

| | | |
|--|----------------|--|
| Congestion and Pre-Congestion Notification | B. Briscoe | |
| Internet-Draft | T. Moncaster | |
| Intended status: Experimental | BT | |
| Expires: September 9, 2010 | March 08, 2010 | |

PCN 3-State Encoding Extension in a single DSCP draft-ietf-pcn-3-in-1-encoding-02

Abstract

The objective of Pre-Congestion Notification (PCN) is to protect the quality of service (QoS) of inelastic flows within a Diffserv domain. The overall rate of the PCN-traffic is metered on every link in the PCN-domain, and PCN-packets are appropriately marked when certain configured rates are exceeded. The level of marking allows the boundary nodes to make decisions about whether to admit or block a new flow request, and (in abnormal circumstances) whether to terminate some of the existing flows, thereby protecting the QoS of previously admitted flows. This document specifies how such marks are to be encoded into the IP header by re-using the Explicit Congestion Notification (ECN) codepoints within this controlled domain. This encoding builds on the baseline encoding and provides for three PCN encoding states: Not-marked, Threshold-marked and Excess-traffic-marked.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and 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 September 9, 2010.

Copyright Notice

Copyright (c) 2010 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 BSD License.

1. Introduction

The objective of Pre-Congestion Notification (PCN) [[RFC5559](#)] ([Eardley, P., "Pre-Congestion Notification \(PCN\) Architecture," June 2009.](#)) 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 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 level of marking allows boundary nodes to make decisions about whether to admit or terminate. This is achieved by marking packets on interior nodes according to some metering function implemented at each node. Threshold-traffic-marking marks all PCN packets once they exceed the threshold-traffic-rate on a link while Excess-traffic-marking marks only those PCN packets that exceed the excess-traffic-rate, which is higher than the threshold-traffic-rate [[RFC5670](#)] ([Eardley, P., "Metering and Marking Behaviour of PCN-Nodes," November 2009.](#)). These marks are monitored by the egress nodes of the PCN domain.

To fully support these two types of marking, three encoding states are needed. The baseline encoding described in [[RFC5696](#)] ([Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," November 2009.](#)) provides for deployment scenarios that only require two PCN encoding states using a single Diffserv codepoint. This document describes an experimental extension to the baseline-encoding that adds a third PCN encoding state in the IP header, still using a single Diffserv codepoint. For brevity it will be called the 3-in-1 PCN Encoding.

General PCN-related terminology is defined in the PCN architecture [\[RFC5559\] \(Eardley, P., "Pre-Congestion Notification \(PCN\) Architecture," June 2009.\)](#), and terminology specific to packet encoding is defined in the PCN baseline encoding [\[RFC5696\] \(Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," November 2009.\)](#). Note that [\[RFC5696\] \(Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," November 2009.\)](#) requires the PCN Working Group to maintain a list of all DSCPs used for PCN experiments.

1.1. Changes in This Version (to be removed by RFC Editor)

From draft-ietf-pcn-3-in-1-encoding-01 to -02: Corrected mistake in introduction, which wrongly stated that the threshold-traffic rate is higher than the excess-traffic rate. Other minor corrections.

*Updated acks & refs.

From draft-ietf-pcn-3-in-1-encoding-00 to -01: Altered the wording to make sense if [\[I-D.ietf-tsvwg-ecn-tunnel\] \(Briscoe, B., "Tunnelling of Explicit Congestion Notification," March 2010.\)](#) moves to proposed standard.

*References updated

From draft-briscoe-pcn-3-in-1-encoding-00 to draft-ietf-pcn-3-in-1-encoding-00:

Filename changed to draft-ietf-pcn-3-in-1-encoding.

*Introduction altered to include new template description of PCN.

*References updated.

*Terminology brought into line with [\[RFC5670\] \(Eardley, P., "Metering and Marking Behaviour of PCN-Nodes," November 2009.\)](#).

*Minor corrections.

2. 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 \(Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," March 1997.\)](#) [RFC2119].

3. The Requirement for Three PCN Encoding States

The PCN architecture [\[RFC5559\] \(Eardley, P., "Pre-Congestion Notification \(PCN\) Architecture," June 2009.\)](#) describes proposed PCN schemes that expect traffic to be metered and marked using both Threshold and Excess Traffic schemes. In order to achieve this it is necessary to allow for three PCN encoding states: one as a Not Marked (NM) state and the other two to distinguish these two levels of marking severity [\[RFC5670\] \(Eardley, P., "Metering and Marking Behaviour of PCN-Nodes," November 2009.\)](#). The way tunnels processed the ECN field before [\[I-D.ietf-tsvwg-ecn-tunnel\] \(Briscoe, B., "Tunnelling of Explicit Congestion Notification," March 2010.\)](#) severely limited how to encode these states.

