will improve the
   chance of successful signaling along the next AS in case of resource
   shortage or unsatisfied constraints on inter-ASBR links and it
   potentially reduces one level of crankback. Note that no topology
   information is flooded and these links are not used in IGP SPF
   computations. Only the TE information for the outgoing links
   directly connected to the ASBR is advertised.

   Note that an Operator may decide to operate a stitched segment or
   1-hop hierarchical LSP for the inter-ASBR link.

4.1. Example with an inter-area TE LSP

   The following example uses figure 1 as a reference.

4.1.1. Case 1: T0 is a contiguous TE LSP

   The Head-end LSR (R0) first determines the next hop ABR (which could
   be manually configured by the user or dynamically determined by using
   auto-discovery mechanism). R0 then computes the path to reach the
   selected next hop ABR (ABR1) and signals the Path message. When the
   Path message reaches ABR1, it first determines the next hop ABR from
   its area 0 along the LSP path (say ABR3), either directly from the
   ERO (if for example the next hop ABR is specified as a loose hop in
   the ERO) or by using the auto-discovery mechanism specified above.

   - Example 1 (set of loose hops): R0-ABR1(loose)-ABR3(loose)-R1(loose)

   - Example 2 (mix of strict and loose hops): R0-X1-ABR1-ABR3(loose)-
   X2-X3-R1

   Note that a set of paths can be configured on the Head-end LSR,
   ordered by priority. Each priority path can be associated with a
   different set of constraints. It may be desirable to systematically
   have a last resort option with no constraint to ensure that the
   inter-area TE LSP could always be set up if at least a TE path exists
   between the inter-area TE LSP source and destination. In case of set
   up failure or when an RSVP PathErr is received indicating the TE LSP
   has suffered a failure, an implementation might support the
   possibility to retry a particular path option configurable amount of
   times (optionally with dynamic intervals between each trial) before
   trying a lower priority path option.

   Once it has computed the path up to the next hop ABR (ABR3), ABR1
   sends the Path message along the computed path. Upon receiving the
   Path message, ABR3 then repeats a similar procedure. If ABR3 cannot
   find a path obeying the set of constraints for the inter-area TE LSP,
   the signaling process stops and ABR3 sends a PathErr message to ABR1.
   Then ABR1 can in turn trigger a new path computation by selecting
   another egress boundary LSR (ABR4 in the example above) if crankback
   is allowed for this inter-area TE LSP (see
   [I-D.ietf-ccamp-crankback]). If crankback is not allowed for that
   inter-area TE LSP or if ABR1 has been configured not to perform
   crankback, then ABR1 MUST stop the signaling process and MUST forward
   a PathErr up to the Head-end LSR (R0) without trying to select
   another ABR.

4.1.2. Case 2: T0 is a stitched or nested TE LSP

   The Head-end LSR (R0) first determines the next hop ABR (which could
   be manually configured by the user or dynamically determined by using
   auto-discovery mechanism). R0 then computes the path to reach the
   selected next hop ABR and signals the Path message. When the Path
   message reaches ABR1, it first determines the next hop ABR from its
   area 0 along the LSP path (say ABR3), either directly from the ERO
   (if for example the next hop ABR is specified as a loose hop in the
   ERO) or by using an auto-discovery mechanism, specified above.

   ABR1 then checks if it has a H-LSP or S-LSP to ABR3 matching the
   constraints carried in the inter-area TE LSP Path message. If not,
   ABR1 computes the path for a H-LSP or S-LSP from ABR1 to ABR3
   satisfying the constraint and sets it up accordingly. Note that the
   H-LSP or S-LSP could have also been pre-configured.

   Once ABR1 has selected the H-LSP/S-LSP for the inter-area LSP, using
   the signaling procedures described in
   [I-D.ietf-ccamp-inter-domain-rsvp-te], ABR1 sends the Path message
   for inter-area TE LSP to ABR3. Note that irrespective of whether
   ABR1 does nesting or stitching, the Path message for the inter-area
   TE LSP is always forwarded to ABR3. ABR3 then repeats the exact same
   procedures. If ABR3 cannot find a path obeying the set of
   constraints for the inter-area TE LSP, ABR3 sends a PathErr message
   to ABR1. Then ABR1 can in turn either select another H-LSP/S-LSP to
   ABR3 if such an LSP exists or select another egress boundary LSR
   (ABR4 in the example above) if crankback is allowed for this inter-
   area TE LSP (see [I-D.ietf-ccamp-crankback]). If crankback is not
   allowed for that inter-area TE LSP or if ABR1 has been configured not
   to perform crankback, then ABR1 forwards the PathErr up to the inter-
   area Head-end LSR (R0) without trying to select another egress LSR.

