

Next Steps in Signaling (nsis)
Internet-Draft
Intended status: Standards Track
Expires: March 7, 2008

M. Arumaithurai
University of Goettingen
September 4, 2007

**NSIS PCN-QoS: A Quality of Service Model for Pre-Congestion
Notification (PCN)
draft-arumaithurai-nsis-pcn-00.txt**

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on March 7, 2008.

Copyright Notice

Copyright (C) The IETF Trust (2007).

Abstract

This document describes a Quality of Service model (QoS), for networks that support the use of the Controlled load service over a Diffserv cloud using pre-congestion notification. This is a technique to perform admission control in a Diffserv network by measuring the congestion level in a domain. New flows are admitted only if the current congestion level is below the allowed threshold. When the controlled load in a domain exceeds a certain threshold, the egress

prompts the ingress to pre-empt certain flows. This proposal is based on the proposal made by B.Briscoe et al., for the extension of RSVP to support the Controlled load service over a Diffserv cloud using pre-congestion notification.

Table of Contents

1.	Introduction	3
2.	Terminology and Abbreviations	3
3.	Overview of the Extensions and Operations	5
3.1.	Overall QoS Architecture	5
3.2.	Overview of Procedures for Admission Control of New Reservations	6
3.2.1.	PCN-QOSM/QoS-NSLP Upstream Signaling	6
3.2.2.	PCN-QOSM/QoS-NSLP Downstream Signaling	8
3.2.3.	REFRESH Messages	10
3.2.4.	ERROR Messages	10
3.2.5.	NOTIFY Messages	11
3.3.	Removal of E2E Reservations	11
3.4.	Overview of Procedures for Preemption of Existing Reservations	11
3.4.1.	Receiver Initiated QoS Reservation	12
4.	PCN-CL-QSPEC Parameter	13
5.	Error Codes	14
5.1.	CL-PCN-Probes-Required-Error-Code	15
5.2.	CL-PCN-Inconsistent-Admission-Control-Behaviour-Error-Code	15
6.	IANA Considerations	15
7.	Security Considerations	15
8.	Acknowledgements	15
9.	References	16
9.1.	Normative References	16
9.2.	Informative References	16
	Author's Address	17
	Intellectual Property and Copyright Statements	18

1. Introduction

GIST [[I-D.ietf-nsis-ntlp](#)] is a signalling protocol that can be used by applications to set up state in routers in a given network. GIST, a successor to RSVP [[RFC2205](#)] can be used to signal NAT / FW etc in addition to QoS signalling. The Quality of Service NSIS Signalling Layer Protocol (QoS-NSLP), as defined in [[I-D.ietf-nsis-qos-nslp](#)], is a generic QoS signalling protocol for carrying Quality of Service (QoS) information between arbitrary nodes in the network and runs on top of the GIST layer. It could be used in various QoS models such as Intserv, Diffserv, RMD etc.

This document describes a 'Next Steps in Signalling (NSIS) QoS model' henceforth known as NSIS-PCN-QoSM for networks. These networks use a controlled load service over a Diffserv cloud using pre-congestion notification as described in CL-PCN [[I-D.ietf-pcn-architecture](#)]. It also elaborates on the corresponding QSPEC [[I-D.ietf-nsis-qspec](#)] parameters and error codes for expressing reservations in a suitable form that can be easily interpreted and processed by the relevant nodes.

CL-PCN [[I-D.ietf-pcn-architecture](#)] outlines a deployment mode to achieve a Controlled Load (CL) service [[RFC2211](#)] by using distributed measurement based admission control edge to edge, which is within a particular region of the Internet. The measurement is made via CL packets that have their Congestion Experienced (CE) codepoint set as they travel across the edge-to-edge region. Setting the CE codepoint, which is under the control of a new pre-congestion marking behaviour, provides an "early warning" of potential congestion. This information is passed on to the ingress node by the egress, thereby instructing the former to decide whether to admit a new CL microflow. CL-PCN [[I-D.ietf-pcn-architecture](#)] also describes how the framework uses rate-based pre-emption to maintain the CL service to as many admitted microflows as possible even after localised failure and routing changes in the interior of the edge-to-edge region.

The edge-to-edge architecture of CL-PCN [[I-D.ietf-pcn-architecture](#)] is a building block in delivering an end-to-end CL service. This approach is similar to implementing an Intserv operation over Diffserv networks, wherein a CL region is viewed as a single hop in achieving an end-to-end reservation. Interior nodes in a CL region do not hold states or process signalling flows.

