Network Working Group
Request for Comments: 4804
Category: Standards Track

F. Le Faucheur, Ed.
Cisco Systems, Inc.

February 2007

Aggregation of Resource ReSerVation Protocol (RSVP) Reservations over MPLS TE/DS-TE Tunnels

Status of This Memo

Copyright Notice

Abstract

RFC 3175 specifies aggregation of Resource ReSerVation Protocol (RSVP) end-to-end reservations over aggregate RSVP reservations. This document specifies aggregation of RSVP end-to-end reservations over MPLS Traffic Engineering (TE) tunnels or MPLS Diffserv-aware MPLS Traffic Engineering (DS-TE) tunnels. This approach is based on RFC 3175 and simply modifies the corresponding procedures for operations over MPLS TE tunnels instead of aggregate RSVP reservations. This approach can be used to achieve admission control of a very large number of flows in a scalable manner since the devices in the core of the network are unaware of the end-to-end RSVP reservations and are only aware of the MPLS TE tunnels.

Table of Contents

1. Introduction

The Integrated Services (Intserv) [INT-SERV] architecture provides a means for the delivery of end-to-end Quality of Service (QoS) to applications over heterogeneous networks.

[RSVP] defines the Resource reSerVation Protocol that can be used by applications to request resources from the network. The network responds by explicitly admitting or rejecting these RSVP requests. Certain applications that have quantifiable resource requirements express these requirements using Intserv parameters as defined in the appropriate Intserv service specifications ([GUARANTEED], [CONTROLLED]).

The Differentiated Services (DiffServ) architecture ([DIFFSERV]) was then developed to support the differentiated treatment of packets in very large scale environments. In contrast to the per-flow orientation of Intserv and RSVP, Diffserv networks classify packets into one of a small number of aggregated flows or "classes", based on the Diffserv codepoint (DSCP) in the packet IP header. At each Diffserv router, packets are subjected to a "per-hop behavior" (PHB), which is invoked by the DSCP. The primary benefit of Diffserv is its scalability. Diffserv eliminates the need for per-flow state and per-flow processing, and therefore scales well to large networks.

However, DiffServ does not include any mechanism for communication between applications and the network. Thus, as detailed in [INT-DIFF], significant benefits can be achieved by using Intserv over Diffserv including resource-based admission control, policy-based admission control, assistance in traffic identification/classification, and traffic conditioning. As discussed in [INT-DIFF], Intserv can operate over Diffserv in multiple ways. For example, the Diffserv region may be statically provisioned or RSVP aware. When it is RSVP aware, several mechanisms may be used to support dynamic provisioning and topology-aware admission control, including aggregate RSVP reservations, per-flow RSVP, or a bandwidth broker. The advantage of using aggregate RSVP reservations is that it offers dynamic, topology-aware admission control over the Diffserv region without per-flow reservations and the associated level of RSVP signaling in the Diffserv core. In turn, this allows dynamic, topology-aware admission control of flows requiring QoS reservations over the Diffserv core even when the total number of such flows carried over the Diffserv core is extremely large.

[RSVP-AGG] and [RSVP-GEN-AGG] describe in detail how to perform such aggregation of end-to-end RSVP reservations over aggregate RSVP reservations in a Diffserv cloud. They establish an architecture

where multiple end-to-end RSVP reservations sharing the same ingress router (Aggregator) and egress router (Deaggregator) at the edges of an "aggregation region" can be mapped onto a single aggregate reservation within the aggregation region. This considerably reduces the amount of reservation state that needs to be maintained by routers within the aggregation region. Furthermore, traffic belonging to aggregate reservations is classified in the data path purely using Diffserv marking.

[MPLS-TE] describes how MPLS Traffic Engineering (TE) tunnels can be used to carry arbitrary aggregates of traffic for the purposes of traffic engineering. [RSVP-TE] specifies how such MPLS TE tunnels can be established using RSVP-TE signaling. MPLS TE uses Constraint-Based Routing to compute the path for a TE tunnel. Then, Admission Control is performed during the establishment of TE tunnels to ensure they are granted their requested resources.

[DSTE-REQ] presents the Service Providers requirements for support of Diffserv-aware MPLS Traffic Engineering (DS-TE). With DS-TE, separate DS-TE tunnels can be used to carry different Diffserv classes of traffic, and different resource constraints can be enforced for these different classes. [DSTE-PROTO] specifies RSVP-TE signaling extensions as well as OSPF and Intermediate System to Intermediate System (IS-IS) extensions for support of DS-TE.

In the rest of this document we will refer to both TE tunnels and DS-TE tunnels simply as "TE tunnels".

TE tunnels have much in common with the aggregate RSVP reservations used in [RSVP-AGG] and [RSVP-GEN-AGG]:

- A TE tunnel is subject to Admission Control and thus is effectively an aggregate bandwidth reservation.

- In the data plane, packet scheduling relies exclusively on Diffserv classification and PHBs.

- Both TE tunnels and aggregate RSVP reservations are controlled by "intelligent" devices on the edge of the "aggregation core" (Head-end and Tail-end in the case of TE tunnels; Aggregator and Deaggregator in the case of aggregate RSVP reservations.

- Both TE tunnels and aggregate RSVP reservations are signaled using the RSVP protocol (with some extensions defined in [RSVP-TE] and [DSTE-PROTO] respectively for TE tunnels and DS-TE tunnels).

This document provides a detailed specification for performing aggregation of end-to-end RSVP reservations over MPLS TE tunnels (which act as aggregate reservations in the core). This document builds on the RSVP Aggregation procedures defined in [RSVP-AGG] and [RSVP-GEN-AGG], and only changes those where necessary to operate over TE tunnels. With [RSVP-AGG] and [RSVP-GEN-AGG], a lot of responsibilities (such as mapping end-to-end reservations to Aggregate reservations and resizing the Aggregate reservations) are assigned to the Deaggregator (which is the equivalent of the tunnel Tail-end) while with TE, the tunnels are controlled by the tunnel Head-end. Hence, the main change over the RSVP Aggregations procedures defined in [RSVP-AGG] and [RSVP-GEN-AGG] is to modify these procedures to reassign responsibilities from the Deaggregator to the Aggregator (i.e., the tunnel Head-end).

[LSP-HIER] defines how to aggregate MPLS TE Label Switched Paths (LSPs) by creating a hierarchy of such LSPs. This involves nesting of end-to-end LSPs into an aggregate LSP in the core (by using the label stack construct). Since end-to-end TE LSPs are themselves signaled with RSVP-TE and reserve resources at every hop, this can be looked at as a form of aggregation of RSVP(-TE) reservations over MPLS TE tunnels. This document capitalizes on the similarities between nesting of TE LSPs over TE tunnels and RSVP aggregation over TE tunnels, and reuses the procedures of [LSP-HIER] wherever possible.

