Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 27, 2010

B. Davie
F. le Faucheur
A. Narayanan
Cisco Systems, Inc.

November 23, 2009

Support for RSVP in Layer 3 VPNs
draft-ietf-tsvwg-rsvp-l3vpn-05.txt

Abstract

RFC 4364 and RFC 4659 define an approach to building provider-provisioned Layer 3 VPNs for IPv4 and IPv6. It may be desirable to use RSVP to perform admission control on the links between Customer Edge (CE) routers and Provider Edge (PE) routers. This document specifies procedures by which RSVP messages travelling from CE to CE across an L3VPN may be appropriately handled by PE routers so that admission control can be performed on PE-CE links. Optionally, admission control across the provider's backbone may also be supported.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of this Memo

This Internet-Draft will expire on May 27, 2010.

Copyright Notice

Table of Contents

1. Introduction

[RFC4364] and [RFC4659] define a Layer 3 VPN service known as BGP/MPLS VPNs for IPv4 and for IPv6 respectively. [RFC2205] defines the Resource Reservation Protocol (RSVP) which may be used to perform admission control as part of the Integrated Services (Int-Serv) architecture [RFC1633][RFC2210].

Customers of a layer 3 VPN service may run RSVP for the purposes of admission control (and associated resource reservation) in their own networks. Since the links between Provider Edge (PE) and Customer Edge (CE) routers in a layer 3 VPN may often be resource constrained, it may be desirable to be able to perform admission control over those links. In order to perform admission control using RSVP in such an environment, it is necessary that RSVP control messages, such as Path messages and Resv messages, are appropriately handled by the PE routers. This presents a number of challenges in the context of BGP/MPLS VPNs:

- RSVP Path message processing depends on routers recognizing the router alert option ([RFC2113], [RFC2711]) in the IP header. However, packets traversing the backbone of a BGP/MPLS VPN are MPLS encapsulated and thus the router alert option is not normally visible to the egress PE.

- BGP/MPLS VPNs support non-unique addressing of customer networks. Thus a PE at the ingress or egress of the provider backbone may be called upon to process Path messages from different customer VPNs with non-unique destination addresses.

- A PE at the ingress of the provider's backbone may receive Resv messages corresponding to different customer VPNs from other PEs, and needs to be able to associate those Resv messages with the appropriate customer VPNs.

This document describes a set of procedures to overcome these challenges and thus to enable admission control using RSVP over the PE-CE links. We note that similar techniques may be applicable to other protocols used for admission control such as the combination of the NSIS Signaling Layer Protocol (NSLP) for QoS Signaling ([I-D.ietf-nsis-qos-nslp]) and General Internet Signaling Transport (GIST) protocol ([I-D.ietf-nsis-ntlp]).

Additionally, it may be desirable to perform admission control over the provider's backbone on behalf of one or more L3VPN customers. Core (P) routers in a BGP/MPLS VPN do not have forwarding entries for customer routes, and thus cannot natively process RSVP messages for customer flows. Also the core is a shared resource that carries

traffic for many customers, so issues of resource allocation among customers and trust (or lack thereof) also ought to be addressed. This document specifies procedures for supporting such a scenario.

This document deals with establishing reservations for unicast flows only. Because the support of multicast traffic in BGP/MPLS VPNs is still evolving, and raises additional challenges for admission control, we leave the support of multicast flows for further study at this point.

1.1. Terminology

This document draws freely on the terminology defined in [RFC2205] and [RFC4364]. For convenience, we provide a few brief definitions here:

o CE (Customer Edge) Router: Router at the edge of a customer site that attaches to the network of the VPN provider.

o PE (Provider Edge) Router: Router at the edge of the service provider's network that attaches to one or more customer sites.

o VPN Label: An MPLS label associated with a route to a customer prefix in a VPN (also called a VPN route label).

o VRF: VPN Routing and Forwarding Table. A PE typically has multiple VRFs, enabling it to be connected to CEs that are in different VPNs.

2. Problem Statement

The problem space of this document is the support of admission control between customer sites when the customer subscribes to a BGP/ MPLS VPN. We subdivide the problem into (a) the problem of admission control on the PE-CE links (in both directions), and (b) the problem of admission control across the provider's backbone.

For the PE-CE link subproblem, the most basic challenge is that RSVP control messages contain IP addresses that are drawn from the customer's address space, and PEs need to deal with traffic from many customers who may have non-unique (or overlapping) address spaces. Thus, it is essential that a PE be able in all cases to identify the correct VPN context in which to process an RSVP control message. Much of this document deals with this issue.

For the case of making reservations across the provider backbone, we observe that BGP/MPLS VPNs do not create any per-customer forwarding

state in the P (provider core) routers. Thus, in order to make reservations on behalf of customer-specified flows, it is clearly necessary to make some sort of aggregated reservation from PE-PE and then map individual, customer-specific reservations onto an aggregate reservation. That is similar to the problem tackled in [RFC3175] and [RFC4804], with the additional complications of handling customer-specific addressing associated with BGP/MPLS VPNs.

Finally, we note that RSVP Path messages are normally addressed to the destination of a session, and contain the router alert option. Routers along the path to the destination that are configured to process RSVP messages need to detect the presence of the router alert option to allow them to intercept Path messages. However, the egress PEs of a network supporting BGP/MPLS VPNs receive packets destined for customer sites as MPLS-encapsulated packets, and possibly forwards those based only on examination of the MPLS label. Hence, a Path message would be forwarded without examination of the IP options and would therefore not receive appropriate processing at the PE. This problem of recognizing and processing Path messages is also discussed below.

Consider the case where an MPLS VPN customer uses RSVP signaling across his sites for resource reservation and admission control. Let's further assume that, initially, RSVP is not processed through the MPLS VPN cloud (i.e RSVP messages from the sender to the receiver travel transparently from CE to CE). In that case, RSVP allows establishment of resource reservations and admission control on a subset of the flow path (from sender to ingress CE; and from the RSVP router downstream of the egress CE to the receiver). If admission control is then activated on any of the CE-PE link, provider's backbone or PE-CE link (as allowed by the present document), the customer will benefit from an extended coverage of admission control and resource reservation: the resource reservation will now span over a bigger subset of (and possibly the whole) flow path, which in turn will increase the quality of service granted to the corresponding flow. Specific flows whose reservation is successful through admission control on the newly enabled segments will indeed benefit from this quality of service enhancement. However, it must be noted that, in case there is not enough resources on one (or more) of the newly enabled segments (e.g. Say admission control is enabled on a given PE-->CE link and there is not enough capacity on that link to admit all reservations for all the flows traversing that link) then some flows will not be able to maintain, or establish, their reservation. While this may appear undesirable for these flows, we observe that this only occurs if there is indeed a lack of capacity on a segment, and that in the absence of admission control all flows would be established but would all suffer from the resulting congestion on the bottleneck segment. We also observe that, in case

of such lack of capacity, admission control allows enforcement of controlled and flexible policies (so that, for example, more important flows can be granted higher priority at reserving resources). We note also that flows are given a chance to establish smaller reservations so that the aggregate load can adapt dynamically to the bottleneck capacity.

2.1. Model of Operation

Figure 1 illustrates the basic model of operation with which this document is concerned.

```
                      --------------------------
                     /       Provider           \
        |----|      |         Backbone           |      |----|
Sender->| CE1|  |-----|                       |-----|   |CE2 |->Receiver
        |    |--|     |   |---|     |---|     |     |---|    |
        |----|  |     |   | P |     | P |     |     |   |----|
                | PE1 |---|   |-----|   |-----| PE2 |
                |     |   |   |     |   |     |     |
                |     |   |---|     |---|     |     |
                |-----|                       |-----|
                    |                            |
                     \                          /
                      --------------------------
```

Figure 1. Model of Operation for RSVP-based admission control over MPLS/BGP VPN

To establish a unidirectional reservation for a point-to-point flow from Sender to Receiver that takes account of resource availability on the CE-PE and PE-CE links only, the following steps need to take place:

1. Sender sends a Path message to an IP address of the Receiver.

2. Path message is processed by CE1 using normal RSVP procedures and forwarded towards the Receiver along the link CE1-PE1.