2. Terminology and Abbreviations

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](#)].

For readability, a number of definitions from CL-PCN [[I-D.ietf-pcn-architecture](#)], and NSIS terminology as defined in QoS-NSLP [[I-D.ietf-nsis-qos-nslp](#)], NSIS-QSPEC [[I-D.ietf-nsis-qspec](#)] are repeated here for easy readability:

Ingress Edge (or Ingress Gateway, or Ingress Node):

A router at an ingress to the CL-region. A CL-region may have several ingress gateways.

Egress Edge (or Egress Gateway, or Egress Node):

A router at an egress from the CL-region. A CL-region may have several egress gateways.

Interior router:

A router that is part of the CL-region, but is not an ingress or egress gateway.

CL-region:

A region of the Internet in which all traffic enters/leaves through an ingress/egress gateway and all routers run Pre-Congestion Notification marking. A CL-region is a DiffServ region (a DiffServ region is either a single DiffServ domain or set of contiguous DiffServ domains), but note that the CL-region does not use the traffic conditioning agreements (TCAs) of the (informational) DiffServ architecture.

CL-region-aggregate:

All the microflows between a specific pair of ingress and egress gateways. Note there is no field in the flow packet headers that uniquely identifies the aggregate.

Congestion-Level-Estimate:

The number of bits in CL packets that are admission marked (or pre-emption marked), divided by the number of bits in all CL packets. It is calculated as an exponentially weighted moving average. It is calculated by an egress gateway for the CL packets from a particular ingress gateway, i.e., there is a Congestion-Level-Estimate for each CL-region-aggregate.

Sustainable-Aggregate-Rate:

The rate of traffic that the network can actually support for a specific CL-region-aggregate. So it is measured by an egress gateway for the CL packets from a particular ingress gateway.

QNE:

An NSIS Entity (NE), which supports the QoS NSLP.

QNI:

The first node in the sequence of QNEs that issues a reservation request for a session.

QNR:

The last node in the sequence of QNEs that receives a reservation request for a session.

3. Overview of the Extensions and Operations

3.1. Overall QoS Architecture

The overall QoS architecture is described in CL-PCN [[I-D.ietf-pcn-architecture](#)] and is reproduced below in Figure 1 for readability and modified to incorporate NSIS terminology.

