Congestion and Pre Congestion
Internet-Draft
Intended status: Experimental
Expires: January 7, 2008

M. Menth
University of Wuerzburg
J. Babiarz
Nortel Networks
T. Moncaster
BT
B. Briscoe
BT & UCL
July 7, 2008

PCN Encoding for Packet-Specific Dual Marking (PSDM)
draft-menth-pcn-psdm-encoding-00

Abstract

This document proposes how PCN marks can be encoded into the IP header. The presented encoding reuses the ECN field of the Voice-Admit DSCP in a single PCN domain. The encoding of unmarked PCN

packets indicates whether they are subject to either excess- or exhaustive-marking. This is useful, e.g., when data and probe packets require different marking mechanisms.

Status

This memo is posted as an Internet-Draft with an intent to eventually be published as an experimental RFC.

Table of Contents

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 [PCN-arch]. 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 exhaustive marking marks all PCN packets on a link when the PCN traffic rate exceeds a reference rate [PCN-marking-behaviour]. 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. The presented encoding reuses the ECN field of the Voice-Admit DSCP in a single PCN domain. The encoding of unmarked PCN packets indicates whether they are subject to either excess- or exhaustive-marking. Therefore, we call this proposal encoding for packet-specific dual marking (PSDM).

PSDM supports exhaustive marking and excess marking as long as individual packets are subject to only one of them. It can be applied in networks implementing

- only AC based on exhaustive marking (reference rate = admissible rate),
- only FT based on excess marking (reference rate = supportable rate),
- both AC and FT based on excess marking (reference rate = admissible rate)
- Probe-based AC based on exhaustive marking (reference rate = admissible rate) and FT based on excess marking (reference rate = supportable rate).

Although the motivation for this encoding scheme is to exhaustive-mark probe packets and to excess-mark data packets, routers do not need to differentiate explicitly between probe and data packets since packets are a priori marked with an appropriate codepoint indicating the marking mechanism applying to them.

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 [PCN-arch]. The following additional terms are defined in this document:

- Exhaustive marking - generalization of threshold and ramp marking
- 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-ExM, not-EhM, or M, 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
- not-excess-marked (not-ExM) - new ECN codepoint for unmarked PCN packets that are subject to excess marking
- not-exhaustive-marked (not-EhM) - new ECN codepoint for unmarked PCN packets that are subject to exhaustive marking
- marked (M) - new ECN codepoint for marked PCN packets regardless whether they were subject to excess or exhaustive marking.

3. Encoding for Packet-Specific Dual Marking

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

3.1. Proposed Encoding and Expected Node Behavior

The encoding reuses the Voice-Admit DSCP [voice-admit] as a PCN-enabled DSCP to indicate packets of PCN-capable flows within a PCN domain. So far, this is the only DSCP considered for that use, but this encoding scheme is easily extensible towards multiple PCN-

enabled DSCPs.

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 Table 1.

DSCP	00	10	01	11
PCN-enabled DSCP	not-PCN	not-ExM	Not-EhM	M

Table 1: Mapping of PCN codepoints into the ECN field

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 exhaustive marking to not-EhM (e.g. probe packets), and all PCN packets that are subject to excess marking to not-ExM (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.

3.1.3. Codepoint Handling by PCN Interfaces

If the meter for excess marking of a PCN node indicates that a PCN packet should be marked, its ECN field is set to marked (M) only if it was not-ExM before. If the meter for exhaustive marking of a PCN node indicates that a PCN packet should be marked, its ECN field is set to marked (M) only if it was not-EhM before.

3.1.4. Codepoint Handling by PCN Egress Nodes

If the egress node of a PCN domain receives a marked PCN packet, it infers somehow whether the packet was not-ExM or not-EhM by the PCN ingress node to interpret the marking. This can be done as probe packets must be distinguishable from PCN data packets. 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).

3.2. Reasons for the Proposed Encoding

3.2.1. Problems with DSCPs

DSCPs are a scarce resource in the IP header such that at most one should be used for PCN. To avoid the requirement for a new DSCP, the Voice-Admit DSCP is reused. To differentiate pure Voice-Admit traffic from PCN traffic within a PCN domain, pure Voice-Admit traffic has its ECN field set to not-PCN within a PCN domain. The encoding should be extensible towards different data plane priorities for PCN traffic in PCN domains which requires different PCN-enabled DSCPs, one for each priority level.

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 such 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 marking as it takes already marked traffic into account and for some required packet drop policies.

Tunnelling with IPSec copies the inner header ECN bits to the outer header ECN bits 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 that can be used to indicate a marked packet beyond tunnelling.

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 Voice-Admit packets must be dropped before entering a PCN domain and the ECN field of all Voice-Admit packets must be set 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 are plans to reuse ECN signals for rate adapatation [ecn-pcn-usecases]. 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. It does however suggest a change to the ([RFC3168]) behaviour for the ECN field for the Voice-Admit [voice-admit] DSCP within a PCN domain.

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.

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

8.2. Informative References

[PCN-arch] Eardley, P., "Pre-Congestion Notification Architecture", draft-ietf-pcn-architecture-03 (work in progress),

February 2008.

[PCN-marking-behaviour]
Eardley, P., "Marking behaviour of PCN-nodes", draft-eardley-pcn-marking-behaviour-01 (work in progress), June 2008.

[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.

[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.

[voice-admit]
Baker, F., Polk, J., and M. Dolly, "DSCPs for Capacity-Admitted Traffic", draft-ietf-tsvwg-admitted-realtime-dscp-04 (work in progress), February 2008.

Authors' Addresses

Michael Menth
University of Wuerzburg
room B206, Institute of Computer Science
Am Hubland
Wuerzburg D-97074
Germany

Phone: +49 931 888 6644
Email: menth@informatik.uni-wuerzburg.de

Jozef Babiarz
Nortel Networks
3500 Carling Avenue
Ottawa K2H 8E9
Canada

Phone: +1-613-763-6098
Email: babiarz@nortel.com

Toby Moncaster
BT
B54/70, Adastral Park
Martlesham Heath
Ipswich IP5 3RE
UK

Phone: +44 1473 648734
Email: toby.moncaster@bt.com
URI: http://www.cs.ucl.ac.uk/staff/B.Briscoe/

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

Phone: +44 1473 645196
Email: bob.briscoe@bt.com

Full Copyright Statement

Intellectual Property

Acknowledgments

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA). This document was produced using xml2rfc v1.32 (of http://xml.resource.org/) from a source in RFC-2629 XML format.