3. PE1 processes Path message and forwards towards the Receiver across the provider backbone.

4. PE2 processes Path message and forwards towards the Receiver along link PE2-CE2.

5. CE2 processes Path message using normal RSVP procedures and forwards towards Receiver.

6. Receiver sends Resv message to CE2.

7. CE2 sends Resv message to PE2.

8. PE2 processes Resv message (including performing admission control on link PE2-CE2) and sends Resv to PE1.

9. PE1 processes Resv message and sends Resv to CE1.

10. CE1 processes Resv using normal RSVP procedures, performs admission control on the link CE1-PE1 and sends Resv message to Sender if successful.

In each of the steps involving Resv messages (6 through 10) the node sending the Resv uses the previously established Path state to determine the "RSVP Previous Hop (PHOP)" and sends a Resv message to that address. We note that establishing that Path state correctly at PEs is one of the challenges posed by the BGP/MPLS environment.

3. Admission Control on PE-CE Links

In the following sections we trace through the steps outlined in Section 2.1 and expand on the details for those steps where standard RSVP procedures need to be extended or modified to support the BGP/MPLS VPN environment. For all the remaining steps described in the preceding section, standard RSVP processing rules apply.

All the procedures described below support both IPv4 and IPv6 addressing. In all cases where IPv4 is referenced, IPv6 can be substituted with identical procedures and results. Object definitions for both IPv4 and IPv6 are provided in Section 8.

3.1. New Objects of Type VPN-IPv4

For RSVP signaling within a VPN, certain RSVP objects need to be extended. Since customer IP addresses need not be unique, the current types of SESSION, SENDER_TEMPLATE and FILTERSPEC objects are no longer sufficient to globally identify RSVP states in P/PE routers, since those are currently based on IP addresses. We propose new types of SESSION, SENDER_TEMPLATE and FILTERSPEC objects, which contain globally unique VPN-IPv4 format addresses. The ingress and egress PE nodes translate between the regular IPv4 addresses for messages to and from the CE, and VPN-IPv4 addresses for messages to and from PE routers. The rules for this translation are described in

later sections.

The RSVP_HOP object in a RSVP message currently specifies an IP address to be used by the neighboring RSVP hop to reply to the message sender. However, MPLS VPN PE routers (especially those separated by Option-B Autonomous System Border Routers -ASBRs) are not required to have direct IP reachability to each other. To solve this issue, we propose the use of label switching to forward RSVP messages between nodes within a MPLS VPN. This is achieved by defining a new VPN-IPv4 RSVP_HOP object. Use of the VPN-IPv4 RSVP_HOP object enables RSVP control plane reachability between any two adjacent RSVP hops in a MPLS VPN (e.g. a PE in AS 1 and a PE in AS2), regardless of whether they have IP reachability to each other.

The VPN-IPv4 RSVP_HOP object carries the IPv4 address of the message sender and a Logical Interface Handle (LIH) as before, but in addition carries a VPN-IPv4 address which also represents the sender of the message. The message sender MUST also advertise this VPN-IPv4 address into BGP, associated with a locally allocated label, and this advertisement MUST be propagated by BGP throughout the VPN and to adjacent ASes if required to provide reachability to this PE. Frames received by the PE marked with this label MUST be given to the local control plane for processing. When a neighboring RSVP hop wishes to reply to a message carrying a VPN-IPv4 RSVP_HOP, it looks for a BGP advertisement of the VPN-IPv4 address contained in that RSVP_HOP. If this address is found and carries an associated label, the neighboring RSVP node MUST encapsulate the RSVP message with this label and send it via MPLS encapsulation to the BGP next-hop associated with the route. The destination IP address of the message is taken from the IP address field of the RSVP_HOP object, as described in [RFC2205]. Additionally, the IPv4 address in the RSVP_HOP object continues to be used for all other existing purposes, including neighbor matching between Path/Resv and SRefresh messages ([RFC2961]), authentication ([RFC2747]), etc.

The VPN-IPv4 address used in the VPN-IPv4 RSVP_HOP object MAY represent an existing address in the VRF that corresponds to the flow (e.g. a local loopback or PE-CE link address within the VRF for this customer), or MAY be created specially for this purpose. In the case where the address is specially created for RSVP signaling (and possibly other control protocols), the BGP advertisement MUST NOT be redistributed to, or reachable by, any CEs outside the MPLS VPN. One way to achieve this is by creating a special "control protocols VPN" with VRF state on every PE/ASBR, carrying route targets not imported into customer VRFs. In the case where a customer VRF address is used as the VPN-IPv4 address, a VPN-IPv4 address in one customer VRF MUST NOT be used to signal RSVP messages for a flow in a different VRF.

If a PE/ASBR is sending a Path message to another PE/ASBR within the VPN, and it has any appropriate VPN-IPv4 address for signaling that satisfies the requirements outlined above, it MUST use a VPN-IPv4 RSVP_HOP object with this address for all RSVP messages within the VPN. If a PE/ASBR does not have any appropriate VPN-IPv4 address to use for signaling, it MAY send the Path message with a regular IPv4 RSVP_HOP object. In this case, the reply will be IP encapsulated. This option is not preferred because there is no guarantee that the neighboring RSVP hop has IP reachability to the sending node. If a PE/ASBR receives or originates a Path message with a VPN-IPv4 RSVP_HOP object, any RSVP_HOP object in corresponding upstream messages for this flow (e.g. Resv, ResvTear) or downstream messages (e.g. ResvError, PathTear) sent by this node within the VPN MUST be a VPN-IPv4 RSVP_HOP.

3.2. Path Message Processing at Ingress PE

When a Path message arrives at the ingress PE (step 3 of Section 2.1) the PE needs to establish suitable Path state and forward the Path message on to the egress PE. In the following paragraphs we described the steps taken by the ingress PE.

The Path message is addressed to the eventual destination (the receiver at the remote customer site) and carries the IP router alert option, in accordance with [RFC2205]. The ingress PE MUST recognize the router alert option, intercept these messages and process them as RSVP signaling messages.

As noted above, there is an issue in recognizing Path messages as they arrive at the egress PE (PE 2 in Figure 1). The approach defined here is to address the Path messages sent by the ingress PE directly to the egress PE, and send it without IP router alert option; that is, rather than using the ultimate receiver's destination address as the destination address of the Path message, we use the loopback address of the egress PE as the destination address of the Path message. This approach has the advantage that it does not require any new data plane capabilities for the egress PE beyond those of a standard BGP/MPLS VPN PE. Details of the processing of this message at the egress PE are described below in Section 3.3. The approach of addressing a Path message directly to an RSVP next hop (that may or may not be the next IP hop) is already used in other environments such as those of [RFC4206] and [RFC4804].

