Congestion and Pre Congestion
Internet-Draft
Intended status: Historic
Expires: September 13, 2012

M. Menth
University of Tuebingen
J. Babiarz
3inova Networks Inc
T. Moncaster
University of Cambridge
B. Briscoe
BT

March 12, 2012

PCN Encoding for Packet-Specific Dual Marking (PSDM Encoding) draft-ietf-pcn-psdm-encoding-02

Abstract

Pre-congestion notification (PCN) is a link-specific and load-dependent packet re-marking mechanism and provides in Differentiated Services networks feedback to egress nodes about load conditions within the domain. It is used to support admission control and flow termination decision in a simple way. This document proposes how PCN marks can be encoded into the IP header for packet-specific dual marking (PSDM). PSDM encoding provides three different codepoints: not-ETM, not-ThM, and PM. Excess-traffic-marking may re-mark not-ETM-packets to PM and threshold-marking may re-mark not-ThM-packets to PM.

Status

Since its original publication, the baseline encoding (RFC5696) on which this document depends has become obsolete. The PCN working GRoup has chosen to publish this as a historical document to preserve the details of the encoding and to allow it to be cited in other documents.

Status of this Memo

material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 13, 2012.

Copyright Notice

Table of Contents

1. Introduction 4
1.1. Requirements notation 5
2. Terminology 5
3. Encoding for Packet-Specific Dual Marking 6
3.1. Proposed Encoding and Expected Node Behavior 6
3.1.1. PCN Codepoints 6
3.1.2. Codepoint Handling by PCN Ingress Nodes 6
3.1.3. Codepoint Handling by PCN Interfaces 7
3.1.4. Codepoint Handling by PCN Egress Nodes 7
3.2. Reasons for the Proposed Encoding 7
3.2.1. Scarcity of DSCPs 7
3.2.2. Problems with Tunneling 7
3.2.3. Problems with the ECN Field 8
3.3. Handling of ECN Traffic 9
4. IANA Considerations 9
5. Security Considerations 9
6. Conclusions 9
7. Comments Solicited 9
8. References 10
8.1. Normative References 10
8.2. Informative References 10
Authors' Addresses 11

1. Introduction

The objective of Pre-Congestion Notification (PCN) [RFC5559] 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 (AC), to decide whether to admit or block a new flow request, and (in abnormal circumstances) flow termination (FT) 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. 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 lower reference rate [RFC5670]. These marks are monitored by the egress nodes of the PCN domain.

This document proposes how PCN marks can be encoded into the IP header when packet-specific dual marking (PSDM) is used to re-mark packets [Menth09f]. That means, both excess-traffic-marking and threshold-marking are activated on the links within a PCN domain, but packets are subject to re-marking by only one of them. The encoding of unmarked PCN packets indicates whether they are subject to either excess-traffic-marking (not-ETM) or threshold-marking (not-ThM) and they may be re-marked to PCN-marked (PM).

PSDM encoding can be applied in networks implementing

- only AC based on threshold-marking (reference rate = PCN-admissible-rate),
- only FT based on excess-traffic-marking (reference rate = PCN-supportable-rate),
- both AC and FT based on excess-traffic-marking (reference rate = PCN-admissible-rate)
- Probe-based AC based on threshold-marking (reference rate = PCN-admissible-rate) and FT based on excess-traffic-marking (reference rate = PCN-supportable-rate)[Menth09f].

The motivation for PSDM encoding is that probe packets are subject only to threshold-marking and that data packets are subject only to excess-traffic-marking. Nevertheless, routers should not need to differentiate explicitly between probe and data packets since packets are a priori marked with an appropriate codepoint (not-ETM, not-ThM) indicating the marking mechanism applying to them.

Following the publication of new rules relating to the tunnelling of ECN marks [RFC6040], the PCN workign group decided to obsolete [RFC5696] in favour of the 3-in-1 encoding [I-D.ietf-pcn-3-in-1-encoding]. A side-effect of this decision was to make the PSDM encoding obsolete. However the PCN working group feels it is useful to have a formal historical record of this encoding. This ensures details of the encoding are not lost and also allows it to be cited in other documents.

1.1. Requirements notation

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].

