Internet Engineering Task Force (IETF)
Request for Comments: 7417
Category: Experimental
ISSN: 2070-1721

G. Karagiannis
Huawei Technologies
A. Bhargava
Cisco Systems, Inc.

December 2014

Extensions to Generic Aggregate RSVP for IPv4 and IPv6 Reservations over Pre-Congestion Notification (PCN) Domains

Abstract

This document specifies extensions to Generic Aggregate RSVP (RFC 4860) for support of the Pre-Congestion Notification (PCN) Controlled Load (CL) and Single Marking (SM) edge behaviors over a Diffserv cloud using PCN.

Status of This Memo

Table of Contents

1. Introduction

1.1. Objective

Pre-Congestion Notification (PCN) can support the Quality of Service (QoS) of inelastic flows within a Diffserv domain in a simple, scalable, and robust fashion. Two mechanisms are used: admission control and flow termination. Admission control is used to decide whether to admit or block a new flow request, while flow termination is used in abnormal circumstances to decide whether to terminate some of the existing flows. To support these two features, the overall rate of PCN-traffic is metered on every link in the domain, and PCN-packets are appropriately marked when certain configured rates are exceeded. These configured rates are below the rate of the link, thus providing notification to boundary nodes about overloads before any congestion occurs (hence "pre-congestion" notification). The PCN-egress-nodes measure the rates of differently marked PCN-traffic in periodic intervals and report these rates to the Decision Points for admission control and flow termination; the Decision Points use these rates to make decisions. The Decision Points may be collocated with the PCN-ingress-nodes, or their function may be implemented in another node. For more details, see [RFC5559], [RFC6661], and [RFC6662].

The main objective of this document is to specify the signaling protocol that can be used within a PCN-domain to carry reports from a PCN-ingress-node to a PCN Decision Point, considering that the PCN Decision Point and PCN-egress-node are collocated.

If the PCN Decision Point is not collocated with the PCN-egress-node, then additional signaling procedures are required that are out of scope for this document. Moreover, as mentioned above, this architecture conforms with Policy-Based Admission Control (PBAC), where the Decision Point is located in a node other than the PCN-ingress-node [RFC2753].

Several signaling protocols can be used to carry information between PCN-boundary-nodes (PCN-ingress-node and PCN-egress-node). However, since (1) both the PCN-egress-node and PCN-ingress-node are located on the data path and (2) the admission control procedure needs to be done at the PCN-egress-node, a signaling protocol that follows the same path as the data path, like RSVP, is more suited for this purpose. In particular, this document specifies extensions to Generic Aggregate RSVP [RFC4860] for support of the PCN Controlled Load (CL) and Single Marking (SM) edge behaviors over a Diffserv cloud using Pre-Congestion Notification.

This document is published as an Experimental document in order to:

- o validate industry interest by allowing implementation and deployment
- o gather operational experience, particularly related to dynamic interactions of RSVP signaling and PCN, and corresponding levels of performance

Support for the techniques specified in this document involves RSVP functionality in boundary nodes of a PCN-domain whose interior nodes forward RSVP traffic without performing RSVP functionality.

1.2. Overview and Motivation

Two main QoS architectures have been specified by the IETF: the Integrated Services (Intserv) [RFC1633] architecture and the Differentiated Services (Diffserv) architecture ([RFC2475]).

Intserv provides methods for the delivery of end-to-end QoS to applications over heterogeneous networks. One of the QoS signaling protocols used by the Intserv architecture is RSVP [RFC2205], which can be used by applications to request per-flow resources from the network. These RSVP requests can be admitted or rejected by the network. Applications can express their quantifiable resource requirements using Intserv parameters as defined in [RFC2211] and [RFC2212]. The Controlled Load (CL) service [RFC2211] is a form of QoS that closely approximates the QoS that the same flow would receive from a lightly loaded network element. The CL service is useful for inelastic flows such as those used for real-time media.

The Diffserv architecture can support the differentiated treatment of packets in very large-scale environments. While Intserv and RSVP classify packets per flow, Diffserv networks classify packets into one of a small number of aggregated flows or "classes", based on the Diffserv Codepoint (DSCP) in the packet IP header. At each Diffserv router, packets are subjected to a "Per Hop Behavior" (PHB), which is invoked by the DSCP. The primary benefit of Diffserv is its scalability, since the need for per-flow state and per-flow processing is eliminated.

However, Diffserv does not include any mechanism for communication between applications and the network. Several solutions have been specified to solve this issue. One of these solutions is Intserv over Diffserv [RFC2998], including Resource-Based Admission Control (RBAC), PBAC, assistance in traffic identification/classification, and traffic conditioning. Intserv over Diffserv can operate over a statically provisioned or an RSVP-aware Diffserv region. When it is RSVP aware, several mechanisms may be used to support dynamic provisioning and topology-aware admission control, including aggregate RSVP reservations, per-flow RSVP, or a bandwidth broker. [RFC3175] specifies aggregation of RSVP end-to-end reservations over aggregate RSVP reservations. In [RFC3175], the RSVP generic aggregate reservation is characterized by an RSVP SESSION object using the 3-tuple <source IP address, destination IP address, Diffserv Codepoint>.

Several scenarios require the use of multiple generic aggregate reservations that are established for a given PHB from a given source IP address to a given destination IP address; see [RFC4923] and [RFC4860]. For example, multiple generic aggregate reservations can be applied in situations where multiple end-to-end (E2E) reservations using different preemption priorities need to be aggregated through a PCN-domain using the same PHB. Using multiple aggregate reservations for the same PHB allows

- o enforcement of the different preemption priorities within the aggregation region
- o more efficient management of Diffserv resources
- o sustainment of a larger number of E2E reservations with higher preemption priorities during periods of resource shortage

In particular, [RFC4923] discusses in detail how end-to-end RSVP reservations can be established in a nested VPN environment through RSVP aggregation.

[RFC4860] provides generic aggregate reservations by extending [RFC3175] to support multiple aggregate reservations for the same source IP address, destination IP address, and PHB (or set of PHBs). In particular, multiple such generic aggregate reservations can be established for a given PHB from a given source IP address to a given destination IP address. This is achieved by adding the concept of a Virtual Destination Port and an Extended Virtual Destination Port in the RSVP SESSION object. In addition to this, the RSVP SESSION object for generic aggregate reservations uses the PHB Identification Code (PHB-ID) defined in [RFC3140] instead of using the Diffserv Codepoint (DSCP) used in [RFC3175]. The PHB-ID is used to identify the PHB, or set of PHBs, from which the Diffserv resources are to be reserved.

The RSVP-like signaling protocol required to carry (1) requests from a PCN-egress-node to a PCN-ingress-node and (2) reports from a PCN-ingress-node to a PCN-egress-node needs to follow the PCN signaling requirements defined in [RFC6663]. In addition to that, the signaling protocol functionality supported by the PCN-ingress-nodes and PCN-egress-nodes needs to maintain logical aggregate constructs (i.e., ingress-egress-aggregate state) and be able to map E2E reservations to these aggregate constructs. Moreover, no actual reservation state is needed to be maintained inside the PCN-domain, i.e., the PCN-interior-nodes are not maintaining any reservation state.