The details of operation at the ingress PE are as follows. When the ingress PE (PE1 in Figure 1) receives a Path message from CE1 that is addressed to the receiver, the VRF that is associated with the incoming interface is identified, just as for normal data path operations. The Path state for the session is stored, and is

associated with that VRF, so that potentially overlapping addresses among different VPNs do not appear to belong to the same session. The destination address of the receiver is looked up in the appropriate VRF, and the BGP Next-Hop for that destination is identified. That next-hop is the egress PE (PE2 in Figure 1). A new VPN-IPv4 SESSION object is constructed, containing the Route Distinguisher (RD) that is part of the VPN-IPv4 route prefix for this destination, and the IPv4 address from the SESSION. In addition, a new VPN-IPv4 SENDER_TEMPLATE object is constructed, with the original IPv4 address from the incoming SENDER_TEMPLATE plus the RD that is used by this PE to advertise that prefix for this customer into the VPN. A new Path message is constructed with a destination address equal to the address of the egress PE identified above. This new Path message will contain all the objects from the original Path message, replacing the original SESSION and SENDER_TEMPLATE objects with the new VPN-IPv4 type objects. The Path message is sent without router alert option and contains a RSVP_HOP object constructed as specified in Section 3.1.

3.3. Path Message Processing at Egress PE

When a Path message arrives at the egress PE, it is addressed to the PE itself, and is handed to RSVP for processing. The router extracts the RD and IPv4 address from the VPN-IPv4 SESSION object, and determines the local VRF context by finding a matching VPN-IPv4 prefix with the specified RD that has been advertised by this router into BGP. The entire incoming RSVP message, including the VRF information, is stored as part of the Path state.

Now the RSVP module can construct a Path message which differs from the Path it received in the following ways:

a. Its destination address is the IP address extracted from the SESSION Object;

b. The SESSION and SENDER_TEMPLATE objects are converted back to IPv4-type by discarding the attached RD

c. The RSVP_HOP Object contains the IP address of the outgoing interface of the egress PE and a Logical Interface Handle (LIH), as per normal RSVP processing.

The router then sends the Path message on towards its destination over the interface identified above. This Path message carries the router alert option as required by [RFC2205].

3.4. Resv Processing at Egress PE

When a receiver at the customer site originates a Resv message for the session, normal RSVP procedures apply until the Resv, making its way back towards the sender, arrives at the "egress" PE (it is "egress" with respect to the direction of data flow, i.e. PE2 in figure 1). On arriving at PE2, the SESSION and FILTER_SPEC objects in the Resv, and the VRF in which the Resv was received, are used to find the matching Path state stored previously. At this stage, admission control can be performed on the PE-CE link.

Assuming admission control is successful, the PE constructs a Resv message to send to the RSVP HOP stored in the Path state, i.e., the ingress PE (PE1 in Figure 1). The IPv4 SESSION object is replaced with the same VPN-IPv4 SESSION object received in the Path. The IPv4 FILTER_SPEC object is replaced with a VPN-IPv4 FILTER_SPEC object, which copies the VPN-IPv4 address from the SENDER_TEMPLATE received in the matching Path message. The RSVP_HOP in the Resv message MUST be constructed as specified in Section 3.1. The Resv message MUST be addressed to the IP address contained within the RSVP_HOP object in the Path message. If the Path message contained a VPN-IPv4 RSVP_HOP object, the Resv MUST be MPLS-encapsulated using the label associated with that VPN-IPv4 address in BGP, as described in Section 3.1. If the Path message contained an IPv4 RSVP_HOP object, the Resv is simply IP-encapsulated and addressed directly to the IP address in the RSVP_HOP object.

If admission control is not successful on the egress PE, a ResvError message is sent towards the receiver as per normal RSVP processing.

3.5. Resv Processing at Ingress PE

Upon receiving a Resv message at the ingress PE (with respect to data flow, i.e. PE1 in Figure 1), the PE determines the local VRF context and associated Path state for this Resv by decoding the received SESSION and FILTER_SPEC objects. It is now possible to generate a Resv message to send to the appropriate CE. The Resv message sent to the ingress CE will contain IPv4 SESSION and FILTER_SPEC objects, derived from the appropriate Path state. Since we assume in this section that admission control over the Provider's backbone is not needed, the ingress PE does not perform any admission control for this reservation.

3.6. Other RSVP Messages

Processing of PathError, PathTear, ResvError, ResvTear and ResvConf messages is generally straightforward and follows the rules of [RFC2205]. These additional rules MUST be observed for messages

transmitted within the VPN (i.e. Between the PEs):

- The SESSION, SENDER_TEMPLATE and FILTER_SPEC objects MUST be converted from IPv4 to VPN-IPv4 form and back in the same manner as described above for Path and Resv messages.

- The appropriate type of RSVP_HOP object (VPN-IPv4 or IPv4) MUST be used as described above.

- Depending on the type of RSVP_HOP object received from the neighbor, the message MUST be MPLS-encapsulated or IP-encapsulated as described above.

- The matching state & VRF MUST be determined by decoding the RD and IPv4 addresses in the SESSION and FILTER_SPEC objects.

- The message MUST be directly addressed to the appropriate PE, without using the router alert option.

4. Admission Control in Provider's Backbone

The preceding section outlines how per-customer reservations can be made over the PE-CE links. This may be sufficient in many situations where the backbone is well engineered with ample capacity and there is no need to perform any sort of admission control in the backbone. However, in some cases where excess capacity cannot be relied upon (e.g., during failures or unanticipated periods of overload) it may be desirable to be able to perform admission control in the backbone on behalf of customer traffic.

Because of the fact that routes to customer addresses are not present in the P routers, along with the concerns of scalability that would arise if per-customer reservations were allowed in the P routers, it is clearly necessary to map the per-customer reservations described in the preceding section onto some sort of aggregate reservations. Furthermore, customer data packets need to be tunneled across the provider backbone just as in normal BGP/MPLS VPN operation.

Given these considerations, a feasible way to achieve the objective of admission control in the backbone is to use the ideas described in [RFC4804]. MPLS-TE tunnels can be established between PEs as a means to perform aggregate admission control in the backbone.

An MPLS-TE tunnel from an ingress PE to an egress PE can be thought of as a virtual link of a certain capacity. The main change to the procedures described above is that when a Resv is received at the ingress PE, an admission control decision can be performed by

checking whether sufficient capacity of that virtual link remains available to admit the new customer reservation. We note also that [RFC4804] uses the IF_ID RSVP_HOP object to identify the tunnel across the backbone, rather than the simple RSVP_HOP object described in Section 3.2. The procedures of [RFC4804] should be followed here as well.

To achieve effective admission control in the backbone, there needs to be some way to separate the data plane traffic that has a reservation from that which does not. We assume that packets that are subject to admission control on the core will be given a particular MPLS EXP value, and that no other packets will be allowed to enter the core with this value unless they have passed admission control. Some fraction of link resources will be allocated to queues on core links for packets bearing that EXP value, and the MPLS-TE tunnels will use that resource pool to make their constraint-based routing and admission control decisions. This is all consistent with the principles of aggregate RSVP reservations described in [RFC3175].

5. Inter-AS operation

[RFC4364] defines three modes of inter-AS operation for MPLS/BGP VPNs, referred to as options A, B and C. In the following sections we describe how the scheme described above can operate in each inter-AS environment.

5.1. Inter-AS Option A

Operation of RSVP in Inter-AS Option A is quite straightforward. Each ASBR operates like a PE, and the ASBR-ASBR links can be viewed as PE-CE links in terms of admission control. If the procedures defined in Section 3 are enabled on both ASBRs, then admission control may be performed on the inter-ASBR links. In addition, the operator of each AS can independently decide whether or not to perform admission control across his backbone. The new objects described in this document MUST NOT be sent in any RSVP message between two Option-A ASBRs.

5.2. Inter-AS Option B

To support inter-AS Option B, we require some additional processing of RSVP messages on the ASBRs. Recall that, when packets are forwarded from one AS to another in option B, the VPN label is swapped by each ASBR as a packet goes from one AS to another. The BGP next hop seen by the ingress PE will be the ASBR, and there need not be IP visibility between the ingress and egress PEs. Hence when the ingress PE sends the Path message to the BGP next hop of the VPN-

IPv4 route towards the destination, it will be received by the ASBR. The ASBR determines the next hop of the route in a similar way as the ingress PE - by finding a matching BGP VPN-IPv4 route with the same RD and a matching prefix.