This document also builds on the "RSVP over Tunnels" concepts of RFC 2746 [RSVP-TUN]. It differs from that specification in the following ways:

- This document describes operation over MPLS tunnels, whereas RFC 2746 describes operation with IP tunnels. One consequence of this difference is the need to deal with penultimate hop popping (PHP).

- MPLS-TE tunnels inherently reserve resources, whereas the tunnels in RFC 2746 do not have resource reservations by default. This leads to some simplifications in the current document.

- This document builds on the fact that there is exactly one aggregate reservation per MPLS-TE tunnel, whereas RFC 2746 permits a model where one reservation is established on the tunnel path for each end-to-end flow.

- We have assumed in the current document that a given MPLS-TE tunnel will carry reserved traffic and nothing but reserved traffic, which negates the requirement of RFC 2746 to distinguish reserved and non-reserved traffic traversing the same tunnel by using distinct encapsulations.

- There may be several MPLS-TE tunnels that share common Head-end and Tail-end routers, with the Head-end policy determining which tunnel is appropriate for a particular flow. This scenario does not appear to be addressed in RFC 2746.

At the same time, this document does have many similarities with RFC 2746. MPLS-TE tunnels are "type 2 tunnels" in the nomenclature of RFC 2746:

> "The (logical) link may be able to promise that some overall level of resources is available to carry traffic, but not to allocate resources specifically to individual data flows".

Aggregation of end-to-end RSVP reservations over TE tunnels combines the benefits of [RSVP-AGG] and [RSVP-GEN-AGG] with the benefits of MPLS, including the following:

- Applications can benefit from dynamic, topology-aware, resource-based admission control over any segment of the end-to-end path, including the core.

- As per regular RSVP behavior, RSVP does not impose any burden on routers where such admission control is not needed (for example, if the links upstream and downstream of the MPLS TE core are vastly over-engineered compared to the core capacity, admission control is not required on these over-engineered links and RSVP need not be processed on the corresponding router hops).

- The core scalability is not affected (relative to the traditional MPLS TE deployment model) since the core remains unaware of end-to-end RSVP reservations and only has to maintain aggregate TE tunnels since the datapath classification and scheduling in the core relies purely on the Diffserv mechanism (or more precisely the MPLS Diffserv mechanisms, as specified in [DIFF-MPLS]).

- The aggregate reservation (and thus the traffic from the corresponding end to end reservations) can be network engineered via the use of Constraint based routing (e.g., affinity, optimization on different metrics) and when needed can take advantage of resources on other paths than the shortest path.

- The aggregate reservations (and thus the traffic from the corresponding end-to-end reservations) can be protected against failure through the use of MPLS Fast Reroute.

This document, like [RSVP-AGG] and [RSVP-GEN-AGG], covers aggregation of unicast sessions. Aggregation of multicast sessions is for further study.

2. Specification of Requirements

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 [KEYWORDS].

3. Definitions

For readability, a number of definitions from [RSVP-AGG] as well as definitions for commonly used MPLS TE terms are provided here:

Aggregator — This is the process in (or associated with) the router at the ingress edge of the aggregation region (with respect to the end-to-end RSVP reservation) and behaving in accordance with [RSVP-AGG]. In this document, it is also the TE tunnel Head-end.

Deaggregator — This is the process in (or associated with) the router at the egress edge of the aggregation region (with respect to the end-to-end RSVP reservation) and behaving in accordance with [RSVP-AGG]. In this document, it is also the TE tunnel Tail-end

E2E — End to end

E2E Reservation — This is an RSVP reservation such that:

(i) corresponding Path messages are initiated upstream of the Aggregator and terminated downstream of the Deaggregator, and

(ii) corresponding Resv messages are initiated downstream of the Deaggregator and terminated upstream of the Aggregator, and

(iii) this RSVP reservation is aggregated over an MPLS TE tunnel between the Aggregator and Deaggregator.

An E2E RSVP reservation may be a per-flow reservation. Alternatively, the E2E reservation may itself be an aggregate reservation of various types (e.g., Aggregate IP reservation, Aggregate IPsec reservation). See Section 5 and 6 for more details on the types of E2E RSVP reservations. As per regular RSVP operations, E2E RSVP reservations are unidirectional.

Head-end This is the Label Switch Router responsible for establishing, maintaining, and tearing down a given TE tunnel.

Tail-end This is the Label Switch Router responsible for terminating a given TE tunnel.

Transit LSR This is a Label Switch Router that is on the path of a given TE tunnel and is neither the Head-end nor the Tail-end.

4. Operations of RSVP Aggregation over TE with Pre-established Tunnels

[RSVP-AGG] and [RSVP-GEN-AGG] support operations both in the case where aggregate RSVP reservations are pre-established and where Aggregators and Deaggregators have to dynamically discover each other and dynamically establish the necessary aggregate RSVP reservations.

Similarly, RSVP Aggregation over TE tunnels could operate both in the case where the TE tunnels are pre-established and where the tunnels need to be dynamically established.

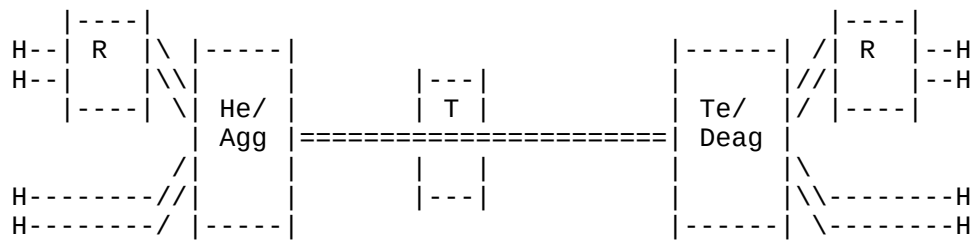
In this document we provide a detailed description of the procedures in the case where TE tunnels are already established. These procedures are based on those defined in [LSP-HIER]. The routing aspects discussed in Section 3 of [LSP-HIER] are not relevant here because those aim at allowing the constraint based routing of end-to-end TE LSPs to take into account the (aggregate) TE tunnels. In the present document, the end-to-end RSVP reservations to be aggregated over the TE tunnels rely on regular SPF routing. However, as already mentioned in [LSP-HIER], we note that a TE tunnel may be advertised into IS-IS or OSPF, to be used in normal SPF by nodes upstream of the Aggregator. This would affect SPF routing and thus routing of end-to-end RSVP reservations. The control of aggregation boundaries discussed in Section 6 of [LSP-HIER] is also not relevant here. This uses information exchanged in GMPLS protocols to dynamically discover the aggregation boundary. In this document, TE tunnels are pre-established, so that the aggregation boundary can be easily inferred. The signaling aspects discussed in Section 6.2 of

[LSP-HIER] apply to the establishment/termination of the aggregate TE tunnels when this is triggered by GMPLS mechanisms (e.g., as a result of an end-to-end TE LSP establishment request received at the aggregation boundary). As this document assumes pre-established tunnels, those aspects are not relevant here. The signaling aspects discussed in Section 6.1 of [LSP-HIER] relate to the establishment/maintenance of the end-to-end TE LSPs over the aggregate TE tunnel. This document describes how to use the same procedures as those specified in Section 6.1 of [LSP-HIER], but for the establishment of end-to-end RSVP reservations (instead of end-to-end TE LSPs) over the TE tunnels. This is covered further in Section 4 of the present document.

