Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: June 19, 2011

A. Charny
Cisco Systems
F. Huang
Huawei Technologies
G. Karagiannis
U. Twente
M. Menth
University of Tuebingen
T. Taylor, Ed.
Huawei Technologies

December 16, 2010

PCN Boundary Node Behaviour for the Controlled Load (CL) Mode of Operation
draft-ietf-pcn-cl-edge-behaviour-08

Abstract

Pre-congestion notification (PCN) is a means for protecting the quality of service for inelastic traffic admitted to a Diffserv domain. The overall PCN architecture is described in RFC 5559. This memo is one of a series describing possible boundary node behaviours for a PCN-domain. The behaviour described here is that for a form of measurement-based load control using three PCN marking states, not-marked, threshold-marked, and excess-traffic-marked. This behaviour is known informally as the Controlled Load (CL) PCN-boundary-node behaviour.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

The objective of Pre-Congestion Notification (PCN) is to protect 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, to decide whether to admit or block a new flow request, and (in abnormal circumstances) flow termination to decide whether to terminate some of the existing flows. To achieve this, the overall rate of PCN-traffic is metered on every link in the PCN-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 PCN-boundary-nodes about incipient overloads before any congestion occurs (hence the "pre" part of "pre-congestion notification"). The level of marking allows decisions to be made about whether to admit or terminate PCN-flows. For more details see [RFC5559].

PCN-boundary-node behaviours specify a detailed set of algorithms and procedures used to implement the PCN mechanisms. Since the algorithms depend on specific metering and marking behaviour at the interior nodes, it is also necessary to specify the assumptions made about PCN-interior-node behaviour. Finally, because PCN uses DSCP values to carry its markings, a specification of PCN-boundary-node behaviour MUST include the per domain behaviour (PDB) template specified in [RFC3086], filled out with the appropriate content. The present document accomplishes these tasks for the Controlled Load (CL) mode of operation.

[RFC EDITOR'S NOTE: you may choose to delete the following paragraph and the "[CL-specific]" tags throughout this document when publishing it, since they are present primarily to aid reviewers. RFCyyyy is the published version of draft-ietf-pcn-sm-edge-behaviour.]

A companion document [RFCyyyy] specifies the Single Marking (SM) PCN-boundary-node behaviour. This document and [RFCyyyy] have a great deal of text in common. To simplify the task of the reader, the text in the present document that is specific to the CL PCN-boundary-node behaviour is preceded by the phrase: "[CL-specific]". A similar distinction for SM-specific text is made in [RFCyyyy].

1.1. 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 the following terms defined in Section 2 of [RFC5559]:

- PCN-domain;
- PCN-ingress-node;
- PCN-egress-node;
- PCN-interior-node;
- PCN-boundary-node;
- PCN-flow;
- ingress-egress-aggregate (IEA);
- [CL-specific] PCN-threshold-rate;
- PCN-excess-rate;
- PCN-admissible-rate;
- PCN-supportable-rate;
- PCN-marked;
- [CL-specific] threshold-marked;
- excess-traffic-marked.

It also uses the following terms, for which the definition is repeated from [RFC5559] because of their importance to the understanding of the text that follows:

PCN-traffic, PCN-packets, PCN-BA

A PCN-domain carries traffic of different Diffserv behaviour 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 and ECN fields.

This document uses the following terms from [RFC5670]:

- [CL-specific] threshold-meter;
- excess-traffic-meter.

To complete the list of borrowed terms, this document reuses the following terms and abbreviations defined in Section 3 of [RFC5696]:

- not-PCN codepoint;

- Not-marked (NM) codepoint;

- PCN-marked (PM) codepoint;

- EXP (experimental) [codepoint].

This document defines the following additional terms:

Decision Point
The node that makes the decision about which flows to admit and to terminate. In a given network deployment, this can be the PCN-ingress-node or a centralized control node. Regardless of the location of the Decision Point, the PCN-ingress-node is the point where the decisions are enforced.

NM-rate
The rate of not-marked PCN-traffic received at a PCN-egress-node for a given ingress-egress-aggregate in octets per second. For further details see Section 3.2.1.

