TSVWG
Internet-Draft
Expires: April 24, 2006

F. Le Faucheur
B. Davie
Cisco Systems
P. Bose
Lockheed Martin
C. Christou
M. Davenport
Booz Allen Hamilton
October 21, 2005

Generic Aggregate RSVP Reservations
draft-lefaucheur-rsvp-ipsec-02.txt

Status of this Memo

Copyright Notice

Abstract

[RSVP-IPSEC] defines RSVP extensions for IPsec which permit support of reservations for individual IPsec flows, but it does not support aggregate reservations between the IPsec devices with Diffserv

[DIFFSERV] classification and scheduling. Conversely, [RSVP-AGG] defines how to aggregate individual RSVP reservations over Aggregate IP reservations when the aggregation region supports Diffserv, but it does not address the case where the Aggregator and Deaggregator use IPsec. Also, [RSVP-AGG] does not address the case where multiple Aggregate reservations are needed for the same DSCP from the same Aggregator to the same Deaggregator. However, there are scenarios requiring aggregate reservations for IPsec tunnels or requiring multiple aggregate reservations for the same DSCP from a given Aggregator to a given Deaggregator. This document specifies the incremental RSVP extensions beyond those defined in [RSVP-IPSEC] and [RSVP-AGG] to support such reservations.

Table of Contents

1. Introduction

[RSVP-IPSEC] defines RSVP extensions for IPsec that permit support of reservations for individual IPsec flows, but it does not support aggregate reservations between IPsec devices with Diffserv [DIFFSERV] classification and scheduling. Conversely, [RSVP-AGG] defines how to aggregate individual RSVP reservations over Aggregate IP reservations when the aggregation region supports Diffserv, but it does not address the case where the Aggregator and Deaggregator use IPsec. Also, [RSVG-AGG] does not address the case where multiple Aggregate reservations are needed for the same DSCP from the same Aggregator to the same Deaggregator. However, there are scenarios requiring aggregate reservations for IPsec tunnels or requiring multiple aggregate reservations for the same DSCP from a given Aggregator to a given Deaggregator. This document specifies the incremental RSVP extensions beyond those defined in [RSVP-IPSEC] and [RSVP-AGG] to support such reservations.

1.1. Aggregate Reservations For IPsec Tunnels

[IPSEC-ARCH] defines the term "security gateway" to refer to an intermediate system that implements IPsec protocols. In this document we refer to a an IP router behaving as a security gateway as an "IPsec-Router".

Consider an environment as depicted in Figure 1. Let us assume that the IPsec-Routers tunnel traffic to each other via IPsec and that the devices within Cloud-1, Cloud-2 and Cloud-3 want to establish RSVP reservations with one another transparently over the IPsec Tunnels. Let us also assume that Cloud-0 supports Diffserv (and not per-flow classification -except perhaps at the edge for policing purposes) and that there is a need to reserve resources over Cloud-0 to achieve the targeted levels of QoS assurance. Then there is a need to establish aggregate reservations within Cloud-0 for the IPsec tunnels transiting through Cloud-0. These aggregate reservations will be used to aggregate the end-to-end RSVP reservations between Cloud-1/2/3. This document concerns itself with establishment of such aggregate reservations for IPsec tunnels.

The reader is referred to [SIG-NESTED] for a description of a more generic nested VPN environment and for discussion and examples of QoS signaling in that environment.

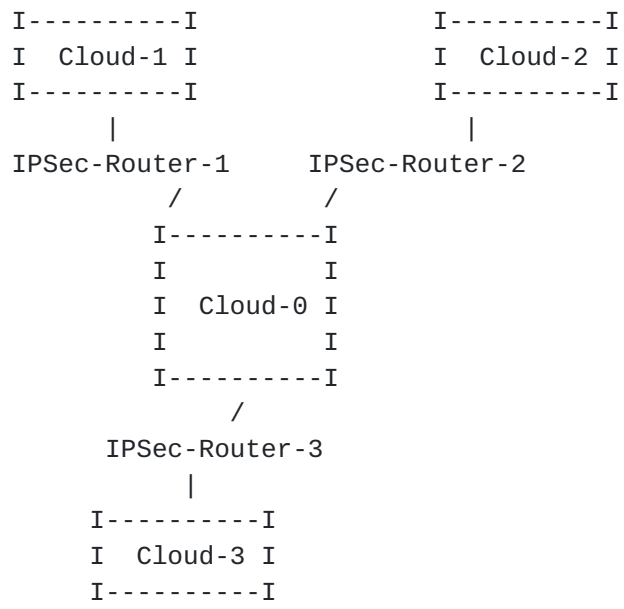


Figure 1: Example Scenario Requiring Aggregate Reservations for IPsec tunnels

[RSVP-AGG] defines a Session Object containing only the Deaggregator IP address and the DSCP, and defines a Filter Spec Object containing only the aggregator IP address. Thus, we observe that it is not possible to convey the IPsec Security Parameter Index (SPI) that is used for a given IPsec tunnel (unlike with [RSVP-IPSEC]). In turn, this means that, if [RSVP-AGG] was used to establish aggregate reservations for IPsec tunnels, it would not be possible for the (edge) routers within Cloud-0 to classify traffic belonging to the reservation corresponding to a given IPsec tunnel (say for the purpose of doing policing on the edge of Cloud-0). It also means that it would not be possible (short of multiplying IP addresses) to setup separate reservations for different IPsec tunnels (using different SPIs) between the same IPsec encryptor and decryptor (which may be used if different types of traffic have different security requirements). Similarly, it would not be possible to set up separate reservations for traffic going over the IPsec tunnel and for traffic that is not encrypted (which is a useful scenario if some traffic has IPsec requirement while the rest doesn't). Moreover, it would not be possible to setup multiple reservations between a given pair of IPsec encryptor and decryptor for transport of flows with different preemptions [RSVP-PREEMP]. These restrictions illustrate why the RSVP extensions defined in [RSVP-AGG] are not sufficient to support aggregate reservations for IPsec tunnels.

[RSVP-IPSEC] defines a Session Object containing several fields including a Virtual Destination Port (VDstPort) which allows support of a different reservation for each IPsec flow , or even of multiple reservations for a given IPsec flow. However, (unlike with [RSVP-

AGG]), the RSVP extensions of [RSVP-IPSEC] do not allow the DSCP to be part of the fields uniquely identifying the reservation (i.e. the Session and Filter Spec objects). The extensions of [RSVP-IPSEC] essentially assume a per-flow classification model instead of Diffserv aggregate classification and scheduling. This is why the RSVP extensions defined in [RSVP-IPSEC] are not sufficient either to support aggregate reservations for IPsec tunnels.