Pre-establishment of the TE tunnels may be triggered by any mechanisms including; for example, manual configuration or automatic establishment of a TE tunnel mesh through dynamic discovery of TE Mesh membership as allowed in [AUTOMESH].

Procedures in the case of dynamically established TE tunnels are for further studies.

4.1. Reference Model



H = Host requesting end-to-end RSVP reservations
R = RSVP router
He/Agg = TE tunnel Head-end/Aggregator
Te/Deag = TE tunnel Tail-end/Deaggregator
T = Transit LSR

-- = E2E RSVP reservation
== = TE tunnel

4.2. Receipt of E2E Path Message by the Aggregator

The first event is the arrival of the E2E Path message at the Aggregator. The Aggregator MUST follow traditional RSVP procedures for the processing of this E2E path message augmented with the extensions documented in this section.

The Aggregator MUST first attempt to map the E2E reservation onto a TE tunnel. This decision is made in accordance with routing information as well as any local policy information that may be available at the Aggregator. Examples of such policies appear in the following paragraphs. Just for illustration purposes, among many other criteria, such mapping policies might take into account the Intserv service type, the Application Identity [RSVP-APPID], and/or the signaled preemption [RSVP-PREEMP] of the E2E reservation (for example, the aggregator may take into account the E2E reservations RSVP preemption priority and the MPLS TE tunnel setup and/or hold priorities when mapping the E2E reservation onto an MPLS TE tunnel).

There are situations where the Aggregator is able to make a final mapping decision. That would be the case, for example, if there is a single TE tunnel toward the destination and if the policy is to map any E2E RSVP reservation onto TE tunnels.

There are situations where the Aggregator is not able to make a final determination. That would be the case, for example, if routing identifies two DS-TE tunnels toward the destination, one belonging to DS-TE Class-Type 1 and one to Class-Type 0, if the policy is to map Intserv Guaranteed Services reservations to a Class-Type 1 tunnel and Intserv Controlled Load reservations to a Class-Type 0 tunnel, and if the E2E RSVP Path message advertises both Guaranteed Service and Controlled Load.

Whether final or tentative, the Aggregator makes a mapping decision and selects a TE tunnel. Before forwarding the E2E Path message toward the receiver, the Aggregator SHOULD update the ADSPEC inside the E2E Path message to reflect the impact of the MPLS TE cloud onto the QoS achievable by the E2E flow. This update is a local matter and may be based on configured information, on the information available in the MPLS TE topology database, on the current TE tunnel path, on information collected via RSVP-TE signaling, or a combination thereof. Updating the ADSPEC allows receivers that take into account the information collected in the ADSPEC within the network (such as delay and bandwidth estimates) to make more informed reservation decisions.

The Aggregator MUST then forward the E2E Path message to the Deaggregator (which is the Tail-end of the selected TE tunnel). In accordance with [LSP-HIER], the Aggregator MUST send the E2E Path message with an IF_ID RSVP_HOP object instead of an RSVP_HOP object. The data interface identification MUST identify the TE tunnel.

To send the E2E Path message, the Aggregator MUST address it directly to the Deaggregator by setting the destination address in the IP Header of the E2E Path message to the Deaggregator address. The Router Alert is not set in the E2E Path message.

Optionally, the Aggregator MAY also encapsulate the E2E Path message in an IP tunnel or in the TE tunnel itself.

Regardless of the encapsulation method, the Router Alert is not set. Thus, the E2E Path message will not be visible to routers along the path from the Aggregator to the Deaggregator. Therefore, in contrast to the procedures of [RSVP-AGG] and [RSVP-GEN-AGG], the IP Protocol number does not need to be modified to "RSVP-E2E-IGNORE"; it MUST be left as is (indicating "RSVP") by the Aggregator.

In some environments, the Aggregator and Deaggregator MAY also act as IPsec Security Gateways in order to provide IPsec protection to E2E traffic when it transits between the Aggregator and the Deaggregator. In that case, to transmit the E2E Path message to the Deaggregator, the Aggregator MUST send the E2E Path message into the relevant IPsec tunnel terminating on the Deaggregator.

E2E PathTear and ResvConf messages MUST be forwarded by the Aggregator to the Deaggregator exactly like Path messages.

4.3. Handling of E2E Path Message by Transit LSRs

Since the E2E Path message is addressed directly to the Deaggregator and does not have Router Alert set, it is hidden from all transit LSRs.

4.4. Receipt of E2E Path Message by the Deaggregator

Upon receipt of the E2E Path message addressed to it, the Deaggregator will notice that the IP Protocol number is set to "RSVP" and will thus perform RSVP processing of the E2E Path message.

As with [LSP-HIER], the IP TTL vs. RSVP TTL check MUST NOT be made. The Deaggregator is informed that this check is not to be made because of the presence of the IF_ID RSVP HOP object.

The Deaggregator MAY support the option to perform the following checks (defined in [LSP-HIER]) by the receiver Y of the IF_ID RSVP_HOP object:

1. Make sure that the data interface identified in the IF_ID RSVP_HOP object actually terminates on Y.

2. Find the "other end" of the above data interface, i.e., X. Make sure that the PHOP in the IF_ID RSVP_HOP object is a control channel address that belongs to the same node as X.

The information necessary to perform these checks may not always be available to the Deaggregator. Hence, the Deaggregator MUST support operations in such environments where the checks cannot be made.

The Deaggregator MUST forward the E2E Path downstream toward the receiver. In doing so, the Deaggregator sets the destination address in the IP header of the E2E Path message to the IP address found in the destination address field of the Session object. The Deaggregator also sets the Router Alert.

An E2E PathErr sent by the Deaggregator in response to the E2E Path message (which contains an IF_ID RSVP_HOP object) SHOULD contain an IF_ID RSVP_HOP object.

4.5. Handling of E2E Resv Message by the Deaggregator

As per regular RSVP operations, after receipt of the E2E Path, the receiver generates an E2E Resv message which travels upstream hop-by-hop towards the sender.