4.2. Example with an inter-AS TE LSP

   The following example uses figure 2 as a reference.

   The path computation procedures for establishing an inter-AS TE LSP
   are very similar to those of an inter-area TE LSP described above.
   The main difference is related to the presence of inter-ASBRs
   link(s).

4.2.1. Case 1: T1 is a contiguous TE LSP

   The inter-AS TE path may be configured on the Head-end LSR as a set
   of strict hops, loose hops or a combination of both.

   - Example 1 (set of loose hops): ASBR4(loose)-ASBR9(loose)-R6(loose)

- Example 2 (mix of strict and loose hops): R2-ASBR3-ASBR2-ASBR1-
ASBR4-ASBR10(loose)-ASBR9-R6

In the example 1 above, a per-AS path computation is performed,
respectively on R0 for AS1, ASBR4 for AS2 and ASBR9 for AS3. 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 is
announced in the IGP TE LSA/LSP originated by its neighboring ASBR.
In the example 1 above, the TE LSP path is defined as: ASBR4(loose)-
ASBR9(loose)-R6(loose). This implies that R0 must compute the path
from R0 to ASBR4, hence the need for R0 to get the TE reservation
state related to the ASBR1-ASBR4 link (flooded in AS1 by ASBR1). In
addition, ASBR1 must also announce the IP address of ASBR4 specified
in the T1's path configuration.

Once it has computed the path up to the next hop ASBR, ASBR1 sends
the Path message for the inter-area TE LSP to ASBR4 (supposing that
ASBR4 is the selected next hop ASBR). ASBR4 then repeats the exact
same procedures. If ASBR4 cannot find a path obeying the set of
constraints for the inter-AS TE LSP, then ASBR4 sends a PathErr
message to ASBR1. Then ASBR1 can in turn either select another ASBR
(ASBR5 in the example above) if crankback is allowed for this
inter-AS TE LSP (see [I-D.ietf-ccamp-crankback]). If crankback is
not allowed for that inter-AS TE LSP or if ASBR1 has been configured
not to perform crankback, ABR1 stops the signaling process and
forwards a PathErr up to the Head-end LSR (R0) without trying to
select another egress LSR. In this case, the Head-end LSR can in
turn select another sequence of loose hops, if configured.
Alternatively, the Head-end 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
Head-end LSR to retry to same sequence of loose hops after having
relaxed some constraint(s).

4.2.2. Case 2: T1 is a stitched or nested TE LSP

The path computation procedures are very similar to the inter-area
LSP setup case described earlier. In this case, the H-LSPs or S-LSPs
are originated by the ASBRs at the entry to the AS.

5. Path optimality/diversity

Since the inter-domain TE LSP is computed on a per domain (area, AS)
basis, one cannot guarantee that the optimal inter-domain path can be
found.

Moreover, computing two diverse paths using a per-domain path
computation approach may not be possible in some topologies (due to
the well-known 'trapping' problem).

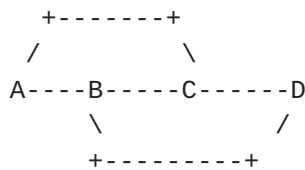For example, consider the following simple topology:


```
     +-------+
    /         \
   A----B-----C------D
        \           /
         +---------+
```

 Figure 4 - Example of "trapping" problem
In the simple topology depicted in figure 3, with a serialized
approach using the per-domain path computation technique specified in
this document, a first TE LSP may be computed following the path
A-B-C-D, in which case no diverse path could be found although two
diverse paths actually exist: A-C-D and A-B-D. The aim of that
simple example that can easily be extended to the inter-domain case
is to illustrate the potential issue of not being able to find
diverse paths using the per-domain path computation approach when
diverse paths exist.

As already pointed out, the required path computation method can be
selected by the Service Provider on a per LSP basis.

If the per-domain path computation technique does not meet the set of
requirements for a particular TE LSP (e.g. path optimality,
requirements for a set of diversely routed TE LSPs, ...) other
techniques such as PCE-based path computation techniques may be used
(see [RFC4655]).

6. Reoptimization of an inter-domain TE LSP

As stated in [RFC4216]and in [RFC4105], the ability to reoptimize an
already established inter-domain TE LSP constitutes a requirement.
The reoptimization process significantly differs based upon the
nature of the TE LSP and the mechanism in use for the TE LSP
computation.