This can be accomplished by two possible approaches:

Approach (1):

- o adapting the aggregation procedures of RFC 4860 to fit the PCN requirements with as little change as possible over the functionality provided in RFC 4860.

- o hence, performing aggregate RSVP signaling (even if it is to be ignored by PCN-interior-nodes).

- o using the aggregate RSVP signaling procedures to carry PCN information between the PCN-boundary-nodes (PCN-ingress-node and PCN-egress-node).

Approach (2):

- o adapting the aggregation procedures of RFC 4860 to fit the PCN requirements with significant changes over RFC 4860 (i.e., the aspect of the procedures that have to do with maintaining aggregate states and mapping the E2E reservations to aggregate constructs are kept, but the procedures that are specific to aggregate RSVP signaling and aggregate reservation establishment/maintenance are dropped).

- o hence not performing aggregate RSVP signaling.

- o piggybacking the PCN information inside the E2E RSVP signaling.

Both approaches are probably viable; however, since the operations of RFC 4860 have been thoroughly studied and implemented, it can be considered that the solution from RFC 4860 can better deal with the more challenging situations (rerouting in the PCN-domain, failure of a PCN-ingress-node, failure of a PCN-egress-node, rerouting towards a different edge, etc.). This is the reason for choosing Approach (1) for the specification of the signaling protocol used to carry PCN information between the PCN-boundary-nodes (PCN-ingress-node and PCN-egress-node).

As noted earlier, this document specifies extensions to Generic Aggregate RSVP [RFC4860] for support of the PCN Controlled Load (CL) and Single Marking (SM) edge behaviors over a Diffserv cloud using Pre-Congestion Notification.

This document follows the PCN signaling requirements defined in [RFC6663] and specifies extensions to Generic Aggregate RSVP [RFC4860] for support of PCN edge behaviors as specified in [RFC6661] and [RFC6662]. Moreover, this document specifies how RSVP aggregation can be used to set up and maintain (1) Ingress-Egress-Aggregate (IEA) states at Ingress and Egress nodes and (2) generic aggregation of end-to-end RSVP reservations over PCN (Congestion and Pre-Congestion Notification) domains.

To comply with this specification, PCN-nodes MUST be able to support the functionality specified in [RFC5670], [RFC5559], [RFC6660], [RFC6661], and [RFC6662]. Furthermore, the PCN-boundary-nodes MUST support the RSVP generic aggregate reservation procedures specified in [RFC4860], which are augmented with procedures specified in this document.

1.3. Requirements Language and Terminology

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 [RFC2119].

This document uses terms defined in [RFC4860], [RFC3175], [RFC5559], [RFC5670], [RFC6661], and [RFC6662].

For readability, a number of definitions from [RFC3175] as well as definitions for terms used in [RFC5559], [RFC6661], and [RFC6662] are provided here, where some of them are augmented with new meanings:

Aggregator
   The process in (or associated with) the router at the ingress edge of the aggregation region (with respect to the end-to-end RSVP reservation) and behaving in accordance with [RFC4860]. In this

document, it is also the PCN-ingress-node. It is important to notice that in the context of this document the Aggregator must be able to determine the Deaggregator using the procedures specified in Section 4 of [RFC4860] and Section 1.4.2 of [RFC3175].

Congestion Level Estimate (CLE)
The ratio of PCN-marked to total PCN-traffic (measured in octets) received for a given ingress-egress-aggregate during a given measurement period. The CLE is used to derive the PCN-admission-state and is also used by the report suppression procedure if report suppression is activated.

Deaggregator
The process in (or associated with) the router at the egress edge of the aggregation region (with respect to the end-to-end RSVP reservation) and behaving in accordance with [RFC4860]. In this document, it is also the PCN-egress-node and Decision Point.

E2E
End to end

E2E Microflow
A microflow where its associated packets are being forwarded on an E2E path.

E2E Reservation
An RSVP reservation such that:

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

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

(3) this RSVP reservation is aggregated over an Ingress-Egress-Aggregate (IEA) between the Aggregator and Deaggregator.

An E2E RSVP reservation may be a per-flow reservation, which in this document is only maintained at the PCN-ingress-node and PCN-egress-node. Alternatively, the E2E reservation may itself be an aggregate reservation of various types (e.g., Aggregate IP reservation, Aggregate IPsec reservation [RFC4860]). As per regular RSVP operations, E2E RSVP reservations are unidirectional.

ETM-Rate
The rate of excess-traffic-marked (ETM) PCN-traffic received at a PCN-egress-node for a given ingress-egress-aggregate in octets per second.

Extended vDstPort (Extended Virtual Destination Port)
An identifier used in the SESSION that remains constant over the life of the generic aggregate reservation. The length of this identifier is 32 bits when IPv4 addresses are used and 128 bits when IPv6 addresses are used.

A sender (or Aggregator) that wishes to narrow the scope of a SESSION to the sender-receiver pair (or Aggregator-Deaggregator pair) should place its IPv4 or IPv6 address here as a network unique identifier. A sender (or Aggregator) that wishes to use a common session with other senders (or Aggregators) in order to use a shared reservation across senders (or Aggregators) must set this field to all zeros. In this document, the Extended vDstPort should contain the IPv4 or IPv6 address of the Aggregator.

Ingress-Egress-Aggregate (IEA)
The collection of PCN-packets from all PCN-flows that travel in one direction between a specific pair of PCN-boundary-nodes. In this document, one RSVP generic aggregate reservation is mapped to only one ingress-egress-aggregate, while one ingress-egress-aggregate is mapped to one or more RSVP generic aggregate reservations. PCN-flows and their PCN-traffic that are mapped into a specific RSVP generic aggregate reservation can also be easily mapped into their corresponding ingress-egress-aggregate.

Microflow (from [RFC2474])
A single instance of an application-to-application flow of packets, which is identified by <source address, destination address, protocol id> and (where applicable) <source port, destination port>.

PCN-Admission-State
The state ("admit" or "block") derived by the Decision Point for a given ingress-egress-aggregate based on statistics about PCN-packet marking. The Decision Point decides to admit or block new flows offered to the aggregate based on the current value of the PCN-admission-state.

PCN-Boundary-Node
A PCN-node that connects one PCN-domain to a node in either another PCN-domain or a non-PCN-domain.

PCN-Domain

A PCN-capable domain; a contiguous set of PCN-enabled nodes that perform Diffserv scheduling [RFC2474]; the complete set of PCN-nodes that in principle can, through PCN-marking packets, influence decisions about flow admission and termination within the domain; includes the PCN-egress-nodes, which measure these PCN-marks, and the PCN-ingress-nodes.

PCN-Egress-Node

A PCN-boundary-node in its role in handling traffic as it leaves a PCN-domain. In this document, the PCN-egress-node also operates as a Decision Point and Deaggregator.