2. Terminology

Most of the terminology used in this document is defined in [RFC5559]. The following additional terms are defined in this document:

- PCN-capable flow - a flow subject to PCN-based admission control and/or flow termination
- PCN-enabled DSCP - DSCP indicating within a PCN domain that packets possibly belong to a PCN-capable flow
- PCN-capable ECN codepoint (PCN codepoint) - DSCP set to a PCN-enabled DSCP and ECN field set to a codepoint indicating that a packet belongs to a PCN-capable flow (not-ThM, not-ETM, or PM, explained below)
- PCN packet - a packet belonging to a PCN capable flow within a PCN domain, must have a PCN-enabled DSCP and a PCN-capable ECN codepoint
- not-PCN capable (not-PCN) - new ECN codepoint for packets of non-PCN-capable flows when a PCN-enabled DSCP is set

- o not-excess-traffic-marked (not-ETM) - new ECN codepoint for unmarked PCN packets that are subject to excess-traffic-marking

- o not-threshold-marked (not-ThM) - new ECN codepoint for unmarked PCN packets that are subject to threshold-marking

- o PCN-marked (PM) - new ECN codepoint for re-marked PCN packets regardless whether they were subject to excess-traffic-marking or threshold-marking.

3. Encoding for Packet-Specific Dual Marking

In this section the encoding for packet-specific dual marking (PSDM) is presented and the reasons for the proposed design are outlined.

3.1. Proposed Encoding and Expected Node Behavior

The encoding reuses a PCN-enabled DSCP to indicate packets of PCN-capable flows within a PCN domain.

3.1.1. PCN Codepoints

The ECN field of packets with a PCN-enabled DSCP is interpreted within a PCN domain as PCN codepoint while it is interpreted as ECN codepoint outside PCN domains. Four new PCN codepoints are defined in Figure 1. PSDM encoding can be seen as an extension of baseline encoding [RFC5696]

	Codepoint in ECN field of IP header			
DSCP	(RFC3168 codepoint name)			
	00 (Not-ECT)	10 (ECT(0))	01 (ECT(1))	11 (CE)
DSCP n	not-PCN	not-ETM	not-ThM	PM

Figure 1: PSDM encoding.

3.1.2. Codepoint Handling by PCN Ingress Nodes

When packets belonging to PCN flows arrive at the ingress router of the PCN domain, the ingress router first drops all CE-marked packets. Then, it sets the DSCP of the remaining PCN packets to an PCN-enabled DSCP and re-marks the ECN field of all PCN packets that are subject to threshold-marking to not-ThM (e.g. probe packets), and all PCN packets that are subject to excess-traffic-marking to not-ETM (e.g.

data packets). If packets with a PCN-enabled DSCP arrive that belong to non-PCN flows, the PCN ingress node re-marks their ECN field to not-PCN or re-marks their DSCP to a different one while preserving the contents of the ECN field.

3.1.3. Codepoint Handling by PCN Interfaces

If the meter for excess-traffic-marking of a PCN node indicates that a PCN packet should be re-marked, its ECN field is set to PCN-marked (PM) only if it was not-ETM before. If the meter for threshold-marking of a PCN node indicates that a PCN packet should be re-marked, its ECN field is set to PCN-marked (PM) only if it was not-ThM before.

3.1.4. Codepoint Handling by PCN Egress Nodes

If the egress node of a PCN domain receives a PM-packet, it infers somehow whether the packet was not-ETM or not-ThM by the PCN ingress node to interpret the marking. This can be done as probe packets must be distinguishable from PCN data packets anyway. The egress node resets the ECN field of all packets with PCN-enabled DSCPs to not-ECT. This breaks the ECN capability for all flows with PCN-enabled DSCPs, regardless whether they are PCN-capable or not. Appropriate tunnelling across a PCN domain can preserve the ECN marking of packets with PCN-enabled DSCPs and the ECN-capability of their flows (see Section 3.3). When the DSCPs in the headers of packets belonging to flows with PCN-enabled DSCPs have been changed to another DSCP, the egress node should reverse that change.

3.2. Reasons for the Proposed Encoding

3.2.1. Scarcity of DSCPs

DSCPs are a scarce resource in the IP header so that at most one should be used for PCN encoding.

3.2.2. Problems with Tunneling

The encoding scheme must cope with tunnelling within PCN domains. However, various tunnelling schemes limit the persistence of ECN marks in the top-most IP header to a different degree. Two IP-in-IP tunnelling modes are defined in [RFC3168] and a third one in [RFC4301] for IP-in-IPsec tunnels.

The limited-functionality option in [RFC3168] requires that the ECN codepoint in the outer header is set to not-ECT so that ECN is disabled for all tunnel routers, i.e., they drop packets instead of mark them in case of congestion. The tunnel egress just decapsulates

the packet and leaves the ECN codepoints of the inner packet header unchanged.

- This mode protects the inner IP header from being PCN-marked upon decapsulation. It can be used to tunnel ECN marks across PCN domains such that PCN marking is applied to the outer header without affecting the inner header.

- This mode is not useful to tunnel PCN traffic with PCN-enabled DSCP and PCN-capable PCN-codepoints within PCN domain because the ECN marking information from the outer ECN fields is lost upon decapsulation.