This document defines incremental RSVP extensions that simply combine the concepts introduced in [RSVP-IPSEC] and in [RSVP-AGG]. This way, their benefits can be obtained simultaneously hence allowing aggregate reservations for IPsec tunnels with Diffserv classification and scheduling.

These extensions can be used in a number of scenarios. They allow aggregation of end-to-end RSVP reservations over aggregate reservations for IPsec tunnels. They also allow multi-level aggregation. For example, end-to-end RSVP reservations may first be aggregated by a router acting as an [RSVP-AGG] aggregator and then the resulting [RSVP-AGG] aggregate reservations may in turn be aggregated by the IPsec encryptor into generic aggregate RSVP reservations. These extensions may also be used to establish an aggregate reservation for an IPsec tunnel between an IPsec encryptor and an IPsec decryptor for transport of other traffic than the one corresponding to end to end RSVP reservations (for example to provide a fixed pipe of Diffserv bandwidth from IPsec encryptor to IPsec decryptor to carry end-to-end Diffserv traffic). Another possible example usage is for establishment of an aggregate reservation end-to-end from an IPsec end-system to another IPsec end-system.

These extensions allow full support of QoS signaling in Nested VPNs as discussed in [SIG-NESTED]. Example usage of these extensions in Nested VPN is described in section 5.

1.2. Multiple Reservations Per DSCP From A Given Aggregator To A Given Deaggregator

Let us consider an environment where E2E RSVP reservations need to be aggregated over an aggregation region. Now imagine that different E2E RSVP reservations (corresponding to the same DSCP) are established with different preemptions [RSVP-PREEMP] and that the corresponding preemption need to be enforced over the aggregation region. One method to achieve this is to establish one Aggregate RSVP reservation per preemption level for a given DSCP and from a given Aggregator to a given Deaggregator.

As mentioned earlier, [RSVP-AGG] defines a Session Object containing only the Deaggregator IP address and the DSCP, and defines a Filter

Spec Object containing only the aggregator IP address. Thus, the extensions defined in [RSVP-AGG] do not allow establishment of multiple Aggregate RSVP reservations for a given <Aggregator/ Deaggregator/DSCP> Tuple (short of multiplying the IP addresses allocated to the Aggregator or Deaggregator).

The extensions defined in this document combine the concept of Virtual Destination Port introduced in [RSVP-IPSEC] (which allows establishment of multiple reservations between same source/ destination) with the inclusion of DSCP in the Session object introduced in [RSVP-AGG]. This allows establishment of multiple Aggregate RSVP reservations for a given <Aggregator/Deaggregator/ DSCP> Tuple.

1.3. Related RFCs and Internet-Drafts

The mechanisms defined in [BW-REDUC] allow an existing reservation to be reduced in allocated bandwidth in lieu of tearing that reservation down. These mechanisms are applicable to the aggregate reservations for IPsec tunnels defined in the present document.

[RSVP-TUNNEL] describes a general approach to running RSVP over various types of tunnels. One of these types of tunnel, referred to as a "type 2 tunnel", is similar to the tunnels described in this draft. The similarity stems from the fact that a single, aggregate reservation is made for the tunnel while many individual flows are carried over that tunnel. However, [RSVP-TUNNEL] does not address the case where data flows are encrypted, and thus does not deal with the use of the SPI to identify flows and sessions. Nor does it address the use of Diffserv-based classification and queuing in the core of a network (between tunnel endpoints), but rather relies on a UDP/IP tunnel header for classification. Thus we require some additional objects and procedures, defined in this draft, beyond those of [RSVP-TUNNEL].

1.4. Organization Of This Document

Section 2 presents an overview of the RSVP extensions defined in this document and how those are used. Section 3 provides specification for the new RSVP objects. The changes to existing RSVP processing rules are identified in Section 4. Section 5 provides example usages of aggregate reservations for IPsec tunnels in a Nested VPN environment as well as of aggregate IP reservations. The IANA Considerations and the Security Considerations are discussed in Section 6 and 7, respectively.

1.5. Change History

1.5.1. Changes From -00 To -01

The most significant change is the broadening of the applicability of the new type of aggregate reservations beyond use for Aggregate reservations for IPsec tunnels (to environments where IPsec is not used). This affects the document in multiple places including the following changes:

- document renamed to "Generic Aggregate RSVP Reservations"
- added a subsection in Introduction to discuss a case where Generic Aggregate RSVP Reservations are needed in non IPsec environments
- added text about the fact that the Generic Aggregate Reservations can be used with IP-in-IP and GRE encapsulation (in addition to with IPsec AH and ESP)
- added example usage under Section 5 for environment where IPsec is not used

The other significant changes are:

- added a subsection on the changes of the [RSVP-AGG] procedures under Section 4
- added explanation about allocation of VDstPort values by Deaggregator, in that same subsection
- added value of Protocol ID in all example generic aggregate reservations in Section 5

1.5.2. Changes From -01 To -02

The most significant changes are :

- added text in section 4.2 about Aggregator/Deaggregator responsibilities with respect to mapping of end-to-end reservations onto aggregate reservations. The text also clarified that DCLASS object is no longer needed in PathErr message requesting new Aggregate Reservations
- Moved the text discussing details of the procedures to handle dynamic update of SPI values from Security Considerations section into a new section 4.4.

- updates to Security Considerations section to start addressing some comments from Security experts review.

2. Overview of Extensions

The extensions defined in this document can be seen as simply the combination of the RSVP extensions defined in [RSVP-IPSEC] and in [RSVP-AGG].

