Network Working Group
Request for Comments: 4216
Category: Informational

R. Zhang, Ed.
Infonet Services Corporation
J.-P. Vasseur, Ed.
Cisco Systems, Inc.

November 2005

MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements

Status of This Memo

Copyright Notice

Abstract

This document discusses requirements for the support of inter-AS MPLS Traffic Engineering (MPLS TE). Its main objective is to present a set of requirements and scenarios which would result in general guidelines for the definition, selection, and specification development for any technical solution(s) meeting these requirements and supporting the scenarios.

Table of Contents

1. Introduction

The MPLS Traffic Engineering (TE) mechanism documented in [TE-RSVP] may be deployed by Service Providers (SPs) to achieve some of the most important objectives of network traffic engineering as described in [TE-OVW]. These objectives are summarized as:

- Supporting end-to-end services requiring Quality of Service (QoS) guarantees
- Performing network resource optimization
- Providing fast recovery

However, this traffic engineering mechanism can only be used within an Autonomous System (AS).

This document discusses requirements for an inter-AS MPLS Traffic Engineering mechanism that may be used to achieve the same set of objectives across AS boundaries within or beyond an SP's administrative domains.

The document will also present a set of application scenarios where the inter-AS traffic engineering mechanism may be required. This mechanism could be implemented based upon the requirements presented in this document.

These application scenarios will also facilitate discussions for a detailed requirements list for this inter-AS Traffic Engineering mechanism.

Please note that there are other means of traffic engineering including Interior Gateway Protocol (IGP); metrics-based (for use within an AS); and Border Gateway Protocol (BGP) attribute-based (for use across ASes, as described in Appendix A), which provide coarser control of traffic paths. However, this document addresses requirements for a MPLS-based, fine-grained approach for inter-AS TE.

This document doesn't make any claims with respect to whether it is possible to have a practical solution that meets all the requirements listed in this document.

1.1. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC-2119].

2. Contributing Authors

The co-authors listed below contributed to the text and content of this document. (The contact information for the editors appears in section 9, and is not repeated below.)

Kenji Kumaki
KDDI Corporation
Garden Air Tower
Iidabashi, Chiyoda-ku,
Tokyo 102-8460, JAPAN
EMail : ke-kumaki@kddi.com

Paul Mabey
Qwest Communications
950 17th Street,
Denver, CO 80202, USA
EMail: pmabey@qwest.com

Nadim Constantine
Infonet Services Corporation
2160 E. Grand Ave.
El Segundo, CA 90025. USA
EMail: nadim_constantine@infonet.com

Pierre Merckx
EQUANT
1041 route des Dolines - BP 347
06906 SOPHIA ANTIPOLIS Cedex, FRANCE
EMail: pierre.merckx@equant.com

Ting Wo Chung
Bell Canada
181 Bay Street, Suite 350
Toronto, Ontario, Canada, M5J 2T3
EMail: ting_wo.chung@bell.ca

Jean-Louis Le Roux
France Telecom
2, avenue Pierre-Marzin
22307 Lannion Cedex, France
EMail: jeanlouis.leroux@francetelecom.com

Yonghwan Kim
SBC Laboratories, Inc.
4698 Willow Road
Pleasanton, CA 94588, USA
EMail: Yonghwan_Kim@labs.sbc.com

3. Definitions and Requirements Statement

3.1. Definitions

The following provides a list of abbreviations and acronyms specifically pertaining to this document:

SP: Service Providers including regional or global providers.

SP Administrative Domain: a single SP administration over a network or networks that may consist of one AS or multiple ASes.

IP-only networks: SP's network where IP routing protocols such as IGP/BGP are activated.

IP/MPLS networks: SP's network where MPLS switching capabilities and signaling controls (e.g., ones described in [MPLS-ARCH]) are activated in addition to IP routing protocols.

Intra-AS TE: A generic definition for traffic engineering mechanisms operating over IP-only and/or IP/MPLS network within an AS.

Inter-AS TE: A generic definition for traffic engineering mechanisms operating over IP-only and/or IP/MPLS network across one or multiple ASes. Since this document only addresses IP/MPLS networks, any reference to Inter-AS TE in this document refers only to IP/MPLS networks and is not intended to address IP-only TE requirements.

TE LSP: MPLS Traffic Engineering Label Switched Path.

Intra-AS MPLS TE: An MPLS Traffic Engineering mechanism where its TE Label Switched Path (LSP), Head-end Label Switching Router (LSR), and Tail-end LSR reside in the same AS for traffic engineering purposes.

Inter-AS MPLS TE: An MPLS Traffic Engineering mechanism where its TE LSPs, Head-end LSR, and Tail-end LSR do not reside within the same AS or both Head-end LSR and Tail-end LSR are in the same AS, but the TE LSP transiting path may be across different ASes.

ASBRs:	Autonomous System Border Routers used to connect to another AS of a different or the same Service Provider via one or more links that interconnect ASes.
Inter-AS TE Path:	A TE path traversing multiple ASes and ASBRs, e.g., AS1-ASBR1-inter-AS link(s)-ASBR2-AS2... ASBRn-ASn.
Inter-AS TE Segment:	A portion of the Inter-AS TE path.
Inter-AS DS-TE:	Diffserv-aware Inter-AS TE.
CE:	Customer Edge Equipment
PE:	Provider Edge Equipment that has direct connections to CEs.
P:	Provider Equipment that has backbone trunk connections only.
VRF:	Virtual Private Network (VPN) Routing and Forwarding Instance.
PoP:	Point of presence or a node in SP’s network.
SRLG:	A set of links may constitute a ’shared risk link group’ (SRLG) if they share a resource whose failure may affect all links in the set as defined in [GMPLS-ROUT].
PCC:	Path Computation Client; any client application requesting a path computation to be performed by the Path Computation Element.
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.

Please note that the terms of CE, PE, and P used throughout this document are generic in their definitions. In particular, whenever such acronyms are used, it does not necessarily mean that CE is connected to a PE in a VRF environment described in such IETF documents as [BGP-MPLSVPN].

3.2. Objectives and Requirements of Inter-AS Traffic Engineering

As mentioned in section 1 above, some SPs have requirements for achieving the same set of traffic engineering objectives as presented in [TE-OVW] across AS boundaries.