PCN-Flow

The unit of PCN-traffic that the PCN-boundary-node admits (or terminates); the unit could be a single E2E microflow (as defined in [RFC2474]) or some identifiable collection of microflows.

PCN-Ingress-Node

A PCN-boundary-node in its role in handling traffic as it enters a PCN-domain. In this document, the PCN-ingress-node also operates as an Aggregator.

PCN-Interior-Node

A node in a PCN-domain that is not a PCN-boundary-node.

PCN-Node

A PCN-boundary-node or a PCN-interior-node.

PCN-Sent-Rate

The rate of PCN-traffic received at a PCN-ingress-node and destined for a given ingress-egress-aggregate in octets per second.

PCN-Traffic, PCN-Packets, PCN-BA

A PCN-domain carries traffic of different Diffserv Behavior Aggregates (BAs) [RFC2474]. The PCN-BA uses the PCN mechanisms to carry PCN-traffic, and the corresponding packets are PCN-packets. The same network will carry traffic of other Diffserv BAs. The PCN-BA is distinguished by a combination of the Diffserv Codepoint (DSCP) and Explicit Congestion Notification (ECN) fields.

PHB-ID (Per Hop Behavior Identification Code)

A 16-bit field containing the Per Hop Behavior Identification Code of the PHB, or of the set of PHBs, from which Diffserv resources are to be reserved. This field must be encoded as specified in Section 2 of [RFC3140].

RSVP Generic Aggregate Reservation
An RSVP reservation that is identified by using the RSVP SESSION object for generic RSVP aggregate reservation. This RSVP SESSION object is based on the RSVP SESSION object specified in [RFC4860], augmented with the following information:

- The IPv4 DestAddress, IPv6 DestAddress should be set to the IPv4 or IPv6 destination addresses, respectively, of the Deaggregator (PCN-egress-node).

- The PHB-ID should be set equal to PCN-compatible Diffserv Codepoint(s).

- The Extended vDstPort should be set to the IPv4 or IPv6 destination addresses, of the Aggregator (PCN-ingress-node).

VDstPort (Virtual Destination Port)
A 16-bit identifier used in the SESSION that remains constant over the life of the generic aggregate reservation.

1.4. Organization of This Document

This document is organized as follows. Section 2 gives an overview of RSVP extensions and operations. The elements of the procedures that are used in this document are specified in Section 3. Section 4 describes the protocol elements. The security considerations are given in Section 5, and the IANA considerations are provided in Section 6.

2. Overview of RSVP Extensions and Operations

2.1. Overview of RSVP Aggregation Procedures in PCN-Domains

The PCN-boundary-nodes (see Figure 1) can support RSVP SESSIONS for generic aggregate reservations [RFC4860], which depend on ingress-egress-aggregates. In particular, one RSVP generic aggregate reservation matches to only one ingress-egress-aggregate.

However, one ingress-egress-aggregate matches to one or more RSVP generic aggregate reservations. In addition, to comply with this specification, the PCN-boundary-nodes need to distinguish and process (1) RSVP SESSIONS for generic aggregate sessions and their messages according to [RFC4860] and (2) E2E RSVP SESSIONS and messages according to [RFC2205].

This document locates all RSVP processing for a PCN-domain at PCN-boundary-nodes. PCN-interior-nodes do not perform any RSVP functionality or maintain RSVP-related state information. Rather, PCN-interior-nodes forward all RSVP messages (for both generic aggregate reservations [RFC4860] and E2E reservations [RFC2205]) as if they were ordinary network traffic.

Moreover, each Aggregator and Deaggregator (i.e., PCN-boundary-nodes) needs to support policies to initiate and maintain, for each pair of PCN-boundary-nodes of the same PCN-domain, one ingress-egress-aggregate.

```
                     --------------------------
                    /       PCN-domain         \
      |----|       |                            |      |----|
   H--| R  |\ |-----|                          |------| /| R  |-->H
   H--|    |\\|     |    |---|     |---|     |      |//|    |-->H
      |----| \|     |    | I |     | I |     |      |/ |----|
              | Agg |======================>| Deag |
             /|     |    |   |     |   |     |      |\
   H--------//|     |    |---|     |---|     |      |\\-------->H
   H--------/ |-----|                          |------| \-------->H
                   |                            |
                    \                          /
                     --------------------------

   H     = Host requesting end-to-end RSVP reservations
   R     = RSVP router
   Agg   = Aggregator (PCN-ingress-node)
   Deag  = Deaggregator (PCN-egress-node)
   I     = Interior Router (PCN-interior-node)
   -->   = E2E RSVP reservation
   ==>   = Aggregate RSVP reservation
```

Figure 1: Aggregation of E2E Reservations over Generic Aggregate RSVP Reservations in PCN-Domains, Based on [RFC4860]

Both the Aggregator and Deaggregator can maintain one or more RSVP generic aggregate reservations, but the Deaggregator is the entity that initiates these RSVP generic aggregate reservations. Note that one RSVP generic aggregate reservation matches to only one ingress-egress-aggregate, while one ingress-egress-aggregate matches to one or more RSVP generic aggregate reservations. This can be accomplished by using for the different RSVP generic aggregate reservations the same combinations of ingress and egress identifiers, but with a different PHB-ID value (see [RFC4860]). The procedures for aggregation of E2E reservations over generic aggregate RSVP reservations are the same as the procedures specified in Section 4 of [RFC4860], augmented with the ones specified in Section 2.5.

One significant difference between this document and [RFC4860] is the fact that in this document the admission control of E2E RSVP reservations over the PCN-core is performed according to the PCN procedures, while in [RFC4860] this is achieved via first admitting aggregate RSVP reservations over the aggregation region and then admitting the E2E reservations over the aggregate RSVP reservations. Therefore, in this document, the RSVP generic aggregate RSVP reservations are not subject to admission control in the PCN-core, and the E2E RSVP reservations are not subject to admission control over the aggregate reservations. In turn, this means that several procedures described in [RFC4860] are significantly simplified in this document:

- o Unlike [RFC4860], the generic aggregate RSVP reservations need not be admitted in the PCN-core.

- o Unlike [RFC4860], the RSVP aggregated traffic does not need to be tunneled between Aggregator and Deaggregator; see Section 2.3.

- o Unlike [RFC4860], the Deaggregator need not perform admission control of E2E reservations over the aggregate RSVP reservations.

- o Unlike [RFC4860], there is no need for dynamic adjustment of the RSVP generic aggregate reservation size; see Section 2.6.

2.2. PCN-Marking, Encoding, and Transport of Pre-congestion Information

The method of PCN-marking within the PCN-domain is specified in [RFC5670]. In addition, the method of encoding and transport of pre-congestion information is specified in [RFC6660]. The PHB-ID (Per Hop Behavior Identification Code) used SHOULD be set equal to PCN-compatible Diffserv Codepoint(s).