The provider(s) who interconnect ASes using option B may or may not desire to perform admission control on the inter-AS links. This choice affects the detailed operation of ASBRs. We describe the two modes of operation - with and without admission control at the ASBRs - in the following sections.

5.2.1. Admission control on ASBR

In this scenario, the ASBR performs full RSVP signaling and admission control. The RSVP database is indexed on the ASBR using the VPN-IPv4 SESSION, SENDER_TEMPLATE and FILTER_SPEC objects (which uniquely identify RSVP sessions and flows as per the requirements of [RFC2205]). These objects are forwarded unmodified in both directions by the ASBR. All other procedures of RSVP are performed as if the ASBR was a RSVP hop. In particular, the RSVP_HOP objects sent in Path and Resv messages contain IP addresses of the ASBR, which MUST be reachable by the neighbor to whom the message is being sent. Note that since the VPN-IPv4 SESSION, SENDER_TEMPLATE and FILTER_SPEC objects satisfy the uniqueness properties required for a RSVP database implementation as per [RFC2209], no customer VRF awareness is required on the ASBR.

5.2.2. No admission control on ASBR

If the ASBR is not doing admission control, it is desirable that per-flow state not be maintained on the ASBR. This requires adjacent RSVP hops (i.e. The ingress and egress PEs of the respective ASes) to send RSVP messages directly between them. This is only possible if they are MPLS-encapsulated. The use of the VPN-IPv4 RSVP_HOP object described in Section 3.1 is REQUIRED in this case.

When an ASBR that is not installing local RSVP state receives a Path message, it looks up the next-hop of the matching BGP route as described in Section 3.2, and sends the Path message to the next-hop, without modifying any RSVP objects (including the RSVP_HOP). This process is repeated at subsequent ASBRs until the Path message arrives at a router that is installing local RSVP state (either the ultimate egress PE, or an ASBR configured to perform admission control). This router receives the Path and processes it as described in Section 3.3 if it is a PE, or Section 5.2.1 if it is an ASBR performing admission control. When this router sends the Resv upstream, it looks up the routing table for a next-hop+label for the VPN-IPv4 address in the PHOP, encapsulates the Resv with that label

and sends it upstream. This message will be received for control processing directly on the upstream RSVP hop (that last updated the RSVP_HOP field in the Path message), without any involvement of intermediate ASBRs.

The ASBR is not expected to process any other RSVP messages apart from the Path message as described above. The ASBR also does not need to store any RSVP state. Note that any ASBR along the path that wishes to do admission control or insert itself into the RSVP signaling flow, may do so by writing its own RSVP_HOP object with IPv4 and VPN-IPv4 address pointing to itself.

If an Option-B ASBR receives a RSVP Path message with an IPv4 RSVP_HOP, does not wish to perform admission control but is willing to install local state for this flow, the ASBR MUST process and forward RSVP signaling messages for this flow as described in Section 5.2.1 (with the exception that it does not perform admission control). If an Option-B ASBR receives a RSVP Path message with an IPv4 RSVP_HOP, but does not wish to install local state or perform admission control for this flow, the ASBR MUST NOT forward the Path message. In addition, the ASBR SHOULD send a PathError message of Error Code "RSVP over MPLS Problem" and Error Value "RSVP_HOP not reachable across VPN" (see Section 9) signifying to the upstream RSVP hop that the supplied RSVP_HOP object is insufficient to provide reachability across this VPN. This failure condition is not expected to be recoverable.

5.3. Inter-AS Option C

Operation of RSVP in Inter-AS Option C is also quite straightforward, because there exists an LSP directly from ingress PE to egress PE. In this case, there is no significant difference in operation from the single AS case described in Section 3. Furthermore, if it is desired to provide admission control from PE to PE, it can be done by building an inter-AS TE tunnel and then using the procedures described in Section 4.

6. Operation with RSVP disabled

It is often the case that RSVP will not be enabled on the PE-CE links. In such an environment, a customer may reasonably expect that RSVP messages sent into the L3 VPN network should be forwarded just like any other IP datagrams. This transparency is useful when the customer wishes to use RSVP within his own sites or perhaps to perform admission control on the CE-PE links (in CE->PE direction only), without involvement of the PEs. For this reason, a PE SHOULD NOT discard or modify RSVP messages sent towards it from a CE when

RSVP is not enabled on the PE-CE links. Similarly a PE SHOULD NOT discard or modify RSVP messages which are destined for one of its attached CEs, even when RSVP is not enabled on those links. Note that the presence of the router alert option in some RSVP messages may cause them to be forwarded outside of the normal forwarding path, but that the guidance of this paragraph still applies in that case. Note also that this guidance applies regardless of whether RSVP-TE is used in some, all, or none of the L3VPN network.

7. Other RSVP procedures

This section describes modifications to other RSVP procedures introduced by MPLS VPNs

7.1. Refresh overhead reduction

The following points ought to be noted regarding RSVP refresh overhead reduction ([RFC2961]) across a MPLS VPN:

- The hop between the ingress and egress PE of a VPN is to be considered as traversing one or more non-RSVP hops. As such, the procedures described in Section 5.3 of [RFC2961] relating to non-RSVP hops SHOULD be followed.

- The source IP address of a SRefresh message MUST match the IPv4 address signalled in the RSVP_HOP object contained in the corresponding Path or Resv message. The IPv4 address in any received VPN-IPv4 RSVP_HOP object MUST be used as the source address of that message for this purpose.

7.2. Cryptographic Authentication

The following points ought to be noted regarding RSVP cryptographic authentication ([RFC2747]) across a MPLS VPN:

- The IPv4 address in any received VPN-IPv4 RSVP_HOP object MUST be used as the source address of that message for purposes of identifying the security association.

- Forwarding of Challenge and Response messages MUST follow the same rules as described above for hop-by-hop messages. Specifically, if the originator of a Challenge/Response message has received a VPN-IPv4 RSVP_HOP object from the corresponding neighbor, it MUST use the label associated with that VPN-IPv4 address in BGP to forward the Challenge/Response message.

7.3. RSVP Aggregation

[RFC3175] and [RFC4860] describe mechanisms to aggregate multiple individual RSVP reservations into a single larger reservation on the basis of a common DSCP/PHB for traffic classification. The following points ought to be noted in this regard:

o The procedures described in this section apply only in the case where the Aggregator and Deaggregator nodes are C/CE devices, and the entire MPLS VPN lies within the Aggregation Region. The case where the PE is also an Aggregator/Deaggregator is more complex and not considered in this document.

o Support of Aggregate RSVP sessions is OPTIONAL. When supported:

  * Aggregate RSVP sessions MUST be treated in the same way as regular IPv4 RSVP sessions. To this end, all the procedures described in Section 3 and Section 4 MUST be followed for aggregate RSVP sessions. The corresponding new SESSION, SENDER_TEMPLATE and FILTERSPEC objects are defined in Section 8.

  * End-To-End (E2E) RSVP sessions are passed unmodified through the MPLS VPN. These RSVP messages SHOULD be identified by their IP protocol (RSVP-E2E-IGNORE, 134). When the ingress PE receives any RSVP message with this IP protocol, it MUST process this frame as if it is regular customer traffic and ignore any router alert option. The appropriate VPN and transport labels are applied to the frame and it is forwarded towards the remote CE. Note that this message will not be received or processed by any other P or PE node.

  * Any SESSION-OF-INTEREST object (defined in [RFC4860]) MUST be conveyed unmodified across the MPLS VPN.