[CL-specific] ThM-rate
The rate of threshold-marked PCN-traffic received at a PCN-egress-node for a given ingress-egress-aggregate in octets per second. For further details see Section 3.2.1.

ETM-rate
The rate of excess-traffic-marked PCN-traffic received at a PCN-egress-node for a given ingress-egress-aggregate in octets per second. For further details see Section 3.2.1.

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. For further details see Section 3.4.

Congestion level estimate (CLE)
A value derived from the measurement of PCN-packets received at a PCN-egress-node for a given ingress-egress-aggregate, representing the ratio of marked to total PCN-traffic (measured in octets) received over a short period. The CLE is used to derive the PCN-admission-state (Section 3.3.1) and also by the report suppression procedure (Section 3.2.3) if report suppression is activated.

PCN-admission-state

The state ("admit" or "block") derived by the Decision Point 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. For further details see Section 3.3.1.

Sustainable aggregate rate (SAR)

The estimated maximum rate of PCN-traffic that can be admitted to a given ingress-egress-aggregate at a given moment without risking degradation of quality of service for the admitted flows. The intention is that if the PCN-sent-rate of every ingress-egress-aggregate passing through a given link is limited to its sustainable aggregate rate, the total rate of PCN-traffic flowing through the link will be limited to the PCN-supportable-rate for that link. An estimate of the sustainable aggregate rate for a given ingress-egress-aggregate is derived as part of the flow termination procedure, and is used to determine how much PCN-traffic needs to be terminated. For further details see Section 3.3.2.

CLE-reporting-threshold

A configurable value against which the CLE is compared as part of the report suppression procedure. For further details, see Section 3.2.3.

CLE-limit

A configurable value against which the CLE is compared in order to derive the PCN-admission-state for a given ingress-egress-aggregate. For further details, see Section 3.3.1.

T-meas

An interval, the value of which is configurable, defining the measurement period at the PCN-egress-node during which statistics relating to PCN-traffic marking are collected. At the end of the interval the values NM-rate, [CL-specific] ThM-rate, and ETM-rate as defined above are calculated and a report is sent to the Decision Point, subject to the operation of the report suppression feature. For further details see Section 3.2.

T-maxsuppress

An interval, the value of which is configurable, 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 is used as a keep-alive mechanism for signalling between the PCN-egress-node and the Decision Point when report suppression is activated. For further details, see Section 3.2.3.

T-fail
An interval, the value of which is configurable, 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. For further details see Section 3.3.3.

2. [CL-Specific] Assumed Core Network Behaviour for CL

This section describes the assumed behaviour for nodes of the PCN-domain when acting in their role as PCN-interior-nodes. The CL mode of operation assumes that:

- PCN-interior-nodes perform both threshold-marking and excess-traffic-marking of packets, according to the rules specified in [RFC5670];

- excess-traffic-marking of packets uses the PCN-Marked (PM) codepoint defined in [RFC5696];

- threshold-marking of packets uses the EXP codepoint defined in [RFC5696];

- the PCN-domain satisfies the conditions specified in [RFC5696];

- on each link the reference rate for the threshold-meter is configured to be equal to the PCN-admissible-rate for the link;

- on each link the reference rate for the excess-traffic-meter is configured to be equal to the PCN-supportable-rate for the link;

- the set of valid codepoint transitions is as shown in Section 4.2 of [RFC5696].

3. Node Behaviours

3.1. Overview

This section describes the behaviour of the PCN-ingress-node, PCN-egress-node, and the Decision Point (which MAY be collocated with the PCN-ingress-node).

The PCN-egress-node collects the rates of not-marked, [CL-specific] threshold-marked, and excess-traffic-marked PCN-traffic for each ingress-egress-aggregate and reports them to the Decision Point. [CL-specific] It MAY also identify and report PCN-flows that have

experienced excess-traffic-marking. For a detailed description, see Section 3.2.