2.3. Traffic Classification within the Aggregation Region

The PCN-ingress marks a PCN-BA using PCN-marking (i.e., a combination of the DSCP and ECN fields), which interior nodes use to classify PCN-traffic. The PCN-traffic (e.g., E2E microflows) belonging to an RSVP generic aggregate reservation can be classified only at the PCN-boundary-nodes (i.e., Aggregator and Deaggregator) by using the RSVP SESSION object for RSVP generic aggregate reservations; see Section 2.1 of [RFC4860]. Note that the DSCP value included in the SESSION object SHOULD be set equal to a PCN-compatible Diffserv Codepoint. Since no admission control procedures over the RSVP generic aggregate reservations in the PCN-core are required, unlike [RFC4860], the RSVP aggregated traffic need not be tunneled between Aggregator and Deaggregator. In this document, one RSVP generic aggregate reservation is mapped to only one ingress-egress-aggregate, while one ingress-egress-aggregate is mapped to one or more RSVP generic aggregate reservations. PCN-flows and their PCN-traffic that are mapped into a specific RSVP generic aggregate reservation can also easily be classified into their corresponding ingress-egress-aggregate. The method of traffic conditioning of PCN-traffic and non-PCN-traffic, as well as the method of PHB configuration, are described in [RFC6661] and [RFC6662].

2.4. Deaggregator (PCN-Egress-Node) Determination

This document assumes the same dynamic Deaggregator determination method as that used in [RFC4860].

2.5. Mapping E2E Reservations onto Aggregate Reservations

To comply with this specification, for the mapping of E2E reservations onto aggregate reservations, the same methods MUST be used as the ones described in Section 4 of [RFC4860], augmented by the following rules:

- An Aggregator (PCN-ingress-node) or Deaggregator (PCN-egress-node and Decision Point) MUST use one or more policies to determine whether an RSVP generic aggregate reservation can be mapped into an ingress-egress-aggregate. This can be accomplished by using for the different RSVP generic aggregate reservations the same combinations of ingress and egress identifiers, but with a different PHB-ID value (see [RFC4860]) corresponding to the PCN specifications -- in particular, the RSVP SESSION object specified in [RFC4860], augmented with the following information:

  - The IPv4 DestAddress, IPv6 DestAddress MUST be set to the IPv4 or IPv6 destination addresses, respectively, of the Deaggregator (PCN-egress-node); see [RFC4860]. Note that the PCN-domain is considered as being only one RSVP hop (for generic aggregate RSVP or E2E RSVP). This means that the next RSVP hop for the Aggregator in the downstream direction is the Deaggregator and the next RSVP hop for the Deaggregator in the upstream direction is the Aggregator.

  - The PHB-ID (Per Hop Behavior Identification Code) SHOULD be set equal to PCN-compatible Diffserv Codepoint(s).

  - The Extended vDstPort SHOULD be set to the IPv4 or IPv6 destination addresses of the Aggregator (PCN-ingress-node); see [RFC4860].

2.6. Size of Aggregate Reservations

Since (1) no admission control of E2E reservations over the RSVP aggregate reservations is required and (2) no admission control of the RSVP aggregate reservation over the PCN-core is required, the size of the generic aggregate reservation is irrelevant and can be set to any arbitrary value by the Deaggregator. The Deaggregator SHOULD set the value of a generic aggregate reservation to a null bandwidth. We also observe that there is no need for dynamic adjustment of the RSVP aggregate reservation size.

2.7. E2E Path ADSPEC Update

To comply with this specification, for the update of the E2E Path ADSPEC, the same methods can be used as the ones described in [RFC4860].

2.8. Intra-domain Routes

The PCN-interior-nodes maintain neither E2E RSVP nor RSVP generic aggregation states and reservations. Therefore, intra-domain route changes will not affect intra-domain reservations, since such reservations are not maintained by the PCN-interior-nodes.

Furthermore, it is considered that by configuration the PCN-interior-nodes can distinguish neither RSVP generic aggregate sessions and their associated messages [RFC4860] nor E2E RSVP SESSIONS and their associated messages [RFC2205].

2.9. Inter-domain Routes

The PCN-charter scope precludes inter-domain considerations. However, for solving inter-domain route changes associated with the operation of the RSVP messages, the same methods SHOULD be used as the ones described in [RFC4860] and in Section 1.4.7 of [RFC3175].

2.10. Reservations for Multicast Sessions

PCN does not consider reservations for multicast sessions.

2.11. Multi-level Aggregation

PCN does not consider multi-level aggregations within the PCN-domain. Therefore, the PCN-interior-nodes do not support multi-level aggregation procedures. However, the Aggregator and Deaggregator SHOULD support the multi-level aggregation procedures specified in [RFC4860] and in Section 1.4.9 of [RFC3175].

2.12. Reliability Issues

To comply with this specification, for solving possible reliability issues, the same methods MUST be used as the ones described in Section 4 of [RFC4860].

3. Elements of Procedures

This section describes the procedures used to implement the aggregate RSVP procedure over PCN. It is considered that the procedures for aggregation of E2E reservations over generic aggregate RSVP reservations are the same as the procedures specified in Section 4 of [RFC4860], except where a departure from these procedures is explicitly described in this section. Please refer to Appendix B of [RFC2205] and Section 3 of [RFC4860] for the processing rules and error handling for the error cases listed below:

- o Message formatting errors, e.g., incomplete message

- o Unknown objects

3.1. Receipt of E2E Path Message by PCN-Ingress-Node (Aggregating Router)

When the E2E Path message arrives at the exterior interface of the Aggregator (PCN-ingress-node), then standard RSVP generic aggregation [RFC4860] procedures are used.

3.2. Handling of E2E Path Message by Interior Routers

The E2E Path messages traverse zero or more PCN-interior-nodes. The PCN-interior-nodes receive the E2E Path message on an interior interface and forward it on another interior interface. It is considered that, by configuration, the PCN-interior-nodes ignore the E2E RSVP signaling messages [RFC2205]. Therefore, the E2E Path messages are simply forwarded as normal IP datagrams.

3.3. Receipt of E2E Path Message by PCN-Egress-Node (Deaggregating Router)

When receiving the E2E Path message, the Deaggregator (PCN-egress-node and Decision Point) performs the regular procedures of [RFC4860], augmented with the following rules:

- o The Deaggregator MUST NOT perform the RSVP-TTL vs. IP TTL-check (TTL = Time To Live) and MUST NOT update the ADSPEC Break bit. This is because the whole PCN-domain is effectively handled by E2E RSVP as a virtual link on which integrated service is indeed supported (and admission control performed) so that the Break bit MUST NOT be set; see also [RSVP-PCN-CL].

The Deaggregator forwards the E2E Path message towards the receiver.

3.4. Initiation of New Aggregate Path Message by PCN-Ingress-Node (Aggregating Router)

To comply with this specification, for the initiation of the new RSVP generic aggregate Path message by the Aggregator (PCN-ingress-node), the same methods MUST be used as the ones described in [RFC4860].

