

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: September 10, 2012

Georgios Karagiannis
University of Twente
Anurag Bhargava
Cisco Systems, Inc.
March 10, 2012

**Generic Aggregation of Resource ReSerVation Protocol (RSVP)
for IPv4 And IPv6 Reservations over PCN domains
draft-ietf-tsvwg-rsvp-pcn-01**

Abstract

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

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 10, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

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](#)].

Table of Contents

1.	Introduction	4
1.1.	Terminology	6
2.	Overview of RSVP extensions and Operations	9
2.1	Overview of RSVP Aggregation Procedures in PCN domains	9
2.1.1	PCN Marking and encoding and transport of pre-congestion Information	11
2.1.2.	Traffic Classification Within The Aggregation Region	11
2.1.3.	Deaggregator (PCN-egress-node) Determination	11
2.1.4.	Mapping E2E Reservations Onto Aggregate Reservations	11
2.1.5.	Size of Aggregate Reservations	12
2.1.6.	E2E Path ADSPEC update	12
2.1.7.	Intra-domain Routes	12
2.1.8.	Inter-domain Routes	12
2.1.9.	Reservations for Multicast Sessions	12
2.1.10.	Multi-level Aggregation	12
2.1.11.	Reliability Issues	12
2.1.12.	Message Integrity and Node Authentication	12
3.	Elements of Procedure	13
3.1.	Receipt of E2E Path Message By PCN-ingress-node (aggregating router)	13
3.2.	Handling Of E2E Path Message By Interior Routers	14
3.3.	Receipt of E2E Path Message By PCN-egress-node (deaggregating router)	14
3.4.	Initiation of new Aggregate Path Message By PCN-ingress node (Aggregating Router)	14
3.5.	Handling Of new Aggregate Path Message By Interior Routers	15

3.6.	Handling of E2E Resv Message by Deaggregating Router	15
3.7.	Handling Of E2E Resv Message By Interior Routers	15
3.8.	Initiation of New Aggregate Resv Message By Deaggregating Router	15

3.9.	Handling of Aggregate Resv Message by Interior Routers	15
3.10.	Handling of E2E Resv Message by Aggregating Router	16
3.11.	Handling of Aggregated Resv Message by Aggregating Router . .	16
3.12.	Removal of E2E Reservation	16
3.13.	Removal of Aggregate Reservation	17
3.14.	Handling of Data On Reserved E2E Flow by Aggregating Router .	17
3.15.	Procedures for Multicast Sessions	17
4.	Protocol Elements	17
4.1	PCN object	18
5.	Security Considerations	24
6.	IANA Considerations	24
7.	Acknowledgments	24
8.	Normative References	24
9.	Informative References	25
10.	Authors' Address	26

1. Introduction

Two main Quality of Service (QoS) architectures have been specified By the IETF. These are the Integrated Services (Intserv) [[RFC1633](#)] architecture and the Differentiated Services (DiffServ) architecture ([[RFC2475](#)]).

Intserv provides methods for the delivery of end-to-end Quality of Service (QoS) to applications over heterogeneous networks. One of the QoS signaling protocols used by the Intserv architecture is the Resource reServation Protocol (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 quality of service (QoS) closely approximating 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, policy-based admission control, assistance in traffic identification/classification, and traffic conditioning. Intserv over Diffserv can operate over a statically provisioned Diffserv region or RSVP aware. When it is RSVP aware, several mechanisms may be used to support dynamic provisioning and topology-Aware admission control, including aggregate RSVP reservations, per-flow RSVP, or a bandwidth broker.

[RFC 3175](#) [[RFC3175](#)] specifies aggregation of Resource ReSerVation Protocol (RSVP) end-to-end reservations over aggregate RSVP reservations. In [[RFC3175](#)] the RSVP aggregated reservation is characterized by a RSVP SESSION object using the 3-tuple <source IP address, destination IP address, Diffserv Code Point>.

[[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 (or set of PHBs) 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 of 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 Code Point (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. This is among others used to specify whether the Diffserv resources belong to a single PHB or to a set of PHBs.

The main objective of Pre-Congestion Notification (PCN) is to 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, based on which they take their decisions. The decision points may be collocated with the PCN-ingress-nodes or their function may be implemented in a centralized node. For more details see [\[RFC5559\]](#), [\[draft-ietf-pcn-cl-edge-behaviour-12\]](#), [\[draft-ietf-pcn-sm-edge-behaviour-09\]](#). In this document it is considered that the decision point is collocated with the PCN-ingress-node.