The basic notion of [RSVP-IPSEC] is to extend RSVP to use the IPsec Security Parameter Index (SPI) in place of the UDP/TCP-like ports. This was achieved via:

- definition of a new FILTER_SPEC object which includes a Generalized Port Identifier (GPI) field which is used to convey the SPI

- definition of a new SESSION object which includes a Virtual Destination Port (VDstPort). The VDstPort effectively allows for the differentiation of multiple IPsec sessions destined to the same IP address. (The VDstPort is used in the Session rather than the SPI because it isn't feasible to force all senders to a session to use the same SPI - which is needed in situations where sharing of reservations across multiple senders is required)

One of the key notions of [RSVP-AGG] is that inside the aggregation region, some RSVP reservation state is maintained per aggregate reservation, while classification and scheduling state (e.g., DSCPs used for classifying traffic) is maintained on a more highly aggregated basis. For example, if Guaranteed Service reservations are mapped to the EF DSCP throughout the aggregation region, there may be a reservation for each Aggregator/Deaggregator pair in each router. However, only the EF DSCP needs to be inspected for classification of the data traffic at each interior interface, and only a single queue is used for all EF traffic. Support for this in [RSVP-AGG] involved:

- definition of a new SESSION object which includes the DSCP

Hence, in order to simultaneously achieve support of per IPsec flow reservations as well as Diffserv aggregate classification and scheduling, this document :

- reuses the FILTER_SPEC object defined in [RSVP-IPSEC] and containing a GPI (which in turn can include the SPI)

- defines a new SESSION object which contains both the VDstPort and the DSCP

The use of the VDstPort field is as specified in [RSVP-IPSEC]. When traffic from the E2E reservations is transported in aggregate IPsec

tunnels using AH or ESP, the use of the GPI field is as specified in [RSVP-IPSEC]. When traffic from the E2E reservations is transported into aggregate IP reservations using IP-in-IP or GRE, the GPI field is not used. In that case, the GPI is to be set to 0 by the sender of the RSVP message and to be ignored by the receiver of the RSVP message. The use of the DSCP field is as specified in [RSVP-AGG].

Where these RSVP extensions are used to perform aggregation of RSVP reservations over generic aggregate RSVP reservations, the aggregation and deaggregation functions are as specified in [RSVP-AGG] unless explicitly spelled out in the following paragraphs.

Like with [RSVP-AGG], it is the Deaggregator which is responsible for mapping E2E reservations onto generic aggregate reservations. In turn, this means the Deaggregator is responsible for requesting the Aggregator to initiate establishment of a new generic aggregate reservation when necessary and also for conveying to the Aggregator information about which generic aggregate reservation a given flow needs to be mapped onto.

Like with [RSVP-AGG], to request establishment of a generic aggregate reservation, the Deaggregator sends an E2E PathErr message with an error code of NEW-AGGREGATE-NEEDED. However, to provide all the necessary information about the needed generic aggregate reservation, this document extends the procedures of [RSVP-AGG] and allows the Deaggregator to include in the E2E PathErr message a new object called AGGREGATION-SESSION. This object contains all the information describing the Session of the needed new generic aggregate reservation, in order to convey those to the Aggregator.

This document also extends the procedures of [RSVP-AGG] to allow the Deaggregator to include the new AGGREGATION-SESSION object in the E2E Resv message, in order to convey to the Aggregator which generic aggregate session to map a given E2E reservation onto.

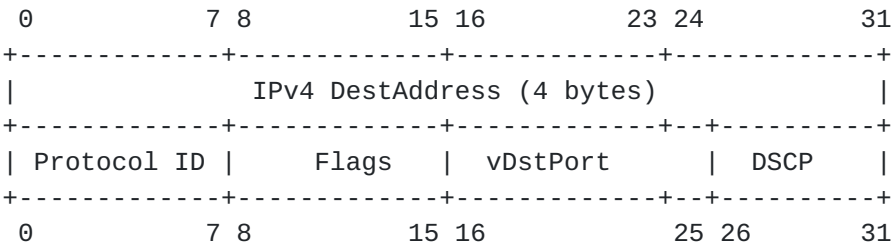
3. Object Definition

This document defines two new objects under the SESSION Class and a new object under a new AGGREGATION SESSION Class.

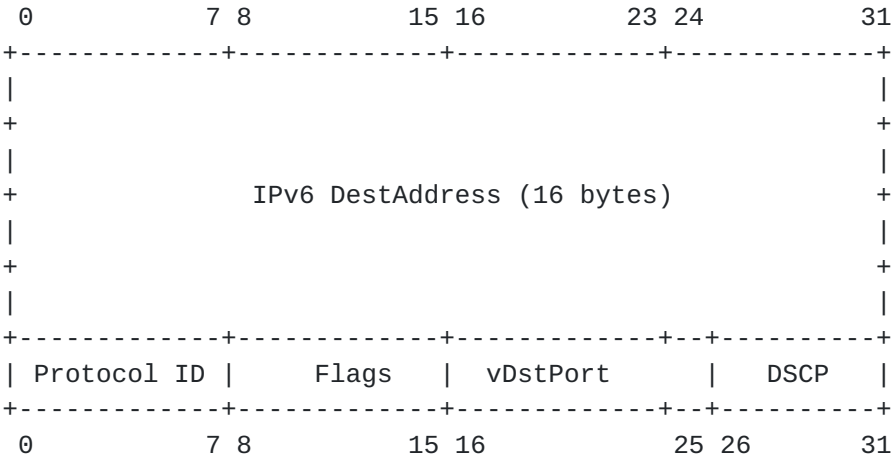
It reuses the IPv4/GPI FILTER_SPEC, IPv6/GPI FILTER_SPEC, IPv4/GPI SENDER_TEMPLATE and IPv6/GPI SENDER_TEMPLATE objects defined in [RSVP-IPSEC].

3.1. SESSION Class

o AGGREGATE-IPv4/GPI SESSION object:
  Class = 1
  C-Type = To be allocated by IANA

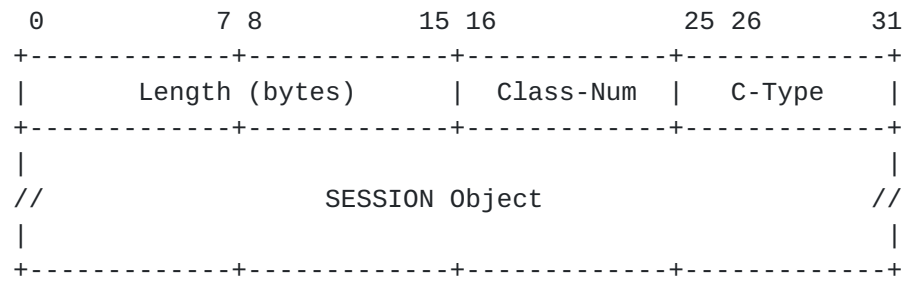


o AGGREGATE-IPv6/GPI SESSION object:
  Class = 1
  C-Type = To be allocated by IANA