The following mechanisms can be used for reoptimization and are
dependent on the nature of the inter-domain TE LSP.

6.1. Contiguous TE LSPs

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

   [RFC4736] proposes a mechanism that allows the Head-end LSR to be
   notified of the existence of a more optimal path in a downstream
   domain. The Head-end LSR may then decide to gracefully reroute the
   TE LSP using the so-called Make-Before-Break procedure. In case of a
   contiguous LSP, the reoptimization process is strictly controlled by
   the Head-end LSR which triggers the make-before-break procedure as
   defined in [RFC3209], regardless of the location of the more optimal
   path.

6.2. Stitched or nested (non-contiguous) TE LSPs

   In the case of a stitched or nested inter-domain TE LSP, the
   reoptimization process is treated as a local matter to any domain.
   The main reason is that the inter-domain TE LSP is a different LSP
   (and therefore different RSVP session) from the intra-domain S-LSP or
   H-LSP in an area or an AS. Therefore, reoptimization in a domain is
   done by locally reoptimizing the intra-domain H-LSP or S-LSP. Since
   the inter-domain TE LSPs are transported using S-LSP or H-LSP across
   each domain, optimality of the inter-domain TE LSP in a domain is
   dependent on the optimality of the corresponding S-LSP or H-LSPs.
   If, after an inter-domain LSP is setup, a more optimal path is
   available within an domain, the corresponding S-LSP or H-LSP will be
   reoptimized using "Make-Before-Break" techniques discussed in
   [RFC3209]. Reoptimization of the H-LSP or S-LSP automatically
   reoptimizes the inter-domain TE LSPs that the H-LSP or the S-LSP
   transports. Reoptimization parameters like frequency of
   reoptimization, criteria for reoptimization like metric or bandwidth
   availability, etc can vary from one domain to another and can be
   configured as required, per intra-domain TE S-LSP or H-LSP if it is
   preconfigured or based on some global policy within the domain.

   Hence, in this scheme, since each domain takes care of reoptimizing
   its own S-LSPs or H-LSPs, and therefore the corresponding inter-
   domain TE LSPs, the Make-Before-Break can happen locally and is not
   triggered by the Head-end LSR for the inter-domain LSP. So, no
   additional RSVP signaling is required for LSP reoptimization and
   reoptimization is transparent to the Head-end LSR of the inter-domain
   TE LSP.

If, however, an operator desires to manually trigger reoptimization
at the Head-end LSR for the inter-domain TE LSP, then this solution
does not prevent that. A manual trigger for reoptimization at the
Head-end LSR SHOULD force a reoptimization thereby signaling a "new"
path for the same LSP (along the more optimal path) making use of the
Make-Before-Break procedure. In response to this new setup request,
the boundary LSR may either initiate new S-LSP setup, in case the
inter-domain TE LSP is being stitched to the intra-domain S-LSP or it
may select an existing or new H-LSP in case of nesting. When the LSP
setup along the current path is complete, the Head-end LSR should
switchover the traffic onto that path and the old path is eventually
torn down. Note that the Head-end LSR does not know a priori whether
a more optimal path exists. Such a manual trigger from the Head-end
LSR of the inter-domain TE LSP is, however, not considered to be a
frequent occurrence.

Procedures described in [RFC4736] MUST be used if the operator does
not desire local reoptimization of certain inter-domain LSPs. In
this case, any reoptimization event within the domain MUST be
reported to the Head-end node. This SHOULD be a configurable policy.

6.3. Path characteristics after reoptimization

Note that in the case of loose hop reoptimization of contiguous
inter-domain TE LSP or local reoptimization of stitched/nested S-LSP
where boundary LSRs are specified as loose hops, the TE LSP may
follow a preferable path within one or more domain(s) but would still
traverse the same set of boundary LSRs. In contrast, in the case of
PCE-based path computation techniques, because end to end optimal
path is computed, the reoptimization process may lead to following a
completely different inter-domain path (including a different set of
boundary LSRs).

7. IANA Considerations

This document makes no request for any IANA action.

8. Security Considerations

Signaling of inter-domain TE LSPs raises security issues (discussed
in section 7 of [I-D.ietf-ccamp-inter-domain-rsvp-te]).