This document follows the PCN signaling requirements defined in [\[draft-ietf-pcn-signaling-requirements-08.txt\]](#) and specifies the extensions to the Generic Aggregated RSVP [\[RFC4860\]](#) for the support of PCN edge behaviors as specified in [\[draft-ietf-pcn-cl-edge-behaviour-12\]](#) and [\[draft-ietf-pcn-sm-edge-behaviour-09\]](#). Moreover, this document specifies how RSVP aggregation can be used to setup and maintain: (1) Ingress Egress Aggregate (IEA) states at Ingress and Egress nodes and (2) generic aggregation of RSVP end-to-end RSVP reservations over PCN (Congestion and Pre-Congestion Notification) domains.

This document, and according to [[RFC4860](#)] MAY also be used end-to-end directly by end-systems attached to a Diffserv network.

Furthermore, this document and according to [\[RFC4860\]](#), in absence of e2e RSVP flows, a variety of policies (not defined in this document) can be used at the Aggregator to set the DSCP of packets passing into the aggregation region and how they are mapped onto generic aggregate reservations. These policies are not described in this document but are a matter of local configuration.

In this document it is considered that the PCN-nodes MUST be able to support the functionality specified in [\[RFC5670\]](#), [\[RFC5559\]](#), [\[RFC5696\]](#), [\[draft-ietf-pcn-cl-edge-behaviour-12\]](#), [\[draft-ietf-pcn-sm-edge-behaviour-09\]](#). Furthermore, the PCN-boundary-nodes MUST support the RSVP generic aggregated reservation procedures specified in [\[RFC4860\]](#) which are augmented with procedures specified in this document.

1.1. Terminology

This document uses terms defined in [\[RFC4860\]](#), [\[RFC3175\]](#), [\[RFC5559\]](#), [\[RFC5670\]](#), [\[draft-ietf-pcn-cl-edge-behaviour-12\]](#), [\[draft-ietf-pcn-sm-edge-behaviour-09\]](#).

For readability, a number of definitions from [\[RFC3175\]](#) as well as definitions for terms used in [\[RFC5559\]](#), [\[draft-ietf-pcn-cl-edge-behaviour-12\]](#), and [\[draft-ietf-pcn-sm-edge-behaviour-09\]](#) are provided here, where some of them are augmented with new meanings:

Aggregator	This is the process in (or associated with) the router at the ingress edge of the aggregation region (with respect to the end-to-end RSVP reservation) and behaving in accordance with [RFC4860] . In this document, it is also the PCN-ingress-node and the decision point.
Deaggregator	This is the process in (or associated with) the router at the egress edge of the aggregation region (with respect to the end-to-end RSVP reservation) and behaving in accordance with [RFC4860] . In this document, it is also the PCN-egress-node.
E2E	End to end
E2E Reservation	This is an RSVP reservation such that: <ul style="list-style-type: none">(i) corresponding RSVP Path messages are initiated upstream of the Aggregator and terminated downstream of the Deaggregator, and(ii) corresponding RSVP Resv messages are initiated downstream of the Deaggregator and terminated upstream of the Aggregator, and

(iii) 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, see [[RFC4860](#)]). As per regular RSVP operations, E2E RSVP reservations are unidirectional.

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 \[RFC3140\]](#).

VDstPort (Virtual Destination Port)

A 16-bit identifier used in the SESSION that remains constant over the life of the generic aggregate reservation.

Extended vDstPort (Extended Virtual Destination Port)

A 32-bit identifier used in the SESSION that remains constant over the life of the generic aggregate reservation. 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.

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 for the PCN-domain; includes the PCN-egress-nodes, which measure these PCN-marks, and the PCN-ingress-nodes.

PCN-boundary-node: a PCN-node that connects one PCN-domain to a node either in another PCN-domain or in a non-PCN-domain.

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-egress-node: a PCN-boundary-node in its role in handling traffic as it leaves a PCN-domain.
- 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 operates also as a Decision Point and aggregator.
- 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 ECN fields.
- PCN-flow: the unit of PCN-traffic that the PCN-boundary-node admits (or terminates); the unit could be a single microflow (as defined in [[RFC2474](#)]) or some identifiable collection of microflows.
- 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. An ingress-egress-aggregate is identified by the combination of (1) fields), (2) IP addresses of the specific pair of PCN-boundary-nodes used by a ingress-egress-aggregate. In this document the ingress-egress-aggregate is associated with a RSVP generic aggregated reservation state [[RFC4860](#)].
- PCN-admission-state
The state ("admit" or "block") derived by the Decision Point (PCN-ingress-node) for a given ingress-egress-aggregate based on PCN packet marking statistics. The Decision Point decides to admit or block new flows offered to the aggregate based on the current value of the PCN-admission-state.
- 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.