7.4. Support for CE-CE RSVP-TE

[I-D.ietf-l3vpn-e2e-rsvp-te-reqts] describes a set of requirements for the establishment for CE-CE MPLS LSPs across networks offering an L3VPN service. The requirements specified in that document are similar to those addressed by this document, in that both address the issue of handling RSVP requests from customers in a VPN context. It is possible that the solution described here could be adapted to meet the requirements of [I-D.ietf-l3vpn-e2e-rsvp-te-reqts]. To the extent that this document uses signaling extensions described in [RFC3473] which have already been used for GMPLS/TE, we expect that CE-CE RSVP/TE will be incremental work built on these extensions. These extensions will be considered in a separate document.

8. Object Definitions

8.1. VPN-IPv4 and VPN-IPv6 SESSION objects

The usage of the VPN-IPv4 (or VPN-IPv6) SESSION Object is described in Section 3.2 to Section 3.6. The VPN-IPv4 (or VPN-IPv6) SESSION object appears in RSVP messages that ordinarily contain a SESSION object and are sent between ingress PE and egress PE in either direction. The object MUST NOT be included in any RSVP messages that are sent outside of the provider's backbone (except in the inter-AS option B and C cases, as described above, when it may appear on inter-AS links).

The VPN-IPv6 SESSION object is analogous to the VPN-IPv4 SESSION object, using an VPN-IPv6 address ([RFC4659]) instead of an VPN-IPv4 address ([RFC4364]).

The formats of the objects are as follows:

o VPN-IPv4 SESSION object: Class = 1, C-Type = TBA

```
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|             VPN-IPv4 DestAddress (12 bytes)           |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
| Protocol Id |    Flags    |          DstPort          |
+-------------+-------------+-------------+-------------+
```

o   VPN-IPv6 SESSION object: Class = 1, C-Type = TBA

```
     +-------------+-------------+-------------+-------------+
     |                                                       |
     +                                                       +
     |                                                       |
     +             VPN-IPv6 DestAddress (24 bytes)           +
     /                                                       /
     .                                                       .
     /                                                       /
     |                                                       |
     +-------------+-------------+-------------+-------------+
     | Protocol Id |     Flags   |          DstPort          |
     +-------------+-------------+-------------+-------------+
```

The VPN-IPv4 DestAddress (respectively VPN-IPv6 DestAddress) field contains an address of the VPN-IPv4 (respectively VPN-IPv6) address family encoded as specified in [RFC4364] (respectively [RFC4659]). The content of this field is discussed in Section 3.2 and Section 3.3.

The protocol ID, flags, and DstPort are identical to the same fields in the IPv4 and IPv6 SESSION objects ([RFC2205]).

8.2. VPN-IPv4 and VPN-IPv6 SENDER_TEMPLATE objects

The usage of the VPN-IPv4 (or VPN-IPv6) SENDER_TEMPLATE Object is described in Section 3.2 and Section 3.3. The VPN-IPv4 (or VPN-IPv6) SENDER_TEMPLATE object appears in RSVP messages that ordinarily contain a SENDER_TEMPLATE object and are sent between ingress PE and egress PE in either direction (such as Path, PathError, and PathTear). The object MUST NOT be included in any RSVP messages that are sent outside of the provider's backbone (except in the inter-AS option B and C cases, as described above, when it may appear on inter-AS links). The format of the object is as follows:

```
         o    VPN-IPv4 SENDER_TEMPLATE object: Class = 11, C-Type = TBA

              +-------------+-------------+-------------+-------------+
              |                                                       |
              +                                                       +
              |             VPN-IPv4 SrcAddress (12 bytes)            |
              +                                                       +
              |                                                       |
              +-------------+-------------+-------------+-------------+
              |          Reserved           |          SrcPort          |
              +-------------+-------------+-------------+-------------+


         o    VPN-IPv6 SENDER_TEMPLATE object: Class = 11, C-Type = TBA

              +-------------+-------------+-------------+-------------+
              |                                                       |
              +                                                       +
              |                                                       |
              +             VPN-IPv6 SrcAddress (24 bytes)            +
              /                                                       /
              .                                                       .
              /                                                       /
              |                                                       |
              +-------------+-------------+-------------+-------------+
              |          Reserved           |          SrcPort          |
              +-------------+-------------+-------------+-------------+
```

The VPN-IPv4 SrcAddress (respectively VPN-IPv6 SrcAddress) field contains an address of the VPN-IPv4 (respectively VPN-IPv6) address family encoded as specified in [RFC4364] (respectively [RFC4659]). The content of this field is discussed in Section 3.2 and Section 3.3.

The SrcPort is identical to the SrcPort field in the IPv4 and IPv6 SENDER_TEMPLATE objects ([RFC2205]).

The Reserved field MUST be set to zero on transmit and ignored on receipt.

8.3. VPN-IPv4 and VPN-IPv6 FILTER_SPEC objects

The usage of the VPN-IPv4 (or VPN-IPv6) FILTER_SPEC Object is described in Section 3.4 and Section 3.5. The VPN-IPv4 (or VPN-IPv6) FILTER_SPEC object appears in RSVP messages that ordinarily contain a FILTER_SPEC object and are sent between ingress PE and egress PE in

either direction (such as Resv, ResvError, and ResvTear). The object MUST NOT be included in any RSVP messages that are sent outside of the provider's backbone (except in the inter-AS option B and C cases, as described above, when it may appear on inter-AS links).

- VPN-IPv4 FILTER_SPEC object: Class = 10, C-Type = TBA

  Definition same as VPN-IPv4 SENDER_TEMPLATE object.

- VPN-IPv6 FILTER_SPEC object: Class = 10, C-Type = TBA

  Definition same as VPN-IPv6 SENDER_TEMPLATE object.

The content of the VPN-IPv4 SrcAddress (or VPN-IPv6 SrcAddress) field is discussed in Section 3.4 and Section 3.5.

The SrcPort is identical to the SrcPort field in the IPv4 and IPv6 SENDER_TEMPLATE objects ([RFC2205]).

The Reserved field MUST be set to zero on transmit and ignored on receipt.

8.4. VPN-IPv4 and VPN-IPv6 RSVP_HOP objects

Usage of the VPN-IPv4 (or VPN-IPv6) RSVP_HOP Object is described in Section 3.1 and Section 5.2.2. The VPN-IPv4 (VPN-IPv6) RSVP_HOP object is used to establish signaling reachability between RSVP neighbors separated by one or more Option-B ASBRs. This object may appear in RSVP messages that carry a RSVP_HOP object, and that travel between the Ingress and Egress PEs. It MUST NOT be included in any RSVP messages that are sent outside of the provider's backbone (except in the inter-AS option B and C cases, as described above, when it may appear on inter-AS links). The format of the object is as follows:

- VPN-IPv4 RSVP_HOP object: Class = 3, C-Type = TBA

```
+-------------+-------------+-------------+-------------+
|       IPv4 Next/Previous Hop Address (4 bytes)        |
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|    VPN-IPv4 Next/Previous Hop Address (12 bytes)      |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|                 Logical Interface Handle              |
+-------------+-------------+-------------+-------------+
```

- VPN-IPv6 RSVP_HOP object: Class = 3, C-Type = TBA

```
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+       IPv6 Next/Previous Hop Address (16 bytes)       +
|                                                       |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+     VPN-IPv6 Next/Previous Hop Address (24 bytes)     +
/                                                       /
.                                                       .
/                                                       /
|                                                       |
+-------------+-------------+-------------+-------------+
|                Logical Interface Handle               |
+-------------+-------------+-------------+-------------+
```

The IPv4 Next/Previous Hop Address, IPv6 Next/Previous Hop Address and the Logical Interface Handle fields are identical to those of the RSVP_HOP object ([RFC2205]).