Upon receipt of the E2E Resv, the Deaggregator MUST follow traditional RSVP procedures on receipt of the E2E Resv message. This includes performing admission control for the segment downstream of the Deaggregator and forwarding the E2E Resv message to the PHOP signaled earlier in the E2E Path message and which identifies the Aggregator. Since the E2E Resv message is directly addressed to the Aggregator and does not carry the Router Alert option (as per traditional RSVP Resv procedures), the E2E Resv message is hidden from the routers between the Deaggregator and the Aggregator which, therefore, handle the E2E Resv message as a regular IP packet.

If the Aggregator and Deaggregator are also acting as IPsec Security Gateways, the Deaggregator MUST send the E2E Resv message into the relevant IPsec tunnel terminating on the Aggregator.

4.6. Handling of E2E Resv Message by the Aggregator

The Aggregator is responsible for ensuring that there is sufficient bandwidth available and reserved over the appropriate TE tunnel to the Deaggregator for the E2E reservation.

On receipt of the E2E Resv message, the Aggregator MUST first perform the final mapping onto the final TE tunnel (if the previous mapping was only a tentative one).

If the tunnel did not change during the final mapping, the Aggregator continues the processing of the E2E Resv as described in the four following paragraphs.

The aggregator calculates the size of the resource request using traditional RSVP procedures. That is, it follows the procedures in [RSVP] to determine the resource requirements from the Sender Tspec and the Flowspec contained in the Resv. Then it compares the resource request with the available resources of the selected TE tunnel.

If sufficient bandwidth is available on the final TE tunnel, the Aggregator MUST update its internal understanding of how much of the TE tunnel is in use and MUST forward the E2E Resv messages to the corresponding PHOP.

As noted in [RSVP-AGG], a range of policies MAY be applied to the re-sizing of the aggregate reservation (in this case, the TE tunnel). For example, the policy may be that the reserved bandwidth of the tunnel can only be changed by configuration. More dynamic policies are also possible, whereby the aggregator may attempt to increase the reserved bandwidth of the tunnel in response to the amount of allocated bandwidth that has been used by E2E reservations. Furthermore, to avoid the delay associated with the increase of the tunnel size, the Aggregator may attempt to anticipate the increases in demand and adjust the TE tunnel size ahead of actual needs by E2E reservations. In order to reduce disruptions, the Aggregator SHOULD use "make-before-break" procedures as described in [RSVP-TE] to alter the TE tunnel bandwidth.

If sufficient bandwidth is not available on the final TE tunnel, the Aggregator MUST follow the normal RSVP procedure for a reservation being placed with insufficient bandwidth to support it. That is, the reservation is not installed and a ResvError is sent back toward the receiver.

If the tunnel did change during the final mapping, the Aggregator MUST first resend to the Deaggregator an E2E Path message with the IF_ID RSVP_HOP data interface identification identifying the final TE tunnel. If needed, the ADSPEC information in this E2E Path message SHOULD be updated. Then the Aggregator MUST

- either drop the E2E Resv message

- or proceed with the processing of the E2E Resv in the same manner as in the case where the tunnel did not change (described above).

In the former case, admission control over the final TE tunnel (and forwarding of E2E Resv message upstream toward the sender) would only occur when the Aggregator received the subsequent E2E Resv message (that will be sent by the Deaggregator in response to the resent E2E Path). In the latter case, admission control over the final tunnel is carried out immediately by the Aggregator, and if successful the E2E Resv message is generated upstream toward the sender.

Upon receipt of an E2E ResvConf from the Aggregator, the Deaggregator MUST forward the E2E ResvConf downstream toward the receiver. In doing so, the Deaggregator sets the destination address in the IP header of the E2E ResvConf message to the IP address found in the RESV_CONFIRM object of the corresponding Resv. The Deaggregator also sets the Router Alert.

4.7. Forwarding of E2E Traffic by the Aggregator

When the Aggregator receives a data packet belonging to an E2E reservations currently mapped over a given TE tunnel, the Aggregator MUST encapsulate the packet into that TE tunnel.

If the Aggregator and Deaggregator are also acting as IPsec Security Gateways, the Aggregator MUST also encapsulate the data packet into the relevant IPsec tunnel terminating on the Deaggregator before transmission into the MPLS TE tunnel.

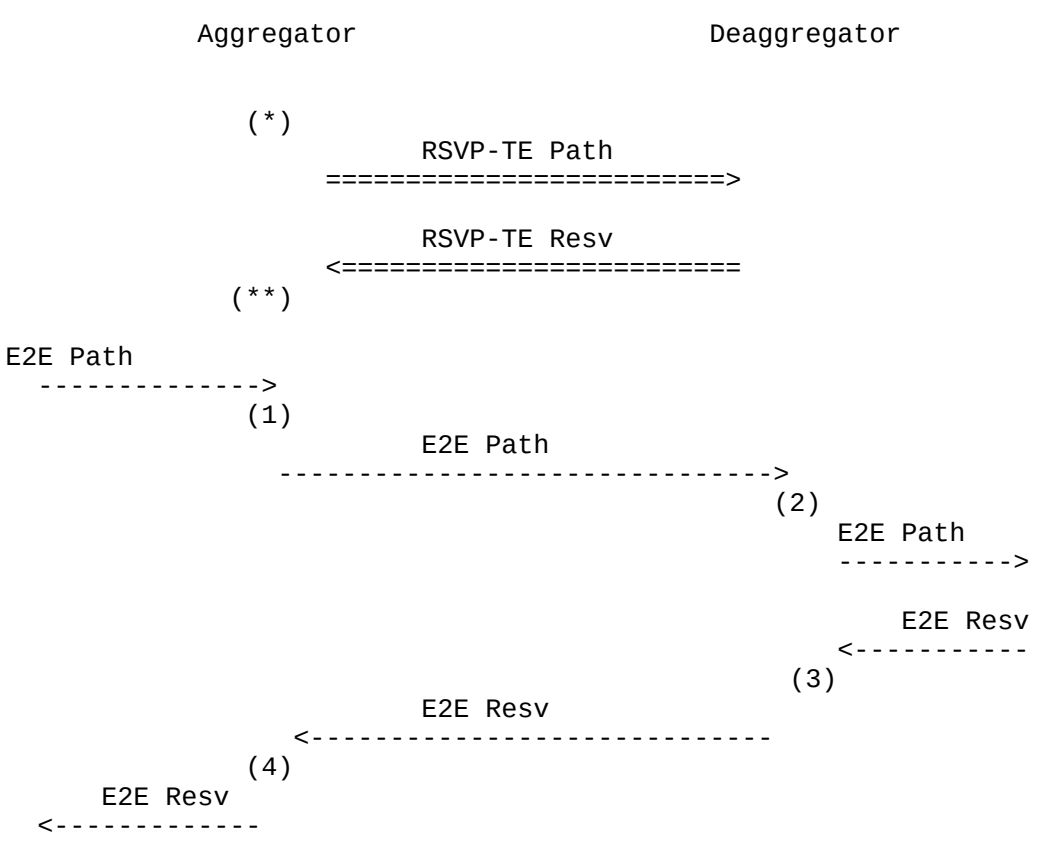
4.8. Removal of E2E Reservations

E2E reservations are removed in the usual way via PathTear, ResvTear, timeout, or as the result of an error condition. When a reservation is removed, the Aggregator MUST update its local view of the resources available on the corresponding TE tunnel accordingly.