T-meas

A configurable time interval that defines the measurement period over which the PCN-egress-node collects statistics relating to PCN-traffic marking.

At the end of the interval the PCN-egress-node calculates the values NM-rate, ThM-rate, and ETM-rate as defined and sends a report to the Decision Point, subject to the operation of the Report suppression feature.

T-maxsuppress

A configurable time interval after which the PCN-egress-node MUST send a report to the Decision Point for a given ingress-egress-aggregate regardless of the most recent values of the CLE. This mechanism provides the Decision Point with a Periodic confirmation of liveness when report suppression is activated.

T-fail

A configurable interval after which the Decision Point Concludes that communication from a given PCN-egress-node has failed if it has received no reports from the PCN-egress-node during that interval.

t-recvFail

An ingress-egress-aggregate timer that is used at The Decision point (in this document at the PCN-ingress-node) which when expires raises an alarm to management, and activates the PCN-ingress-node to block the admission of new PCN-flows. This timer expires when its value is equal to T-fail and is reset when a report, i.e., RSVP aggregated RESV message, is received for the ingress-egress-aggregate.

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 aggregated reservations [RFC4860], which are depending on ingress-egress-aggregates. In particular, an ingress-egress-aggregate matches to only one RSVP SESSION for generic aggregated reservations. However, a RSVP SESSION for generic aggregated reservations can match to one or more than one ingress-egress-aggregates. This can be

accomplished by using for the different ingress-egress-aggregates the same combinations of ingress and egress identifiers, but with a different PHB-ID value (see [[RFC4860](#)]).

Moreover, each Aggregator and Deaggregator (i.e., PCN-boundary-nodes) MUST support policies to initiate and maintain for each combination of the PCN-boundary-node and all other PCN-boundary-nodes of the same PCN-domain one RSVP SESSION for generic aggregated reservations. Note that RSVP SESSION for generic aggregated reservations can match to one or more than one ingress-egress-aggregates. This can be accomplished by using for the different ingress-egress-aggregates the same combinations of ingress and egress identifiers, but with a different PHB-ID value (see [\[RFC4860\]](#)). Depending on a policy the Aggregator SHOULD be able to decide whether an e2e RSVP session can be mapped into one ingress-egress-aggregate maintained by the

Aggregator (i.e., PCN-ingress-node).

The RSVP SESSION object for generic aggregate reservations, maintains the mapping and association between the PCN ingress-egress-aggregate and the PCN-flows (e2e RSVP reservation session) that travel in one direction between the specific pair of PCN-boundary-nodes specified by the ingress-egress-aggregate. Note that in this document the PCN ingress-egress-aggregate is identified by using the RSVP SESSION object for generic aggregate reservation, see [[RFC4860](#)], by using the following:

- o) the IPv4 DestAddress, IPv6 DestAddress SHOULD be set to the IPv4 or IPv6 destination addresses, respectively, of the Deaggregator (PCN-egress-node)
- o) PHB-ID (Per Hop Behavior Identification Code) SHOULD be set equal to PCN-compatible Diffserv codepoint(s).
- o) Extended vDstPort SHOULD be set to the IPv4 or IPv6 destination addresses, of the Aggregator (PCN-ingress-node)

2.1.1 PCN Marking and encoding and transport of pre-congestion information

The method of PCN marking within the PCN domain is based on [[RFC5670](#)]. In addition, the method of encoding and transport of pre-congestion information is based [[RFC5696](#)]. The PHB-ID (Per Hop Behavior Identification Code) used, SHOULD be set equal to PCN-compatible Diffserv codepoint(s).

2.1.2. Traffic Classification Within The Aggregation Region

The PCN-traffic is marked using PCN-marking and is classified using The PCN-BA (i.e., combination of the DSCP and ECN fields).

The PCN-traffic belonging to an PCN aggregated session can be classified only at the PCN-boundary-nodes using the combination of (1) PCN-BA (i.e., combination of the DSCP and ECN fields), (2) IP addresses of the specific pair of PCN-boundary-nodes used by a ingress-egress-aggregate.

The method of classification and traffic conditioning of PCN-traffic and non-PCN traffic and PHB configuration is described in [[draft-ietf-pcn-cl-edge-behaviour-12](#)] and [[draft-ietf-pcn-sm-edge-behaviour-09](#)].

2.1.3. Deaggregator (PCN-egress-node) Determination

In this document it is considered that for the determination of the Deaggregator, the same methods can be used as the ones described in [[RFC4860](#)].