The PCN-ingress-node enforces flow admission and termination decisions. It also reports the rate of PCN-traffic sent to a given ingress-egress-aggregate when requested by the Decision Point. For details, see Section 3.4.

Finally, the Decision Point makes flow admission decisions and selects flows to terminate based on the information provided by the PCN-ingress-node and PCN-egress-node for a given ingress-egress-aggregate. For details, see Section 3.3.

3.2. Behaviour of the PCN-Egress-Node

3.2.1. Data Collection

The PCN-egress-node MUST meter received PCN-traffic in order to derive periodically the following rates for each ingress-egress-aggregate passing through it:

- o NM-rate: octets per second of PCN-traffic in PCN-packets that are not-marked (i.e., marked with the NM codepoint);

- o [CL-specific] ThM-rate: octets per second of PCN-traffic in PCN-packets that are threshold-marked (i.e., marked with the PM codepoint);

- o [CL-specific] ETM-rate: octets per second of PCN-traffic in PCN-packets that are excess-traffic-marked (i.e., marked with the EXP codepoint).

The PCN-traffic SHOULD be metered continuously and the measurement intervals themselves SHOULD be of equal length, to minimize the statistical variance introduced by the measurement process itself. The starting and ending times of the measurement intervals for different ingress-egress-aggregates MAY be the same or MAY be different.

[CL-specific] As a configurable option, the PCN-egress-node MAY record flow identifiers of the PCN-flows for which excess-traffic-marked packets have been observed. These can be used by the Decision Point when it selects flows for termination.

> In networks using multipath routing it is possible that congestion is not occurring on all paths carrying a given ingress-egress-aggregate. Assuming that specific PCN-flows are routed via specific paths, identifying the PCN-flows that are experiencing

excess-traffic-marking helps to avoid termination of PCN-flows not contributing to congestion.

3.2.2. Reporting the PCN Data

If the report suppression option described in the next sub-section is not activated, the PCN-egress-node MUST report the latest values of NM-rate, [CL-specific] ThM-rate, and ETM-rate to the Decision Point each time that it calculates them.

[CL-specific] If so configured (e.g., because multipath routing is being used, as explained in the previous section), the PCN-egress-node MUST also report the set of flow identifiers of PCN-flows for which excess-traffic-marking was observed in the most recent measurement interval. If this set is large, the PCN-egress-node MAY report only the most recently excess-traffic-marked PCN-flows rather than the complete set.

3.2.3. Optional Report Suppression

Report suppression MUST be provided as a configurable option, along with two configurable parameters, the CLE-reporting-threshold and the maximum report suppression interval T-maxsuppress. The default value of the CLE-reporting-threshold is zero. T-maxsuppress functions as a keep-alive mechanism for signalling between the PCN-egress-node and the Decision Point.

If the report suppression option is enabled, the PCN-egress-node MUST apply the following procedure to decide whether to send a report to the Decision Point, rather than sending a report automatically at the end of each measurement interval.

1. As well as the quantities NM-rate, [CLE-specific] ThM-rate, and ETM-rate, the PCN-egress-node MUST calculate the congestion level estimate (CLE) for each measurement interval. The CLE is computed as:

   [CL-specific]
   CLE = (ThM-rate + ETM-rate) / (NM-rate + ThM-rate + ETM-rate)

   if any PCN-traffic was observed, or CLE = 0 if all the rates are zero.

2. If the calculated CLE for the latest measurement interval is greater than the CLE-reporting-threshold and/or the calculated CLE for the immediately previous interval was greater than the CLE-reporting-threshold, then the PCN-egress-node MUST send a report to the Decision Point. The contents of the report are

described below.

3. If an interval T-maxsuppress has elapsed since the last report was sent to the Decision Point, then the PCN-egress-node MUST send a report to the Decision Point regardless of the CLE value.

4. If neither of the preceding conditions holds, the PCN-egress-node MUST NOT send a report for the latest measurement interval.

Each report sent to the Decision Point when report suppression has been activated MUST contain the values of NM-rate, [CL-specific] ThM-rate, ETM-rate, and CLE that were calculated for the most recent measurement interval. [CL-specific] If so configured, the PCN-egress-node MUST also report the set of flow identifiers of PCN-flows for which excess-traffic-marking was observed in the most recent measurement interval.