4.9. Removal of the TE Tunnel

Should a TE tunnel go away (presumably due to a configuration change, route change, or policy event), the Aggregator behaves much like a conventional RSVP router in the face of a link failure. That is, it may try to forward the Path messages onto another tunnel, if routing and policy permit, or it may send Path_Error messages to the sender if a suitable tunnel does not exist. In case the Path messages are forwarded onto another tunnel, which terminates on a different Deaggregator, or the reservation is torn down via Path Error messages, the reservation state established on the router acting as the Deaggregator before the TE tunnel went away, will time out since it will no longer be refreshed.

4.10. Example Signaling Flow



(*) Aggregator is triggered to pre-establish the TE tunnel(s)

(**) TE tunnel(s) are pre-established

(1) Aggregator tentatively selects the TE tunnel and forwards E2E path to Deaggregator

(2) Deaggregator forwards the E2E Path toward the receiver

(3) Deaggregator forwards the E2E Resv to the Aggregator

(4) Aggregator selects final TE tunnel, checks that there is sufficient bandwidth on TE tunnel, and forwards E2E Resv to PHOP. If final tunnel is different from tunnel tentatively selected, the Aggregator re-sends an E2E Path with an updated IF_ID RSVP_HOP and possibly an updated ADSPEC.

5. IPv4 and IPv6 Applicability

The procedures defined in this document are applicable to all the following cases:

(1) Aggregation of E2E IPv4 RSVP reservations over IPv4 TE tunnels.

(2) Aggregation of E2E IPv6 RSVP reservations over IPv6 TE tunnels.

(3) Aggregation of E2E IPv6 RSVP reservations over IPv4 TE tunnels, provided a mechanism such as [6PE] is used by the Aggregator and Deaggregator for routing of IPv6 traffic over an IPv4 MPLS core.

(4) Aggregation of E2E IPv4 RSVP reservations over IPv6 TE tunnels, provided a mechanism is used by the Aggregator and Deaggregator for routing IPv4 traffic over IPv6 MPLS.

6. E2E Reservations Applicability

The procedures defined in this document are applicable to many types of E2E RSVP reservations including the following cases:

(1) The E2E RSVP reservation is a per-flow reservation where the flow is characterized by the usual 5-tuple

(2) The E2E reservation is an aggregate reservation for multiple flows as described in [RSVP-AGG] or [RSVP-GEN-AGG] where the set of flows is characterized by the <source address, destination address, DSCP>

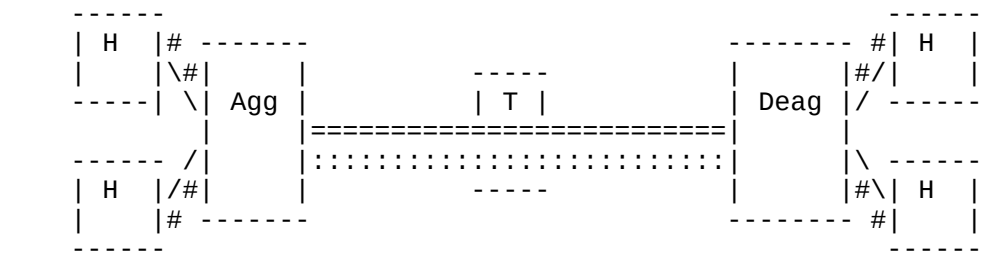
(3) The E2E reservation is a reservation for an IPsec protected flow. For example, where the flow is characterized by the <source address, destination address, SPI> as described in [RSVP-IPSEC].

7. Example Deployment Scenarios

7.1. Voice and Video Reservations Scenario

An example application of the procedures specified in this document is admission control of voice and video in environments with a very high number of hosts. In the example illustrated below, hosts generate E2E per-flow reservations for each of their video streams associated with a video-conference, each of their audio streams associated with a video-conference and each of their voice calls.

These reservations are aggregated over MPLS DS-TE tunnels over the packet core. The mapping policy defined by the user may be that all the reservations for audio and voice streams are mapped onto DS-TE tunnels of Class-Type 1, while reservations for video streams are mapped onto DS-TE tunnels of Class-Type 0.

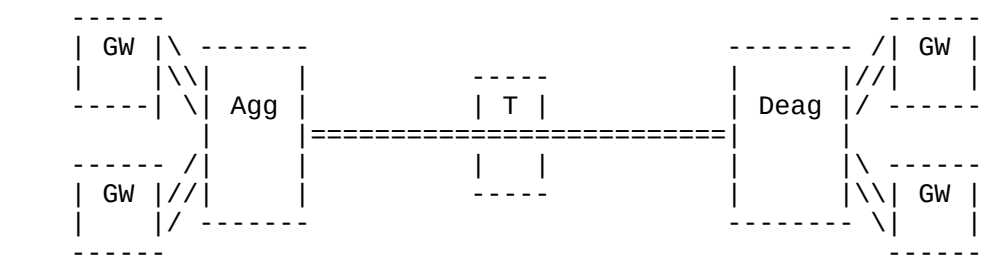


H = Host
Agg = Aggregator (TE tunnel Head-end)
Deagg = Deaggregator (TE tunnel Tail-end)
T = Transit LSR

/ = E2E RSVP reservation for a Voice flow
= E2E RSVP reservation for a Video flow
== = DS-TE tunnel from Class-Type 1
:: = DS-TE tunnel from Class-Type 0

7.2. PSTN/3G Voice Trunking Scenario

An example application of the procedures specified in this document is voice call admission control in large-scale telephony trunking environments. A Trunk VoIP Gateway may generate one aggregate RSVP reservation for all the calls in place toward another given remote Trunk VoIP Gateway (with resizing of this aggregate reservation in a step function depending on the current number of calls). In turn, these reservations may be aggregated over MPLS TE tunnels over the packet core so that tunnel Head-ends act as Aggregators and perform admission control of Trunk Gateway reservations into MPLS TE tunnels. The MPLS TE tunnels may be protected by MPLS Fast Reroute. This scenario is illustrated below:



GW = VoIP Gateway
Agg = Aggregator (TE tunnel Head-end)
Deagg = Deaggregator (TE tunnel Tail-end)
T = Transit LSR

/ = Aggregate Gateway to Gateway E2E RSVP reservation
== = TE tunnel

8. Security Considerations

In the environments concerned by this document, RSVP messages are used to control resource reservations for E2E flows outside the MPLS region as well as to control resource reservations for MPLS TE tunnels inside the MPLS region. To ensure the integrity of the associated reservation and admission control mechanisms, the mechanisms defined in [RSVP-CRYPTO1] and [RSVP-CRYPTO2] can be used. The mechanisms protect the integrity of RSVP messages hop-by-hop and provide node authentication, thereby protecting against corruption and spoofing of RSVP messages. These hop-by-hop integrity mechanisms can naturally be used to protect the RSVP messages used for E2E reservations outside the MPLS region, to protect RSVP messages used for MPLS TE tunnels inside the MPLS region, or for both. These hop-by-hop RSVP integrity mechanisms can also be used to protect RSVP messages used for E2E reservations when those transit through the MPLS region. This is because the Aggregator and Deaggregator behave as RSVP neighbors from the viewpoint of the E2E flows (even if they are not necessarily IP neighbors nor RSVP-TE neighbors). In that case, the Aggregator and Deaggregator need to use a pre-shared secret.