The VPN-IPv4 Next/Previous Hop Address (respectively VPN-IPv6 Next/Previous Hop Address) field contains an address of the VPN-IPv4 (respectively VPN-IPv6) address family encoded as specified in [RFC4364] (respectively [RFC4659]). The content of this field is discussed in Section 3.1.

8.5. Aggregated VPN-IPv4 and VPN-IPv6 SESSION objects

The usage of Aggregated VPN-IPv4 (or VPN-IPv6) SESSION object is described in Section 7.3. The AGGREGATE-VPN-IPv4 (respectively AGGREGATE-IPv6-VPN) SESSION object appears in RSVP messages that ordinarily contain a AGGREGATE-IPv4 (respectively AGGREGATE-IPv6) SESSION object as defined in [RFC3175] and are sent between ingress PE and egress PE in either direction. The GENERIC-AGGREGATE-VPN-IPv4 (respectively AGGREGATE-VPN-IPv6) SESSION object should appear in all RSVP messages that ordinarily contain a GENERIC-AGGREGATE-IPv4 (respectively GENERIC-AGGREGATE-IPv6) SESSION object as defined in [RFC4860] and are sent between ingress PE and egress PE in either direction. These objects MUST NOT be included in any RSVP messages that are sent outside of the provider's backbone (except in the inter-AS option B and C cases, as described above, when it may appear on inter-AS links). The processing rules for these objects are otherwise identical to those of the VPN-IPv4 (respectively VPN-IPv6) SESSION object defined in Section 8.1. The format of the object is as follows:

- AGGREGATE-VPN-IPv4 SESSION object: Class = 1, C-Type = TBA

```
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|             VPN-IPv4 DestAddress (12 bytes)           |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|  Reserved   |    Flags    |  Reserved   |    DSCP     |
+-------------+-------------+-------------+-------------+
```

o AGGREGATE-VPN-IPv6 SESSION object: Class = 1, C-Type = TBA

```
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+             VPN-IPv6 DestAddress (24 bytes)           +
/                                                       /
.                                                       .
/                                                       /
|                                                       |
+-------------+-------------+-------------+-------------+
|   Reserved  |    Flags    |   Reserved  |     DSCP    |
+-------------+-------------+-------------+-------------+
```

The VPN-IPv4 DestAddress (respectively VPN-IPv6 DestAddress) field contains an address of the VPN-IPv4 (respectively VPN-IPv6) address family encoded as specified in [RFC4364] (respectively [RFC4659]). The content of this field is discussed in Section 3.2 and Section 3.3.

The flags and DSCP are identical to the same fields of the AGGREGATE-IPv4 and AGGREGATE-IPv6 SESSION objects ([RFC3175]).

The Reserved field MUST be set to zero on transmit and ignored on receipt.

o GENERIC-AGGREGATE-VPN-IPv4 SESSION object:
  Class = 1, C-Type = TBA

```
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|             VPN-IPv4 DestAddress (12 bytes)           |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|  Reserved   |    Flags    |           PHB-ID          |
+-------------+-------------+-------------+-------------+
|          Reserved         |          vDstPort         |
+-------------+-------------+-------------+-------------+
|                    Extended vDstPort                  |
+-------------+-------------+-------------+-------------+
```

```
      o    GENERIC-AGGREGATE-VPN-IPv6 SESSION object:
             Class = 1, C-Type = TBA

           +-------------+-------------+-------------+-------------+
           |                                                       |
           +                                                       +
           |                                                       |
           +             VPN-IPv6 DestAddress (24 bytes)           +
           /                                                       /
           .                                                       .
           /                                                       /
           |                                                       |
           +-------------+-------------+-------------+-------------+
           |  Reserved   |    Flags    |           PHB-ID          |
           +-------------+-------------+-------------+-------------+
           |          Reserved         |          vDstPort         |
           +-------------+-------------+-------------+-------------+
           |                    Extended vDstPort                  |
           +-------------+-------------+-------------+-------------+
```

The VPN-IPv4 DestAddress (respectively VPN-IPv6 DestAddress) field contains an address of the VPN-IPv4 (respectively VPN-IPv6) address family encoded as specified in [RFC4364] (respectively [RFC4659]). The content of this field is discussed in Section 3.2 and Section 3.3.

The flags, PHB-ID, vDstPort and Extended vDstPort are identical to the same fields of the GENERIC-AGGREGATE-IPv4 and GENERIC-AGGREGATE-IPv6 SESSION objects ([RFC4860]).

The Reserved field MUST be set to zero on transmit and ignored on receipt.

8.6. AGGREGATE-VPN-IPv4 and AGGREGATE-VPN-IPv6 SENDER_TEMPLATE objects

The usage of Aggregated VPN-IPv4 (or VPN-IPv6) SENDER_TEMPLATE object is described in Section 7.3. The AGGREGATE-VPN-IPv4 (respectively AGGREGATE-VPN-IPv6) SENDER_TEMPLATE object appears in RSVP messages that ordinarily contain a AGGREGATE-IPv4 (respectively AGGREGATE-IPv6) SENDER_TEMPLATE object as defined in [RFC3175] and [RFC4860], and are sent between ingress PE and egress PE in either direction. These objects MUST NOT be included in any RSVP messages that are sent outside of the provider's backbone (except in the inter-AS option B and C cases, as described above, when it may appear on inter-AS links). The processing rules for these objects are otherwise identical to those of the VPN-IPv4 (respectively VPN-IPv6) SENDER_TEMPLATE object defined in Section 8.2. The format of the

object is as follows:

```
      o    AGGREGATE-VPN-IPv4 SENDER_TEMPLATE object:
             Class = 11, C-Type = TBA

           +-------------+-------------+-------------+-------------+
           |                                                       |
           +                                                       +
           |          VPN-IPv4 AggregatorAddress (12 bytes)        |
           +                                                       +
           |                                                       |
           +-------------+-------------+-------------+-------------+


      o    AGGREGATE-VPN-IPv6 SENDER_TEMPLATE object:
             Class = 11, C-Type = TBA

           +-------------+-------------+-------------+-------------+
           |                                                       |
           +                                                       +
           |                                                       |
           +          VPN-IPv6 AggregatorAddress (24 bytes)        +
           /                                                       /
           .                                                       .
           /                                                       /
           |                                                       |
           +-------------+-------------+-------------+-------------+
```

The VPN-IPv4 AggregatorAddress (respectively VPN-IPv6 AggregatorAddress) field contains an address of the VPN-IPv4 (respectively VPN-IPv6) address family encoded as specified in [RFC4364] (respectively [RFC4659]). The content and processing rules for these objects are similar to those of the VPN-IPv4 SENDER_TEMPLATE object defined in Section 8.2.

The flags and DSCP are identical to the same fields of the AGGREGATE-IPv4 and AGGREGATE-IPv6 SESSION objects.