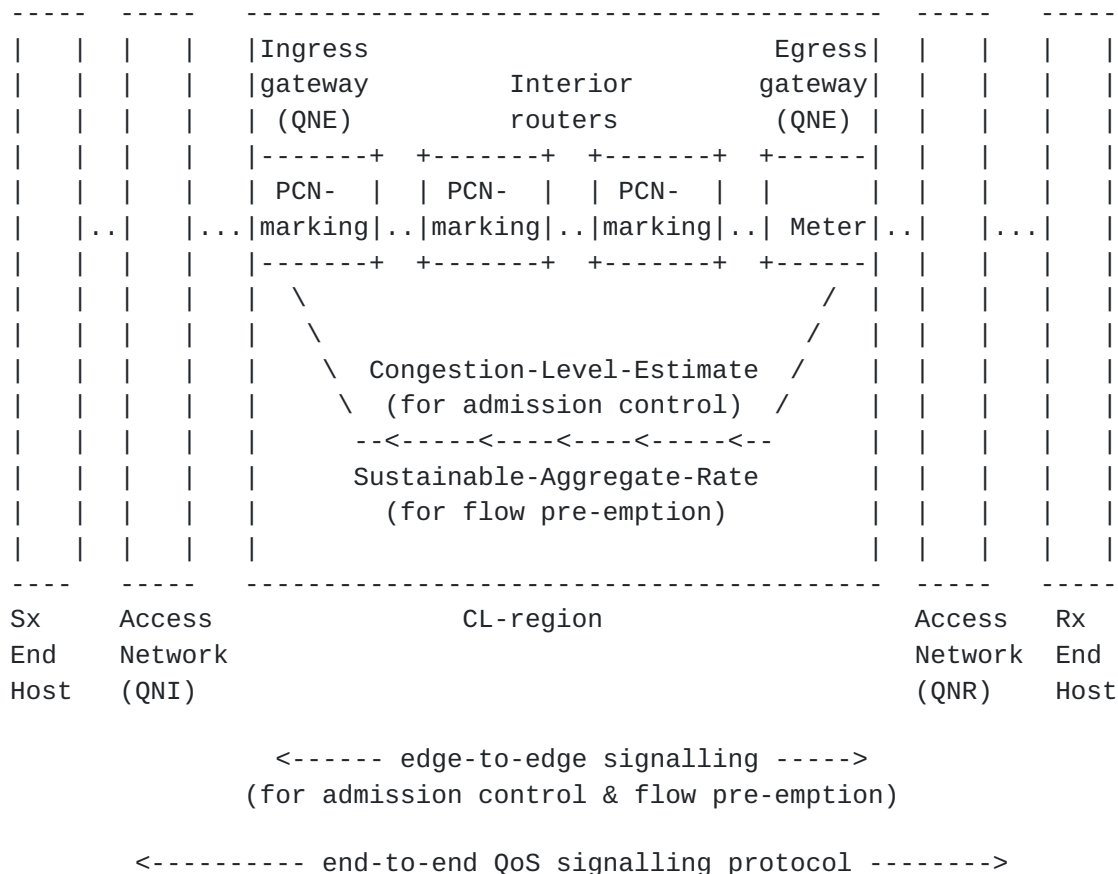


Figure 1: Overall QoS Architecture

3.2. Overview of Procedures for Admission Control of New Reservations

As mentioned earlier, CL-PCN [[I-D.ietf-pcn-architecture](#)] describes a framework to achieve a Controlled Load (CL) service by using distributed measurement-based admission control edge-to-edge, i.e., within a particular region of the Internet. This section keys out NSIS operations to support such an admission control scheme relying on Pre-Congestion Notification in the edge-to-edge region.

3.2.1. PCN-QOSM/QoS-NSLP Upstream Signaling

The basic PCN-QOSM/QoS-NSLP signaling is shown in Figure 2. When a new request for QoS arrives at the QNI from an End-Host, the QNI performs regular QoS-NSLP processing by establishing a hop by hop connection with other QoS-NSLP aware nodes along the path to the destination. It creates a RESERVE message with an initiator QSPEC parameter describing the reservation and then forwards it.

When this end-to-end RESERVE message reaches an Ingress node, it performs the following operation:

- o The ingress node continues the hop-by-hop QoS-RESERVE message by sending it to the next QoS-NSLP aware node. The PCN-capable Interior nodes in the CL region are not QoS-NSLP-capable (or have QoS/NSLP processing disabled) and thus simply ignore the RESERVE message. Therefore, the Egress Edge becomes the next hop of the Ingress node.

The Egress node on receiving the RESERVE message does the following(not necessarily in the order mentioned below):

- o If the Egress node does not have a valid value for Congestion-Level-Estimate for the aggregate flow of traffic for this Ingress-Egress pair, it sends an asynchronous NOTIFY message back to the Ingress with a INFO-SPEC (as described in [Section 5.2.3](#) of QSPEC [I-D.ietf-nsis-qspec]), which has the "CL-PCN-Probes-Required-error-code" error code. The error code mentioned is defined in this document in [Section 5.1](#). Note that the probe packets that are sent by the Ingress SHOULD be in accordance with the CL-PCN [I-D.ietf-pcn-architecture] requirements.
- o The Egress node sends the original RESERVE message with the End-to-End(E2E) QSPEC to the next hop in the upstream direction, i.e., towards the destination.