This section examines these requirements in each of the key corresponding areas: 1) Inter-AS bandwidth guarantees; 2) Inter-AS Resource Optimization and 3) Fast Recovery across ASes, i.e., Recovery of Inter-AS Links/SRLG and ASBR Nodes.

3.2.1. Inter-AS Bandwidth Guarantees

The Diffserv IETF working group has defined a set of mechanisms described in [DIFF_ARCH], [DIFF_AF], and [DIFF_EF] or [MPLS-Diff]. These mechanisms can be activated at the edge of or over a Diffserv domain to contribute to the enforcement of a QoS policy (or a set of QoS policies), which can be expressed in terms of maximum one-way transit delay, inter-packet delay variation, loss rate, etc.

Many SPs have partial or full deployment of Diffserv implementations in their networks today, either across the entire network or minimally on the edge of the network across CE-PE links.

In situations where strict QoS bounds are required, admission control inside the backbone of a network is in some cases required in addition to current Diffserv mechanisms.

When the propagation delay can be bounded, the performance targets, such as maximum one-way transit delay, may be guaranteed by providing bandwidth guarantees along the Diffserv-enabled path.

One typical example of this requirement is to provide bandwidth guarantees over an end-to-end path for VoIP traffic classified as EF (Expedited Forwarding [DIFF_EF]) class in a Diffserv-enabled network. When the EF path is extended across multiple ASes, inter-AS bandwidth guarantee is then required.

Another case for inter-AS bandwidth guarantee is the requirement for guaranteeing a certain amount of transit bandwidth across one or multiple ASes.

Several application scenarios are presented to further illustrate this requirement in section 4 below.

3.2.2. Inter-AS Resource Optimization

In Service Provider (SP) networks, the BGP protocol [BGP] is deployed to exchange routing information between ASes. The inter-AS capabilities of BGP may also be employed for traffic engineering purposes across the AS boundaries. Appendix A provides a brief description of the current BGP-based inter-AS traffic engineering practices.

SPs have managed to survive with this coarse set of BGP-based traffic engineering facilities across inter-AS links in a largely best-effort environment. Certainly, in many cases, ample bandwidth within an SP's network and across inter-AS links reduces the need for more elaborate inter-AS TE policies.

However, in the case where a SP network is deployed over multiple ASes (for example, as the number of inter-AS links grows), the complexity of the inter-AS policies and the difficulty in inter-AS TE path optimization increase to a level such that it may soon become unmanageable.

Another example is where inter-AS links are established between different SP administrative domains. Nondeterministic factors such as uncoordinated routing and network changes, as well as sub-optimum traffic conditions, would potentially lead to a complex set of inter-AS traffic engineering policies where current traffic engineering mechanisms would probably not scale well.

In these situations where resource optimization is required and/or specific routing requirements arise, the BGP-based inter-AS facilities will need to be complemented by a more granular inter-AS traffic engineering mechanism.

3.2.3. Fast Recovery across ASes

When extending services such as VoIP across ASes, customers often require SPs to maintain the same level of performance targets, such as packet loss and service availability, as achieved within an AS. As a consequence, fast convergence in a stable fashion upon link/SRLG/node failures becomes a strong requirement. This is clearly difficult to achieve with current inter-domain techniques, especially in cases of link/SRLG failures between ASBRs or ASBR node failures.

3.3. Inter-AS Traffic Engineering Requirements Statement

Just as in the applicable case of deploying MPLS TE in an SP's network, an inter-AS TE method in addition to BGP-based traffic engineering capabilities needs to be deployed across inter-AS links where resource optimization, bandwidth guarantees and fast recovery are required.

This is especially critical in a Diffserv-enabled, multi-class environment described in [PSTE] where statistical performance targets must be maintained consistently over the entire path across different ASes.

The approach of extending current intra-AS MPLS TE capabilities [TE-RSVP] across inter-AS links for IP/MPLS networks is considered here because of already available implementations and operational experiences.

Please note that the inter-AS traffic engineering over an IP-only network is for future consideration since there is not sufficient interest for similar requirements to those of IP/MPLS networks at this time. More specifically, this document only covers the inter-AS TE requirements for packet-based IP/MPLS networks.

4. Application Scenarios

The following sections present a few application scenarios over IP/MPLS networks where requirements cannot be addressed with the current intra-AS MPLS TE mechanism and give rise to considerations for inter-AS MPLS traffic engineering requirements.

Although not explicitly noted in the following discussions, fast recovery of traffic path(s) crossing multiple ASes in a stable fashion is particularly important in the case of link/SRLG/node failures at AS boundaries for all application scenarios presented here.

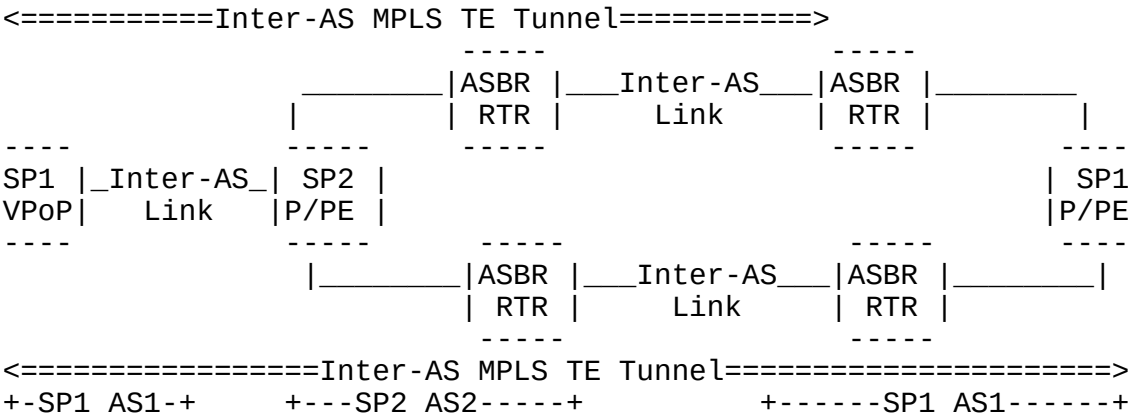
4.1. Application Scenarios Requiring Inter-AS Bandwidth Guarantees

4.1.1. Scenario I - Extended or Virtual PoP (VPoP)

A global service provider (SP1) would like to expand its reach into a region where a regional service provider's (SP2) network has already established a denser network presence.