The two bit ECN field seems to offer four possible encoding states, but one (00) is set aside for traffic controlled by transports that do not understand PCN marking [\[RFC5696\] \(Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," November 2009.\)](#), so it would be irregular and risky to use it as a PCN encoding state. Of the three remaining ECN codepoints, only one (11) can be introduced by a congested node within a tunnel and still survive the decapsulation behaviour of a tunnel egress not updated to comply with [\[I-D.ietf-tsvwg-ecn-tunnel\] \(Briscoe, B., "Tunnelling of Explicit Congestion Notification," March 2010.\)](#). The two remaining codepoints are (10) and (01). But if a node within the tunnel used either of these two remaining codepoints to try to mark packets with a second severity level, a tunnel not updated to comply with [\[I-D.ietf-tsvwg-ecn-tunnel\] \(Briscoe, B., "Tunnelling of Explicit Congestion Notification," March 2010.\)](#) would remove this marking on decapsulation. The ECN field was constrained to two marking states in this way irrespective of which earlier ECN tunnelling specification the tunnel complied with, whether regular IP in IP tunnelling [\[RFC3168\] \(Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification \(ECN\) to IP," September 2001.\)](#) or IPsec tunnelling [\[RFC4301\] \(Kent, S. and K. Seo, "Security Architecture for the Internet Protocol," December 2005.\)](#).

One way to provide another encoding state that survives tunnelling is to use a second Diffserv codepoint [\[I-D.ietf-pcn-3-state-encoding\] \(Briscoe, B., Moncaster, T., and M. Menth, "A PCN encoding using 2](#)

[DSCPs to provide 3 or more states," February 2010.](#)). Instead, to avoid wasting scarce Diffserv codepoints, a network operator can require tunnels in a PCN region to comply with [\[I-D.ietf-tsvwg-ecn-tunnel\] \(Briscoe, B., "Tunnelling of Explicit Congestion Notification," March 2010.\)](#), thus removing the constraints imposed by earlier tunnelling specifications.

Therefore this document presupposes tunnels in the PCN region comply with the newly proposed decapsulation rules defined in [\[I-D.ietf-tsvwg-ecn-tunnel\] \(Briscoe, B., "Tunnelling of Explicit Congestion Notification," March 2010.\)](#). Then the constraints of standard tunnels no longer apply so this document can define a 3-state encoding for PCN within one Diffserv codepoint.

4. The 3-in-1 PCN Encoding

The 3-in-1 PCN Encoding scheme is based closely on the baseline encoding defined in [\[RFC5696\] \(Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," November 2009.\)](#) so that there will be no compatibility issues if a PCN-domain evolves from using the baseline encoding scheme to the experimental scheme described here. The exact manner in which the PCN encoding states are carried in the IP header is shown in [Figure 1 \(3-in-1 PCN Encoding\)](#).

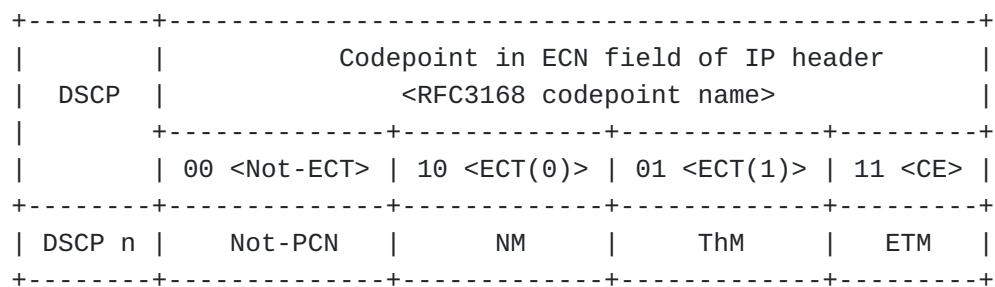


Figure 1: 3-in-1 PCN Encoding

In [Figure 1 \(3-in-1 PCN Encoding\)](#) the 3 PCN states are encoded in the ECN field [\[RFC3168\] \(Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification \(ECN\) to IP," September 2001.\)](#) of an IP packet with its Diffserv field [\[RFC2474\] \(Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field \(DS Field\) in the IPv4 and IPv6 Headers," December 1998.\)](#) set to DSCP n, which is any PCN-Compatible DiffServ

codepoint as defined in Section 4.2 of the PCN baseline encoding [\[RFC5696\] \(Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," November 2009.\)](#)). The PCN codepoint of a packet defines its marking state as follows:

Not-PCN: The packet is controlled by a transport that does not understand PCN marking, therefore the only valid action to notify congestion is to drop the packet;

NM: Not marked. A packet in the NM state has not (yet) had its marking state changed to the ThM or ETM states, but it may be changed to one of these states by a node experiencing congestion or pre-congestion;

ThM: Threshold-marked. Such a packet has had its marking state changed by the threshold-meter function [\[RFC5670\] \(Eardley, P., "Metering and Marking Behaviour of PCN-Nodes," November 2009.\)](#);

ETM: Excess-traffic-marked. Such a packet has had its marking state changed by the excess-traffic-meter function [\[RFC5670\] \(Eardley, P., "Metering and Marking Behaviour of PCN-Nodes," November 2009.\)](#).