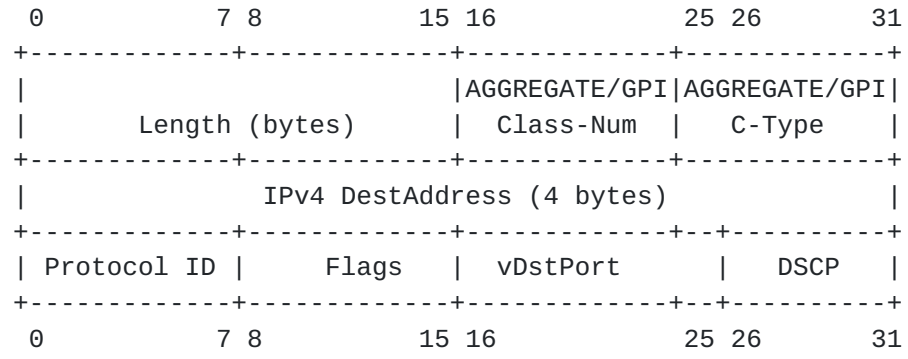


3.2. AGGREGATION-SESSION Class

o AGGREGATION-SESSION object:
  Class = To be allocated by IANA
  C-Type = To be allocated by IANA



The Length, Class-Num and C-Type are those of the Session object which is included inside the AGGREGATION-SESSION object. For example, if the AGGREGATION-SESSION object is used to indicate that the Aggregate Session needed is an AGGREGATE-IPv4/GPI SESSION then the AGGREGATION-SESSION will be encoded like this:



4. Processing Rules

This section presents additions to the Processing Rules presented in [RSVP-PROCESS] and in [RSVP-IPSEC]. These additions are required in order to properly process the AGGREGATE-IPv4/GPI (resp. AGGREGATE-IPv6/GPI) SESSION object and the IPv4/GPI (resp. IPv4-6/GPI) FILTER_SPEC object. Values for referenced error codes can be found in [RSVP]. As with the other RSVP documents, values for internally reported (API) errors are not defined.

When referring to the new AGGREGATE-IPv4/GPI and AGGREGATE-IPv6/GPI SESSION objects, IP version will not be included and they will be referred to simply as AGGREGATE/GPI SESSION, unless a specific distinction between IPv4 and IPv6 is being made.

Similarly, as per the convention used in [RSVP-IPSEC], when referring to the objects defined in [RSVP-IPSEC], IP version will not be included unless a specific distinction between IPv4 and IPv6 is being made.

4.1. Required Changes to Path and Resv Processing

Both RESV and PATH processing will need to be changed to support the new objects.

The following PATH message processing changes are required:

- When a session is defined using the AGGREGATE/GPI SESSION object, only the GPI SENDER_TEMPLATE may be used. When this condition is violated, RSVP end-stations should report a "Conflicting C-Type" API error to the application and routers should consider this as a message formatting error.

- For PATH messages that contain the AGGREGATE/GPI SESSION object, RSVP end-stations must verify that the protocol ID corresponds to a protocol known to use the AGGREGATE/GPI SESSION object. Protocol ID known to use the AGGREGATE/GPI SESSION are values 4 (IP-in-IP), 47 (GRE), 51 (AH) and 50 (ESP). If a router receives such a Path message with a protocol ID which doesn't correspond to a protocol known to use the AGGREGATE/GPI SESSION object, the router should consider this as a message formatting error. If an end-systems receives a Path message with an unknown protocol ID, then the API on the RSVP end-system should report an "API Error" to the application.

- For PATH messages that contain the AGGREGATE/GPI SESSION object, the VDstPort value and the DSCP value should be recorded. These values form part of the recorded state of the session. Only the

DSCP needs be passed to traffic control, since the vDstPort is not contained in data packets.

The changes to RESV message processing are:

- When a RESV message contains a GPI FILTER_SPEC, the session must be defined using either the GPI SESSION object (as per [RSVP-IPSEC]) or the AGGREGATE/GPI SESSION object (as per this document). Otherwise, this is a message formatting error.

- The GPI contained in the GPI FILTER_SPEC must match the GPI contained in the SENDER_TEMPLATE. Otherwise, a "No sender information for this Resv message" error is generated.

- When the GPI FILTER_SPEC is used and the SESSION type is AGGREGATE/GPI, each node must have a data classifier installed for the flow:

  * If the node needs to perform fine-grain classification (for example to perform fine-grain policing on ingress at a trust boundary) then the node must create a data classifier described by

    + the 5-tuple <DestAddress, protocol ID, SrcAddress, GPI, DSCP> if the Protocol ID is AH or ESP. The data classifier will need to look for the four byte GPI at transport header offset +4 for AH, and at transport header offset +0 for ESP (see [RSVP-IPSEC], [IPSEC-AG] and [IPSEC-ESP]). Note that if multiple reservations are established with different Virtual Destination Ports but with the same <DestAddress, protocol ID, SrcAddress, GPI, DSCP>, then those cannot be distinguished by the classifier. If the router is using the classifier for policing purposes, the router will therefore police those together and must program the policing rate to the sum of the reserved rate across all the corresponding reservations.

    + the 4-tuple <DestAddress, protocol ID, SrcAddress, DSCP> if the Protocol ID is IP-in-IP or GRE. Note that if multiple reservations are established with different Virtual Destination Ports but with the same <DestAddress, protocol ID, SrcAddress, DSCP>, then those cannot be distinguished by the classifier. If the router is using the classifier for policing purposes, the router will therefore police those together and must program the policing rate to the sum of the reserved rate across all the corresponding reservations.

* If the node only needs to perform Diffserv classification (for example inside the aggregation domain downstream of the trust boundary) then the node must rely on the Diffserv data classifier based on the DSCP only.

4.2. Required Changes to Aggregator/Deaggregator Processing

As specified in [RSVP-AGG], the Deaggregator requests establishment of the corresponding Aggregate Path by sending an E2E PathErr message with an error code of NEW-AGGREGATE-NEEDED and the desired DSCP encoded in the respective DCLASS Object. This document modifies and extends this procedure by allowing the Deaggregator to include in the E2E PathErr message an AGGREGATION-SESSION object which contains the Session to be used for establishment of the Aggregate Path. Since the AGGREGATION-SESSION object contains the DSCP, the DCLASS object need not be included in the PathErr message. Note that the AGGREGATION-SESSION object provides a very convenient mechanism to ensure that different Aggregators use different sessions for their Aggregate Path towards a given Deaggregator. This is because the Deaggregator can easily select VDstPort numbers which are different for each Aggregator and communicate those inside the AGGREGATION-SESSION object. This provides an easy solution to establish separate reservations from every Aggregator to a given Deaggregator. Conversely, if reservation sharing was needed across multiple Aggregators, the Deaggregator could facilitate this by allocating the same VDstPort to the multiple Aggregators and thus including the same AGGREGATION-SESSION object in the E2E PathErr messages sent to these Aggregators. The Aggregators could then all establish an Aggregate Path with the same Session.