8.7. AGGREGATE-VPN-IPv4 and AGGREGATE-VPN-IPv6 FILTER_SPEC objects

The usage of Aggregated VPN-IPv4 FILTER_SPEC object is described in Section 7.3. The AGGREGATE-VPN-IPv4 FILTER_SPEC object appears in RSVP messages that ordinarily contain a AGGREGATE-IPv4 FILTER_SPEC object as defined in [RFC3175] and [RFC4860], and are sent between ingress PE and egress PE in either direction. These objects MUST NOT

be included in any RSVP messages that are sent outside of the provider's backbone (except in the inter-AS option B and C cases, as described above, when it may appear on inter-AS links). The processing rules for these objects are otherwise identical to those of the VPN-IPv4 FILTER_SPEC object defined in Section 8.3. The format of the object is as follows:

```
   o    AGGREGATE-VPN-IPv4 FILTER_SPEC object:
          Class = 10, C-Type = TBA

        Definition same as AGGREGATE-VPN-IPv4 SENDER_TEMPLATE object.


   o    AGGREGATE-VPN-IPv6 FILTER_SPEC object:
          Class = 10, C-Type = TBA

        Definition same as AGGREGATE-VPN-IPv6 SENDER_TEMPLATE object.
```

9. IANA Considerations

Section 8 defines new objects. Therefore, this document requests IANA to modify the RSVP parameters registry, 'Class Names, Class Numbers, and Class Types' subregistry, and:

- o assign six new C-Types under the existing SESSION Class (Class number 1), as suggested below:

Class Number	Class Name	Reference
1	SESSION	[RFC2205]

Class Types or C-Types:

..	...	...
aa	VPN-IPv4	[RFCXXXX]
bb	VPN-IPv6	[RFCXXXX]
cc	AGGREGATE-VPN-IPv4	[RFCXXXX]
dd	AGGREGATE-VPN-IPv6	[RFCXXXX]
ee	GENERIC-AGGREGATE-VPN-IPv4	[RFCXXXX]
ff	GENERIC-AGGREGATE-VPN-IPv6	[RFCXXXX]

[Note to IANA and the RFC Editor: Please replace RFCXXXX with the RFC number of this specification. Suggested values: aa-ff=19-24]

- o assign four new C-Types under the existing SENDER_TEMPLATE Class (Class number 11), as suggested below:

Class Number	Class Name	Reference
11	SENDER_TEMPLATE	[RFC2205]

Class Types or C-Types:

..	...	...
aa	VPN-IPv4	[RFCXXXX]
bb	VPN-IPv6	[RFCXXXX]
cc	AGGREGATE-VPN-IPv4	[RFCXXXX]
dd	AGGREGATE-VPN-IPv6	[RFCXXXX]

[Note to IANA and the RFC Editor: Please replace RFCXXXX with the RFC number of this specification. Suggested values: aa-dd=14-17]

- o assign four new C-Types under the existing FILTER_SPEC Class (Class number 10), as suggested below:

Class Number	Class Name	Reference
10	FILTER_SPEC	[RFC2205]

Class Types or C-Types:

..	...	...
aa	VPN-IPv4	[RFCXXXX]
bb	VPN-IPv6	[RFCXXXX]
cc	AGGREGATE-VPN-IPv4	[RFCXXXX]
dd	AGGREGATE-VPN-IPv6	[RFCXXXX]

[Note to IANA and the RFC Editor: Please replace RFCXXXX with the RFC number of this specification. Suggested values: aa-dd=14-17]

- o assign two new C-Types under the existing RSVP_HOP Class (Class number 3), as suggested below:

Class Number	Class Name	Reference
3	RSVP_HOP	[RFC2205]

Class Types or C-Types:

..	...	...
aa	VPN-IPv4	[RFCXXXX]
bb	VPN-IPv6	[RFCXXXX]

[Note to IANA and the RFC Editor: Please replace RFCXXXX with the RFC number of this specification. Suggested values: aa-bb=5-6]

In addition, a new PathError code/value is required to identify a signaling reachability failure and the need for a VPN-IPv4 or VPN-IPv6 RSVP_HOP object as described in Section 5.2.2. Therefore, this document requests IANA to modify the RSVP parameters registry, 'Error Codes and Globally-Defined Error Value Sub-Codes' subregistry, and:

- o assign a new Error Code and sub-code, as suggested below:

```
  aa  RSVP over MPLS Problem                     [RFCXXXX]

      This Error Code has the following globally-defined Error
      Value sub-codes:

        1 = RSVP_HOP not reachable across VPN     [RFCXXXX]
```

[Note to IANA and the RFC Editor: Please replace RFCXXXX with the RFC number of this specification. Suggested values: aa=34]

10. Security Considerations

[RFC4364] addresses the security considerations of BGP/MPLS VPNs in general. General RSVP security considerations are discussed in [RFC2205]. To ensure the integrity of RSVP, the RSVP Authentication mechanisms defined in [RFC2747] and [RFC3097] SHOULD be supported. Those protect RSVP message integrity hop-by-hop and provide node authentication as well as replay protection, thereby protecting against corruption and spoofing of RSVP messages. [I-D.ietf-tsvwg-rsvp-security-groupkeying] discusses applicability of various keying approaches for RSVP Authentication. First, we note that the discussion about applicability of group keying to an intra-provider environment where RSVP hops are not IP hops is relevant to securing of RSVP among PEs of a given Service Provider deploying the solution specified in the present document. We note that the RSVP signaling in MPLS VPN is likely to spread over multiple administrative domains (e.g. The service provider operating the VPN service, and the customers of the service). Therefore the considerations in [I-D.ietf-tsvwg-rsvp-security-groupkeying] about inter-domain issues are likely to apply.

Since RSVP messages travel through the L3VPN cloud directly addressed to PE or ASBR routers (without IP router alert option), P routers remain isolated from RSVP messages signaling customer reservations. Providers MAY choose to block PEs from sending datagrams with the router alert option to P routers as a security practice, without impacting the functionality described herein.

Beyond those general issues, four specific issues are introduced by this document: resource usage on PEs, resource usage in the provider backbone, PE route advertisement outside the AS, and signaling exposure to ASBRs and PEs. We discuss these in turn.

A customer who makes resource reservations on the CE-PE links for his sites is only competing for link resources with himself, as in

standard RSVP, at least in the common case where each CE-PE link is dedicated to a single customer. Thus, from the perspective of the CE-PE links, the present document does not introduce any new security issues. However, because a PE typically serves multiple customers, there is also the possibility that a customer might attempt to use excessive computational resources on a PE (CPU cycles, memory etc.) by sending large numbers of RSVP messages to a PE. In the extreme this could represent a form of denial-of-service attack. In order to prevent such an attack, a PE SHOULD support mechanisms to limit the fraction of its processing resources that can be consumed by any one CE or by the set of CEs of a given customer. For example, a PE might implement a form of rate limiting on RSVP messages that it receives from each CE. We observe that these security risks and measures related to PE resource usage are very similar for any control plane protocol operating between CE and PE (e.g. RSVP, routing, multicast).

The second concern arises only when the service provider chooses to offer resource reservation across the backbone, as described in Section 4. In this case, the concern may be that a single customer might attempt to reserve a large fraction of backbone capacity, perhaps with a co-ordinated effort from several different CEs, thus denying service to other customers using the same backbone. [RFC4804] provides some guidance on the security issues when RSVP reservations are aggregated onto MPLS tunnels, which are applicable to the situation described here. We note that a provider MAY use local policy to limit the amount of resources that can be reserved by a given customer from a particular PE, and that a policy server could be used to control the resource usage of a given customer across multiple PEs if desired. It is RECOMMENDED that an implementation of this specification support local policy on the PE to control the amount of resources that can be reserved by a given customer/CE.

Use of the VPN-IPv4 RSVP_HOP object requires exporting a PE VPN-IPv4 route to another AS, and potentially could allow unchecked access to remote PEs if those routes were indiscriminately redistributed. However, as described in Section 3.1, no route which is not within a customer's VPN should ever be advertised to (or reachable from) that customer. If a PE uses a local address already within a customer VRF (like PE-CE link address), it MUST NOT send this address in any RSVP messages in a different customer VRF. A "control plane" VPN MAY be created across PEs and ASBRs and addresses in this VPN can be used to signal RSVP sessions for any customers, but these routes MUST NOT be advertised to, or made reachable from, any customer. An implementation of the present document MAY support such operation using a "control plane" VPN. Alternatively, ASBRs MAY implement the signaling procedures described in Section 5.2.1, even if admission control is not required on the inter-AS link, as these procedures do

not require any direct P/PE route advertisement out of the AS.