The full-functionality option in [RFC3168] requires that the ECN codepoint in the outer header is copied from the inner header unless the inner header codepoint is CE. In this case, the outer header codepoint is set to ECT(0). This choice has been made to disable the ECN fields of the outer header as a covert channel. Upon decapsulation, the ECN codepoint of the inner header remains unchanged unless the outer header ECN codepoint is CE. In this case, the inner header codepoint is also set to CE. This preserves outer header information if it is CE. However, the fact that CE marks of the inner header are not visible in the outer header may be a problem for excess-traffic-marking as it takes already marked traffic into account and for some required packet drop policies.

Tunnelling with IPsec copies the inner header ECN field to the outer header ECN field RFC4301, Sect. 5.1.2.1 [RFC4301] upon encapsulation. Upon decapsulation, CE-marks of the outer header are copied into the inner header, the other marks are ignored. With this tunnelling mode, CE marks of the inner header become visible to all meters, markers, and droppers for tunnelled traffic. In addition, limited information from the outer header is propagated into the inner header. Therefore, only IPsec tunnels should be used inside PCN domains when ECN bits are reused for PCN encoding. Another consequence is that CE is the only codepoint to which packets can be re-marked along a tunnel within a PCN domain so that the changed codepoint survives decapsulation.

3.2.3. Problems with the ECN Field

The guidelines in [RFC4774] describe how the ECN bits can be reused while being compatible with [RFC3168]. A CE mark of a packet must never be changed to another ECN codepoint. Furthermore, a not-ECT mark of a packet must never be changed to one of the ECN-capable codepoints ECT(0), ECT(1), or CE. Care must be taken that this rule is enforced when PCN packets leave the PCN domain. As a consequence, all CE-marked PCN packets must be dropped before entering a PCN

domain and the ECN field of all PCN packets must be reset to not-ECT when leaving a PCN domain.

3.3. Handling of ECN Traffic

ECN is intended to control elastic traffic as TCP reacts to ECN marks. Inelastic real-time traffic is mostly not transmitted over TCP such that this application of ECN is not appropriate. However, there have been proposals made that would re-use the PCN signals for rate adaptation. Therefore, two different options might be useful.

- preserve ECN marks from outside a PCN domain, i.e. CE-marked packets should not be dropped. To handle this case, ECN packets should be tunnelled through a PCN domain such that the ECN marking is hidden from the PCN control and PCN marking is applied only to the outer header.

- add PCN markings to the ECN field if applications wish to receive the PCN markings for whatever purpose. In that case IPsec tunnels should be used for tunnelling. This, however, must be done only if end systems are ECN capable and signal that they wish to receive this additional PCN marking information. If this is useful, the required signalling needs to be defined.

Both options are an independent of the way how PCN marks are encoded. Therefore, they are not in the scope of this document.

4. IANA Considerations

This document makes no request to IANA.

5. Security Considerations

{ToDo}

6. Conclusions

This document describes an encoding scheme with the following benefits: {ToDo}

7. Comments Solicited

Comments and questions are encouraged and very welcome. They can be addressed to the IETF PCN working group mailing list <pcn@ietf.org>,

and/or to the authors.

8. References

8.1. Normative References

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

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, September 2001.

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

[RFC4774] Floyd, S., "Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) Field", BCP 124, RFC 4774, November 2006.

[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.

8.2. Informative References

[I-D.ietf-pcn-3-in-1-encoding]
Briscoe, B., Moncaster, T., and M. Menth, "Encoding 3 PCN-States in the IP header using a single DSCP", draft-ietf-pcn-3-in-1-encoding-09 (work in progress), March 2012.

[Menth09f]
Menth, M., Babiarz, J., and P. Eardley, "Pre-Congestion Notification Using Packet-Specific Dual Marking", Proceedings of the International Workshop on the Network of the Future (Future-Net), IEEE, Dresden, Germany, June 2009.

[RFC6040] Briscoe, B., "Tunnelling of Explicit Congestion Notification", RFC 6040, November 2010.

Authors' Addresses

Michael Menth
University of Tuebingen
Department of Computer Science
Sand 13
Tuebingen D-72076
Germany

Phone: +49 07071 29 70505
Email: menth@informatik.uni-tuebingen.de
URI: http://www.kn.inf.uni-tuebingen.de

Jozef Babiarz
3inova Networks Inc
CRC Innovation Centre
Bldg 94 Room 216D
3701 Carling Avenue
Ottawa K2H 8S2
Canada

Phone: +1-613-298-0438
Email: j.babiarz@3inovanetworks.com

Toby Moncaster
University of Cambridge
Computer Laboratory
JJ Thomson Avenue
Cambridge CB3 0FD
UK

Phone: +44 1223 763654
Email: toby@moncaster.com

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.bobbriscoe.net