The above procedure ensures that at least one report is sent per interval (T-maxsuppress + T-meas). This provides some protection against loss of egress reports and also demonstrates to the Decision Point that both the PCN-egress-node and the communication path between that node and the Decision Point are in operation.

3.3. Behaviour at the Decision Point

Operators can choose to use PCN procedures just for flow admission, or just for flow termination, or for both. A compliant Decision Point MUST implement both mechanisms, but configurable options MUST be provided to activate or deactivate PCN-based flow admission and flow termination independently of each other at a given Decision Point.

If PCN-based flow termination is enabled but PCN-based flow admission is not, flow termination operates as specified in this document. Logically, some other system of flow admission control is in operation, but the description of such a system is out of scope of this document and depends on local arrangements.

3.3.1. Flow Admission

The Decision Point determines the PCN-admission-state for a given ingress-egress-aggregate each time it receives a report from the egress node. It makes this determination on the basis of the congestion level estimate (CLE). If the CLE is provided in the egress node report, the Decision Point SHOULD use the reported value. If the CLE was not provided in the report, the Decision Point MUST calculate it based on the other values provided in the report, using the formula:

[CL-specific]
CLE = (ThM-rate + ETM-rate) / (NM-rate + ThM-rate + ETM-rate)

if any PCN-traffic was observed, or CLE = 0 if all the rates are zero.

The Decision Point MUST compare the reported or calculated CLE to a configurable value, the CLE-limit. If the CLE is less than the CLE-limit, the PCN-admission-state for that aggregate MUST be set to "admit"; otherwise it MUST be set to "block".

[CL-specific] The outcome of the comparison is not very sensitive to the value of the CLE-limit in practice, because when threshold-marking occurs it tends to persist long enough that threshold-marked traffic becomes a large proportion of the received traffic in a given interval.

If the PCN-admission-state for a given ingress-egress-aggregate is "admit", the Decision Point SHOULD allow new flows to be admitted to that aggregate. If the PCN-admission-state for a given ingress-egress-aggregate is "block", the Decision Point SHOULD NOT allow new flows to be admitted to that aggregate. These actions MAY be modified by policy in specific cases, but such policy intervention risks defeating the purpose of using PCN.

3.3.2. Flow Termination

[CL-specific] When the report from the PCN-egress-node includes a non-zero value of the ETM-rate for some ingress-egress-aggregate, the Decision Point MUST request the PCN-ingress-node to provide an estimate of the rate (PCN-sent-rate) at which the PCN-ingress-node is receiving PCN-traffic that is destined for the given ingress-egress-aggregate.

If the Decision Point is collocated with the PCN-ingress-node, the request and response are internal operations.

The Decision Point MUST then wait, for both the requested rate from the PCN-ingress-node and the next report from the PCN-egress-node for the ingress-egress-aggregate concerned. If this next egress node report also includes a non-zero value for the ETM-rate, the Decision Point MUST determine an amount of flow to terminate using the following steps:

1. [CL-specific] The sustainable aggregate rate (SAR) for the given ingress-egress-aggregate is estimated by the sum:

SAR = NM-rate + ThM-rate

for the latest reported interval.

2. The amount of traffic to be terminated is the difference:

PCN-sent-rate - SAR,

where PCN-sent-rate is the value provided by the PCN-ingress-node.

If the difference calculated in the second step is positive, the Decision Point SHOULD select PCN-flows to terminate, until it determines that the PCN-traffic admission rate will no longer be greater than the estimated sustainable aggregate rate. If the Decision Point knows the bandwidth required by individual PCN-flows (e.g., from resource signalling used to establish the flows), it MAY choose to complete its selection of PCN-flows to terminate in a single round of decisions.

Alternatively, the Decision Point MAY spread flow termination over multiple rounds to avoid over-termination. If this is done, it is RECOMMENDED that enough time elapse between successive rounds of termination to allow the effects of previous rounds to be reflected in the measurements upon which the termination decisions are based (see [IEEE-Satoh] and sections 4.2 and 4.3 of [MeLe10]).