Finally, certain operations described herein (Section 3) require an ASBR or PE to receive and locally process a signaling packet addressed to the BGP next-hop address advertised by that router. This requirement does not strictly apply to MPLS/BGP VPNs [RFC4364]. This could be viewed as opening ASBRs and PEs to being directly addressable by customer devices where they were not open before, and could be considered a security issue. If a provider wishes to mitigate this situation, the implementation MAY support the "control protocol VPN" approach described above. That is, whenever a signaling message is to be sent to a PE or ASBR, the address of the router in question would be looked up in the "control protocol VPN", and the message would then be sent on the LSP that is found as a result of that lookup. This would ensure that the router address is not reachable by customer devices.

[RFC4364] mentions use of IPsec both on a CE-CE basis and PE-PE basis: "Cryptographic privacy is not provided by this architecture, nor by Frame Relay or ATM VPNs. These architectures are all compatible with the use of cryptography on a CE-CE basis, if that is desired. The use of cryptography on a PE-PE basis is for further study."

The procedures specified in the present document for admission control on the PE-CE links (Section 3) are compatible with the use of IPsec on a PE-PE basis. The optional procedures specified in the present document for admission control in the Service Provider's backbone (Section 4) are not compatible with the use of IPsec on a PE-PE basis, since those procedures depend on the use of PE-PE MPLS TE Tunnels to perform aggregate reservations through the Service Provider's backbone.

[RFC4923] describes a model for RSVP operation through IPsec Gateways. In a nutshell, a form of hierarchical RSVP reservation is used where an RSVP reservation is made for the IPsec tunnel and then individual RSVP reservations are admitted/aggregated over the tunnel reservation. This model applies to the case where IPsec is used on a CE-CE basis. In that situation, the procedures defined in the present document would simply apply "as is" to the reservation established for the IPsec tunnel(s).

11. Acknowledgments

Thanks to Ashwini Dahiya, Prashant Srinivas, Yakov Rekhter, Eric Rosen, Dan Tappan and Lou Berger for their many contributions to solving the problems described in this document. Thanks to Ferit

Yegenoglu for his useful comments. We also thank Stefan Santesson Vijay Gurbani and Alexey Melnikov for their review comments. We thank Richard Woundy for his very thorough review and comments including those that resulted in additional text discussing scenarios of admission control reject in the MPLVS VPN cloud.

Appendix A. Alternatives Considered

At this stage a number of alternatives to the approach described above have been considered. We document some of the approaches considered here to assist future discussion. None of these has been shown to improve upon the approach described above, and the first two seem to have significant drawbacks relative to the approach described above.

Appendix A.1. GMPLS UNI approach

[RFC4208] defines the GMPLS UNI. In Section 7 the operation of the GMPLS UNI in a VPN context is briefly described. This is somewhat similar to the problem tackled in the current document. The main difference is that the GMPLS UNI is primarily aimed at the problem of allowing a CE device to request the establishment of an LSP across the network on the other side of the UNI. Hence the procedures in [RFC4208] would lead to the establishment of an LSP across the VPN provider's network for every RSVP request received, which is not desired in this case.

To the extent possible, the approach described in this document is consistent with [RFC4208], while filling in more of the details and avoiding the problem noted above.

Appendix A.2. VRF label approach

Another approach to solving the problems described here involves the use of label switching to ensure that Path, Resv, and other RSVP messages are directed to the appropriate VRF. One challenge with such an approach is that [RFC4364] does not require labels to be allocated for VRFs, only for customer prefixes, and that there is no simple, existing method for advertising the fact that a label is bound to a VRF. If, for example, an ingress PE sent a Path message labelled with a VPN label that was advertised by the egress PE for the prefix that matches the destination address in the Path, there is a risk that the egress PE would simply label-switch the Path directly on to the CE without performing RSVP processing.

A second challenge with this approach is that an IP address needs to be associated with a VRF and used as the PHOP address for the Path

message sent from ingress PE to egress PE. That address needs to be reachable from the egress PE, and to exist in the VRF at the ingress PE. Such an address is not always available in today's deployments, so this represents at least a change to existing deployment practices.

Appendix A.3. VRF label plus VRF address approach

It is possible to create an approach based on that described in the previous section which addresses the main challenges of that approach. The basic approach has two parts: (a) define a new BGP Extended Community to tag a route (and its associated MPLS label) as pointing to a VRF; (b) allocate a "dummy" address to each VRF, specifically to be used for routing RSVP messages. The dummy address (which could be anything, e.g. a loopback of the associated PE) would be used as a PHOP for Path messages and would serve as the destination for Resv messages but would not be imported into VRFs of any other PE.

12. References

12.1. Normative References

[RFC2113] Katz, D., "IP Router Alert Option", RFC 2113, February 1997.

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

[RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC2711] Partridge, C. and A. Jackson, "IPv6 Router Alert Option", RFC 2711, October 1999.

[RFC3175] Baker, F., Iturralde, C., Le Faucheur, F., and B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, September 2001.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC4659] De Clercq, J., Ooms, D., Carugi, M., and F. Le Faucheur, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", RFC 4659, September 2006.

[RFC4804] Le Faucheur, F., "Aggregation of Resource ReSerVation Protocol (RSVP) Reservations over MPLS TE/DS-TE Tunnels", RFC 4804, February 2007.

12.2. Informative References

[I-D.ietf-l3vpn-e2e-rsvp-te-reqts]
Kumaki, K., Kamite, Y., and R. Zhang, "Requirements for supporting Customer RSVP and RSVP-TE over a BGP/MPLS IP-VPN", draft-ietf-l3vpn-e2e-rsvp-te-reqts-04 (work in progress), August 2009.

[I-D.ietf-nsis-ntlp]
Schulzrinne, H. and M. Stiemerling, "GIST: General Internet Signalling Transport", draft-ietf-nsis-ntlp-20 (work in progress), June 2009.

[I-D.ietf-nsis-qos-nslp]
Manner, J., Karagiannis, G., and A. McDonald, "NSLP for Quality-of-Service Signaling", draft-ietf-nsis-qos-nslp-17 (work in progress), October 2009.

[I-D.ietf-tsvwg-rsvp-security-groupkeying]
Behringer, M. and F. Faucheur, "Applicability of Keying Methods for RSVP Security", draft-ietf-tsvwg-rsvp-security-groupkeying-05 (work in progress), June 2009.

[RFC1633] Braden, B., Clark, D., and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.

[RFC2209] Braden, B. and L. Zhang, "Resource ReSerVation Protocol (RSVP) -- Version 1 Message Processing Rules", RFC 2209, September 1997.

[RFC2210] Wroclawski, J., "The Use of RSVP with IETF Integrated Services", RFC 2210, September 1997.

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

[RFC2961] Berger, L., Gan, D., Swallow, G., Pan, P., Tommasi, F., and S. Molendini, "RSVP Refresh Overhead Reduction Extensions", RFC 2961, April 2001.

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

April 2001.

[RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

[RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, October 2005.

[RFC4208] Swallow, G., Drake, J., Ishimatsu, H., and Y. Rekhter, "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support for the Overlay Model", RFC 4208, October 2005.

[RFC4860] Le Faucheur, F., Davie, B., Bose, P., Christou, C., and M. Davenport, "Generic Aggregate Resource ReSerVation Protocol (RSVP) Reservations", RFC 4860, May 2007.

[RFC4923] Baker, F. and P. Bose, "Quality of Service (QoS) Signaling in a Nested Virtual Private Network", RFC 4923, August 2007.

Authors' Addresses

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

Email: bsd@cisco.com

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

Email: flefauch@cisco.com

Ashok Narayanan
Cisco Systems, Inc.
1414 Mass. Ave.
Boxborough, MA 01719
USA

Email: ashokn@cisco.com