3.5. Handling of Aggregate Path Message by Interior Routers

The Aggregate Path messages traverse zero or more PCN-interior-nodes. The PCN-interior-nodes receive the Aggregate Path message on an interior interface and forward it on another interior interface. It is considered that, by configuration, the PCN-interior-nodes ignore the Aggregate Path signaling messages. Therefore, the Aggregate Path messages are simply forwarded as normal IP datagrams.

3.6. Handling of Aggregate Path Message by Deaggregating Router

When receiving the Aggregate Path message, the Deaggregator (PCN-egress-node and Decision Point) performs the regular procedures of [RFC4860], augmented with the following rules:

- o When the received Aggregate Path message by the Deaggregator contains the RSVP-AGGREGATE-IPv4-PCN-response or RSVP-AGGREGATE-IPv6-PCN-response PCN objects, which carry the PCN-sent-rate, then the procedures specified in Section 3.18 of this document MUST be followed.

3.7. Handling of E2E Resv Message by Deaggregating Router

When the E2E Resv message arrives at the exterior interface of the Deaggregator (PCN-egress-node and Decision Point), then standard RSVP aggregation procedures of [RFC4860] are used, augmented with the following rules:

- o The E2E RSVP SESSION associated with an E2E Resv message that arrives at the external interface of the Deaggregator is mapped/matched with an RSVP generic aggregate and with a PCN ingress-egress-aggregate.

- o Depending on the type of the PCN edge behavior supported by the Deaggregator, the PCN admission control procedures specified in Section 3.3.1 of [RFC6661] or [RFC6662] MUST be followed. Since no admission control procedures over the RSVP aggregate reservations in the PCN-core are required, unlike [RFC4860], the Deaggregator does not perform any admission control of the E2E reservation over the mapped generic aggregate RSVP reservation. If the PCN-based admission control procedure is successful, then the Deaggregator MUST allow the new flow to be admitted onto the associated RSVP generic aggregation reservation and onto the PCN ingress-egress-aggregate; see [RFC6661] and [RFC6662]. If the PCN-based admission control procedure is not successful, then the E2E Resv MUST NOT be admitted onto the associated RSVP generic aggregate reservation and onto the PCN ingress-egress-aggregation. The E2E Resv message is further processed according to [RFC4860].

How the PCN-admission-state is maintained is specified in [RFC6661] and [RFC6662].

3.8. Handling of E2E Resv Message by Interior Routers

The E2E Resv messages traversing the PCN-core are IP addressed to the Aggregating router and are not marked with Router Alert; therefore, the E2E Resv messages are simply forwarded as normal IP datagrams.

3.9. Initiation of New Aggregate Resv Message by Deaggregating Router

To comply with this specification, for the initiation of the new RSVP generic aggregate Resv message by the Deaggregator (PCN-egress-node and Decision Point), the same methods MUST be used as the ones described in Section 4 of [RFC4860], augmented with the following rules:

- The size of the generic aggregate reservation is irrelevant (see Section 2.6) and can be set to any arbitrary value by the PCN-egress-node. The Deaggregator SHOULD set the value of an RSVP generic aggregate reservation to a null bandwidth. We also observe that there is no need for dynamic adjustment of the RSVP generic aggregate reservation size.

- When [RFC6661] is used and the ETM-rate measured by the Deaggregator contains a non-zero value for some ingress-egress-aggregate (see [RFC6661] and [RFC6662]), the Deaggregator MUST request the PCN-ingress-node to provide an estimate of the rate (PCN-sent-rate) at which the Aggregator (PCN-ingress-node) is receiving PCN-traffic that is destined for the given ingress-egress-aggregate.

- When [RFC6662] is used and the PCN-admission-state computed by the Deaggregator on the basis of the CLE is "block" for the given ingress-egress-aggregate, the Deaggregator MUST request the PCN-ingress-node to provide an estimate of the rate (PCN-sent-rate) at which the Aggregator is receiving PCN-traffic that is destined for the given ingress-egress-aggregate.

- In the above two cases and when the PCN-sent-rate needs to be requested from the Aggregator, the Deaggregator MUST generate and send to the Aggregator a (refresh) Aggregate Resv message that MUST carry one of the following PCN objects (see Section 4.1), depending on whether IPv4 or IPv6 is supported:

    - RSVP-AGGREGATE-IPv4-PCN-request

    - RSVP-AGGREGATE-IPv6-PCN-request

3.10. Handling of Aggregate Resv Message by Interior Routers

The Aggregate Resv messages traversing the PCN-core are IP addressed to the Aggregating router and are not marked with Router Alert; therefore, the Aggregate Resv messages are simply forwarded as normal IP datagrams.

3.11. Handling of E2E Resv Message by Aggregating Router

When the E2E Resv message arrives at the interior interface of the Aggregator (PCN-ingress-node), then standard RSVP aggregation procedures of [RFC4860] are used.

3.12. Handling of Aggregate Resv Message by Aggregating Router

When the Aggregate Resv message arrives at the interior interface of the Aggregator (PCN-ingress-node), then standard RSVP aggregation procedures of [RFC4860] are used, augmented with the following rules:

- The Aggregator SHOULD use the information carried by the PCN objects (see Section 4) and follow the steps specified in Section 3.4 of [RFC6661] and [RFC6662]. If the "R" flag carried by the RSVP-AGGREGATE-IPv4-PCN-request or RSVP-AGGREGATE-IPv6-PCN-request PCN objects is set to ON (see Section 4.1), then the Aggregator follows the steps described in Section 3.4 of [RFC6661] and [RFC6662] on calculating the PCN-sent-rate. In particular, the Aggregator MUST provide the estimated current rate of PCN-traffic received at that node and destined for a given ingress-egress-aggregate in octets per second (the PCN-sent-rate). The way this rate estimate is derived is a matter of implementation; see [RFC6661] or [RFC6662].

- The Aggregator initiates an Aggregate Path message. In particular, when the Aggregator receives an Aggregate Resv message that carries one of the following PCN objects -- RSVP-AGGREGATE-IPv4-PCN-request or RSVP-AGGREGATE-IPv6-PCN-request -- with the "R" flag set to ON (see Section 4.1), the Aggregator initiates an Aggregate Path message and includes the calculated PCN-sent-rate in the RSVP-AGGREGATE-IPv4-PCN-response or RSVP-AGGREGATE-IPv6-PCN-response PCN objects (see Section 4.1), which MUST be carried by the Aggregate Path message. This Aggregate Path message is sent towards the Deaggregator (PCN-egress-node and Decision Point) that requested the calculation of the PCN-sent-rate.

3.13. Removal of E2E Reservation

To comply with this specification, for the removal of E2E reservations, the same methods MUST be used as the ones described in Section 4 of [RFC4860] and Section 5 of [RFC4495].

3.14. Removal of Aggregate Reservation

To comply with this specification, for the removal of RSVP generic aggregate reservations, the same methods MUST be used as the ones described in Section 4 of [RFC4860] and Section 2.10 of [RFC3175]. In particular, should an aggregate reservation go away (presumably due to a configuration change, route change, or policy event), the E2E reservations it supports are no longer active. They MUST be treated accordingly.

