PCN
Internet-Draft
Intended status: Informational
Expires: October 12, 2011

G. Karagiannis
University of Twente
K. Chan
Huawei Technologies
T. Moncaster
Moncaster.com
M. Menth
University of Tuebingen
P. Eardley
B. Briscoe
BT

April 12, 2011

Overview of Pre-Congestion Notification Encoding
draft-ietf-pcn-encoding-comparison-05

Abstract

The objective of Pre-Congestion Notification (PCN) is to protect the quality of service (QoS) of inelastic flows within a Diffserv domain. On every link in the PCN domain, the overall rate of the PCN-traffic is metered, and PCN-packets are appropriately marked when certain configured rates are exceeded. Egress nodes provide decision points with information about the PCN-marks of PCN-packets which allows them to take decisions about whether to admit or block a new flow request, and to terminate some already admitted flows during serious pre-congestion.

The PCN Working Group explored a number of approaches for encoding this pre-congestion information into the IP header. This document provides details of all those approaches along with an explanation of the constraints that had to be met by any solution.

Status of this Memo

This Internet-Draft will expire on October 12, 2011.

Copyright Notice

Table of Contents

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, to decide whether to admit or block a new flow request, and flow termination to terminate some existing flows during serious pre-congestion. 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 boundary nodes are notified of a potential overload before

any real congestion occurs (hence "pre-congestion notification").

[RFC5670] provides for two metering and marking functions that are configured with reference rates. Threshold-marking marks all PCN packets once their traffic rate on a link exceeds the configured reference rate (PCN-threshold-rate). Excess-traffic-marking marks only those PCN packets that exceed the configured reference rate (PCN-excess-rate).

Egress nodes monitor the PCN-marks of received PCN-packets and provide information about the PCN-marks to decision points which take decisions about flow admission and termination on this basis [I-D.ietf-pcn-cl-edge-behaviour], [I-D.ietf-pcn-sm-edge-behaviour].

This PCN information has to be encoded into the IP header. This requires at least three different codepoints for not-marked PCN traffic, PCN traffic re-marked by the threshold-marker, and PCN traffic re-marked by the excess-traffic-marker. Since unused codepoints are not available for that purpose in the IP header (version 4 and 6), already used codepoints must be re-used which imposes additional constraints on design and applicability of PCN-based AC and FT. This document summarizes these issues for historical purposes.

In Section 2, we briefly point out PCN encoding requirement imposed by metering and marking algorithms, and by special packet drop strategies. The Differentiated Services Codepoint (6 bits) and the ECN field (2 bits) have been selected to be re-used for encoding of PCN marks (PCN encoding). In Section 3, we briefly explain the constraints imposed by this decision. In Section 4, we review different PCN encodings supported by the PCN working group that allow different implementations of PCN-based admission control and flow termination which have different pros and cons.

2. General PCN Encoding Requirements

The choice of metering and marking algorithms and the way they are applied to PCN-based AC (Admission Control) and FT (Flow Termination) impose certain requirements on PCN encoding.

2.1. Metering and Marking Algorithms

Two different metering and marking algorithms are defined in [RFC5670]: excess-traffic-marking and threshold-marking. They are both configured with reference rates which are termed PCN-excess-rate and PCN-threshold-rate, respectively. When traffic for PCN flows enter a PCN domain, the PCN ingress node sets a codepoint in the IP header indicating that the packet is subject to PCN metering and marking and that it is not-marked (NM). The two metering and marking algorithms possibly re-mark PCN packets as PCN and excess-traffic-marked (ETM) or threshold-marked (ThM).

Excess-traffic-marking leaves a rate of PCN traffic equal to the PCN-excess-rate to be not-ETM marked if possible. To that end, the algorithm needs to know whether a PCN packet has already been ETM marked or not. Threshold-marking re-marks all not-marked PCN traffic to ThM when the rate of PCN traffic exceeds the PCN-threshold-rate. Therefore, it does not need knowledge of the prior marking state of the packet for metering, but it needs it for packet re-marking.

2.2. Approaches for PCN-Based Admission Control and Flow Termination

We briefly review three different approaches to implement PCN-based AC and FT and derive their requirements for PCN encoding.