As discussed in Section 6 of [RSVP-TE], filtering of traffic associated with an MPLS TE tunnel can only be done on the basis of an MPLS label, instead of the 5-tuple of conventional RSVP reservation as per [RSVP]. Thus, as explained in [RSVP-TE], an administrator may wish to limit the domain over which TE tunnels (which are used for aggregation of RSVP E2E reservations as per this specification) can be established. See Section 6 of [RSVP-TE] for a description of how

filtering of RSVP messages associated with MPLS TE tunnels can be deployed to that end.

This document is based in part on [RSVP-AGG], which specifies aggregation of RSVP reservations. Section 5 of [RSVP-AGG] raises the point that because many E2E flows may share an aggregate reservation, if the security of an aggregate reservation is compromised, there is a multiplying effect in the sense that it can in turn compromise the security of many E2E reservations whose quality of service depends on the aggregate reservation. This concern applies also to RSVP Aggregation over TE tunnels as specified in the present document. However, the integrity of MPLS TE tunnels operation can be protected using the mechanisms discussed in the previous paragraphs. Also, while [RSVP-AGG] specifies RSVP Aggregation over dynamically established aggregate reservations, the present document restricts itself to RSVP Aggregation over pre-established TE tunnels. This further reduces the security risks.

In the case where the Aggregators dynamically resize the TE tunnels based on the current level of reservation, there are risks that the TE tunnels used for RSVP aggregation hog resources in the core, which could prevent other TE tunnels from being established. There are also potential risks that such resizing results in significant computation and signaling as well as churn on tunnel paths. Such risks can be mitigated by configuration options allowing control of TE tunnel dynamic resizing (maximum TE tunnel size, maximum resizing frequency, etc.), and/or possibly by the use of TE preemption.

Section 5 of [RSVP-AGG] also discusses a security issue specific to RSVP aggregation related to the necessary modification of the IP Protocol number in RSVP E2E Path messages that traverses the aggregation region. This security issue does not apply to the present document since aggregation of RSVP reservation over TE tunnels does not use this approach of changing the protocol number in RSVP messages.

Section 7 of [LSP-HIER] discusses security considerations stemming from the fact that the implicit assumption of a binding between data interface and the interface over which a control message is sent is no longer valid. These security considerations are equally applicable to the present document.

If the Aggregator and Deaggregator are also acting as IPsec Security Gateways, the Security Considerations of [SEC-ARCH] apply.

9. Acknowledgments

This document builds on the [RSVP-AGG], [RSVP-TUN], and [LSP-HIER] specifications. We would like to thank Tom Phelan, John Drake, Arthi Ayyangar, Fred Baker, Subha Dhesikan, Kwok-Ho Chan, Carol Iturralde, and James Gibson for their input into this document.

10. Normative References

[CONTROLLED] Wroclawski, J., "Specification of the Controlled-Load Network Element Service", RFC 2211, September 1997.

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

[DSTE-PROTO] Le Faucheur, F., "Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering", RFC 4124, June 2005.

[GUARANTEED] Shenker, S., Partridge, C., and R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212, September 1997.

[INT-DIFF] Bernet, Y., Ford, P., Yavatkar, R., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J., and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks", RFC 2998, November 2000.

[INT-SERV] Braden, R., Clark, D., and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.

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

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

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

[RSVP] Braden, R., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RSVP-AGG] Baker, F., Iturralde, C., Le Faucheur, F., and B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, September 2001.

[RSVP-CRYPTO1] Baker, F., Lindell, B., and M. Talwar, "RSVP Cryptographic Authentication", RFC 2747, January 2000.

[RSVP-CRYPTO2] Braden, R. and L. Zhang, "RSVP Cryptographic Authentication -- Updated Message Type Value", RFC 3097, April 2001.

[RSVP-TE] 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.

[SEC-ARCH] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

11. Informative References

[6PE] De Clercq, J., Ooms, D., Prevost, S., and F. Le Faucheur, "Connecting IPv6 Islands over IPv4 MPLS using IPv6 Provider Edge Routers (6PE)", RFC 4798, February 2007.

[AUTOMESH] Vasseur and Leroux, "Routing extensions for discovery of Multiprotocol (MPLS) Label Switch Router (LSR) Traffic Engineering (TE) mesh membership", Work in Progress.

[DIFF-MPLS] 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.

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

[L-RSVP] Manner, et al., Localized RSVP for Controlling RSVP Proxies, Work in Progress.

[RSVP-APPID] Yadav, S., Yavatkar, R., Pabbati, R., Ford, P., Moore, T., Herzog, S., and R. Hess, "Identity Representation for RSVP", RFC 3182, October 2001.

[RSVP-GEN-AGG] Le Faucheur, R., Davie, B., Bose, P., Martin, L., Christou, C., Davenport, M., and A. Hamilton, "Generic Aggregate Resource ReSerVation Protocol (RSVP) Reservations", Work in Progress, January 2007.

[RSVP-IPSEC] Berger, L. and T. O'Malley, "RSVP Extensions for IPSEC Data Flows", RFC 2207, September 1997.

[RSVP-PREEMP] Herzog, S., "Signaled Preemption Priority Policy Element", RFC 3181, October 2001.

[RSVP-PROXY1] Gai, et al., RSVP Proxy, Work in Progress.

[RSVP-PROXY2] Le Faucheur, et al., RSVP Proxy Approaches, Work in Progress.

[RSVP-TUN] Terzis, A., Krawczyk, J., Wroclawski, J., and L. Zhang, "RSVP Operation Over IP Tunnels", RFC 2746, January 2000.

[SIP-RSVP] Camarillo, G., Marshall, W., and J. Rosenberg, "Integration of Resource Management and Session Initiation Protocol (SIP)", RFC 3312, October 2002.