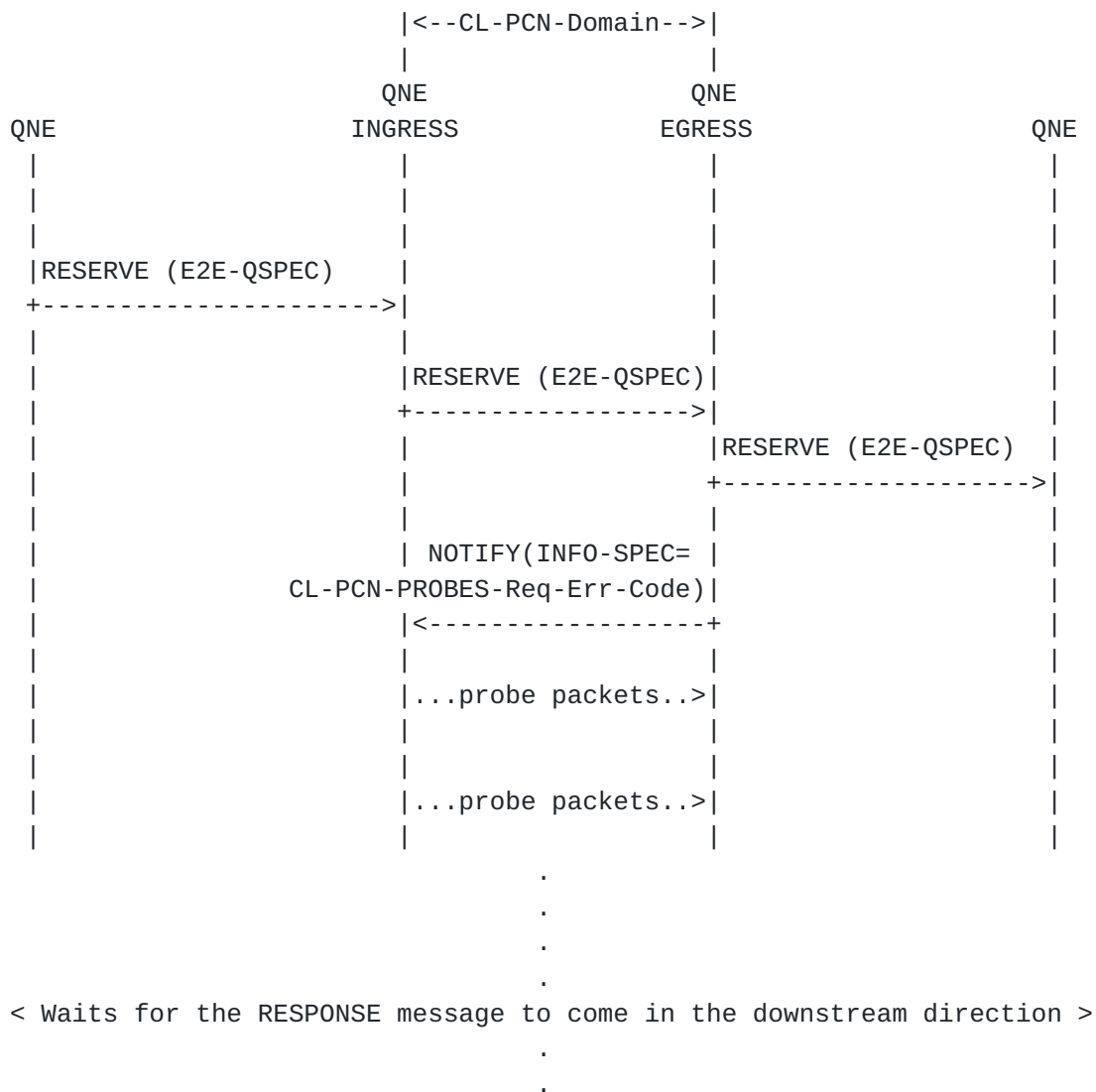


Figure 2: PCN-QoSM Upstream Signalling

3.2.2. PCN-QoSM/QoS-NSLP Downstream Signaling

Figure 3 shows the RESPONSE message received in the Downstream direction. When the RESPONSE message is received by the Egress Edge (from the downstream side), the latter performs regular QoS-NSLP processing (including performing admission control for the segment downstream of the Egress Edge) augmented with the procedures described in this section:

- o If the Egress node has a valid Congestion-Level-Estimate value, it forwards the RESPONSE message to the ingress node with the PCN-CL-QSPEC parameter (as described in [Section 4](#)) having the Congestion-

Level-Estimate value and the E2E-QSPEC parameter already present in the received RESPONSE message. Details for computing the Congestion-Level-estimate can be found in CL-PCN

[[I-D.ietf-pcn-architecture](#)] and PCN-MARKING

[[I-D.briscoe-tsvwg-cl-phb](#)].

- o If the Egress node does not have a current Congestion-Level-Estimate value(because there was no traffic received by the Egress Edge from that Ingress Edge), it does the following:
 - * Triggers a timer(Tx) and puts the RESPONSE message on hold.
 - * If it is still receiving probe packets from ingress, it does the Congestion-Level-Estimate after receiving the probe packets.
 - * If it isn't receiving probe packets(Due to some error situation), it sends a NOTIFY message with the INFO-SPEC set to the new Error Code of "CL-PCN-Probes-Required" specified in this document in [Section 5.1](#).
 - * When the timer Tx expires, if it has a valid Congestion-Level-Estimate, it sends the RESPONSE message with the PCN-CL-QSPEC and the E2E-QSPEC. If the Congestion-Level-estimate value is still not available, it may loop a few times through the previous steps, and if the situation continues it must send a RESPONSE with an error code.