2.2.1. Dual Marking (DM)

The intuitive approach for PCN-based AC and FT requires that threshold and excess-traffic-marking are simultaneously activated on all links of a PCN domain and their reference rate is configured with the PCN-admissible-rate (AR) and the PCN-supportable-rate (SR), respectively. Threshold-marking meters all PCN traffic, but re-marks only not-marked traffic (NM) to ThM. Excess-traffic-marking meters only non-ETM traffic and re-marks either not-marked (NM) or threshold-marked (ThM) PCN traffic to ETM. Thus, both meters and markers need to identify PCN packets and their exact PCN codepoint. We call this marking behavior dual marking (DM) and Figure 1 illustrates all possible re-marking actions.

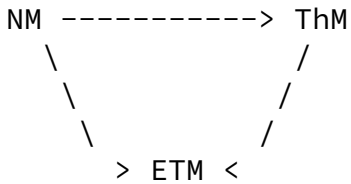


Figure 1: PCN Codepoint Re-Marking Diagram for Dual Marking (DM)

Dual marking is used to support the Controlled-Load PCN (CL-PCN) edge behavior [I-D.ietf-pcn-cl-edge-behaviour]. We briefly summarize the concept. All actions are performed on per ingress-egress-aggregate basis. The egress node measures the rate of NM-, ThM-, and ETM-traffic in regular intervals and sends them as PCN egress reports to

the AC and FT decision point.

If the proportion of re-marked (ThM- and ETM-) PCN traffic is larger than a CLE-limit, the decision point blocks new flow requests until new PCN egress reports are received, otherwise it admits them. With CL-PCN, AC is rather robust with regard to the value chosen for the CLE-limit. FT works as follows. If the ETM-traffic rate is positive, the decision point triggers the ingress node to send a newly measured rate of the sent PCN traffic. The decision point calculates the rate of PCN traffic that needs to be terminated by:

termination-rate = PCN-ingress-rate - (rate-of-NM-traffic + rate-of-ThM-traffic)

and terminates an appropriate set of flows. CL-PCN is accurate enough for most application scenarios and its implementation complexity is acceptable, therefore, it is a preferred implementation option for PCN-based AC and FT.

2.2.2. Single Marking (SM)

Single-marking uses only excess-traffic-marking whose reference rate is set to the PCN-admissible-rate (AR) on all links of the PCN domain. Figure 2 illustrates all possible re-marking actions.

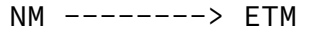


Figure 2: PCN Codepoint Re-Marking Diagram for Single Marking (SM)

Single marking is used to support the single-marking PCN (SM-PCN) edge behavior [I-D.ietf-pcn-sm-edge-behaviour]. We briefly summarize the concept. AC works essentially in the same way as with CL-PCN but AC is sensitive to the value of the CLE-limit. Also FT works similarly to CL-PCN. The PCN-supportable-rate (SR) is not configured on any link, but is implicitly:

SR=u*AR

in the PCN domain using a network-wide constant u. The decision point triggers FT only if the rate-of-NM-traffic * u < rate-of-NM-traffic + rate-of-ETM-traffic, requests the PCN-sent-rate from the corresponding PCN-ingress-node, calculates the amount of PCN traffic

to be terminated by

termination-rate = PCN-sent-rate - rate-of-NM-traffic * u,

and terminates an appropriate set of flows.

SM-PCN has two major benefits: it requires only two PCN codepoints and only excess-traffic-marking is needed which means that it might be earlier to the market than CL-PCN since some chipsets do not yet support threshold-marking. However, it only works well when ingress-egress-aggregates have a high PCN packet rate which is not always the case. Otherwise, over-admission and over-termination may occur [Menth08-Sub-8] [Menth10q].

2.2.3. Packet Specific Dual Marking (PSDM)

Packet-specific dual marking (PSDM) uses threshold-marking and excess-traffic-marking whose reference rates are configured with the PCN-admissible-rate and the PCN-supportable-rate, respectively. There are two different types of not-marked packets: those that are subject to threshold-marking (not-ThM) and those that are subject to excess-traffic-marking (not-ETM). Both not-Thm and not-ETM have the same NM-marking and are distinguished by higher layer information (see below). Threshold-marking meters all PCN traffic and re-marks only not-ThM packets to PCN-marked (PM). In contrast, excess-traffic-marking meters only not-ETM packets and possibly re-marks them to PM, too. Again, both meters and markers need to identify PCN packets and their exact PCN codepoint. Figure 3 illustrates all possible re-marking actions.

