

Congestion and Pre-Congestion
Notification
Internet-Draft
Intended status: Experimental
Expires: April 30, 2009

B. Briscoe
BT
October 27, 2008

**PCN 3-State Encoding Extension in a single DSCP
draft-briscoe-pcn-3-in-1-encoding-00**

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 April 30, 2009.

Abstract

Pre-congestion notification (PCN) is a mechanism designed to protect the quality of service of inelastic flows. It does this by marking packets when traffic load on a link is approaching or has exceeded a threshold below the physical link rate. This document specifies an extension to the two-state PCN baseline encoding that enables three encoding states to be carried in the IP header without using more than one Diffserv codepoint. It presupposes a standards action has removed the limit of two encoding states in current tunnelling mechanisms.

1. Introduction

Pre-congestion notification provides information to support admission control and flow termination at the boundary nodes of a Diffserv region in order to protect the quality of service (QoS) of inelastic flows [[I-D.ietf-pcn-architecture](#)]. This is achieved by marking packets on interior nodes according to some metering function implemented at each node. Excess traffic marking marks PCN packets that exceed a certain reference rate on a link while threshold marking marks all PCN packets on a link when the PCN traffic rate exceeds a higher reference rate [[I-D.ietf-pcn-marking-behaviour](#)]. 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: one encoding for packets that are not marked plus two encodings for the two types of marking. The baseline encoding described in [[I-D.ietf-pcn-baseline-encoding](#)] 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.

If more than three states are required, e.g. to support end-to-end ECN as well as edge-to-edge PCN [[I-D.sarker-pcn-ecn-pcn-usecases](#)], end-to-end ECN would have to be encapsulated in the inner header of a tunnel through the PCN region, as outlined in [[I-D.ietf-pcn-architecture](#)].

General PCN-related terminology is defined in the PCN architecture [[I-D.ietf-pcn-architecture](#)], and terminology specific to packet encoding is defined in the PCN baseline encoding [[I-D.ietf-pcn-baseline-encoding](#)].

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](#) [[RFC2119](#)].

3. The Requirement for Three PCN Encoding States

The PCN architecture [[I-D.ietf-pcn-architecture](#)] 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. The way tunnels process the ECN field severely limits 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, so it would be irregular 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 as currently standardised. 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, this marking would be removed on decapsulation. The ECN field is constrained to two marking states in this way irrespective of whether regular IP in IP tunnelling [[RFC3168](#)] or IPsec tunnelling [[RFC4301](#)] is used.

One way to provide another encoding state that survives tunnelling is to use a second Diffserv codepoint [[I-D.moncaster-pcn-3-state-encoding](#)]. Instead, to avoid wasting scarce Diffserv codepoints, we could modify standard tunnels in the PCN region to remove the constraints imposed by standard tunnelling.

Therefore this document presupposes tunnels in the PCN region comply with the newly proposed Comprehensive Decapsulation Rules defined in [Appendix C](#) of [[I-D.ietf-tsvwg-ecn-tunnel](#)]. Then the constraints of standard tunnels no longer apply so this document can define a 3-state encoding for PCN within one Diffserv codepoint.

Note that [[I-D.ietf-tsvwg-ecn-tunnel](#)] only records the Comprehensive Decapsulation Rules in an appendix, solely to allow us to discuss whether they should be brought on to the standards track. So, if they are not, the offending tunnelling constraints might not be removed. However, [[I-D.ietf-tsvwg-ecn-tunnel](#)] carefully establishes that the constraints imposed by standard tunnelling are actually unnecessary; they are merely the result of an unfortunate sequence of historical events. Then the analysis in [Appendix C](#) of [[I-D.ietf-tsvwg-ecn-tunnel](#)] shows that the proposed new rules would not introduce any compatibility issues; they only affect one combination of inner and outer header which has never occurred under any legal transitions of any IETF tunnelling schemes. Therefore it is a reasonable working assumption for the purposes of this document that tunnelling constraints will one day be removed.