The Ingress node on receiving a NOTIFY message processes it as per regular QoS-NSLP signalling with the following exceptions:

- o If it received an INFO-SPEC parameter with the Error Code of "CL-PCN-Probes-Required-error-code" specified in this document, it starts sending probe packets to the Egress node.

The Ingress node on receiving a RESPONSE message processes it as per regular QoS-NSLP signalling with the following exceptions:

- o If it received the PCN-CL-QSPEC parameter with a valid Congestion-Level-Estimate, it performs admission control and sends a RESPONSE message upstream indicating the success or failure of the reservation attempt. It must carry out the admission control decision (for admission of the reservation over the path from Ingress Edge to Egress Edge through the PCN cloud) taking into account the congestion information provided in the PCN-CL-QSPEC parameter of the RESPONSE message in accordance with the

procedures of PCN-CL-QSPEC and PCN-MARKING

[[I-D.briscoe-tsvwg-cl-phb](#)]. For example, if the Congestion level Estimate conveyed in the PCN-CL-QSPEC parameter exceeds a configured threshold, the Ingress Edge may decide to reject this new reservation. Once the admission control decision is taken by the Ingress Edge, regular NSIS procedures are followed to either proceed with the reservation (and forward the RESPONSE towards the sender) or tear down the reservation (and, in particular, send a error RESPONSE with the appropriate error code.

- o In case the Ingress Edge forwards the RESPONSE message upstream, the Ingress Edge MUST remove the PCN-CL-QSPEC parameter.

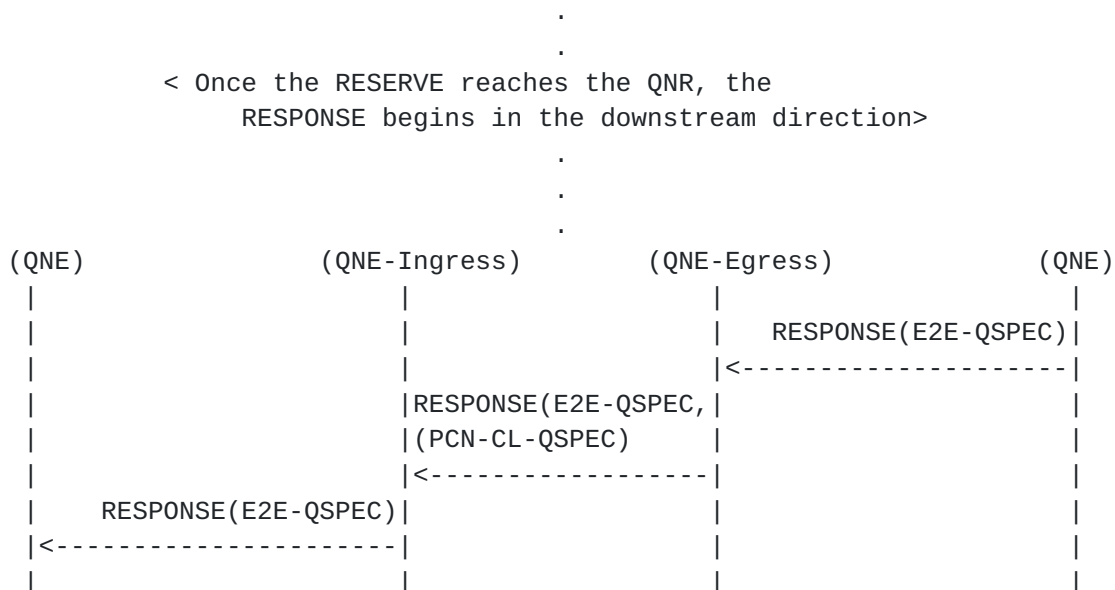


Figure 3: PCN-QoSM Downstream signalling

3.2.3. REFRESH Messages

Since the set up states in the routers are in soft state, refreshing is done by sending RESERVE messages in the REFRESH mode. The Egress can respond to this RESERVE by sending a NOTIFY message with a new value of Congestion-level-estimate to keep the ingress informed.