2.1.4. Mapping E2E Reservations Onto Aggregate Reservations

In this document it is considered that for the mapping of e2e

reservations onto aggregate reservations, the same methods can be used as the ones described in [[RFC4860](#)], augmented by the following rules:

- o) PCN-ingress-node MUST use one or more policies to estimate whether an e2e RSVP reservation session associated with an e2e Path message that arrives at the external interface of the PCN-ingress-node can be mapped onto an existing RSVP generic aggregation reservation state, i.e., PCN ingress-egress-aggregate.

2.1.5. Size of Aggregate Reservations

In this document it is considered that for the determination of the size of the aggregate reservations, the same methods can be used as the ones described in [[RFC4860](#)].

2.1.6. E2E Path ADSPEC update

In this document it is considered that for the update of the e2e Path ADSPEC, the same methods can be used as the ones described in [[RFC4860](#)].

2.1.7. Intra-domain Routes

The PCN-interior-nodes are neither maintaining 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.

2.1.8. Inter-domain Routes

In this document it is considered that for the solving the issues caused by the inter-domain route changes, the same methods can be used as the ones described in [[RFC4860](#)].

2.1.9. Reservations for Multicast Sessions

PCN does not consider reservations for multicast sessions.

2.1.10. Multi-level Aggregation

PCN does not consider multi-level aggregations within the PCN domain.

2.1.11. Reliability Issues

In this document it is considered that for solving possible reliability issues, the same methods can be used as the ones described in [[RFC4860](#)].

2.1.12. Message Integrity and Node Authentication

In this document it is considered that for message integrity and node

authentication, the same methods can be used as the ones described in [\[RFC4860\]](#) and [\[RFC5559\]](#).

3. Elements of Procedure

This section describes the procedures used to implement the aggregated RSVP procedure over PCN.

3.1. Receipt of E2E Path Message By PCN-ingress-node (aggregating router)

When the e2e RSVP message arrives at the exterior interface of the aggregator, i.e., PCN-ingress-node, then standard RSVP generic aggregation [[RFC4860](#)] procedures are used, augmented with the following rules:

- o) The e2e RSVP reservation session associated with an e2e Path message that arrives at the external interface of the PCN-ingress-node is mapped onto an existing RSVP generic aggregation reservation state (i.e., PCN ingress-egress-aggregate).
- o) If the timer t-recvFail expires for a given PCN-egress-node, the Decision Point (i.e., PCN-ingress-node) SHOULD NOT allow the e2e RSVP flow to be admitted to that ingress-egress-aggregate. This procedure is defined in detail in: [[draft-ietf-pcn-cl-edge-behaviour-12](#)] and [[draft-ietf-pcn-sm-edge-behaviour-09](#)].

Depending on a local policy the Aggregator SHOULD decide whether this situation is considered of being an error, or whether the e2e reservation session SHOULD be mapped to another ingress-egress-aggregate maintained by the same RSVP SESSION for aggregated reservations.

If the Aggregator is not able to map the requesting e2e RSVP session into another ingress-egress-aggregate, then the Aggregator SHOULD NOT admit the e2e RSVP session and it SHOULD generate an e2e PathErr message using standard e2e RSVP procedures [[RFC2205](#)]. This e2e PathErr message is sent to the originating sender of the e2e Path message.

- o) If the timer t-recvFail does NOT expire for a given PCN-egress-node, then:
 - *) If the PCN-admission state for the ingress-egress-aggregate associated with the received e2e Path is "admit", the Decision Point (i.e., PCN-ingress-node) SHOULD allow new flows to be admitted to that aggregate. The e2e Path message is then forwarded towards destination.

- *) If the PCN-admission-state for the same PCN aggregation state is "block", the Aggregator using the same policy as mentioned above SHOULD either map the incoming e2e RSVP session to another ingress-egress-aggregate associated with the same generic aggregated RSVP session, or the flow SHOULD NOT be admitted and an e2e PathErr message SHOULD be generated, using standard e2e RSVP procedures [[RFC2205](#)], [[RFC4495](#)].

This e2e PathErr message is sent to the originating sender of the e2e Path message, using standard e2e RSVP procedures [[RFC2205](#)], [[RFC4495](#)]. A new error code "PCN-domain rejects e2e reservation" MUST be augmented to the RSVP error codes to inform the sender that a PCN domains rejects the e2e reservation request.

The way of how the PCN-admission-state is maintained is specified in [[draft-ietf-pcn-cl-edge-behaviour-12](#)] and [[draft-ietf-pcn-sm-edge-behaviour-09](#)].