Similarly, the [RSVP-AGG] procedures for processing of an E2E PathErr message by the Aggregator are extended so that the Aggregator uses the Session provided in the AGGREGATION-SESSION object to establish the Aggregate Path.

This document also extends the procedures of [RSVP-AGG] to allow the Deaggregator to include the new AGGREGATION-SESSION object in the E2E Resv message, in order to convey to the Aggregator which aggregate session to map a given E2E reservation onto. Again, since the AGGREGATION-SESSION object contains the DSCP, the DCLASS object need not be included in the E2E Resv message.

As discussed in [RSVP-AGG]:

o the rules for mapping E2E reservations onto aggregate reservations are policy decisions which are outside the scope of [RSVP-AGG] and of this specification. Suffice to know that such a policy is somehow accessible to the Aggregators/Deaggregators.

- Regardless of the actual policy, a range of options are conceivable for where the decision to map an E2E reservation onto an aggregate reservation is taken and how this decision is communicated between Aggregator and Deaggregator.

For simplicity and reliability, [RSVP-AGG] assigns the responsibility of the mapping decision entirely to the Deaggregator. In that mode, the Aggregator is notified of the selected mapping by the Deaggregator and follows this decision. The Deaggregator was chosen rather than the Aggregator because the Deaggregator is the first to have access to all the information required to make such a decision (in particular receipt of the E2E Resv which indicates the requested Int-Serv service type and includes information signaled by the receiver). This allows faster operations such as set-up or size adjustment of an Aggregate Reservation in a number of situations resulting in faster E2E reservation establishment.

This document retains this approach of assigning the responsibility of the mapping decision entirely to the Deaggregator. However, in one specific case, this document imposes one specific constraint on the mapping decision made by the Deaggregator:

- The specific case is where multiple security associations are used between the Aggregator and the Deaggregator.

- The specific constraint is that the aggregate reservation selected for the mapping of the E2E reservation must correspond to the Security Association that is to be used for transport of the E2E reservation traffic.

Note that, in IPsec environments, when multiple Security Associations are used from an IPsec router to another IPsec router, the selection of which security association is to be used for transport of traffic (or even the decision to use IPsec at all for a transport of a given flow) is normally controlled by the encryptor side (i.e. the Aggregator in our environment). However, the Deaggregator is made aware of which Security Association is selected by the Aggregator for a particular E2E flow because the E2E path message is received by the Deaggregator encrypted using that Security Association. The Deaggregator can store that information (i.e. which Security Association the E2E Path was received on) and take that into account when selecting the Aggregate reservation on which to map the E2E reservation in order to comply with the specific constraint stated above.

When IPsec is not used from the Aggregator to the Deaggregator, the Deaggregator has full responsibility of the mapping decision as per [RSVP-AGG] and this document does not impose any additional

constraints on that mapping.

4.3. Merging Rules

When using the extensions defined in this draft, RSVP sessions are defined by the 4-tuple: (DestAddress, protocol Id, vDstPort, DSCP). Similarly, a sender is defined by the tuple: (SrcAddress, GPI) when the protocol is AH or ESP, where the GPI field will be a four byte representation of a generalized source port, or by the tuple (SrcAddress) when the protocol is IP-in-IP or GRE . These extensions have some ramifications depending upon the reservation style.

We note that VDstPorts can be communicated by Deaggregators to Aggregators via the AGGREGATION-SESSION object included in the E2E PathErr. This can be used to facilitate various sharing scenarios as needed (e.g. the Deaggregator can convey the same VDstPort to different Aggregators which need to share a reservation; or conversely, the Deaggregator can communicate different VDstPorts to different Aggregators which need to have separate reservations). Policies followed by the Deaggregator to determine which aggregators need shared or separate reservations are beyond the scope of this document.

4.3.1. FF and SE Styles

In the FF and SE Styles, the FILTER_SPEC object contains the (SrcAddress) or the (SrcAddress, SPI) pair. When the SPI is used, this allows the receiver to uniquely identify senders based on both elements of the pair. When merging explicit sender descriptors, the senders may only be considered identical when both elements are identical.

4.3.2. WF Styles

As with [RSVP-IPSEC], WF style is not well supported with these extensions. Because there are no FILTER_SPEC objects for a WF reservation, any data packets with the session's destination IP address, protocol ID and DSCP will match the reservation (even if it is not carried inside the relevant IPsec tunnels because the SPI is not signaled and cannot be used for data classification). This limitation is considered acceptable because of the expectation that WF reservations are not likely to be used in this environment.

4.4. Handling SPI Value Changes

Changes in SPI values for a given IPsec tunnel will affect associated generic aggregate RSVP reservations. Changes will happen whenever that IPsec tunnel updates its Security Association. Such changes

will occur when a tunnel is re-keyed (i.e. to use a new key). Re-keying intervals are typically set based on traffic levels, key size, threat environment, and crypto algorithm in use. Implementations of this specification need to take the possibility of changes of SPI into account to ensure proper reservation behavior. This section discusses how generic aggregate RSVP reservations associated with IPsec tunnels can be adjusted in line with SPI value change.

When an SPI change occurs it will, in most cases, be necessary to update (send) the corresponding SENDER_TEMPLATEs and FILTER_SPECs.

The impact of sending new PATH and RESV messages corresponding to aggregate reservations will vary based on the reservation style being used. Builders of such applications may want to select reservation style based on interaction with SPI changes.

The least impact of an SPI change would be to WF style reservations. For such reservations, a new SENDER_TEMPLATE will need to be sent, but no new RESV is required. However, as mentioned earlier, WF reservations are not likely to be used in this environment. This is because of the fact that any data packets with the session's destination IP address, protocol ID and DSCP would match the reservation (regardless of any other field in the packet header such as source address or SPI).

For SE style reservations, both a new SENDER_TEMPLATE and a new RESV will need to be sent. This will result in changes to state, but should not affect data packet delivery or actual resource allocation in any way.