Appendix A - Optional Use of RSVP Proxy on RSVP Aggregator

A number of approaches ([RSVP-PROXY1],[RSVP-PROXY2], [L-RSVP]) have been, or are being, discussed in the IETF in order to allow a network node to behave as an RSVP proxy which:

- originates the Resv Message (in response to the Path message) on behalf of the destination node

- originates the Path message (in response to some trigger) on behalf of the source node.

We observe that such approaches may optionally be used in conjunction with the aggregation of RSVP reservations over MPLS TE tunnels as specified in this document. In particular, we consider the case where the RSVP Aggregator/Deaggregator also behaves as the RSVP proxy.

The information in this Appendix is purely informational and illustrative.

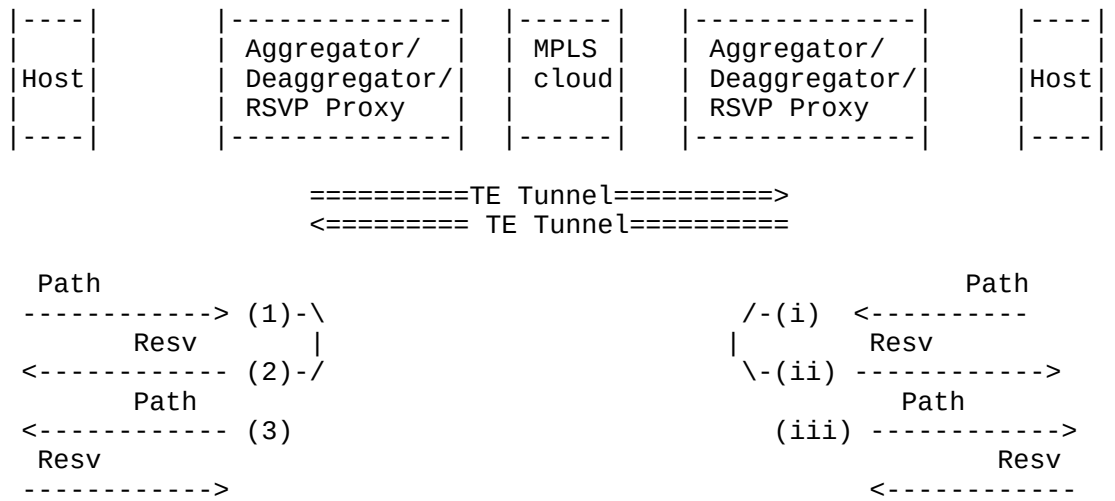
As discussed in [RSVP-PROXY1]:

"The proxy functionality does not imply merely generating a single Resv message. Proxying the Resv involves installing state in the node doing the proxy i.e. the proxying node should act as if it had received a Resv from the true endpoint. This involves reserving resources (if required), sending periodic refreshes of the Resv message and tearing down the reservation if the Path is torn down."

Hence, when behaving as the RSVP Proxy, the RSVP Aggregator may effectively perform resource reservation over the MPLS TE tunnel (and hence over the whole segment between the RSVP Aggregator and the RSVP Deaggregator) even if the RSVP signaling only takes place upstream of the MPLS TE tunnel (i.e., between the host and the RSVP aggregator).

Also, the RSVP Proxy can generate the Path message on behalf of the remote source host in order to achieve reservation in the return direction (i.e., from RSVP aggregator/Deaggregator to host).

The resulting Signaling Flow is illustrated below, covering reservations for both directions:



(1)(i) : Aggregator/Deaggregator/Proxy receives Path message, selects the TE tunnel, performs admission control over the TE tunnel. (1) and (i) happen independently of each other.

(2)(ii) : Aggregator/Deaggregator/Proxy generates the Resv message toward Host. (2) is triggered by (1) and (ii) is triggered by (i). Before generating this Resv message, the Aggregator/Proxy performs admission control of the corresponding reservation over the TE tunnel that will eventually carry the corresponding traffic.

(3)(iii) : Aggregator/Deaggregator/Proxy generates the Path message toward Host for reservation in return direction. The actual trigger for this depends on the actual RSVP proxy solution. As an example, (3) and (iii) may simply be triggered respectively by (1) and (i).

Note that the details of the signaling flow may vary slightly depending on the actual approach used for RSVP Proxy. For example, if the [L-RSVP] approach was used instead of [RSVP-PROXY1], an additional PathRequest message would be needed from host to Aggregator/Deaggregator/Proxy in order to trigger the generation of the Path message for return direction.

But regardless of the details of the call flow and of the actual RSVP Proxy approach, RSVP proxy may optionally be deployed in combination with RSVP Aggregation over MPLS TE tunnels, in such a way that

ensures (when used on both the Host-Aggregator and Deaggregator-Host sides, and when both end systems support RSVP):

(i) admission control and resource reservation is performed on every segment of the end-to-end path (i.e., between source host and Aggregator, over the TE tunnel between the Aggregator and Deaggregator that itself has been subject to admission control by MPLS TE, between Deaggregator and destination host).

(ii) this is achieved in both directions.

(iii) RSVP signaling is localized between hosts and Aggregator/Deaggregator, which may result in significant reduction in reservation establishment delays (and in turn in post-dial delay in the case where these reservations are pre-conditions for voice call establishment), particularly in the case where the MPLS TE tunnels span long distances with high propagation delays.

Appendix B - Example Usage of RSVP Aggregation over DSTE Tunnels for VoIP Call Admission Control (CAC)

This Appendix presents an example scenario where the mechanisms described in this document are used, in combination with other mechanisms specified by the IETF, to achieve Call Admission Control (CAC) of Voice over IP (VoIP) traffic over the packet core.

The information in this Appendix is purely informational and illustrative.

Consider the scenario depicted in Figure B1. VoIP Gateways GW1 and GW2 are both signaling and media gateways. They are connected to an MPLS network via edge routers PE1 and PE2, respectively. In each direction, a DSTE tunnel passes from the Head-end edge router, through core network P routers, to the Tail-end edge router. GW1 and GW2 are RSVP-enabled. The RSVP reservations established by GW1 and GW2 are aggregated by PE1 and PE2 over the DS-TE tunnels. For reservations going from GW1 to GW2, PE1 serves as the Aggregator/Head-end and PE2 serves as the Deaggregator/Tail-end. For reservations going from GW2 to GW2, PE2 serves as the Aggregator/Head-end and PE1 serves as the Deaggregator/Tail-end.

To determine whether there is sufficient bandwidth in the MPLS core to complete a connection, the originating and destination GWs each send for each connection a Resource Reservation Protocol (RSVP) bandwidth request to the network PE router to which it is connected. As part of its Aggregator role, the PE router effectively performs

admission control of the bandwidth request generated by the GW onto the resources of the corresponding DS-TE tunnel.