[RFC4726] provides an overview of the requirements for security in an
MPLS-TE or GMPLS multi-domain environment. In particular, when
signaling an inter-domain RSVP-TE LSP, an operator may make use of
the security features already defined for RSVP-TE ([RFC3209]). This

may require some coordination between the domains to share the keys
(see [RFC2747] and [RFC3097]), and care is required to ensure that
the keys are changed sufficiently frequently. Note that this may
involve additional synchronization, should the domain border nodes be
protected with Fast Rerotue ([RFC4090], since the Merge Point (MP)
and Point of Local Repair (PLR) should also share the key. For an
inter-domain TE LSP, especially when it traverses different
administrative or trust domains, the following mechanisms SHOULD be
provided to an operator (also see [RFC4216]):

1) A way to enforce policies and filters at the domain borders to
process the incoming inter-domain TE LSP setup requests (Path
messages) based on certain agreed trust and service levels/contracts
between domains. Various LSP attributes such as bandwidth, priority,
etc. could be part of such a contract. 2) A way for the operator to
rate-limit LSP setup requests or error notifications from a
particular domain. 3) A mechanism to allow policy-based outbound RSVP
message processing at the domain border node, which may involve
filtering or modification of certain addresses in RSVP objects and
messages.

This document relates to inter-domain path computation. It must be
noted that the process for establishing paths described in this
document does not increase the information exchanged between ASes and
preserves topology confidentiality, in compliance with [RFC4105] and
[RFC4216]. That being said, the signaling of inter-domain TE LSP
according to the procedure defined in this document requires path
computation on boundary nodes that may be exposed to various attacks.
Thus it is RECOMMENDED to support policy decisions to reject the ERO
expansion of an inter-domain TE LSP if not allowed.

9. Acknowledgements

We would like to acknowledge input and helpful comments from Adrian
Farrel, Jean-Louis Le Roux, Dimitri Papadimitriou and Faisal Aslam.

10. References

10.1. Normative References

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

[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.

10.2. Informative References

[I-D.ietf-ccamp-crankback]
            Farrel, A., "Crankback Signaling Extensions for MPLS and
            GMPLS RSVP-TE", draft-ietf-ccamp-crankback-06 (work in
            progress), January 2007.

[I-D.ietf-ccamp-inter-domain-rsvp-te]
            Ayyangar, A., "Inter domain Multiprotocol Label Switching
            (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering -
            RSVP-TE extensions",
            draft-ietf-ccamp-inter-domain-rsvp-te-07 (work in
            progress), September 2007.

[I-D.ietf-ccamp-lsp-stitching]
            Ayyangar, A., "Label Switched Path Stitching with
            Generalized Multiprotocol Label Switching Traffic
            Engineering (GMPLS TE)", draft-ietf-ccamp-lsp-stitching-06
            (work in progress), April 2007.

[RFC2702]   Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J.
            McManus, "Requirements for Traffic Engineering Over MPLS",
            RFC 2702, September 1999.

[RFC2747]   Baker, F., Lindell, B., and M. Talwar, "RSVP Cryptographic
            Authentication", RFC 2747, January 2000.

[RFC3097]   Braden, R. and L. Zhang, "RSVP Cryptographic
            Authentication -- Updated Message Type Value", RFC 3097,
            April 2001.

[RFC3630]   Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering
            (TE) Extensions to OSPF Version 2", RFC 3630,
            September 2003.

[RFC3784]   Smit, H. and T. Li, "Intermediate System to Intermediate
            System (IS-IS) Extensions for Traffic Engineering (TE)",
            RFC 3784, June 2004.

[RFC4090]   Pan, P., Swallow, G., and A. Atlas, "Fast Reroute
            Extensions to RSVP-TE for LSP Tunnels", RFC 4090,
            May 2005.

[RFC4105]   Le Roux, J., Vasseur, J., and J. Boyle, "Requirements for

Inter-Area MPLS Traffic Engineering", RFC 4105, June 2005.

[RFC4203]   Kompella, K. and Y. Rekhter, "OSPF Extensions in Support
of Generalized Multi-Protocol Label Switching (GMPLS)",
RFC 4203, October 2005.

[RFC4205]   Kompella, K. and Y. Rekhter, "Intermediate System to
Intermediate System (IS-IS) Extensions in Support of
Generalized Multi-Protocol Label Switching (GMPLS)",
RFC 4205, October 2005.

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

[RFC4655]   Farrel, A., Vasseur, J., and J. Ash, "A Path Computation
Element (PCE)-Based Architecture", RFC 4655, August 2006.

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

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

Authors' Addresses

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

Email: jpv@cisco.com

Arthi Ayyangar (editor)
Juniper Networks, Inc
1194 N.Mathilda Ave
Sunnyvale, CA 94089
USA

Email: arthi@juniper.net

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

Email: raymond.zhang@bt.com