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. 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 PCN-egress-node (deaggregating router) performs main regular [[RFC4860](#)] procedures, augmented with the following rules, see also [[draft-lefaucheur-rsvp-ecn-01](#)]:

- o) 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.

The PCN-egress-nodes forwards the e2e Path message towards the receiver.

3.4. Initiation of new Aggregate Path Message By PCN-ingress node (Aggregating Router)

In this document it is considered that for the initiation of the new RSVP aggregated Path message by the PCN-ingress-node (Aggregation Router), the same methods can be used as the ones described in

[[RFC4860](#)].

Karagiannis, et al. Expires September 10, 2012

[Page 14]

3.5. Handling Of new Aggregate Path Message By Interior Routers

The Aggregate 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. The Aggregated Path messages are simply forwarded as normal IP datagrams.

3.6. Handling of E2E Resv Message by Deaggregating Router

When the e2e Resv message arrives at the exterior interface of the Deaggregating router, i.e., PCN-egress-node, then standard RSVP aggregation [[RFC4860](#)] procedures are used.

3.7. Handling Of E2E Resv Message By Interior Routers

The e2e Resv messages traverse zero or more PCN-interior-nodes. The PCN-interior-nodes receive the e2e Resv message on an interior interface and forward it on another interior interface. The e2e Resv messages are simply forwarded as normal IP datagrams.

3.8. Initiation of New Aggregate Resv Message By Deaggregating Router

In this document it is considered that for the initiation of the new RSVP aggregated Resv message by the PCN-ingress-node (Aggregation Router), the same methods can be used as the ones described in [[RFC4860](#)] augmented with the following rules:

- o) At the end of each t-meas measurement interval, or less frequently if "optional report suppression" is activated, see [[draft-ietf-pcn-cl-edge-behaviour-12](#)], and [[draft-ietf-pcn-sm-edge-behaviour-09](#)], the PCN-egress-node MUST include the new PCN object that will be sent to the associated Decision Point (i.e., PCN-ingress-node). The PCN object is specified in this document and is used to report of the data measured by the PCN-egress-node, for a particular ingress-egress-aggregate, see [[draft-ietf-pcn-cl-edge-behaviour-12](#)], and [[draft-ietf-pcn-sm-edge-behaviour-09](#)]. The address of the PCN-ingress-node is the one specified in the same ingress-egress-aggregate.

3.9. Handling of Aggregate Resv Message by Interior Routers

The Aggregated Resv messages traverse zero or more PCN-interior-nodes. The PCN-interior-nodes receive the Aggregated Resv message on an interior interface and forward it on another interior interface. The Aggregated Resv messages are simply forwarded as normal IP datagrams.

3.10. Handling of E2E Resv Message by Aggregating Router

When the e2e Resv message arrives at the interior interface of the Aggregating router, i.e., PCN-ingress-node, then standard RSVP aggregation [[RFC4860](#)] procedures are used.

3.11. Handling of Aggregated Resv Message by Aggregating Router

When the Aggregated Resv message arrives at the interior interface of the Aggregating router, i.e., PCN-ingress-node, then standard RSVP aggregation [[RFC4860](#)] procedures are used, augmented with the following rules:

- o) the Decision Point (i.e., the PCN-ingress-node) SHOULD use the information carried by the PCN objects as specified in [[draft-ietf-pcn-cl-edge-behaviour-12](#)], [[draft-ietf-pcn-sm-edge-behaviour-09](#)].

When the Aggregator (i.e., PCN-ingress-node) needs to terminate an amount of traffic associated to one ingress-egress-aggregate (see bullet 2 in Section 3.3.2 of [[draft-ietf-pcn-cl-edge-behaviour-12](#)] and [[draft-ietf-pcn-sm-edge-behaviour-09](#)]), then the following procedure is followed. Based on a local policy, the Aggregator SHOULD select one of the following options:

- o) for the same ingress-egress-aggregate, select a number of e2e RSVP sessions 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, see above. In this situation the same mechanisms for terminating an e2e RSVP flow can be followed as specified in [[RFC4495](#)].
- o) for the same ingress-egress-aggregate, select a number of e2e RSVP sessions 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, see above. In this situation the same mechanisms for terminating an e2e RSVP flow or reducing bandwidth associated with an e2e RSVP flow can be followed as specified in [[RFC4495](#)].

3.12. Removal of E2E Reservation

In this document it is considered that for the removal of e2e reservations, the same methods can be used as the ones described in [[RFC4860](#)] and [[RFC4495](#)].