The FF style will be impacted the most. Like with SE, both PATH and RESV messages will need to be sent. But, since FF style reservations result in sender receiving its own resource allocation, resources will be allocated twice for a period of time. Or, even worse, there won't be enough resources to support the new flow without first freeing the old flow. To address this issue, it is recommended that applications that want FF style reservations (in other words that want separate reservations) actually use multiple SE reservations. Each Aggregator would have a separate SESSION definition thanks to a different VDstPort value. This is facilitated by the ability of the Deaggregator to distribute different VDstPorts to each Aggregator (through the AGGREGATION-SESSION object in the E2E PathErr as discussed above). When it came time to switch SPIs, a shared reservation could be made for the new SPI while the old SPI was still active. Once the new SPI was in use, the old reservation could be torn down. This will provide uninterrupted service over the aggregate reservations for IPsec tunnels.

In conclusion, it is possible to ensure uninterrupted QoS service (and avoid duplicate resource allocation) during SPI value change, both in the case where shared reservations across Aggregators are required and in the case where separate reservations per Aggregator are required.

5. Example Usages

5.1. Example Usage Of Generic Aggregate Reservations in Nested VPNs

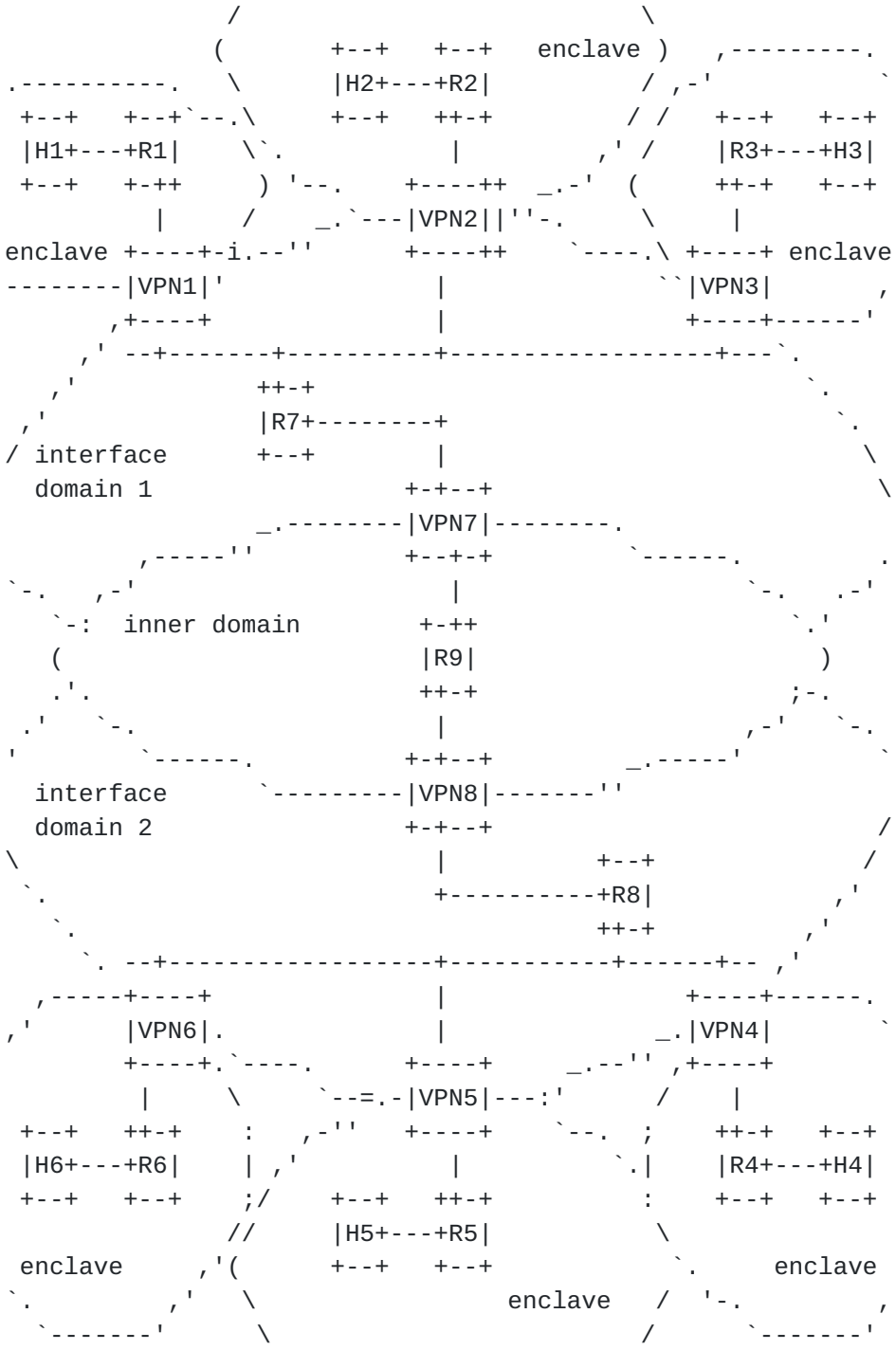


Figure 2: Reservations in a Nested VPN

For clarity we will only consider a subset of the traffic flows and will only consider:

- the flows from the VPN1 enclave to the VPN5 enclave (e.g. flows from Host H1 to Host H5)
- the flows from the VPN2 enclave to the VPN5 enclave (e.g. flows from Host H2 to Host H5)

Let us assume that:

- there is one security association between VPN1 and VPN5 (SPI1)
- there are two security associations between VPN2 and VPN5 (SPI2 and SPI3)
- the reservations from VPN1 enclave to VPN5 enclave have a preemption P1
- the reservations from VPN2 enclave to VPN5 have a preemption of either P1 or P2
- the reservations are either Voice (which needs to be treated in the aggregation region using the EF PHB) or Video (which needs to be treated in the aggregation region using the AF41 PHB)
- there is one security association between VPN7 and VPN8 (SPI4)
- that AH is used for IPsec tunneling

Then, the following generic aggregate RSVP reservations may be established from VPN1 to VPN5 for aggregation of the lower level RSVP reservations:

1. Reservation for aggregation of Voice reservations from VPN1 enclave to VPN5 enclave, requiring use of SPI1 and preemption P1:
    * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V1/EF
    * STYLE=FF or SE
    * IPv4/GPI FILTER_SPEC=VPN1/SPI1
    * POLICY_DATA (PREEMPTION_PRI)=P1

2. Reservation for aggregation of Video reservations from VPN1 enclave to VPN5 enclave, requiring use of SPI1 and preemption P1:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V2/AF41

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN1/SPI1

   * POLICY_DATA (PREEMPTION_PRI)=P1

where V1 and V2 are arbitrary VDstPort values picked by VPN5 within the range set aside for dynamic allocation (see section 6).