In this scenario, the SP1 may establish interconnections with SP2 in one or multiple points in that region. In their customer-dense regions, SP1 may utilize SP2's network as an extended transport by co-locating aggregation routers in SP2's PoPs.

In order to ensure bandwidth capacity provided by SP2 and to achieve some degrees of transparency to SP2's network changes in terms of capacity and network conditions, one or more inter-AS MPLS TE LSPs can be built between SP1's ASBR or PE router inside AS1 and SP1's PE routers co-located in SP2's PoPs, as illustrated in the diagram below:



In situations where end-to-end Diffserv paths must be maintained, both SPs' networks may need to provision Diffserv PHB at each hop in order to support a set of traffic classes with compatible performance targets. The subsequent issues regarding Service Level Agreement (SLA) boundaries, reporting and measuring system interoperability and support demarcations are beyond the scope of this document and are not discussed further.

If either SP1's or SP2's network is not a Diffserv-aware network, the scenario would still apply to provide bandwidth guarantees.

The SP2, on the other hand, can similarly choose to expand its reach beyond its servicing region over SP1's network via inter-AS MPLS TE tunnels.

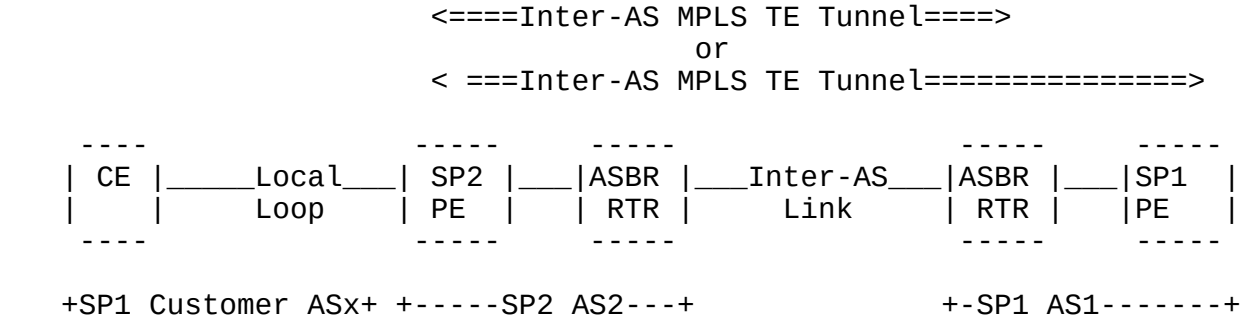
It is worth mentioning that these remote aggregation routers co-located in another SP's network are unlikely to host SP1's IGP and BGP routing planes and will more likely maintain their own AS or be part of the SP1's AS. In this case, such TE tunnels may cross several ASes, but the Head-end and Tail-end LSRs of TE tunnel may have the same AS number, as shown in the diagram above.

4.1.2. Scenario II - Extended or Virtual Trunk

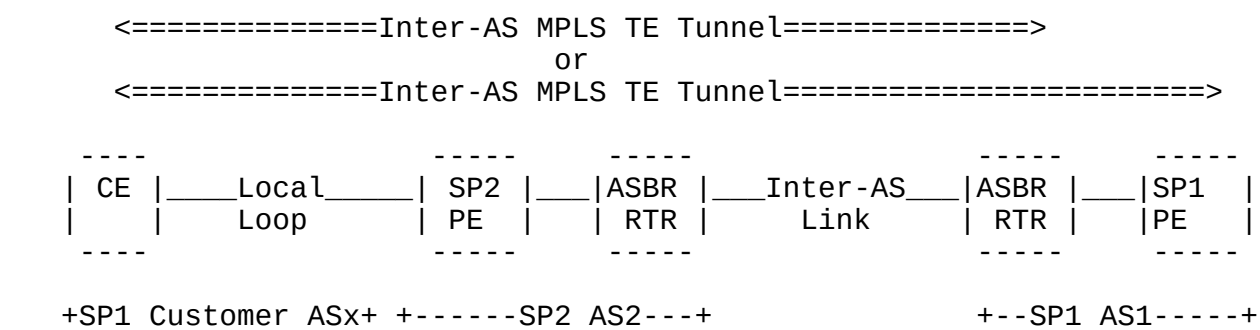
Instead of co-locating a PE router in SP2's PoP, SP1 may also choose to aggregate customer VPN sites onto a SP2's PE router where inter-AS TE tunnels can be built and signaled through SP2's MPLS network between the SP2 PoP (to which SP1 and customer CEs are directly connected) and SP1's ASBR or PE routers inside SP1's network. This allows SP1's customers connected to SP2 PE router to receive a guaranteed bandwidth service up to the TE LSP tail-end router located in SP1's network.

In this scenario, there could be two applicable cases:

Case 1 - the inter-AS MPLS TE tunnel functions as an extended or virtual trunk aggregating SP1's CE's local-loop access circuits on SP2's MPLS network over which the bandwidth can be guaranteed to the TE LSP tail-end router located in SP1's network, as shown in the diagram below:



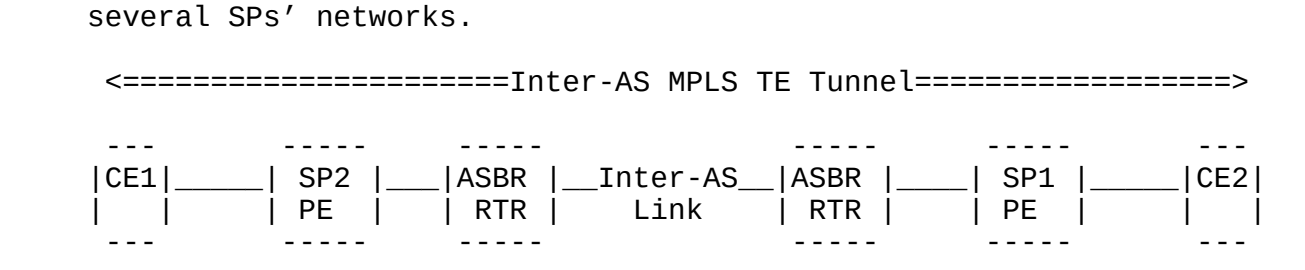
Case 2 - the inter-AS MPLS TE tunnel in this case functions as an extended or virtual local access link from SP1's CE on SP2's network to the SP1's ASBR or PE:



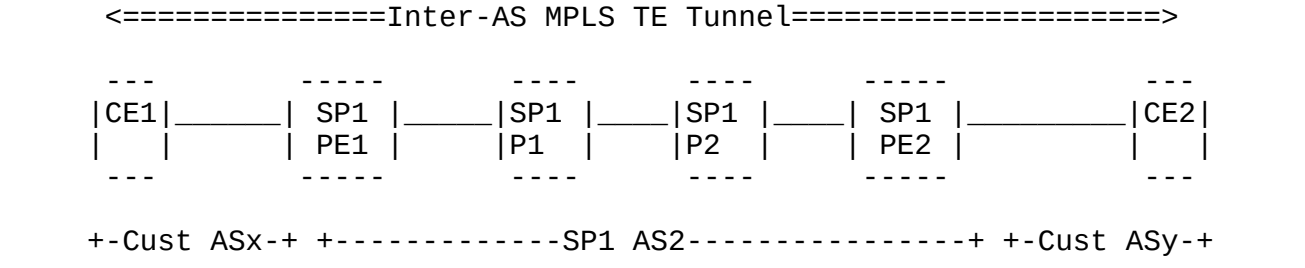
In Case 2 above, SP2 may elect to establish an aggregating or hierarchical intra-AS MPLS TE tunnel between the transiting P or PE router and SP2's ASBR router just to reduce the number of tunnel states signaled from the SP2 PE to where SP1's CEs are connected.

4.1.3. Scenario III - End-to-End Inter-AS MPLS TE from CE to CE

In this scenario as illustrated below, customers require the establishment of MPLS TE tunnel from CE1 to CE2 end-to-end across several SPs' networks.



The diagram below illustrates another example where CE1 and CE2 are customers of SP1 with external BGP (eBGP) peering relationships established across the CE-PE links. An inter-AS MPLS TE tunnel may then be established from CE1 in ASx to CE2, which may belong to the same AS or a different AS than that of CE1 across SP1's network in AS2.



The above example shows that SP1's network has a single AS. Obviously, there may be multiple ASes between CE1 and CE2, as well as in the SP1's network.

In addition, where both CE1 and CE2 reside in the same AS, they will likely share the same private AS number.

However, Scenario III will not scale well if there is a greater number of inter-AS TE MPLS tunnels in some degrees of partial mesh or full mesh. Therefore, it is expected that this scenario will have few deployments, unless some mechanisms such as hierarchical intra-AS TE-LSPs are used to reduce the number of signaling states.

4.2. Application Scenarios Requiring Inter-AS Resource Optimization

The scenarios presented in this section mainly deal with inter-AS resource optimization.

4.2.1. Scenario IV - TE across multi-AS within a Single SP Administrative Domain

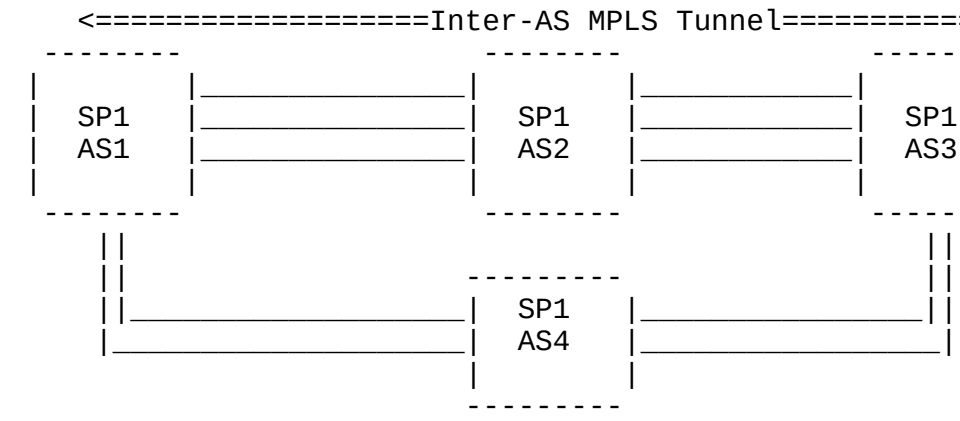
As mentioned in [TE-APP], SPs have generally admitted that the current MPLS TE mechanism provides a great deal of tactical and strategic value in areas of traffic path optimization [TE-RSVP] and rapid local repair capabilities [TE-FRR] via a set of on-line or off-line constraint-based path computation algorithms.

From a service provider's perspective, another way of stating the objectives of traffic engineering is to utilize available capacity in the network for delivering customer traffic without violating performance targets, and/or to provide better QoS services via an improved network utilization, more likely operating below congestion thresholds.

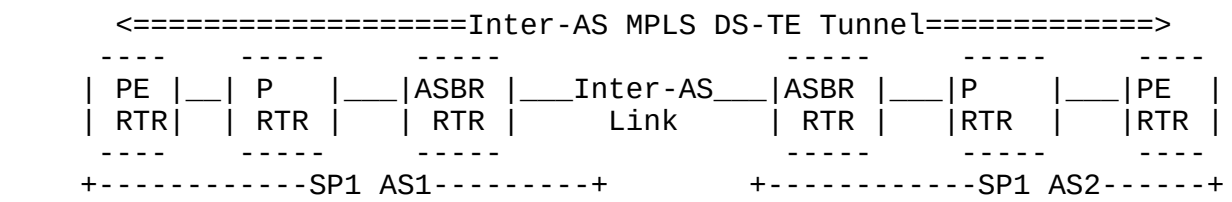
It is worth noting that situations where resource provisioning is not an issue (e.g., low density in inter-AS connectivity or ample inter-AS capacity), it may not require more scalable and granular TE facilities beyond BGP routing policies. This is because such policies can be rather simple and because inter-AS resource optimization is not an absolute requirement.

However many SPs, especially those with networks across multiple continents, as well as those with sparsely connected networks, have designed their multi-AS routing policies along or within the continental or sub-continental boundaries where the number of ASes can range from a very few to dozens. Generally, inter-continent or sub-continent capacity is very expensive. Some Service Providers have multiple ASes in the same country and would like to optimize resources over their inter-region links. This would demand a more scalable degree of resource optimization, which warrants the consideration of extending current intra-AS MPLS TE capabilities across inter-AS links.