3.13. Removal of Aggregate Reservation

In this document it is considered that for the removal of aggregated reservations, the same methods can be used as the ones described in [\[RFC4860\]](#).

3.14. Handling of Data On Reserved E2E Flow by Aggregating Router

The handling of data on the reserved e2e Flow by Aggregating Router is using the procedures described in [\[RFC4860\]](#) augmented with:

- o) Regarding, PCN marking and traffic classification the procedures defined in [Section 2.1.1](#) and 2.1.3 of this document are used.

3.15. Procedures for Multicast Sessions

In this document no multicast sessions are considered.

4. Protocol Elements

The protocol elements in this document are using the protocol Elements defined in [\[RFC4860\]](#), augmented with the following rules:

- o) A PCN-egress-node (i.e., deaggregator) SHOULD send periodically and at the end of each t-meas measurement interval, or less frequently if "optional report suppression" is activated, an (refresh) aggregated RSVP message to the PCN-ingress-node (i.e. aggregator).
- o) the DSCP value included in the SESSION object, SHOULD be set equal to a PCN-compatible Diffserv codepoint.
- o) An aggregated Resv message MUST carry one or more PCN objects, see [Section 4.1](#), to report the data measured by an PCN-egress-node (i.e., Deaggregator).
- o) As described in [\[draft-ietf-pcn-cl-edge-behaviour-12\]](#), [\[draft-ietf-pcn-signaling-requirements-08\]](#), PCN reports from the PCN-egress-node (Deaggregator) to the decision point may contain flow identifiers for individual flows within an ingress-egress-aggregate that have recently experienced excess-marking. Hence, the PCN report messages used by the PCN CL edge behavior MUST be capable of carrying sequences of octet strings constituting such identifiers. When the PCN CL edge behavior is used, the individual flow identifiers need to be included in specific PCN objects, see [Section 4.1](#) (C-Type = RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs,

= RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs)

Karagiannis, et al. Expires September 10, 2012

[Page 17]

Open issue:

=====

There are at least two possible options of carrying the PCN objects of C-Type: RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs or RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs:

- o) Option 1: The PCN objects of C-Type: RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs or RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs MUST be carried by the aggregated Resv message together with the other PCN object C-Types. The advantage of this object is that no additional message needs to be supported by this signaling protocol. The drawback of this option is that the PCN objects of C-Type: RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs or RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs can become larger than the maximum transmission unit (MTU) along a path to the Aggregator.

- o) Option 2: The PCN objects of C-Type: RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs or RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs MUST be carried by NOTIFY messages, see [[RFC3473](#)]. In particular, the NOTIFY <flow descriptor list> field could carry the flow IDs. The advantage of this option is that the total list of the flow IDs that need to be sent to the Aggregator can be divided in smaller sets. Each of these sets can be then carried by one NOTIFY message. The number of flow IDs that are included in such a set MUST be such that the length of any NOTIFY message will not become larger than the maximum transmission unit (MTU) along a path to the Aggregator. The main disadvantage is the signaling protocol needs to use an additional message type. If this option is chosen then the format of the PCN objects of C-Type: RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs or RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs may need modifications. The same holds for the procedures on handling the NOTIFY message by the Interior nodes and by the Aggregator.

4.1 PCN object

The PCN object reports data measured by an PCN-egress-node.

PCN objects are defined for different PCN edge behavior drafts. This document defines several types of PCN objects.

- o) Single Marking (SM) PCN object, when IPv4 addresses are used:
 - Class = PCN
 - C-Type = RSVP-AGGREGATE-IPv4-PCN-SM


```

+-----+-----+-----+-----+
| IPv4 PCN-ingress-node Address (4 bytes) |
+-----+-----+-----+-----+
| IPv4 PCN-egress-node Address (4 bytes) |
+-----+-----+-----+-----+
| rate of not marked PCN-traffic (NM-rate) |
+-----+-----+-----+-----+
| rate of PCN-marked PCN-traffic (PM-rate) |
+-----+-----+-----+-----+

```

- o) Single Marking (SM) PCN object, when IPv6 addresses are used:

Class = PCN

C-Type = RSVP-AGGREGATE-IPv6-PCN-SM

```

+-----+-----+-----+-----+
|
+
|
+   IPv6 PCN-ingress-node Address (16 bytes)
+
|
+
+-----+-----+-----+-----+
|
+
|
+   IPv6 PCN-egress-node Address (16 bytes)
+
|
+
+-----+-----+-----+-----+
| rate of not marked PCN-traffic (NM-rate) |
+-----+-----+-----+-----+
| rate of PCN-marked PCN-traffic (PM-rate) |
+-----+-----+-----+-----+