In general, the selection of flows for termination MAY be guided by policy. [CL-specific] If the egress node has supplied a list of identifiers of PCN-flows that experienced excess-traffic-marking (Section 3.2), the Decision Point SHOULD first consider terminating PCN-flows in that list.

3.3.3. **Decision Point Action For Missing** PCN-Boundary-Node Reports

If the Decision Point fails to receive any report from a given PCN-egress-node for a configurable interval T-fail, it SHOULD raise an alarm to management. A Decision Point collocated with a PCN-ingress-node SHOULD cease to admit PCN-flows to the ingress-egress-aggregate passing from the PCN-ingress-node to the given PCN-egress-node, until it again receives a report from that node. A centralized Decision Point MAY cease to admit PCN-flows to all ingress-egress-aggregates destined to the PCN-egress-node concerned, until it again receives a report from that node.

If a centralized Decision Point fails to receive a reply within a reasonable period of time to a request for a PCN-sent-rate value sent to a given PCN-ingress-node, it SHOULD raise an alarm to management.

3.4. Behaviour of the Ingress Node

The PCN-ingress-node 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) when the Decision Point requests it. The way this rate estimate is derived is a matter of implementation.

> For example, the rate that the PCN-ingress-node supplies MAY be based on a quick sample taken at the time the information is required. It is RECOMMENDED that such a sample be based on observation of at least thirty PCN-packets to achieve reasonable statistical reliability.

3.5. Summary of Timers

Table 1 summarizes the timers implied by the preceding procedures. The three configurable limits T-meas, T-maxsuppress, and T-fail apply to the three timers t-meas, t-maxsuppress, and t-fail respectively. t-meas and t-maxsuppress are reset upon expiry. t-fail is reset by management action or by receipt of a report from the PCN-egress-node concerned.

IEA = ingress-egress-aggregate.

Limit	Where	Incidence	Action on Expiry
T-meas	Egress node	One per node	Calculate and possibly report NM-rate, ThM-rate*, ETM-rate and CLE for each IEA.
T-maxsuppress	Egress node	One per IEA if report suppression is enabled.	Send a report for that IEA at the next expiry of t-meas.
T-fail	Decision Point	One per egress node	Assume failure and cease to admit flows passing through that egress node.

* ThM-rate is [CL-specific].

Table 1: Timers Used For the CL Boundary Node Behaviour

The value of T-meas SHOULD be configurable, and is RECOMMENDED to be of the order of 100 to 500 ms to provide a reasonable tradeoff between signalling demands on the network and the time taken to react to impending congestion.

t-maxsuppress is active only when report suppression is enabled. The value of T-maxsuppress SHOULD be configurable. The appropriate value for T-maxsuppress depends on whether the transport protocol between the PCN-egress-node and the Decision Point is reliable, and whether it implements its own keep-alive procedures. At the time of writing, that transport protocol has not yet been specified. This specification therefore requires that any transport protocol specification for carrying PCN reports MUST specify an appropriate default value for T-maxsuppress.

The value of T-fail MUST be configurable. As for T-maxsuppress, the appropriate value of T-fail depends on the transport protocol between the PCN-boundary-nodes and the Decision Point. It is RECOMMENDED that the default value for T-fail be three times the default value for T-maxsuppress as proposed by the transport protocol specification. The transport protocol specification MAY propose a different default value for T-fail in view of the particular characteristics of that protocol.

4. Identifying Ingress and Egress Nodes For PCN Traffic

The operation of PCN depends on the ability of the PCN-ingress-node to identify the ingress-egress-aggregate to which each new PCN-flow belongs and the ability of the egress node to identify the ingress-egress-aggregate to which each received PCN-packet belongs. If the Decision Point is collocated with the PCN-ingress-node, the PCN-egress-node also needs to associate each ingress-egress-aggregate with the address of the PCN-ingress-node to which it MUST send its reports.