3.2.4. ERROR Messages

On receiving a NOTIFY message with the INFO-SPEC having the "CL-PCN-Probes-Required-error-code" error code flag set, the Ingress Edge MUST generate probe packets, as described in CL-PCN

[[I-D.ietf-pcn-architecture](#)], towards the Egress Edge which sent the NOTIFY Message with the Error Code. The Ingress Node must not send this NOTIFY message further upstream.

3.2.5. NOTIFY Messages

NOTIFY messages are asynchronous messages specified in QoS-NSLP [[I-D.ietf-nsis-qos-nslp](#)]. They can be send upstream and downstream, by the ingress to the Egress and viceversa, with the PCN-CL-QSPEC parameter or the INFO-SPEC Error Codes to notify one another, when required. The NOTIFY message is mainly used by the egress to notify the ingress of pre-emption of flows. A detailed explanation is given in [Section 3.4](#). It is the responsibility of the QoS-NSLP to take care of the sure delivery of the NOTIFY message.

3.3. Removal of E2E Reservations

E2E reservations are removed in the usual QoS-NSLP way via sending the RESERVE message with the Tear flag set or when a timeout occurs, since fresh REFRESH messages were not sent or as the result of an error condition. This does not directly affect CL-PCN [[I-D.ietf-pcn-architecture](#)] operations.

3.4. Overview of Procedures for Preemption of Existing Reservations

As mentioned earlier, CL-PCN [[I-D.ietf-pcn-architecture](#)] describes how rate-based pre-emption can be used to maintain the CL service to as many admitted microflows as possible, even after localised failure and routing changes in the interior of the edge-to-edge region. This requires the egress to check on the CL-region aggregate and alerting the ingress of any abnormality by sending it the measured Sustainable-Aggregate-Rate. Based on this information, the ingress edge node might decide to pre-empt certain admitted flows to maintain the CL of the admitted flows. CL-PCN [[I-D.ietf-pcn-architecture](#)] identifies that this information needs to be transferred reliably.

This section describes the corresponding NSIS extensions.

As shown in Figure 4, let us assume that a number of reservations are established and transit through a given Ingress Edge and a given Egress Edge and that the sustainable-aggregate-level, that is monitored by the Egress Edge exceeds a certain threshold level. Therefore the Egress Edge sends this measured Sustainable-Aggregate-Rate for the CL-region-aggregate from Ingress Edge to Egress Edge. To do this, the Egress Edge sends an asynchronous NOTIFY message (described in [Section 5.4.4](#) of the QoS-NSLP [[I-D.ietf-nsis-qos-nslp](#)]) with the PCN-CL-QSPEC parameter containing the current Sustainable-Aggregate-Rate.

To avoid the risk that this NOTIFY message gets lost and in turn that the Ingress Edge is not made aware in a timely manner that preemption may be needed, the NSIS reliable messaging procedures specified for reliable NOTIFY messages mentioned in QoS-NSLP MUST be used.

On receipt of the NOTIFY message, Ei on noting the value of Sustainable-Aggregate-Rate will identify that pre-emption of flows should take place and will trigger its pre-emption trigger. This will assess whether some reservations need to be dropped in accordance with the CL-PCN [[I-D.ietf-pcn-architecture](#)] and PCN-MARKING [[I-D.briscoe-tsvwg-cl-phb](#)] scheme. In case some do, those will be torn down as per regular NSIS procedures.

3.4.1. Receiver Initiated QoS Reservation

TODO

Editor's note: We will have to consider the case of the Receiver initiated QoS signalling w.r.t the CL-PCN architecture.