In this example, in addition to behaving as Aggregator/Deaggregator, PE1 and PE2 behave as RSVP proxy. So when a PE receives a Path message from a GW, it does not propagate the Path message any further. Rather, the PE performs admission control of the bandwidth signaled in the Path message over the DSTE tunnel toward the destination. Assuming there is enough bandwidth available on that tunnel, the PE adjusts its bookkeeping of remaining available bandwidth on the tunnel and generates a Resv message back toward the GW to confirm resources have been reserved over the DSTE tunnel.

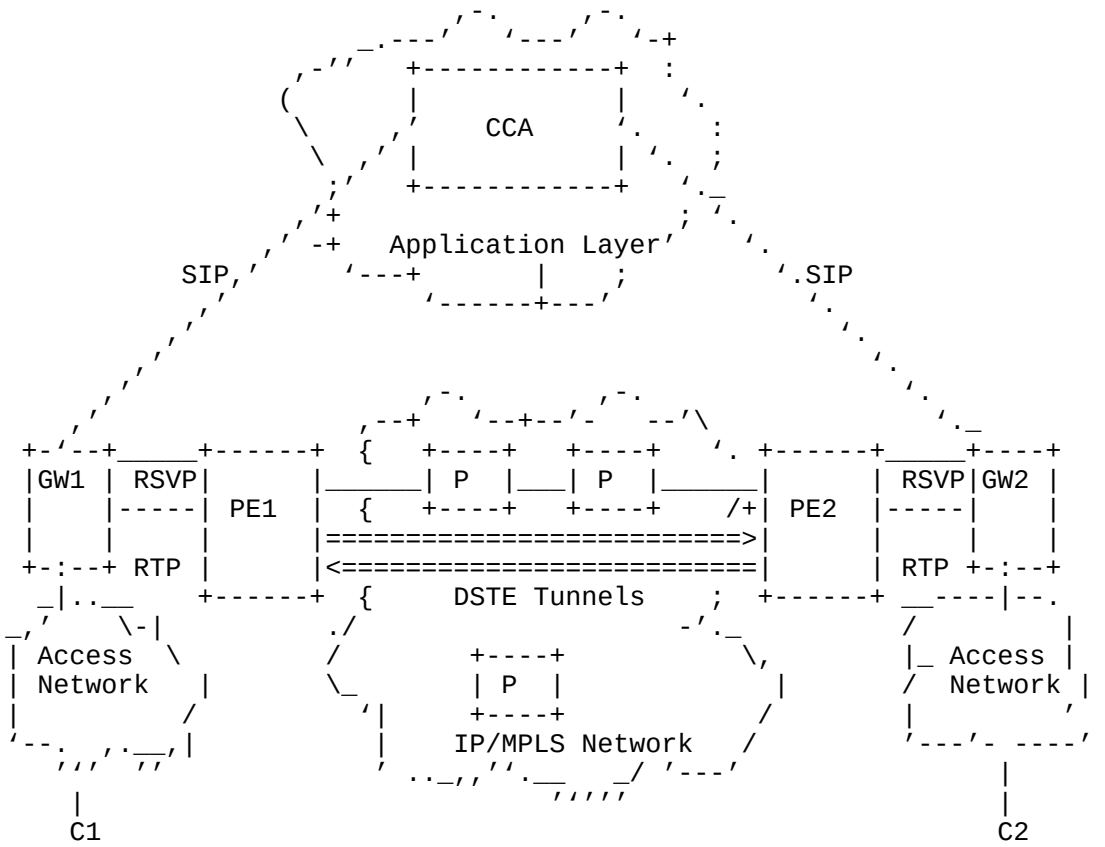


Figure B1. Integration of SIP Resource Management and RSVP Aggregation over MPLS TE Tunnels

[SIP-RSVP] discusses how network quality of service can be made a precondition for establishment of sessions initiated by the Session

Initiation Protocol (SIP). These preconditions require that the participant reserve network resources before continuing with the session. The reservation of network resources are performed through a signaling protocol such as RSVP.

Through the collaboration between SIP resource management, RSVP signaling, RSVP Aggregation and DS-TE as described above, we see that:

a) the PE and GW collaborate to determine whether there is enough bandwidth on the tunnel between the calling and called GWs to accommodate the connection,

b) the corresponding accept/reject decision is communicated to the GWs on a connection-by-connection basis, and

c) the PE can optimize network resources by dynamically adjusting the bandwidth of each tunnel according to the load over that tunnel. For example, if a tunnel is operating at near capacity, the network may dynamically adjust the tunnel size within a set of parameters.

We note that admission Control of voice calls over the core network capacity is achieved in a hierarchical manner whereby:

- DSTE tunnels are subject to Admission Control over the resources of the MPLS TE core

- Voice calls are subject to CAC over the DSTE tunnel bandwidth

This hierarchy is a key element in the scalability of this CAC solution for voice calls over an MPLS Core.

It is also possible for the GWs to use aggregate RSVP reservations themselves instead of per-call RSVP reservations. For example, instead of setting one reservation for each call GW1 has in place toward GW2, GW1 may establish one (or a small number of) aggregate reservations as defined in [RSVP-AGG] or [RSVP-GEN-AGG], which is used for all (or a subset of all) the calls toward GW2. This effectively provides an additional level of hierarchy whereby:

- DSTE tunnels are subject to Admission Control over the resources of the MPLS TE core

- Aggregate RSVP reservations (for the calls from one GW to another GW) are subject to Admission Control over the DSTE tunnels (as per the "RSVP Aggregation over TE Tunnels" procedures defined in this document)

- Voice calls are subject to CAC by the GW over the aggregate reservation toward the appropriate destination GW.

This pushes even further the scalability limits of this voice CAC architecture.

Contributing Authors

This document was the collective work of several authors. The text and content were contributed by the editor and the co-authors listed below.

Michael DiBiasio
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
USA
EMail: dibiasio@cisco.com

Bruce Davie
Cisco Systems, Inc.
300 Beaver Brook Road
Boxborough, MA 01719
USA
EMail: bdavie@cisco.com

Christou Christou
Booz Allen Hamilton
8283 Greensboro Drive
McLean, VA 22102
USA
EMail: christou_chris@bah.com

Michael Davenport
Booz Allen Hamilton
8283 Greensboro Drive
McLean, VA 22102
USA
EMail: davenport_michael@bah.com

Jerry Ash
AT&T
200 Laurel Avenue
Middletown, NJ 07748
USA
EMail: gash@att.com

Bur Goode
AT&T
32 Old Orchard Drive
Weston, CT 06883
USA
EMail: bgoode@att.com

Editor’s Address

Francois Le Faucheur
Cisco Systems, Inc.
Village d’Entreprise Green Side - Batiment T3
400, Avenue de Roumanille
06410 Biot Sophia-Antipolis
France

EMail: flefauch@cisco.com

Full Copyright Statement

Intellectual Property

Acknowledgement

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