The means by which this is done depends on the packet routing technology in use in the network. The procedure to provide the required information is out of the scope of this document.

5. Specification of Diffserv Per-Domain Behaviour

This section provides the specification required by [RFC3086] for a per-domain behaviour.

5.1. Applicability

This section draws heavily upon points made in the PCN architecture document, [RFC5559].

The PCN CL boundary node behaviour specified in this document is applicable to inelastic traffic (particularly video and voice) where quality of service for admitted flows is protected primarily by admission control at the ingress to the domain. In exceptional circumstances (e.g., due to network failures) already-admitted flows MAY be terminated to protect the quality of service of the remaining flows. [CL-specific] The CL boundary node behaviour is less likely to terminate too many flows under such circumstances than the SM boundary node behaviour [RFCyyyy].

[RFC EDITOR'S NOTE: please replace RFCyyyy above by the reference to the published version of draft-ietf-pcn-sm-edge-behaviour.]

5.2. Technical Specification

5.2.1. Classification and Traffic Conditioning

This section paraphrases the applicable portions of Sections 3.6 and 4.2 of [RFC5559].

Packets at the ingress to the domain are classified as either PCN or non-PCN. Non-PCN packets MAY share the network with PCN packets within the domain. Because the encoding specified in [RFC5696] and used in this document requires the use of the ECN fields, PCN-ingress-nodes MUST block ECN-capable traffic that uses the same DSCP as PCN from entering the PCN-domain directly. "Blocking" means it is dropped or downgraded to a lower-priority behaviour aggregate. Alternatively such traffic MAY be tunnelled through the PCN-domain.

PCN packets are further classified as belonging or not belonging to an admitted flow. PCN packets not belonging to an admitted flow are dropped. (This assumes that requests for flow admission are signalled in advance of the arrival of the flows themselves.) Packets belonging to an admitted flow are policed to ensure that they adhere to the agreed rate or flowspec.

5.2.2. PHB Configuration

The PCN SM and CL boundary node behaviours are metering and marking behaviours rather than scheduling behaviours. As a result, they are not tied to the selection of a specific DSCP value. The PCN working group suggests using admission control for the following service classes (defined in [RFC4594]):

- o Telephony (EF)
- o Real-time interactive (CS4)
- o Broadcast Video (CS3)
- o Multimedia Conferencing (AF4)

For a fuller discussion, see Section A.1 of Appendix A of [RFC5696].

5.3. Attributes

The purpose of this per-domain behaviour is to achieve low loss and jitter for the target class of traffic. The design requirement for PCN was that recovery from overloads through the use of flow termination SHOULD happen within 1-3 seconds. PCN probably performs better than that.

5.4. Parameters

In the list that follows, note that most PCN-ingress-nodes are also PCN-egress-nodes, and vice versa. Furthermore, the PCN-ingress-nodes MAY be collocated with Decision Points.

Parameters at the PCN-ingress-node:

- o Filters for distinguishing PCN from non-PCN inbound traffic.
- o The markings to be applied to PCN-traffic.
- o Reference rates on each inward link for the [CL-specific] threshold-meter and the excess-traffic-meter; see Section 2.
- o The information needed to distinguish PCN-traffic belonging to a given ingress-egress-aggregate.

Parameters at the PCN-egress-node:

- o The measurement interval T-meas.
- o Whether report suppression is enabled and, if so, the values of the CLE-reporting-threshold and T-maxsuppress.
- o [CL-specific] Whether individual flow identifiers will be reported for excess-traffic-marked PCN-traffic.

- o The information needed to distinguish PCN-traffic belonging to a given ingress-egress-aggregate.

- o The marking rules for re-marking PCN-traffic leaving the PCN domain.

Parameters at each interior node:

- o Reference rates on each link for the [CL-specific] threshold-meter and the excess-traffic-meter; see Section 2.

- o The markings to be applied to PCN-traffic, including the identification of PCN-packets and the encodings to indicate [CL-specific] threshold-marking and excess-traffic-marking.

Parameters at the Decision Point:

- o Activation/deactivation of PCN-based flow admission.