3.15. Handling of Data on Reserved E2E Flow by Aggregating Router

The handling of data on the reserved E2E flow by the Aggregator (PCN-ingress-node) uses the procedures described in [RFC4860], augmented with the following:

o Regarding PCN-marking and traffic classification, the procedures defined in Sections 2.2 and 2.3 of this document are used.

3.16. Procedures for Multicast Sessions

No multicast sessions are considered in this document.

3.17. Misconfiguration of PCN-Node

In an event where a PCN-node is misconfigured within a PCN-domain, the desired behavior is the same as that described in Section 3.10.

3.18. PCN-Based Flow Termination

When the Deaggregator (PCN-egress-node and Decision Point) needs to terminate an amount of traffic associated with one ingress-egress-aggregate (see Section 3.3.2 of [RFC6661] and [RFC6662]), then several procedures for terminating E2E microflows can be deployed. The default procedure for terminating E2E microflows (i.e., PCN-flows) is as follows; see, for example, [RFC6661] and [RFC6662].

For the same ingress-egress-aggregate, select a number of E2E microflows to be terminated in order to decrease the total incoming amount of bandwidth associated with one ingress-egress-aggregate by the amount of traffic to be terminated. In this situation, the same mechanisms for terminating an E2E microflow can be followed as the mechanisms specified in [RFC2205]. However, based on a local policy, the Deaggregator could use other ways of selecting which microflows should be terminated. For example, for the same ingress-egress-aggregate, select a number of E2E microflows to be terminated or to reduce their reserved bandwidth in order to decrease the total incoming amount of bandwidth associated with one ingress-egress-aggregate by the amount of traffic to be terminated. In this situation, the same mechanisms for terminating an E2E microflow or reducing bandwidth associated with an E2E microflow can be followed as the mechanisms specified in Section 5 of [RFC4495].

4. Protocol Elements

The protocol elements in this document are using the elements defined in Section 4 of [RFC4860] and Section 3 of [RFC3175], augmented with the following rules:

- The DSCP value included in the SESSION object SHOULD be set equal to a PCN-compatible Diffserv Codepoint.

- The Extended vDstPort SHOULD be set to the IPv4 or IPv6 destination addresses of the Aggregator (PCN-ingress-node); see [RFC4860].

- When the Deaggregator (PCN-egress-node and Decision Point) needs to request the PCN-sent-rate from the PCN-ingress-node (see Section 3.9 of this document), the Deaggregator MUST generate and send a (refresh) Aggregate Resv message to the Aggregator that MUST carry one of the following PCN objects (see Section 4.1), depending on whether IPv4 or IPv6 is supported:

    - RSVP-AGGREGATE-IPv4-PCN-request

    - RSVP-AGGREGATE-IPv6-PCN-request

- When the Aggregator receives an Aggregate Resv message that carries one of the following PCN objects -- RSVP-AGGREGATE-IPv4-PCN-request or RSVP-AGGREGATE-IPv6-PCN-request, with the "R" flag set to ON (see Section 4.1) -- then the Aggregator MUST generate and send to the Deaggregator an Aggregate Path message that carries one of the following PCN objects (see Section 4.1), depending on whether IPv4 or IPv6 is supported:

    - RSVP-AGGREGATE-IPv4-PCN-response

    - RSVP-AGGREGATE-IPv6-PCN-response

4.1. PCN Objects

This section describes four types of PCN objects that can be carried by the (refresh) Aggregate Path or the (refresh) Aggregate Resv messages specified in [RFC4860].

These objects are as follows:

- o RSVP-AGGREGATE-IPv4-PCN-request
- o RSVP-AGGREGATE-IPv6-PCN-request
- o RSVP-AGGREGATE-IPv4-PCN-response
- o RSVP-AGGREGATE-IPv6-PCN-response

o RSVP-AGGREGATE-IPv4-PCN-request: PCN request object, when IPv4 addresses are used:

Class = 248 (PCN)
C-Type = 1 (RSVP-AGGREGATE-IPv4-PCN-request)

```
+-------------+-------------+-------------+-------------+
|     IPv4 PCN-ingress-node Address (4 bytes)           |
+-------------+-------------+-------------+-------------+
|     IPv4 PCN-egress-node Address (4 bytes)            |
+-------------+-------------+-------------+-------------+
|     IPv4 Decision Point Address (4 bytes)             |
+-------------+-------------+-------------+-------------+
|R|     Reserved                                        |
+-------------+-------------+-------------+-------------+
```

o RSVP-AGGREGATE-IPv6-PCN-request: PCN object, when IPv6 addresses are used:

Class = 248 (PCN)
C-Type = 2 (RSVP-AGGREGATE-IPv6-PCN-request)

```
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+     IPv6 PCN-ingress-node Address (16 bytes)          +
|                                                       |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+     IPv6 PCN-egress-node Address (16 bytes)           +
|                                                       |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+     Decision Point Address (16 bytes)                 +
|                                                       |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|R|     Reserved                                        |
+-------------+-------------+-------------+-------------+
```

o RSVP-AGGREGATE-IPv4-PCN-response: PCN object, IPv4 addresses are used:

Class = 248 (PCN)
C-Type = 3 (RSVP-AGGREGATE-IPv4-PCN-response)

```
+-------------+-------------+-------------+-------------+
|      IPv4 PCN-ingress-node Address (4 bytes)          |
+-------------+-------------+-------------+-------------+
|      IPv4 PCN-egress-node Address (4 bytes)           |
+-------------+-------------+-------------+-------------+
|      IPv4 Decision Point Address (4 bytes)            |
+-------------+-------------+-------------+-------------+
| PCN-sent-rate                                         |
+-------------+-------------+-------------+-------------+
```

o RSVP-AGGREGATE-IPv6-PCN-response: PCN object, IPv6 addresses are used:

Class = 248 (PCN)
C-Type = 4 (RSVP-AGGREGATE-IPv6-PCN-response)

```
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+     IPv6 PCN-ingress-node Address (16 bytes)          +
|                                                       |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+     IPv6 PCN-egress-node Address (16 bytes)           +
|                                                       |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
|                                                       |
+                                                       +
|                                                       |
+     Decision Point Address (16 bytes)                 +
|                                                       |
+                                                       +
|                                                       |
+-------------+-------------+-------------+-------------+
| PCN-sent-rate                                         |
+-------------+-------------+-------------+-------------+
```

The fields carried by the PCN object are specified in [RFC6663], [RFC6661], and [RFC6662]:

o The IPv4 or IPv6 address of the PCN-ingress-node (Aggregator) and the IPv4 or IPv6 address of the PCN-egress-node (Deaggregator): together, they specify the ingress-egress-aggregate to which the report refers. According to [RFC6663], the report should carry the identifier of the PCN-ingress-node (Aggregator) and the identifier of the PCN-egress-node (Deaggregator) (typically their IP addresses).

o Decision Point Address: specifies the IPv4 or IPv6 address of the Decision Point. In this document, this field MUST contain the IP address of the Deaggregator.