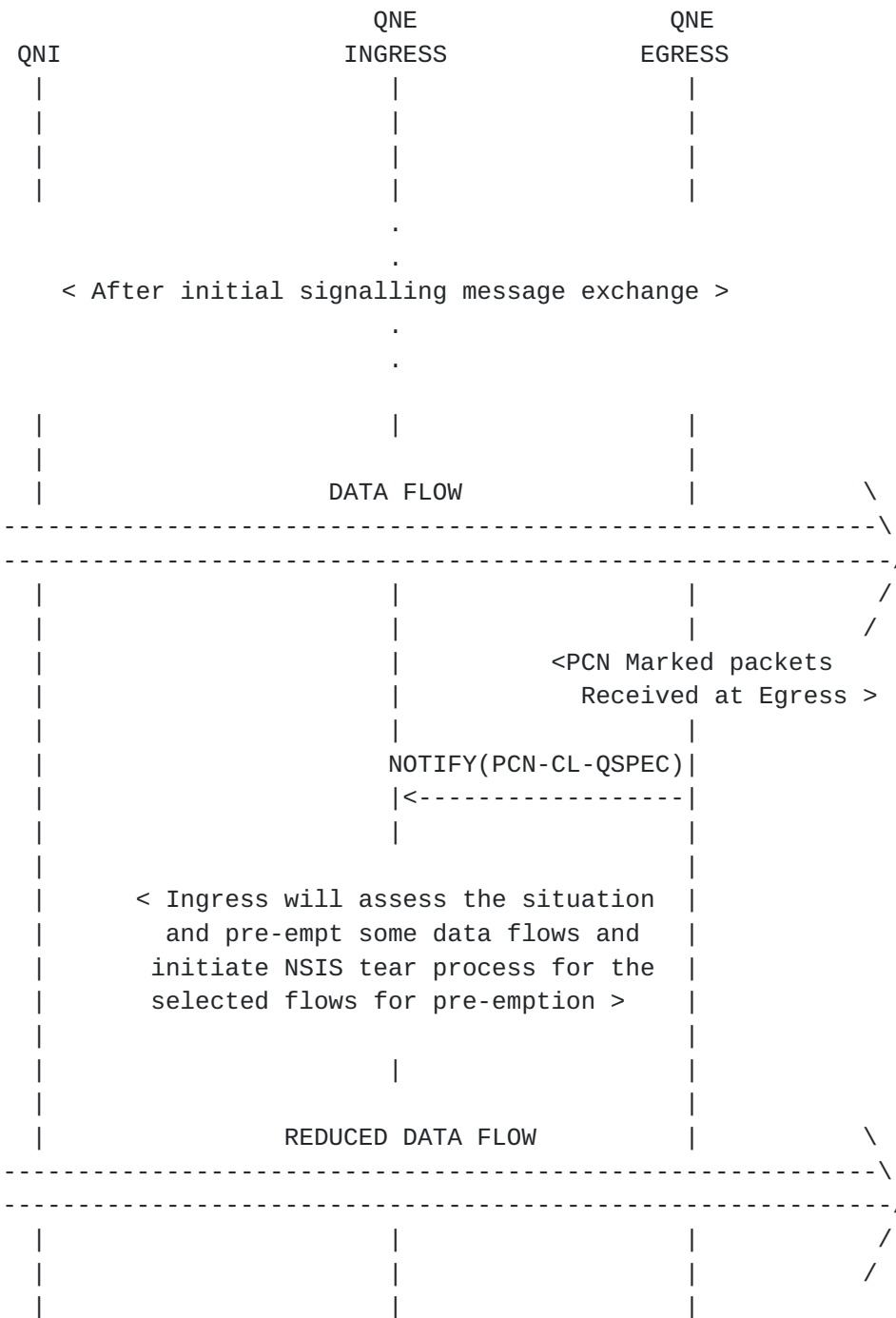


Figure 4: Admission Control by Pre-emption of Flows

4. PCN-CL-QSPEC Parameter

This document defines a new QSPEC parameter.

The PCN-QOSM uses the QSpec format specified in QSPEC [[I-D.ietf-nsis-qspec](#)].

The < I > flag is set to "Local" (i.e., "1"). There is no necessity for a QoS-Desired or a QoS-Reserved object since there is no state being set in the local domain. Moreover there is no need to indicate the message sequence as sender initiated or receiver initiated.

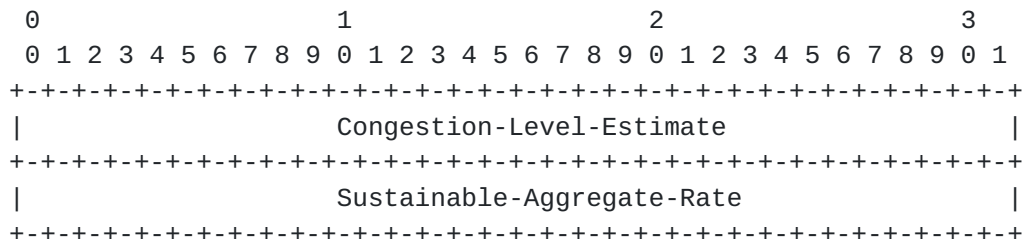


Figure 5: PCN-CL-QSPEC Parameter

The PCN-CL-QSPEC parameter can appear in any message that can carry a QSPEC. The fields inside the PCN-CL-QSPEC parameter have the following semantic (Encoding details will be provided in future versions):

Congestion-Level-Estimate:

This contains the value of the Congestion-Level-Estimate (defined in CL-PCN [[I-D.ietf-pcn-architecture](#)] and PCN-MARKING [[I-D.briscoe-tsvwg-cl-phb](#)]) computed by Egress edge for the CL-region-aggregate from the ingress edge to the egress edge.