In addition, one may only realize higher efficiency in conducting traffic optimization and path protection/restoration planning when coordinating all network resources as a whole, rather than partially. For a network which may consist of many ASes, this could be realized via the establishment of inter-AS TE LSPs, as shown in the diagram below:



The motivation for inter-AS MPLS TE is even more prominent in a Diffserv-enabled network over which statistical performance targets are to be maintained from any point to any point of the network as illustrated in the diagram below with an inter-AS DS-TE LSP:

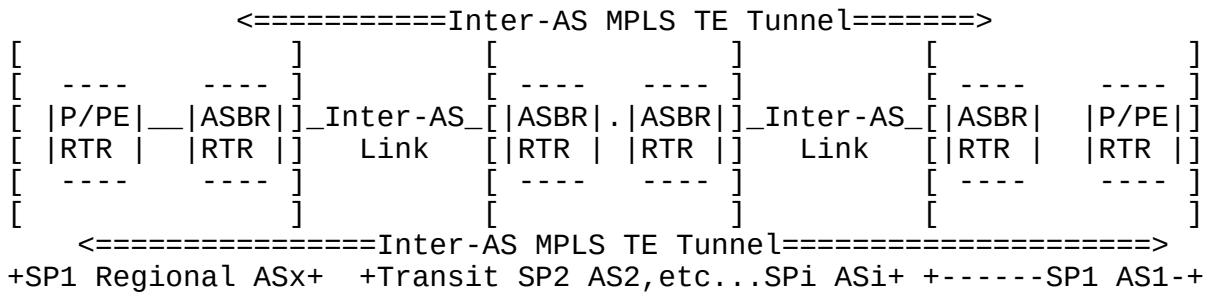


For example, the inter-AS MPLS DS-TE LSP shown in the diagram above could be used to transport a set of L2 Pseudo Wires or VoIP traffic with corresponding bandwidth requirement.

Furthermore, fast recovery in case of ASBR-ASBR link failure or ASBR node failure is a strong requirement for such services.

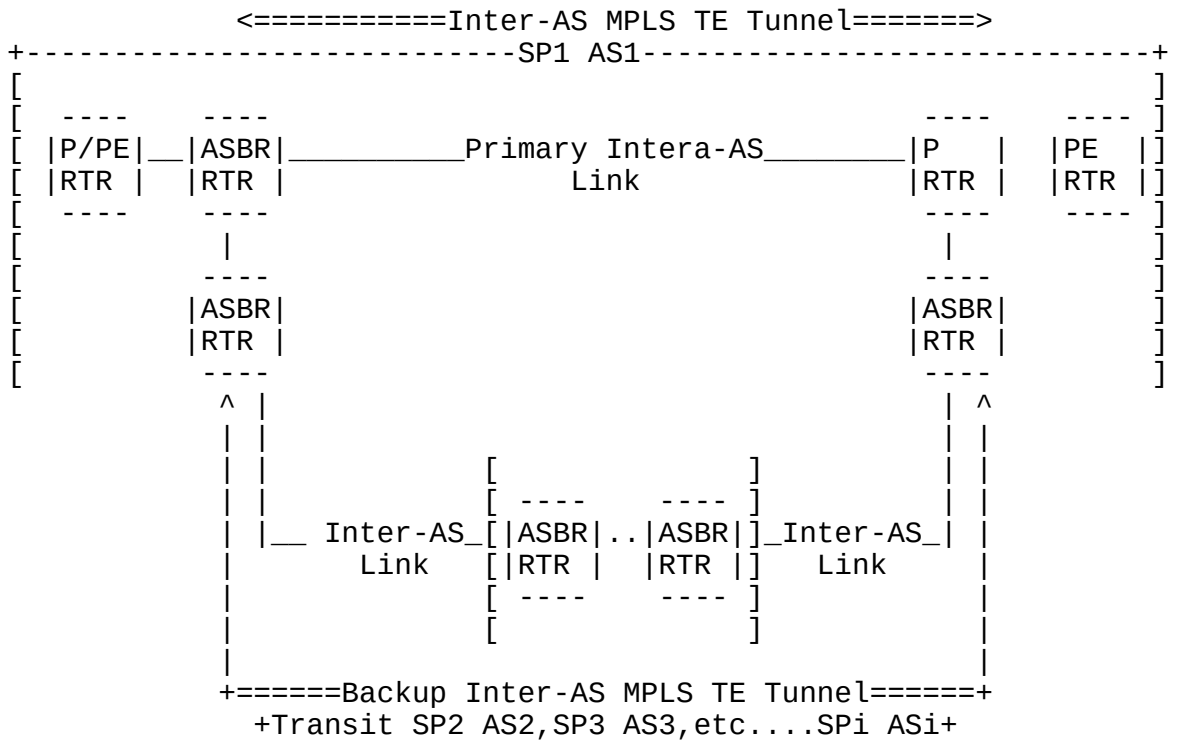
4.2.2. Scenario V - Transit ASes as Primary and Redundant Transport

Scenario V presents another possible deployment case. SP1 with AS1 wants to link a regional network to its core backbone by building an inter-AS MPLS TE tunnel over one or multiple transit ASes belonging to SP2, SP3, etc., as shown in the following diagram:



This scenario can be viewed as a broader case of Scenario I shown in section 4.1.1 where the "VPoP" could be expanded into a regional network of SP1. By the same token, the AS number for SP1's regional network ASx may be the same as or different from AS1.

The inter-AS MPLS TE LSP in this case may also be used to backup an internal path, as depicted in the diagram below, although this could introduce routing complexities:



5. Detailed Requirements for Inter-AS MPLS Traffic Engineering

This section discusses detailed requirements for inter-AS MPLS TE in two principal areas: 1) requirements for inter-AS MPLS TE in the same SP administrative domain and 2) requirements for inter-AS MPLS TE across different SP administrative domains.

5.1. Requirements within One SP Administrative Domain

This section presents detailed requirements for inter-AS MPLS TE within the same SP administrative domain.

5.1.1. Inter-AS MPLS TE Operations and Interoperability

The inter-AS MPLS TE solution SHOULD be consistent with requirements discussed in [TE-REQ] and the derived solution MUST be such that it will interoperate seamlessly with the current intra-AS MPLS TE mechanism and inherit its capability sets from [TE-RSVP].

The proposed solution SHOULD allow the provisioning of a TE LSP at the Head/Tail-end with end-to-end Resource Reservation Protocol (RSVP) signaling (eventually with loose paths) traversing across the interconnected ASBRs, without further provisioning required along the transit path.