The following generic aggregate RSVP reservations may be established from VPN2 to VPN5 for aggregation of the lower level RSVP reservations:

1. Reservation for aggregation of Voice reservations from VPN2 enclave to VPN5 enclave, requiring use of SPI2 and preemption P1:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V3/EF

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN2/SPI2

   * POLICY_DATA (PREEMPTION_PRI)=P1

2. Reservation for aggregation of Video reservations from VPN2 enclave to VPN5 enclave, requiring use of SPI2 and preemption P1:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V4/AF41

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN2/SPI2

   * POLICY_DATA (PREEMPTION_PRI)=P1

3. Reservation for aggregation of Voice reservations from VPN2 enclave to VPN5 enclave, requiring use of SPI2 and preemption P2:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V5/EF

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN2/SPI2

   * POLICY_DATA (PREEMPTION_PRI)=P2

4. Reservation for aggregation of Video reservations from VPN2 enclave to VPN5 enclave, requiring use of SPI2 and preemption P2:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V6/AF41

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN2/SPI2

   * POLICY_DATA (PREEMPTION_PRI)=P2

5. Reservation for aggregation of Voice reservations from VPN2 enclave to VPN5 enclave, requiring use of SPI3 and preemption P1:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V7/EF

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN2/SPI3

   * POLICY_DATA (PREEMPTION_PRI)=P1

6. Reservation for aggregation of Video reservations from VPN2 enclave to VPN5 enclave, requiring use of SPI3 and preemption P1:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V8/AF41

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN2/SPI3

   * POLICY_DATA (PREEMPTION_PRI)=P1

7. Reservation for aggregation of Voice reservations from VPN2 enclave to VPN5 enclave, requiring use of SPI3 and preemption P2:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V9/EF

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN2/SPI3

   * POLICY_DATA (PREEMPTION_PRI)=P2

8. Reservation for aggregation of Video reservations from VPN2 enclave to VPN5 enclave, requiring use of SPI3 and preemption P2:

   * AGGREGATE-IPv4/GPI SESSION=VPN5/AH/V10/AF41

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN2/SPI3

   * POLICY_DATA (PREEMPTION_PRI)=P2

where V3 to V10 are arbitrary VDstPort values picked by VPN5 within the range set aside for dynamic allocation (see section 6).

The following generic aggregate RSVP reservations may be established from VPN7 to VPN8 for aggregation of the lower level RSVP reservations (i.e. the reservations from VPN1 to VPN5 and from VPN2 to VPN5):

1. Reservation for aggregation of Voice reservations from interface domain 1 to interface domain 2, requiring use of SPI4 and preemption P1:

   * AGGREGATE-IPv4/GPI SESSION=VPN8/AH/V1/EF

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN7/SPI4

   * POLICY_DATA (PREEMPTION_PRI)=P1

2. Reservation for aggregation of Video reservations from interface domain 1 to interface domain 2, requiring use of SPI4 and preemption P1:

   * AGGREGATE-IPv4/GPI SESSION=VPN8/AH/V2/AF41

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=VPN7/SPI4

   * POLICY_DATA (PREEMPTION_PRI)=P1

5.2. Example Usage Of Multiple Generic Aggregate Reservations Per DSCP From a Given Aggregator to a Given Deaggregator

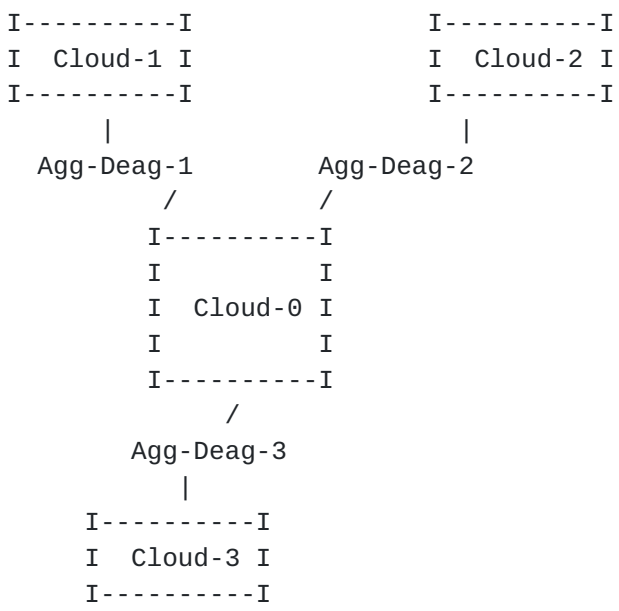


Figure 3: Example Scenario Requiring Multiple Generic Aggregate IP Reservations

Let us assume that:

- the E2E reservations from Cloud-1 to Cloud-3 have a preemption of either P1 or P2
- the E2E reservations from Cloud-2 to Cloud-3 have a preemption of either P1 or P2
- the E2E reservations are only for Voice (which needs to be treated in the aggregation region using the EF PHB)
- traffic from the E2E reservations is encapsulated in Aggregate IP reservations from Aggregator to Deaggregator using GRE

Then, the following generic aggregate RSVP reservations may be established from Agg-Deag-1 to Agg-Deag-3 for aggregation of the end-to-end RSVP reservations:

1. Reservation for aggregation of Voice reservations from Cloud-1 to Cloud-3 requiring use of P1:
    * AGGREGATE-IPv4/GPI SESSION=Agg-Deag-3/GRE/V1/EF
    * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=Agg-Deag-1/0

   * POLICY_DATA (PREEMPTION_PRI)=P1

2. Reservation for aggregation of Voice reservations from Cloud-1 to Cloud-3 requiring use of P2:

   * AGGREGATE-IPv4/GPI SESSION=Agg-Deag-3/GRE/V2/EF

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=Agg-Deag-1/0

   * POLICY_DATA (PREEMPTION_PRI)=P2

where V1 and V2 are arbitrary VDstPort values picked by Agg-Deag-3 within the range set aside for dynamic allocation (see section 6).