4. The 3-in-1 PCN Encoding

The 3-in-1 PCN Encoding scheme is based closely on that defined in [I-D.ietf-pcn-baseline-encoding] 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 Table 1.

Codepoint in ECN field of IP header
 <[RFC3168](#) codepoint name>

+-----+-----+-----+-----+-----+				
DSCP	00 <Not-ECT>	10 <ECT(0)>	01 <ECT(1)>	11 <CE>
+-----+-----+-----+-----+-----+				
DSCP n	Not-PCN	NM	ThM	ETM
+-----+-----+-----+-----+-----+				

Table 1: 3-in-1 PCN Encoding

In Table 1 the 3 PCN states are encoded in the ECN field ([RFC3168]) of an IP packet with its Diffserv field ([RFC2474]) set to DSCP n, which is any PCN-Compatible DiffServ codepoint as defined in Section 4.2 of the PCN baseline encoding [I-D.ietf-pcn-baseline-encoding]). 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 marking algorithm [I-D.ietf-pcn-marking-behaviour];

ETM: Excess traffic marking. Such a packet has had its marking state changed by the excess rate marking algorithm [I-D.ietf-pcn-marking-behaviour].

Packets marked NM, ThM or ETM are termed PCN-Enabled packets because they are controlled by edge nodes that understand how to process PCN markings.

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:

- o It MUST NOT change Not-PCN to a PCN-Enabled codepoint and MUST NOT change a PCN-Enabled codepoint to Not-PCN;
- o It MUST NOT change ThM to NM;
- o 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 essentially the same as those in [[I-D.ietf-pcn-baseline-encoding](#)].

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 the encoding presupposes that any tunnels in the PCN region have been updated to use proposed Comprehensive Decapsulation Rules because standard tunnel decapsulation rules unnecessarily constrain PCN marking.

9. Acknowledgements

Thanks to Phil Eardley for reviewing this.

10. Comments Solicited

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-pcn-marking-behaviour]
Eardley, P., "Marking behaviour of PCN-nodes",
[draft-ietf-pcn-marking-behaviour-01](#) (work in progress),
October 2008.
- [I-D.ietf-tsvwg-ecn-tunnel]
Briscoe, B., "Layered Encapsulation of Congestion
Notification", [draft-ietf-tsvwg-ecn-tunnel-01](#) (work in
progress), October 2008.
- [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.
- [RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition
of Explicit Congestion Notification (ECN) to IP",
[RFC 3168](#), September 2001.

11.2. Informative References

- [I-D.ietf-pcn-architecture]
Eardley, P., "Pre-Congestion Notification (PCN)
Architecture", [draft-ietf-pcn-architecture-08](#) (work in
progress), October 2008.
- [I-D.ietf-pcn-baseline-encoding]
Moncaster, T., Briscoe, B., and M. Menth, "Baseline
Encoding and Transport of Pre-Congestion Information",
[draft-ietf-pcn-baseline-encoding-01](#) (work in progress),
October 2008.
- [I-D.moncaster-pcn-3-state-encoding]
Moncaster, T., Briscoe, B., and M. Menth, "A three state

extended PCN encoding scheme",
[draft-moncaster-pcn-3-state-encoding-00](#) (work in progress), June 2008.

[I-D.sarker-pcn-ecn-pcn-usecases]

Sarker, Z. and I. Johansson, "Usecases and Benefits of end to end ECN support in PCN Domains",
[draft-sarker-pcn-ecn-pcn-usecases-01](#) (work in progress), May 2008.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", [RFC 4301](#), December 2005.

Author's Address

Bob Briscoe
BT
B54/77, Adastral Park
Martlesham Heath
Ipswich IP5 3RE
UK

Phone: +44 1473 645196

Email: bob.briscoe@bt.com

URI: <http://www.cs.ucl.ac.uk/staff/B.Briscoe/>

Full Copyright Statement

Copyright (C) The IETF Trust (2008).

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

This document was produced using xml2rfc v1.33 (of <http://xml.resource.org/>) from a source in [RFC-2629](#) XML format.