Packets marked NM, ThM or ETM are termed PCN-packets. Their entry into the pcn-domain is controlled by edge nodes that understand how to process PCN markings [\[RFC5559\] \(Eardley, P., "Pre-Congestion Notification \(PCN\) Architecture," June 2009.\)](#).

5. Behaviour of a PCN Node Compliant with the 3-in-1 PCN Encoding

To be compliant with the 3-in-1 PCN Encoding, an PCN interior node behaves as follows:

*Except where explicitly stated otherwise, it MUST comply with the baseline encoding specified in [\[RFC5696\] \(Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," November 2009.\)](#)

*It MUST change NM to ThM if the threshold-meter function indicates to mark the packet.

*It MUST change NM or ThM to ETM if the excess-traffic-meter function indicates to mark the packet.

*It MUST NOT change Not-PCN to a PCN-Enabled codepoint and MUST NOT change a PCN-Enabled codepoint to Not-PCN;

*It MUST NOT change ThM to NM;

*It MUST NOT change ETM to ThM or to NM;

In other words, a PCN interior node may increase the severity of packet marking but it MUST NOT decrease it, where the order of severity increases from NM through ThM to ETM.

6. IANA Considerations

This memo includes no request to IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

7. Security Considerations

The security concerns relating to this extended PCN encoding are the same as those in [\[RFC5696\] \(Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," November 2009.\)](#).

8. Conclusions

The 3-in-1 PCN Encoding provides three states to encode PCN markings in the ECN field of an IP packet using just one Diffserv codepoint. One state is for not marked packets while the two others are for PCN nodes to mark packets with increasing levels of severity. Use of this encoding presupposes that any tunnels in the PCN region have been updated to comply with [\[I-D.ietf-tsvwg-ecn-tunnel\] \(Briscoe, B., "Tunnelling of Explicit Congestion Notification," March 2010.\)](#).

9. Acknowledgements

Thanks to Phil Eardley, Teco Boot, Kwok Ho Chan and Michael Menth for reviewing this.

10. Comments Solicited

To be removed by RFC Editor: Comments and questions are encouraged and very welcome. They can be addressed to the IETF Congestion and Pre-Congestion working group mailing list <pcn@ietf.org>, and/or to the authors.

11. References

11.1. Normative References

| | |
|-----------------------------|---|
| [I-D.ietf-tsvwg-ecn-tunnel] | Briscoe, B., " Tunnelling of Explicit Congestion Notification ," draft-ietf-tsvwg-ecn-tunnel-08 (work in progress), March 2010 (TXT). |
| [RFC2119] | Bradner, S. , " Key words for use in RFCs to Indicate Requirement Levels ," BCP 14, RFC 2119, March 1997 (TXT , HTML , XML). |
| [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 (TXT , HTML , XML). |
| [RFC3168] | Ramakrishnan, K., Floyd, S., and D. Black, " The Addition of Explicit Congestion Notification (ECN) to IP ," RFC 3168, September 2001 (TXT). |
| [RFC5559] | Eardley, P., " Pre-Congestion Notification (PCN) Architecture ," RFC 5559, June 2009 (TXT). |
| [RFC5670] | Eardley, P., " Metering and Marking Behaviour of PCN-Nodes ," RFC 5670, November 2009 (TXT). |
| [RFC5696] | Moncaster, T., Briscoe, B., and M. Menth, " Baseline Encoding and Transport of Pre-Congestion Information ," RFC 5696, November 2009 (TXT). |

11.2. Informative References

| | |
|---------------------------------|---|
| [I-D.ietf-pcn-3-state-encoding] | Briscoe, B., Moncaster, T., and M. Menth, " A PCN encoding using 2 DSCPs to provide 3 or more states ," draft-ietf-pcn-3-state-encoding-01 (work in progress), February 2010 (TXT). |
| [RFC4301] | Kent, S. and K. Seo, " Security Architecture for the Internet Protocol ," RFC 4301, December 2005 (TXT). |

Authors' Addresses

| | |
|--------|--|
| | Bob Briscoe |
| | BT |
| | B54/77, Adastral Park |
| | Martlesham Heath |
| | Ipswich IP5 3RE |
| | UK |
| Phone: | +44 1473 645196 |
| Email: | bob.briscoe@bt.com |
| URI: | http://bobbriscoe.net/ |
| | |
| | Toby Moncaster |
| | BT |
| | c/o B54/70, Adastral Park |
| | Martlesham Heath |
| | Ipswich IP5 3RE |
| | UK |
| Phone: | +44 1206 332805 |
| Email: | toby.moncaster@bt.com |