And the following generic aggregate RSVP reservations may be established from Agg-Deag-2 to Agg-Deag-3 for aggregation of the end-to-end RSVP reservations:

1. Reservation for aggregation of Voice reservations from Cloud-2 to Cloud-3 requiring use of P1:

   * AGGREGATE-IPv4/GPI SESSION=Agg-Deag-3/GRE/V3/EF

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=Agg-Deag-2/0

   * POLICY_DATA (PREEMPTION_PRI)=P1

2. Reservation for aggregation of Voice reservations from Cloud-2 to Cloud-3 requiring use of P2:

   * AGGREGATE-IPv4/GPI SESSION=Agg-Deag-3/GRE/V4/EF

   * STYLE=FF or SE

   * IPv4/GPI FILTER_SPEC=Agg-Deag-2/0

   * POLICY_DATA (PREEMPTION_PRI)=P2

where V3 and V4 are arbitrary VDstPort values picked by Agg-Deag-3 within the range set aside for dynamic allocation (see section 6).

6. IANA Considerations

This document requests that IANA allocates two new C-Types under the Class 1 for the two new RSVP objects defined in section 3.1.

This document requests that IANA allocates a new Class-Num and a new C-Type for the two new RSVP object defined in section 3.2.

This document defines in section 3.1 a "Virtual Destination Port (VDstPort)" field of 8 bits within the new Session objects defined in this document. The range of possible vDstPort values is broken down into sections, in a fashion similar to the VDstPort range of [RSVP-IPSEC] (but not identical since the VDstPort field of [RSVP-IPSEC] has 16 bits):

0	Illegal Value
1 - 63	Assigned by IANA
64 - 255	Dynamic

IANA is requested to create and maintain this new name space. The IANA guidelines for assignments for this field are as follows: o values in the range 1-63 are to be assigned according to the "XXX" policy defined in [IANA-CONS].

7. Security Considerations

The security considerations associated with the RSVP protocol [RSVP] apply to this document as it relies on RSVP.

This document assumes that meaningful DSCP values are visible between the Aggregator and the Deaggregator. This needs to be considered in high assurance IPsec environments. Note that this considerations would also apply if a pure Diffserv service was used between IPsec VPN-routers without RSVP reservations, for example as would be the case if the uniform model defined in [DS-TUNNEL] was used for Diffserv support over IPsec tunnels. We also note that in some environments (such as some wireless environments), visibility of the IP traffic (and hence of the DSCP) may be reduced through the use of link layer encryption.

Where IPsec is used, all the E2E RSVP signaling is encrypted inside the IPsec tunnel and thus is not visible in the encrypted cloud.

This document assumes that aggregate RSVP signaling is used between Aggregator and Deaggregator. The mechanisms defined in [RSVP-CRYPTO] and [RSVP-CRYPTO2] can be used to provide hop-by-hop integrity and authentication of RSVP messages related to the aggregate reservations discussed in this document.

This document assumes that, by default, the information signaled in RSVP for aggregate reservations is visible. This may include such information as the preemption priority of an aggregate reservation. This needs to be considered in high assurance IPsec environments. Where justified, RSVP signaling could be encrypted hop-by-hop (for example via hop-by-hop IPsec tunnels). Also, we note again that in some environments (such as some wireless environments), visibility of the IP traffic (and hence of RSVP) may be reduced through the use of link layer encryption.

IPsec common practices involve regular IPsec tunnel key updates. This document does not impact these practices as it allows uninterrupted QoS service (and avoids duplicate resource allocation) during SPI value change. Details of the corresponding procedures are provided in section 4.4.

8. Acknowledgments

This document borrows heavily from [RSVP-IPSEC] and [RSVP-AGG]. Also, we thank Fred Baker, Roger Levesque, Carol Iturralde, Daniel Voce and Anil Agarwal for their input into the content of this document. Thanks to Steve Kent for insightful comments on security aspects.

9. References

9.1. Normative References

[RSVP] "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", Braden et al, RFC2205

[RSVP-IPSEC] "RSVP Extensions for IPsec Data Flows", Berger et al, RFC2207

[RSVP-AGG] "Aggregation of RSVP for IPv4 and IPv6 Reservations", Baker et al, RFC3175

[SIG-NESTED] "QoS Signaling in a Nested Virtual Private Network", Baker et al, draft-baker-tsvwg-vpn-signaled-preemption-04.txt, work in progress

[RSVP-PROCESS] "Resource ReSerVation Protocol (RSVP) -- Version 1 Message Processing Rules", Braden et al, RFC2209

[IPSEC-ARCH] "Security Architecture for the Internet Protocol", Kent et al, RFC2401

[IPSEC-AH] "IP Authentication Header", Kent et al, RFC2402

[IPSEC-ESP] "IP Encapsulating Security Payload (ESP)", Kent et al, RFC2406

[RSVP-CRYPTO] "RSVP Cryptographic Authentication", Baker et al, RFC2747

[RSVP-CRYPTO2] "RSVP Cryptographic Authentication", Braden et al, RFC3097

[DS-TUNNEL] "Differentiated Services and Tunnels", Black, RFC2983

9.2. Informative References

[BW-REDUC] "A Resource Reservation Extension for the Reduction of Bandwidth of a Reservation Flow", Polk et al, draft-polk-tsvwg-rsvp-bw-reduction-01.txt, work in progress

[RSVP-TUNNEL] "RSVP Operation Over IP Tunnels", Terzis et al., RFC 2746, January 2000.

Authors' Addresses

Francois Le Faucheur
Cisco Systems
Greenside - 400 Avenue de Roumanille
Sophia Antipolis, 06410
France

Phone: +33-4-97-23-26-19
Fax: +33-4-97-23-26-26
Email: flefauch@cisco.com

Bruce Davie
Cisco Systems
1414 Massachusetts Ave.
Boxborough, MA 01719
USA

Phone:
Fax:
Email: bdavie@cisco.com

Pratik Bose
Lockheed Martin
22300 Comsat Drive Clarksburg, MD 20814 USA

Phone: +1 301 428 4215
Fax: +1 301 428 5147
Email: pratik.bose@lmco. com

Chris Christou
Booz Allen Hamilton
8283 Greensboro Drive
McLean, VA 22102,
USA

Phone:
Fax:
Email: christou_chris@bah.com

Mike Davenport
Booz Allen Hamilton
Suite 390
5220 Pacific Concourse Drive, Los Angeles, CA 90045
USA

Phone:
Fax:
Email: davenport_michael@bah.com

Intellectual Property Statement

Disclaimer of Validity

Copyright Statement

Acknowledgment

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