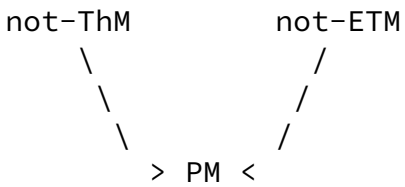


Figure 3: PCN Codepoint Re-Marking Diagram for Packet Specific Dual Marking (PSDM)

An edge behavior for PSDM has been presented in [Menth09f]. We call it PSDM-PCN. In contrast to CL-PCN and SM-PCN, AC is realized by re-using marked signaling messages for probing. The assumption is that

admission requests are triggered by an external end-to-end signaling protocol, e.g. RSVP (RFC2205). Signaling traffic for a flow is also labeled as PCN traffic and if an initial signaling traverses the PCN domain and is re-marked, then the corresponding flow is blocked. This is a light-weight probing mechanism which does not generate extra traffic and does not introduce probing delay [draft-menth-pcn-marked-signaling-ac]. In PSDM-PCN, PCN-ingress-nodes label initial signaling messages as not-ThM and threshold-marking configured with admissible rates possibly re-marks them to PM. Data packets are labeled with not-ETM and excess-traffic-marking configured with supportable rates possibly re-marks them to PM, too, so that the same algorithms for FT may be used as for CL-PCN and SM-PCN.

Disadvantages of this approach are that every end-to-end signaling protocol, e.g. RSVP, needs to be adapted that it denies admission if initial request messages are re-marked to PM.
Advantages are that the AC algorithm is more accurate than the one of CL-PCN and SM-PCN [Menth08-Sub-8], that only a single DSCP is needed, and that the new tunneling rules in RFC6040 are not needed for deployment.

.

2.2.4. Preferential Packet Dropping

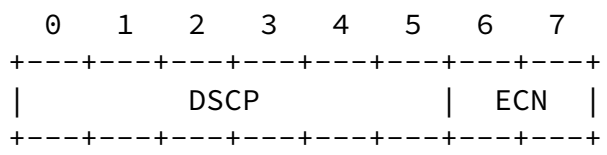
The termination algorithms proposed in the standards require preferential dropping of ETM-marked packets to avoid over-termination in case of packet loss [I-D.ietf-pcn-cl-edge-behaviour], [I-D.ietf-pcn-sm-edge-behaviour]. An analysis explaining this phenomenon can be found in Section 4 of [Menth10q]. Thus, preferential dropping of ETM-marked packets is RECOMMENDED in [RFC5670]. As a consequence, droppers must have access to the exact marking information of PCN-packets.

3. Encoding Constraints

The Differentiated Services (DS) field is chosen for the encoding of PCN marks. This section briefly reviews the DS field and the constraints imposed by its reuse are summarized.

3.1. Structure of the DS Field

Figure 4 shows the structure of the DS field. [RFC0793] defined the 8 bit ToS field and [RFC2474] redefined it as DS field. It consists of a 6 bit DS codepoint (DSCP, see [RFC2474]) and the 2 bit ECN field (see [RFC3168]).



DSCP: Differentiated Services codepoint [RFC2474]
ECN: ECN field [RFC3168]

Figure 4: The Structure of the DS Field

3.2. Constraints from the DSCP

The Differentiated Services codepoint (DSCP) indicates the per-hop behavior (PHB), i.e., the treatment IP packets receive from nodes in a DS domain. Multiple DSCPs may indicate the same PHB. PCN traffic is high-priority traffic and requires a special DSCP that indicate a PHB with preferred treatment.

3.2.1. General Scarcity of DSCPs

As the number of unused DSCPs is small, PCN encoding should use only a single DSCP if possible, in any case not more than two DSCPs. Therefore, the DSCP should be used to indicate that traffic is subject to PCN metering and marking, but not to differentiate different PCN markings.

3.2.2. Tunneling Rules for the Handling of the DSCP

PCN encoding must be chosen in such a way that PCN traffic can be tunneled within a PCN domain without any impact on PCN metering and re-marking. In the following, the "inner header" refers to the header of the encapsulated packet and the "outer header" refers to the encapsulating header.

[RFC2983] provides two tunneling modes for Differentiated Services networks. The uniform model copies the DSCP from the inner header to the outer header upon encapsulation and it copies the DSCP from the outer header to the inner header upon decapsulation. This assures that changes applied to the DSCP field survive encapsulation and decapsulation. In contrast, the pipe model ignores the content of the DSCP field in the outer header upon decapsulation. Therefore, decapsulation erases changes applied to the DSCP along the tunnel. As a consequence, only the uniform model may be used for tunneling PCN traffic within a PCN domain, if PCN encoding uses more than a single DSCP.