Sustainable-Aggregate-Rate:

This contains the Sustainable-Aggregate-Rate for the relevant CL-region-aggregate from the Ingress to the Egress defined in CL-PCN [[I-D.ietf-pcn-architecture](#)] and PCN-MARKING [[I-D.briscoe-tsvwg-cl-phb](#)]) computed by the Egress for this CL-region-aggregate. It has a NULL value when the Egress wants to inform that pre-emption is not required.

5. Error Codes

This section defines two error codes to be added to the INFO-Spec mentioned in [Section 5.2.3](#) of QSPEC [[I-D.ietf-nsis-qspec](#)].

5.1. CL-PCN-Probes-Required-Error-Code

This error code is present in a RESPONSE message from a Egress edge to the Ingress edge, informing the ingress that probe packets need to be sent.

Error code = [[To be allocated by IANA]]

5.2. CL-PCN-Inconsistent-Admission-Control-Behaviour-Error-Code

The "CL-PCN-Inconsistent-Admission-Control-Behaviour-Error-code" may appear only in RESERVE messages to indicate error to the receiver or in a RESPONSE message to indicate to the sender that the Egress Edge node is not CL-PCN [[I-D.ietf-pcn-architecture](#)] capable.

Error code = [[To be allocated by IANA]]

Error value = [[To be allocated by IANA]]

6. IANA Considerations

PCN-QoSM requires a new IANA registry for the CL-PCN QoS Model Identifier. It is a 8-bit value, carried in the <QSPEC Type> field of the QSpec object [[I-D.ietf-nsis-qspec](#)].

PCN-QoSM defines a new QSPEC parameter called PCN-CL-QSPEC parameter (as described in [Section 4](#)) to the QPSEC draft.

PCN-QoSM also defines two new error codes to be added to the QSPEC INFO-SPEC:

- o "CL-PCN-Probes-Required-Error-Code" error code as described in [Section 5](#).
- o "CL-PCN-Inconsistent-Admission-Control-Behaviour-Error-code" error code as described in [Section 5](#).

7. Security Considerations

To be added

8. Acknowledgements

To be added

9. References

9.1. Normative References

- [I-D.briscoe-tsvwg-cl-phb]
Briscoe, B., "Pre-Congestion Notification marking",
[draft-briscoe-tsvwg-cl-phb-03](#) (work in progress),
October 2006.
- [I-D.ietf-nsis-ntlp]
Schulzrinne, H. and R. Hancock, "GIST: General Internet
Signalling Transport", [draft-ietf-nsis-ntlp-14](#) (work in
progress), July 2007.
- [I-D.ietf-nsis-qos-nslp]
Manner, J., "NSLP for Quality-of-Service Signaling",
[draft-ietf-nsis-qos-nslp-15](#) (work in progress), July 2007.
- [I-D.ietf-nsis-qspec]
Ash, J., "QoS NSLP QSPEC Template",
[draft-ietf-nsis-qspec-17](#) (work in progress), July 2007.
- [I-D.ietf-pcn-architecture]
Eardley, P., "Pre-Congestion Notification Architecture",
[draft-ietf-pcn-architecture-00](#) (work in progress),
August 2007.
- [I-D.lefaucheur-rsvp-ecn]
Faucheur, F., "RSVP Extensions for Admission Control over
Diffserv using Pre-congestion Notification (PCN)",
[draft-lefaucheur-rsvp-ecn-01](#) (work in progress),
June 2006.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2211] Wroclawski, J., "Specification of the Controlled-Load
Network Element Service", [RFC 2211](#), September 1997.

9.2. Informative References

- [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S.
Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1
Functional Specification", [RFC 2205](#), September 1997.
- [RFC2212] Shenker, S., Partridge, C., and R. Guerin, "Specification
of Guaranteed Quality of Service", [RFC 2212](#),
September 1997.

Author's Address

Mayutan Arumaithurai
University of Goettingen

Email: arumaithurai@informatik.uni-goettingen.de

Full Copyright Statement

Copyright (C) The IETF Trust (2007).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).