- "R": 1-bit flag that, when set to ON, signifies, according to [RFC6661] and [RFC6662], that the PCN-ingress-node (Aggregator) MUST provide an estimate of the rate (PCN-sent-rate) at which the PCN-ingress-node (Aggregator) is receiving PCN-traffic that is destined for the given ingress-egress-aggregate.

- "Reserved": 31 bits that are currently not used by this document and are reserved. These SHALL be set to 0 and SHALL be ignored on reception.

- PCN-sent-rate: the estimate of the rate at which the PCN-ingress-node (Aggregator) is receiving PCN-traffic that is destined for the given ingress-egress-aggregate. It is expressed in octets/second; its format is a 32-bit IEEE floating-point number. The PCN-sent-rate is specified in [RFC6661] and [RFC6662].

5. Security Considerations

The security considerations specified in [RFC2205], [RFC4860], and [RFC5559] apply to this document. In addition, [RFC4230] and [RFC6411] provide useful guidance on RSVP security mechanisms.

Security within a PCN-domain is fundamentally based on the controlled environment trust assumption stated in Section 6.3.1 of [RFC5559] -- in particular, that all PCN-nodes are PCN-enabled and are trusted to perform accurate PCN-metering and PCN-marking.

In the PCN-domain environments addressed by this document, Generic Aggregate RSVP messages specified in [RFC4860] are used for support of the PCN Controlled Load (CL) and Single Marking (SM) edge behaviors over a Diffserv cloud using Pre-Congestion Notification. Hence, the security mechanisms discussed in [RFC4860] are applicable. Specifically, the INTEGRITY object [RFC2747] [RFC3097] can be used to provide hop-by-hop RSVP message integrity, node authentication, and replay protection, thereby protecting against corruption and spoofing of RSVP messages and PCN feedback conveyed by RSVP messages.

For these reasons, this document does not introduce significant additional security considerations beyond those discussed in [RFC5559] and [RFC4860].

6. IANA Considerations

IANA has modified the "Class Names, Class Numbers, and Class Types" subregistry of the "Resource Reservation Protocol (RSVP) Parameters" registry, to add a new Class Number and assign four new C-Types under this new Class Number, as described below; see Section 4.1:

Class Number	Class Name	Reference
248	PCN	RFC 7417

Class Types or C-Types - 248 PCN

Value	Description	Reference
1	RSVP-AGGREGATE-IPv4-PCN-request	RFC 7417
2	RSVP-AGGREGATE-IPv6-PCN-request	RFC 7417
3	RSVP-AGGREGATE-IPv4-PCN-response	RFC 7417
4	RSVP-AGGREGATE-IPv6-PCN-response	RFC 7417

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997, <http://www.rfc-editor.org/info/rfc2205>.

[RFC3140] Black, D., Brim, S., Carpenter, B., and F. Le Faucheur, "Per Hop Behavior Identification Codes", RFC 3140, June 2001, <http://www.rfc-editor.org/info/rfc3140>.

[RFC3175] Baker, F., Iturralde, C., Le Faucheur, F., and B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, September 2001, <http://www.rfc-editor.org/info/rfc3175>.

[RFC4495] Polk, J. and S. Dhesikan, "A Resource Reservation Protocol (RSVP) Extension for the Reduction of Bandwidth of a Reservation Flow", RFC 4495, May 2006, <http://www.rfc-editor.org/info/rfc4495>.

[RFC4860] Le Faucheur, F., Davie, B., Bose, P., Christou, C., and M. Davenport, "Generic Aggregate Resource ReSerVation Protocol (RSVP) Reservations", RFC 4860, May 2007, <http://www.rfc-editor.org/info/rfc4860>.

[RFC5670] Eardley, P., Ed., "Metering and Marking Behaviour of PCN-Nodes", RFC 5670, November 2009, <http://www.rfc-editor.org/info/rfc5670>.

[RFC6660] Briscoe, B., Moncaster, T., and M. Menth, "Encoding Three Pre-Congestion Notification (PCN) States in the IP Header Using a Single Diffserv Codepoint (DSCP)", RFC 6660, July 2012, <http://www.rfc-editor.org/info/rfc6660>.

[RFC6661] Charny, A., Huang, F., Karagiannis, G., Menth, M., and T. Taylor, Ed., "Pre-Congestion Notification (PCN) Boundary-Node Behavior for the Controlled Load (CL) Mode of Operation", RFC 6661, July 2012, <http://www.rfc-editor.org/info/rfc6661>.

[RFC6662] Charny, A., Zhang, J., Karagiannis, G., Menth, M., and T. Taylor, Ed., "Pre-Congestion Notification (PCN) Boundary-Node Behavior for the Single Marking (SM) Mode of Operation", RFC 6662, July 2012, <http://www.rfc-editor.org/info/rfc6662>.

[RFC6663] Karagiannis, G., Taylor, T., Chan, K., Menth, M., and P. Eardley, "Requirements for Signaling of Pre-Congestion Information in a Diffserv Domain", RFC 6663, July 2012, <http://www.rfc-editor.org/info/rfc6663>.

7.2. Informative References

[RFC1633] Braden, R., Clark, D., and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994, <http://www.rfc-editor.org/info/rfc1633>.

[RFC2211] Wroclawski, J., "Specification of the Controlled-Load Network Element Service", RFC 2211, September 1997, <http://www.rfc-editor.org/info/rfc2211>.

[RFC2212] Shenker, S., Partridge, C., and R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212, September 1997, <http://www.rfc-editor.org/info/rfc2212>.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998, <http://www.rfc-editor.org/info/rfc2474>.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998, <http://www.rfc-editor.org/info/rfc2475>.

[RFC2747] Baker, F., Lindell, B., and M. Talwar, "RSVP Cryptographic Authentication", RFC 2747, January 2000, <http://www.rfc-editor.org/info/rfc2747>.

[RFC2753] Yavatkar, R., Pendarakis, D., and R. Guerin, "A Framework for Policy-based Admission Control", RFC 2753, January 2000, <http://www.rfc-editor.org/info/rfc2753>.

[RFC2998] Bernet, Y., Ford, P., Yavatkar, R., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J., and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks", RFC 2998, November 2000, <http://www.rfc-editor.org/info/rfc2998>.

[RFC3097] Braden, R. and L. Zhang, "RSVP Cryptographic Authentication -- Updated Message Type Value", RFC 3097, April 2001, <http://www.rfc-editor.org/info/rfc3097>.

[RFC4230] Tschofenig, H. and R. Graveman, "RSVP Security Properties", RFC 4230, December 2005, <http://www.rfc-editor.org/info/rfc4230>.

[RFC4923] Baker, F. and P. Bose, "Quality of Service (QoS) Signaling in a Nested Virtual Private Network", RFC 4923, August 2007, <http://www.rfc-editor.org/info/rfc4923>.

[RFC5559] Eardley, P., Ed., "Pre-Congestion Notification (PCN) Architecture", RFC 5559, June 2009, <http://www.rfc-editor.org/info/rfc5559>.