5.1.2. Protocol Signaling and Path Computations

One can conceive that an inter-AS MPLS TE tunnel path signaled across inter-AS links consists of a sequence of ASes, ASBRs, and inter-AS links.

The proposed solution SHOULD provide the ability either to select explicitly or to auto-discover the following elements when signaling the inter-AS TE LSP path:

- a set of AS numbers as loose hops and/or
- a set of LSRs including ASBRs

It should also specify the above elements in the Explicit Route Object (ERO) and record them in the Record Route Object (RRO) of the Resv message just to keep track of the set of ASes or ASBRs traversed by the inter-AS TE LSP.

In the case of establishing inter-AS TE LSP traversing multiple ASes within the same SP networks, the solution SHOULD also allow the Head-end LSR to explicitly specify the hops across any one of the transiting ASes and the TE tunnel Head-end SHOULD also check the explicit segment to make sure that the constraints are met.

In addition, the proposed solution SHOULD provide the ability to specify and signal that certain loose or explicit nodes (e.g., AS numbers, etc.) and resources are to be explicitly excluded in the inter-AS TE LSP path establishment, such as one defined in [EXCLUDE-ROUTE].

5.1.3. Optimality

The solution SHOULD allow the set-up of an inter-AS TE LSP that complies with a set of TE constraints defined in [TE-REQ]) and follows an optimal path.

An optimal path is defined as a path whose end-to-end cost is minimal, based upon either an IGP or a TE metric. Note that in the case of an inter-AS path across several ASes having completely different IGP metric policies, the notion of minimal path might require IGP metric normalization.

The solution SHOULD provide mechanism(s) to compute and establish an optimal end-to-end path for the inter-AS TE LSP and SHOULD also allow for reduced optimality (or sub-optimality) since the path may not remain optimal for the lifetime of the LSP.

5.1.4. Support of Diversely Routed Inter-AS TE LSP

Setting up multiple inter-AS TE LSPs between a pair of LSRs might be desirable when:

(1) a single TE LSP satisfying the required set of constraints cannot be found, in which case it may require load sharing;

(2) multiple TE paths may be required to limit the impact of a network element failure to a portion of the traffic (as an example, two VoIP gateways may load balance the traffic among a set of inter-AS TE LSPs);

(3) path protection (e.g., 1:1 or 1:N) as discussed in [MPLS-Recov].

In the examples above, being able to set up diversely routed TE LSPs becomes a requirement for inter-AS TE.

The solution SHOULD be able to set up a set of link/SRLG/Node diversely routed inter-AS TE LSPs.

5.1.5. Re-Optimization

Once an inter-AS TE LSP has been established, and should there be any resource or other changes inside anyone of the ASes, the solution MUST be able to re-optimize the LSP accordingly and non-disruptively, either upon expiration of a configurable timer or upon being triggered by a network event or a manual request at the TE tunnel Head-End.

The solution SHOULD provide an option for the Head-End LSRs to control if re-optimizing or not should there exist a more optimal path in one of the ASes.

In the case of an identical set of traversed paths, the solution SHOULD provide an option for the Head-End LSRs to control whether re-optimization will occur because there could exist a more optimal path in one of the transit ASes along the inter-AS TE LSP path.

Furthermore, the solution MUST provide the ability to reject re-optimization at AS boundaries.

5.1.6. Fast Recovery Support Using MPLS TE Fast Reroute

There are, in general, two or more inter-AS links between multiple pairs of ASBRs for redundancy. The topological density between ASes in a SP network with multi-ASes is generally much higher. In the event of an inter-AS link failure, rapid local protection SHOULD also be made available and SHOULD interoperate with the current intra-AS MPLS TE fast re-route mechanism from [TE-FRR].

The traffic routed onto an inter-AS TE tunnel SHOULD also be fast protected against any node failure where the node could be internal to an AS or at the AS boundary.

5.1.7. DS-TE Support

The proposed inter-AS MPLS TE solution SHOULD satisfy core requirements documented in [DS-TE].

It is worth pointing out that the compatibility clause in section 4.1 of [DS-TE] SHOULD also be faithfully applied to the solution development.

5.1.8. Scalability and Hierarchical LSP Support

The proposed solution(s) MUST have a minimum impact on network scalability from both intra- and inter-AS perspectives.

This requirement applies to all of the following:

- IGP (impact in terms of IGP flooding, path computation, etc.)
- BGP (impact in terms of additional information carried within BGP, number of routes, flaps, overload events, etc.)
- RSVP TE (impact in terms of message rate, number of retained states, etc.)

It is also conceivable that there would potentially be scalability issues as the number of required inter-AS MPLS TE tunnels increases. In order to reduce the number of tunnel states to be maintained by each transiting PoP, the proposed solution SHOULD allow TE LSP aggregation such that individual tunnels can be carried onto one or more aggregating LSP(s). One such mechanism, for example, is described in [MPLS-LSPHIE].

5.1.9. Mapping of Traffic onto Inter-AS MPLS TE Tunnels

There SHOULD be several possibilities to map particular traffic to a particular destination onto a specific inter-AS TE LSP.

For example, static routing could be used if IP destination addresses are known. Another example is to utilize static routing using recursive BGP route resolution.

The proposed solution SHOULD also provide the ability to "announce" the inter-AS MPLS TE tunnels as a link into the IGPs (ISIS or OSPF) with the link’s cost associated with it. By doing so, PE routers that do not participate in the inter-AS TE path computation can take into account such links in its IGP-based SPF computation.

5.1.10. Inter-AS MPLS TE Management

5.1.10.1. Inter-AS MPLS TE MIB Requirements

An inter-AS TE Management Information Base (MIB) is required for use with network management protocols by SPs to manage and configure inter-AS traffic engineering tunnels. This new MIB SHOULD extend (and not reinvent) the existing MIBs to accommodate this new functionality.