3.2.3. Restoration of Original DSCPs at the Egress Node

It is desirable that the egress node restores the original DSCP when PCN traffic leaves the PCN domain. This may be achieved through various options.

1. PCN-marking does not alter the original DSCP; it only uses the ECN field. Therefore the traffic leaves the PCN-domain with its original DSCP.

   However, if the PCN-marking may alter the DSCP, then some other technique is needed to restore the original DSCP:

2. Each Diffserv class using PCN uses a different set of DSCPs. Therefore, if there are M DSCPs using PCN and PCN encoding uses N different DSCPs, N*M DSCPs are needed. This solution may work well in IP networks. However, when PCN is applied to MPLS networks or other layers restricted to 8 QoS classes and codepoints, this solution fails due to the extreme shortage of available DSCPs.

3. The original DSCP for the packets of a flow is signaled to the egress node. No suitable signaling protocol has been developed and therefore, it is not clear whether this approach could work.

4. PCN-traffic is tunneled across the PCN-domain. The pipe tunneling model is applied and so the original DSCP is restored after decapsulation. However, tunneling across a PCN domain adds an additional IP header and reduces the maximum transfer unit (MTU) from the perspective of the user. GRE, MPLS, or Ethernet using Pseudo-Wires are potential solutions that scale well also in Backbone networks.

Based on the above discussion, it is concluded that option (1) is the most efficient solution used to maintain the original DSCP when the PCN traffic leaves the PCN domain. Therefore, option (1) is selected.

3.3. Constraints from the ECN Field

This section briefly reviews the structure and use of the ECN field, the constraints imposed by the redefinition of the ECN field and its impact on PCN deployment, as well as the constraints imposed by various tunneling rules on the persistence of PCN marks after decapsulation and its impact on possible re-marking actions.

3.3.1. Structure and Use of the ECN Field

Some transport protocols, like TCP, can typically use packet drops as an indication of congestion in the Internet. The idea of Explicit Congestion Notification (ECN) [RFC3168] is that routers provide a congestion indication for incipient congestion, where the notification can sometimes be through ECN marking (and re-marking) packets rather than dropping them. Figure 5 summarizes the ECN codepoints defined [RFC3168].

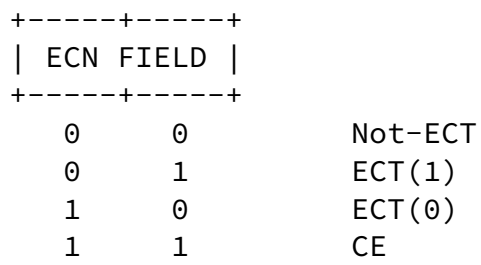


Figure 5: ECN Codepoints within the ECN field

ECT stands for "ECN-capable transport" and indicates that the sender and receivers of a flow understand ECN semantics. Packets of other flows are labeled with not-ECT. To indicate congestion to a receiver, routers may re-mark ECT(1) or ECT(0) labeled packets to CE which stands for "congestion experienced". Two different ECT codepoints were introduced "to protect against accidental or malicious concealment of marked packets from the TCP sender" which may be the case with cheating receivers [RFC3540].

3.3.2. Tunneling Rules for the Handling of the ECN Field

When packets are encapsulated, the ECN field of the inner header may or may not be copied to the ECN field of the outer header and upon decapsulation, the ECN field of the outer header may or may not be copied from the ECN field of the outer header to the ECN field of the inner header. Various tunneling rules with different treatment of the ECN field exist. Two different modes are defined in [RFC3168] for IP-in-IP tunnels and a third one in [RFC4301] for IP-in-IPsec

tunnels.

3.3.2.1. Limited Functionality Option

The limited-functionality option has been defined in [RFC3168]. Upon encapsulation, the ECN field of the outer header is generally set to not-ECT. Upon decapsulation, the ECN field of the inner header remains unchanged.

Since this tunneling mode loses information upon encapsulation and decapsulation, it cannot be used for tunneling PCN traffic within a PCN domain. However, the PCN ingress may use this mode to tunnel traffic with ECN semantics to the PCN egress to preserve the ECN field in the inner header while the ECN field of the outer header is used with PCN semantics within the PCN domain.