```

- o) Controlled (CL) PCN object, IPv4 addresses are used:

Class = PCN

C-Type = RSVP-AGGREGATE-IPv4-PCN-CL


```

+-----+-----+-----+-----+
| IPv4 PCN-ingress-node Address (4 bytes) |
+-----+-----+-----+-----+
| IPv4 PCN-egress-node Address (4 bytes) |
+-----+-----+-----+-----+
| rate of not marked PCN-traffic (NM-rate) |
+-----+-----+-----+-----+
| rate of threshold-marked PCN-traffic (ThM-rate) |
+-----+-----+-----+-----+
| rate of excess-traffic-marked PCN-traffic (ETM-rate) |
+-----+-----+-----+-----+

```

o) Controlled (CL) PCN object, IPv6 addresses are used:

Class = PCN

C-Type = RSVP-AGGREGATE-IPv6-PCN-CL

```

+-----+-----+-----+-----+
|
+
|
+ IPv6 PCN-ingress-node Address (16 bytes)
+
|
+-----+-----+-----+-----+
|
+
|
+ IPv6 PCN-egress-node Address (16 bytes)
+
|
+-----+-----+-----+-----+
| rate of not marked PCN-traffic (NM-rate) |
+-----+-----+-----+-----+
| rate of threshold-marked PCN-traffic (ThM-rate) |
+-----+-----+-----+-----+
| rate of excess-traffic-marked PCN-traffic (ETM-rate) |
+-----+-----+-----+-----+

```

The fields carried by the PCN object are specified in [\[draft-ietf-pcn-signaling-requirements-08.txt\]](#), [\[draft-ietf-pcn-cl-edge-behaviour-12\]](#) and [\[draft-ietf-pcn-sm-edge-behaviour-09\]](#):

- o the IPv4 or IPv6 address of the PCN-ingress-node and the IPv4 or IPv6 address of the PCN-egress-node; together they specify the ingress-egress-aggregate to which the report refers;

- o rate of not-marked PCN-traffic (NM-rate) in octets/second; its format is a 32-bit IEEE floating point number;
 - o rate of PCN-marked traffic (PM-rate) in octets/second; its format is a 32-bit IEEE floating point number;
 - o rate of threshold-marked PCN traffic (ThM-rate) in octets/second; its format is a 32-bit IEEE floating point number;
 - o rate of excess-traffic-marked traffic (ETM-rate) in octets/second; its format is a 32-bit IEEE floating point number;
- o) Controlled (CL) PCN CL Flow IDs object, IPv4 addresses are used:
 Class = PCN
 C-Type = RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs

```

0          1          2          3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
                                     +-+-+-+-+
                                     | Length |
+-+-+-+-+
|                                     |
|          Source Address          |
+-+-+-+-+
|                                     |
|          Destination Address     |
+-+-+-+-+
|          Source Port             |          Destination Port      |
+-+-+-+-+
| Protocol |          Reserved     |
+-+-+-+-+
//                                     //
+                                     +
|                                     |
|          Source Address          |
+-+-+-+-+
|                                     |
|          Destination Address     |
+-+-+-+-+
|          Source Port             |          Destination Port      |
+-+-+-+-+
| Protocol |          Reserved     |
+-+-+-+-+

```


- o) Length (1 byte): the length of the RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs object in units of 16 bytes. This field is used to specify the number of IPv4 flow IDs carried by this object. Each flow ID is represented by the combination of each subsequent 5 tuple: Source address, Destination address, Source Port, Destination Port and Protocol number. If Length is 0 then the RSVP-AGGREGATE-IPv4-PCN-CL-FLIDs is empty.
- o) Source address (4 bytes): The IPv4 source address.
- o) Destination address (4 bytes): The IPv4 destination address.
- o) Protocol (1 byte): The IP protocol number. It refers to the true upper layer protocol carried by the packets.
- o) Source Port (2 bytes): contains the source port number.
- o) Destination Port (2 bytes): contains the destination port number.
- o) Controlled (CL) PCN CL Flow IDs object, IPv6 addresses are used:
Class = PCN
C-Type = RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs



- o) Length (1 byte): the length of the RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs object in units of 40 bytes. This field is used to specify the number of flow IDs carried by this object. Each flow ID is represented by the combination of each subsequent 5 tuple fields: Source address, Destination address, Source Port, Destination Port and Protocol number. If Length is 0 then the RSVP-AGGREGATE-IPv6-PCN-CL-FLIDs object is empty.