An inter-AS TE MIB should have features that include:

- The setup of inter-AS TE tunnels with associated constraints (e.g., resources).
- The collection of traffic and performance statistics not only at the tunnel head-end, but any other points of the TE tunnel.
- The inclusion of both IPv4/v6 + AS# or AS# subobjects in the ERO in the path message, e.g.:

```
EXPLICIT_ROUTE class object:
address1 (loose IPv4 Prefix, /AS1)
address2 (loose IPv4 Prefix, /AS1)
AS2      (AS number)
address3 (loose IPv4 prefix, /AS3)
address4 (loose IPv4 prefix, /AS3) - destination

or

address1 (loose IPv4 Prefix, /AS1)
address2 (loose IPv4 Prefix, /AS1)
address3 (loose IPv4 Prefix, /AS2)
address4 (loose IPv4 Prefix, /AS2)
address5 (loose IPv4 prefix, /AS3)
address6 (loose IPv4 prefix, /AS3) - destination
```

- Similarly, the inclusion of the RRO object in the Resv message recording sub-objects such as interface IPv4/v6 address (if not hidden), AS number, a label, a node-id (when required), etc.
- Inter-AS specific attributes as discussed in section 5 of this document including, for example, inter-AS MPLS TE tunnel accounting records across each AS segment.

5.1.10.2. Inter-AS MPLS TE Fault Management Requirements

In a MPLS network, an SP wants to detect both control plane and data plane failures. But tools for fault detection over LSPs haven’t been widely developed so far. SPs today manually troubleshoot such failures in a hop-by-hop fashion across the data path. If they detect an error on the data plane, they have to check the control plane in order to determine where the faults come from.

The proposed solution SHOULD be able to interoperate with fault detection mechanisms of intra-AS TE and MAY or MAY NOT require the inter-AS TE tunnel ending addresses to be known or routable across IGP areas (OSPF) or levels (IS-IS) within the transiting ASes with working return paths.

For example, [LSPPING] is being considered as a failure detection mechanism over the data plane against the control plane and could be used to troubleshoot intra-AS TE LSPs. Such facilities, if adopted, SHOULD then be extended to inter-AS TE paths.

However, the above example depicts one such mechanism that does require a working return path such that diagnostic test packets can return via an alternate data plane, such as a global IPv4 path in the event that the LSP is broken.

[MPLS-TTL] presents how TTL may be processed across hierarchical MPLS networks, and such a facility as this SHOULD also be extended to inter-AS TE links.

5.1.11. Extensibility

The solution(s) MUST allow extensions as both inter-AS MPLS TE and current intra-AS MPLS TE specifications evolve.

5.1.12. Complexity and Risks

The proposed solution(s) SHOULD NOT introduce unnecessary complexity to the current operating network to such a degree that it would affect the stability and diminish the benefits of deploying such a solution over SP networks.

5.1.13. Backward Compatibility

The deployment of inter-AS MPLS TE SHOULD NOT impact existing BGP-based traffic engineering or MPLS TE mechanisms, but allow for a smooth migration or co-existence.

5.1.14. Performance

The solution SHOULD be evaluated taking into account various performance criteria:

- Degree of path optimality of the inter-AS TE LSP path
- TE LSP setup time
- Failure and restoration time
- Impact and scalability of the control plane due to added overheads, etc.
- Impact and scalability of the data/forwarding plane due to added overheads, etc.

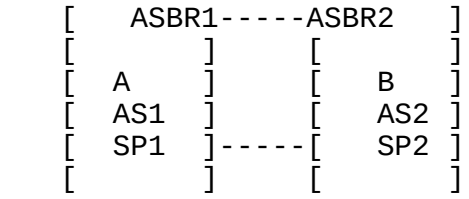
5.2. Requirements for Inter-AS MPLS TE across Multiple SP Administrative Domains

The requirements for inter-AS MPLS TE across multiple SP admin domains SHOULD include all requirements discussed in section 5.1 above in addition to those that are presented in this section here.

Please note that the SP with multi-AS networks may choose not to turn on the features discussed in the following two sections when building TE tunnels across ASes in its own domain.

5.2.1. Confidentiality

Since an inter-AS TE LSP may span multiple ASes belonging to different SPs, the solution MIGHT allow hiding the set of hops used by the TE LSP within an AS, as illustrated in the following example:



Suppose there is an inter-AS TE LSP from A (within AS1 of SP1) to B (within AS2 of SP2). When computing an inter-AS TE LSP path, the set of hops within AS2 might be hidden to AS1. In this case, the solution will allow A to learn that the more optimal TE LSP path to B (that complies with the set of constraints) traverses ASBR2, without a detailed knowledge of the lists of hops used within AS2.

Optionally, the TE LSP path cost within AS2 could be provided to A via, for example, PCC-PCE communication, such that A (PCC) could use this information to compute an optimal path, even if the computed path is not provided by AS2. (See [PCE-COM] for PCC-PCE communication and [PCE] for a description of the PCE-based path computation architecture.)

In addition, the management requirements discussed in section 5.1.10 above, when used across different SP admin domains, SHOULD include similar confidentiality requirements discussed here in terms of "hiding" intermediate hops or interface address and/or labels in the transiting or peering SPs.

5.2.2. Policy Control

In some cases, policy control might be necessary at the AS boundaries, namely ingress policy controls enabling SPs to enforce the inter-AS policies per interconnect agreements or to modify some requested parameters conveyed by incoming inter-AS MPLS TE signaling requests.

It is worth noting that such a policy control mechanism may also be used between ASes within a SP.

This section discusses only the elements that may be used to form a set of ingress control policies, but exactly how SPs establish bilateral or multilateral agreements upon which the control policies can be built is beyond the scope of this document.

5.2.2.1. Inter-AS TE Agreement Enforcement Polices

The following provides a set of TE-LSP parameters in the inter-AS TE Requests (RSVP Path Message) that could be enforced at the AS boundaries:

- RSVP-TE session attributes: affinities and preemption priorities
- Per AS or SP bandwidth admission control to ensure that RSVP-TE messages do not request for bandwidth resources over their allocation
- Request origins which can be represented by Head-End tunnel ending IP address, originating AS#, neighbor AS#, neighbor ASBR interface IP address, etc.
- DS-TE TE-Class <Class-Type, Preemption>
- FRR attribute: local protection desired bit, node protection desired bit, and bandwidth protection desired bit carried in the
- SESSION ATTRIBUTE or the FAST-REROUTE objects in the RSVP Path message as defined in [TE-FRR]
- Optimization allowed or not allowed

In some cases, a TE policy server could also be used for the enforcement of inter-AS TE policies. Implementations SHOULD allow the use of a policy enforcement server. This requirement could allow SPs to make the inter-AS TE policies scale better.