- o Activation/deactivation of PCN-based flow termination.

- o The value of CLE-limit.

- o The maximum interval T-fail between reports from a given PCN-egress-node, for detecting failure of communications with that node.

- o The information needed to map between each ingress-egress-aggregate and the corresponding PCN-ingress-node and PCN-egress-node.

5.5. Assumptions

Assumed that a specific portion of link capacity has been reserved for PCN-traffic. Assumed that the Decision Point receives requests for admission of PCN-flows before the packets in the PCN-flows arrive. This is not a critical assumption, but in its absence, packets will be dropped by the PCN-ingress-node until it obtains the admission decision from the Decision Point.

5.6. Example Uses

The PCN CL behaviour MAY be used to carry real-time traffic, particularly voice and video.

5.7. Environmental Concerns

The PCN CL per-domain behaviour can interfere with the use of end-to-end ECN due to reuse of ECN bits for PCN marking. See Appendix B of [RFC5696] for details.

5.8. Security Considerations

Please see the security considerations in Section 6 as well as those in [RFC2474] and [RFC2475].

6. Security Considerations

[RFC5559] provides a general description of the security considerations for PCN. This memo introduces no new considerations.

7. IANA Considerations

This memo includes no request to IANA.

8. Acknowledgements

The content of this memo bears a family resemblance to [ID.briscoe-CL]. The authors of that document were Bob Briscoe, Philip Eardley, and Dave Songhurst of BT, Anna Charny and Francois Le Faucheur of Cisco, Jozef Babiarz, Kwok Ho Chan, and Stephen Dudley of Nortel, Giorgios Karagiannis of U. Twente and Ericsson, and Attila Bader and Lars Westberg of Ericsson.

Ruediger Geib, Philip Eardley, and Bob Briscoe have helped to shape the present document with their comments. Toby Moncaster gave a careful review to get it into shape for Working Group Last Call.

Amongst the authors, Michael Menth deserves special mention for his constant and careful attention to both the technical content of this document and the manner in which it was expressed.

9. References

9.1. Normative References

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

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[RFC3086] Nichols, K. and B. Carpenter, "Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification", RFC 3086, April 2001.

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

[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.

9.2. Informative References

[ID.briscoe-CL]
Briscoe, B., "An edge-to-edge Deployment Model for Pre-Congestion Notification: Admission Control over a DiffServ Region (expired Internet Draft)", 2006.

[IEEE-Satoh]
Satoh, D. and H. Ueno, ""Cause and Countermeasure of Overtermination for PCN-Based Flow Termination", Proceedings of IEEE Symposium on Computers and Communications (ISCC '10), pp. 155-161, Riccione, Italy", June 2010.

[MeLe10] Menth, M. and F. Lehrieder, "PCN-Based Measured Rate Termination", Computer Networks Journal (Elsevier) vol. 54, no. 13, pages 2099 - 2116, September 2010.

[RFC4594] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, August 2006.

[RFCyyyy] Charny, A., Zhang, J., Karagiannis, G., Menth, M., and T. Taylor, "PCN Boundary Node Behaviour for the Single Marking (SM) Mode of Operation (Work in progress)",

December 2010.

Authors' Addresses

Anna Charny
Cisco Systems
300 Apollo Drive
Chelmsford, MA 01824
USA

Email: acharny@cisco.com

Fortune Huang
Huawei Technologies
Section F, Huawei Industrial Base,
Bantian Longgang, Shenzhen 518129
P.R. China

Phone: +86 15013838060
Email: fqhuang@huawei.com

Georgios Karagiannis
U. Twente

Phone:
Email: karagian@cs.utwente.nl

Michael Menth
University of Tuebingen
Sand 13
Tuebingen D-97074
Germany

Phone: +49-7071-2970505
Email: menth@informatik.uni-tuebingen.de

Tom Taylor (editor)
Huawei Technologies
1852 Lorraine Ave
Ottawa, Ontario K1H 6Z8
Canada

Phone: +1 613 680 2675
Email: tom111.taylor@bell.net