- o) Source address (16 bytes): The IPv6 source address.

o) Destination address (16 bytes): The IPv6 destination address.

o) Protocol (1 byte): The IP protocol number. It refers to the true upper layer protocol carried by the packets.

- o) Source Port (2 bytes): contains the source port number.
- o) Destination Port (2 bytes): contains the destination port number.

5. Security Considerations

The same security considerations specified in [[RFC4860](#)] and [[RFC5559](#)] apply also to this document.

6. IANA Considerations

This document makes the following requests to the IANA:

- o allocate a new Object Class (PCN Object), see [Section 4.1](#).
- o allocate a "PCN-domain rejects e2e reservation" Error Code that may appear only in e2e PathErr messages, see [Section 3.1](#).

Error Value for "PCN-domain rejects e2e reservation" = To be allocated by IANA

7. Acknowledgments

We would like to thank the authors of [[draft-lefaucheur-rsvp-ecn-01.txt](#)], since some ideas used in this document are based on the work initiated in [[draft-lefaucheur-rsvp-ecn-01.txt](#)]. Moreover, we would like to thank Tom Taylor, Francois Le Faucheur and James Polk for the comments provided on the 00 version of this draft.

8. Normative References

[[draft-ietf-pcn-cl-edge-behaviour-12](#)] T. Taylor, A. Charny, F. Huang, G. Karagiannis, M. Menth, "PCN Boundary Node Behaviour for the Controlled Load (CL) Mode of Operation (Work in progress)", February 2012.

[[draft-ietf-pcn-sm-edge-behaviour-09](#)] A. Charny, J. Zhang, G. Karagiannis, M. Menth, T. Taylor, "PCN Boundary Node Behaviour for the Single Marking (SM) Mode of Operation (Work in progress)", February 2012.

[[draft-ietf-pcn-signaling-requirements-08](#)] G. Karagiannis, T. Taylor, K. Chan, M. Menth, P. Eardley, "Requirements for Signaling of (Pre-) Congestion Information in a DiffServ Domain(Work in progress)", February 2012.

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

- [RFC2205] Braden, R., ed., et al., "Resource ReSerVation Protocol (RSVP)- Functional Specification", [RFC 2205](#), September 1997.
- [RFC3140] Black, D., Brim, S., Carpenter, B., and F. Le Faucheur, "Per Hop Behavior Identification Codes", [RFC 3140](#), June 2001.
- [RFC3175] Baker, F., Iturralde, C., Le Faucheur, F., and B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations", [RFC 3175](#), September 2001.
- [RFC3473] L. Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReSerVation Protocol-Traffic Engineering (RSVP-TE) Extensions", [RFC 3473](#), January 2003.
- [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.
- [RFC4860] F. Le Faucheur, B. Davie, P. Bose, C. Christou, M. Davenport, "Generic Aggregate Resource ReSerVation Protocol (RSVP) Reservations", [RFC4860](#), May 2007.
- [RFC5670] Eardley, P., "Metering and Marking Behaviour of PCN-Nodes", [RFC 5670](#), November 2009.
- [RFC5696] Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information", [RFC 5696](#), November 2009.

9. Informative References

- [[draft-lefaucheur-rsvp-ecn-01.txt](#)] Le Faucheur, F., Charny, A., Briscoe, B., Eardley, P., Chan, K., and J. Babiarz, "RSVP Extensions for Admission Control over Diffserv using Pre-congestion Notification (PCN) (Work in progress)", June 2006.
- [RFC1633] Braden, R., Clark, D., and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", [RFC 1633](#), June 1994.
- [RFC2211] J. Wroclawski, Specification of the Controlled-Load Network Element Service, September 1997
- [RFC2212] S. Shenker et al., Specification of Guaranteed Quality of Service, September 1997
- [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.

Karagiannis, et al. Expires September 10, 2012

[Page 25]

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "A framework for Differentiated Services", [RFC 2475](#), December 1998.

[RFC2998] Bernet, Y., Yavatkar, R., Ford, P., 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.

[RFC5559] Eardley, P., "Pre-Congestion Notification (PCN) Architecture", [RFC 5559](#), June 2009.

[10.](#) Authors' Address

Georgios Karagiannis
University of Twente
P.O. Box 217
7500 AE Enschede,
The Netherlands
EMail: g.karagiannis@utwente.nl

Anurag Bhargava
Cisco Systems, Inc.
7100-9 Kit Creek Road
PO Box 14987
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709-4987
USA
Email: anuragb@cisco.com