The signaling of a non-policy-compliant request SHOULD trigger the generation of a RSVP Path Error message by the policy enforcing node towards the Head-end LSR, indicating the cause. The Head-end LSR SHOULD take appropriate actions, such as re-route, upon receipt of such a message.

5.2.2.2. Inter-AS TE Rewrite Policies

In some situations, SPs may need to rewrite some attributes of the incoming inter-AS TE signaling requests due to a lack of resources for a particular TE-Class, non-compliant preemption, or mutual agreements. The following provides a non-exhaustive list of the parameters that can potentially be rewritten at the AS boundaries:

- RSVP-TE session attributes: affinities and preemption priorities
- DS-TE TE-Class <Class-Type, Preemption>
- ERO expansion requests

Similarly, the rewriting node SHOULD generate a RSVP Path Error Message towards the Head-end LSR indicating the cause in terms of types of changes made so as to maintain the end-to-end integrity of the inter-AS TE LSP.

5.2.2.3. Inter-AS Traffic Policing

The proposed solution SHOULD also provide a set of policing mechanisms which could be configured on the inter-AS links to ensure that traffic routed through the tunnel does not exceed the bandwidth negotiated during LSP signaling.

For example, an ingress policer could be configured to enforce the traffic contract on the mutually agreed resource requirements of the established inter-AS TE LSP (i.e., RSVP bandwidth) on the interface to which the inter-AS link is connected.

6. Security Considerations

The proposed solution(s) MUST address security issues across multiple SP administrative domains. Although inter-AS MPLS TE is not expected to add specific security extensions beyond those of current intra-AS TE, greater considerations MUST be given in terms of how to establish a trusted model across AS boundaries. SPs SHOULD have a means to authenticate (such as using RSVP INTEGRITY Object), to allow, and to possibly deny inter-AS signaling requests. Also, SPs SHOULD be protected from DoS attacks.

7. Acknowledgements

We would like to thank Yuichi Ikejiri, David Allan, Kurt Erik Lindqvist, Dave McDysan, Christian Jacquenet, Kireeti Kompella, Ed Kern, Jim Boyle, Thomas Nadeau, Yakov Rekhter, and Bert Wijnen for their suggestions and helpful comments during the discussions of this document.

8. Normative References

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

[TE-RSVP] 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.

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

9. Informative References

[MPLS-ARCH] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[BGP-MPLSVPN] Rosen, E. and Y. Rekhter, "BGP/MPLS IP VPNs", Work in Progress, October 2004.

[DIFF_ARCH] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Service", RFC 2475, December 1998.

[DIFF_AF] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[DIFF_EF] Davie, B., Charny, A., Bennet, J.C., Benson, K., Le Boudec, J., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, March 2002.

[MPLS-Diff] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, May 2002.

[TE-OVW] Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and X. Xiao, "Overview and Principles of Internet Traffic Engineering", RFC 3272, May 2002.

[PSTE] Li, T. and Y. Rekhter, "A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)", RFC 2430, October 1998.

[TE-APP] Boyle, J., Gill, V., Hannan, A., Cooper, D., Awduche, D., Christian, B., and W. Lai, "Applicability Statement for Traffic Engineering with MPLS", RFC 3346, August 2002.

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

[BGP] Rekhter, Y. and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March 1995.

[LSPPING] Kompella, K. and G. Swallow, "Detecting MPLS Data Plane Failures", Work in Progress, May 2005.

[MPLS-TTL] Agarwal, P. and B. Akyol, "Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks", RFC 3443, January 2003.

[DS-TE] Le Faucheur, F. and W. Lai, "Requirements for Support of Differentiated Services-aware MPLS Traffic Engineering", RFC 3564, July 2003.

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

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

[MPLS-Recov] Sharma, V. and F. Hellstrand, "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", RFC 3469, February 2003.

[EXCLUDE-ROUTE] Lee, CY., Farrel, A., and S. De Cnodder, "Exclude Routes - Extension to RSVP-TE", Work in Progress, August 2005.

[PCE] Farrel, A., Vasseur, J.-P., and J. Ash, "Path Computation Element (PCE) Architecture", Work in Progress, September 2005.

[PCE-COM] Vasseur, J.-P., et al., "Path Computation Element (PCE) communication Protocol (PCEP) - Version 1", Work in Progress, September 2005.

Appendix A. Brief Description of BGP-based Inter-AS Traffic Engineering

In today's Service Provider (SP) network, BGP is deployed to meet two different sets of requirements:

- Establishing a scalable exterior routing plane separate from the data forwarding plane within SP's administrative domain
- Exchanging network reachability information with different BGP autonomous systems (ASes) that could belong to a different SP or simply, a different AS within a SP network

Over connections across the AS boundaries, traffic engineering may also be accomplished via a set of BGP capabilities by appropriately enforcing BGP-based inter-AS routing policies. The current BGP-based inter-AS traffic engineering practices may be summarized as follows:

- "Closest exit" routing where egress traffic from one SP to another follows the path defined by the lowest IGP or intra-AS MPLS TE tunnel metrics of the BGP next-HOP of exterior routes learned from other ASes over the inter-AS links
- "BGP path attribute"-based routing selection mechanism where the egress traffic path is determined by interconnect (peering or transit) policies based upon one or a combination of BGP path attributes, like AS_PATH, MULTI_EXIT_DISC (MED), and Local_Pref.

SPs have often faced a number of nondeterministic factors in the practices of inter-AS traffic engineering employing the methods mentioned above:

- Sub-optimum traffic distribution across inter-AS links
- Nondeterministic traffic condition changes due to uncoordinated IGP routing policies or topology changes within other AS and uncoordinated BGP routing policy changes (MED or as-prepend, etc.)

In addition, to achieve some degrees of granularity, SPs may choose to enforce BGP inter-AS policies. These policies are specific to one inter-AS link or to a set of inter-AS links for ingress traffic. By tagging certain sets of routes with a specific attribute when announcing to another AS, the ingress traffic is destined to certain PoPs or to regions within SP's network from another AS. Of course, this operates on the assumption that the other AS permits automated egress policy by matching the predefined attribute from incoming routes.

Editors' Addresses

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

EMail: raymond_zhang@infonet.com

J.-P. Vasseur
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
USA

EMail: jpv@cisco.com

Full Copyright Statement

Intellectual Property

Acknowledgement

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