[RFC6411] Behringer, M., Le Faucheur, F., and B. Weis, "Applicability of Keying Methods for RSVP Security", RFC 6411, October 2011, <http://www.rfc-editor.org/info/rfc6411>.

[RSVP-PCN-CL] Le Faucheur, F., Charny, A., Briscoe, B., Eardley, P., Babiarz, J., and K. Chan, "RSVP Extensions for Admission Control over Diffserv using Pre-congestion Notification (PCN)", Work in Progress, draft-lefaucheur-rsvp-ecn-01, June 2006.

Appendix A. Example Signaling Flow

This appendix is based on Appendix A of [RFC4860]. In particular, it provides an example signaling flow of the specifications detailed in Sections 3 and 4.

This signaling flow assumes an environment where E2E reservations are aggregated over generic aggregate RSVP reservations and applied over a PCN-domain. In particular, the Aggregator (PCN-ingress-node) and Deaggregator (PCN-egress-node) are located at the boundaries of the PCN-domain. The PCN-interior-nodes are located within the PCN-domain, between the PCN-boundary-nodes, but are not shown in the diagram below. It illustrates a possible RSVP message flow that could take place in the successful establishment of a unicast E2E reservation that is the first between a given Aggregator-Deaggregator pair.

```
        Aggregator (PCN-ingress-node)     Deaggregator (PCN-egress-node)

      E2E Path
     ----------->
                  (1)
                             E2E Path
                     ------------------------------->
                                                          (2)
                      E2E PathErr(NEW-AGGREGATE-NEEDED,SOI=GApcn)
                     <----------------------------------------
   (3)
                           AggPath(Session=GApcn)
                     ------------------------------->
   (4)
                                                            E2E Path
                                                           ----------->
                                                    (5)
                           AggResv (Session=GApcn) (PCN object)
                     <-------------------------------
   (6)
                       AggResvConfirm (Session=GApcn)
                     ------------------------------>
   (7)
                                                            E2E Resv
                                                           <---------
                                                    (8)
                             E2E Resv (SOI=GApcn)
                     <-----------------------------
                  (9)
        E2E Resv
     <-----------
```

(1) The Aggregator forwards E2E Path into the aggregation region after modifying its IP protocol number to RSVP-E2E-IGNORE.

(2) Let's assume that no Aggregate Path exists. To be able to accurately update the ADSPEC of the E2E Path, the Deaggregator needs the ADSPEC of Aggregate Path. In this example, the Deaggregator elects to instruct the Aggregator to set up an Aggregate Path state for the PCN PHB-ID. To do that, the Deaggregator sends an E2E PathErr message with a NEW-AGGREGATE-NEEDED PathErr code.

The PathErr message also contains a SESSION-OF-INTEREST (SOI) object. The SOI contains a GENERIC-AGGREGATE SESSION (GApcn) whose PHB-ID is set to the PCN PHB-ID. The GENERIC-AGGREGATE SESSION contains an interface-independent Deaggregator address inside the DestAddress and appropriate values inside the vDstPort and Extended vDstPort fields. In this document, the Extended vDstPort SHOULD contain the IPv4 or IPv6 address of the Aggregator.

(3) The Aggregator follows the request from the Deaggregator and signals an Aggregate Path for the GENERIC-AGGREGATE SESSION (GApcn).

(4) The Deaggregator takes into account the information contained in the ADSPEC from both Aggregate Paths and updates the E2E Path ADSPEC accordingly. The PCN-egress-node MUST NOT perform the RSVP-TTL vs. IP TTL-check and MUST NOT update the ADSPEC Break bit. This is because the whole PCN-domain is effectively handled by E2E RSVP as a virtual link on which integrated service is indeed supported (and admission control performed) so that the Break bit MUST NOT be set; see also [RSVP-PCN-CL]. The Deaggregator also modifies the E2E Path IP protocol number to RSVP before forwarding it.

(5) In this example, the Deaggregator elects to immediately proceed with establishment of the generic aggregate reservation. In effect, the Deaggregator can be seen as anticipating the actual demand of E2E reservations so that the generic aggregate reservation is in place when the E2E Resv request arrives, in order to speed up establishment of E2E reservations. Here it is also assumed that the Deaggregator includes the optional ResvConfirm Request in the Aggregate Resv message.

(6) The Aggregator merely complies with the received ResvConfirm Request and returns the corresponding Aggregate ResvConfirm.

(7) The Deaggregator has explicit confirmation that the generic aggregate reservation is established.

(8) On receipt of the E2E Resv, the Deaggregator applies the mapping policy defined by the network administrator to map the E2E Resv onto a generic aggregate reservation. Let’s assume that this policy is such that the E2E reservation is to be mapped onto the generic aggregate reservation with the PCN PHB-ID=x. After the previous step (7), the Deaggregator knows that a generic aggregate reservation (GApcn) is in place for the corresponding PHB-ID. At this step, the Deaggregator maps the generic aggregate reservation onto one ingress-egress-aggregate maintained by the Deaggregator (as a PCN-egress-node); see Section 3.7. The Deaggregator performs admission control of the E2E Resv onto the generic aggregate reservation for the PCN PHB-ID (GApcn). The Deaggregator also takes into account the PCN admission control procedure as specified in [RFC6661] and [RFC6662]; see Section 3.7. If one or both of the admission control procedures (the PCN-based admission control procedure described in Section 3.3.1 of [RFC6661] or [RFC6662], and the admission control procedure specified in [RFC4860]) are not successful, then the E2E Resv is not admitted onto the associated RSVP generic aggregate reservation for the PCN PHB-ID (GApcn). Otherwise, assuming that the generic aggregate reservation for the PCN (GApcn) had been established with sufficient bandwidth to support the E2E Resv, the Deaggregator adjusts its counter, tracking the unused bandwidth on the generic aggregate reservation. Then it forwards the E2E Resv to the Aggregator, including a SESSION-OF-INTEREST object conveying the selected mapping onto GApcn (and hence onto the PCN PHB-ID).

(9) The Aggregator records the mapping of the E2E Resv onto GApcn (and onto the PCN PHB-ID). The Aggregator removes the SOI object and forwards the E2E Resv towards the sender.

Acknowledgments

We would like to thank the authors of [RSVP-PCN-CL], since some ideas used in this document are based on the work initiated in [RSVP-PCN-CL]. Moreover, we would like to thank Bob Briscoe, David Black, Ken Carlberg, Tom Taylor, Philip Eardley, Michael Menth, Toby Moncaster, James Polk, Scott Bradner, Lixia Zhang, and Robert Sparks for the provided comments. In particular, we would like to thank Francois Le Faucheur for contributing a significant amount of text, in addition to his comments.

Authors' Addresses

Georgios Karagiannis
Huawei Technologies
Hansaallee 205
40549 Dusseldorf
Germany

EMail: Georgios.Karagiannis@huawei.com

Anurag Bhargava
Cisco Systems, Inc.
7100-9 Kit Creek Road
PO Box 14987
Research Triangle Park, NC 27709-4987
United States

EMail: anuragb@cisco.com