3.3.2.2. Full Functionality Option

The full-functionality option has been defined in [RFC3168]. Upon encapsulation, the ECN field of the inner header is copied to the outer header unless the ECN field of the inner header carries CE. In that case, the ECN field of the outer header is set to ECT(0). This choice has been made for security reasons, to disable the ECN fields of the outer header as a covert channel. Upon decapsulation, the ECN field of the inner header remains unchanged unless the ECN field of the outer header carries CE. In that case, the ECN field of the inner header is also set to CE.

This mode imposes the following constraints on PCN metering and marking.
First, PCN must re-mark the ECN field only to CE because any other information is not copied to the inner header upon decapsulation and will be lost.
Second, CE information in encapsulated packet headers is invisible for routers along a tunnel. Threshold marking does not require information about whether PCN packets have already been marked and would work when CE denotes that packets are marked. In contrast, excess-traffic-marking requires information about already excess-traffic-marked packets and cannot be supported with this tunneling mode.
Furthermore, this tunneling mode cannot be used when marked or not-marked packets should be preferentially dropped because the PCN marking information is possibly not visible in the outer header of a packet.

3.3.2.3. Tunneling with IPSec

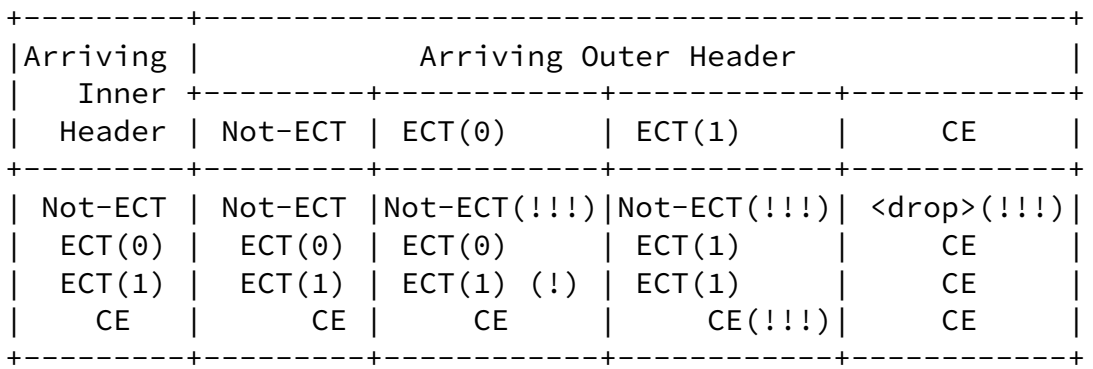
Tunneling has been defined in Section 5.1.2.1 of [RFC4301]. Upon encapsulation, the ECN field of the inner header is copied to the ECN field of the outer header. Decapsulation works as for the full-functionality option in Section 3.3.2.2. Tunneling with IPsec also requires that PCN re-marks the ECN field only to CE because any other information is not copied to the inner header upon decapsulation and lost. In contrast to Section 3.3.2.2, with IPsec tunnels, CE marks of tunneled PCN traffic remain visible for routers along the tunnel and to their meters, markers, and droppers.

3.3.2.4. ECN Tunneling

[RFC6040] imposes new tunneling rules for ECN and updates [RFC3168] and [RFC4301]. These rules provide a consistent and rational approach to encapsulation and decapsulation.

With the normal mode, the ECN field of the inner header is copied to the ECN field of the outer header on encapsulation (like the limited functionality option in Section 3.3.2.1). In compatibility mode, the ECN field of the outer header is reset to not-ECT.

Upon decapsulation, the scheme specified in [RFC6040] and shown in Figure 6 is applied. Thus, re-marking encapsulated not-ECT packets to any other codepoint would not survive decapsulation. Therefore, not-ECT cannot be used for PCN encoding. Furthermore, re-marking encapsulated ECT(0) packets to ECT(1) or CE survives decapsulation, but not vice-versa, and re-marking encapsulated ECT(1) packets to CE also survives decapsulation, but not vice-versa.



The ECN field in the outgoing header is set to the codepoint at the intersection of the appropriate arriving inner header (row) and arriving outer header (column), or the packet is dropped where indicated. Currently unused combinations are indicated by '(!!!)' or '(!)'

Figure 6: New IP in IP Decapsulation Behaviour (from [RFC6040])

3.3.3. Restoration of the Original ECN Field at the PCN-Egress-Node

As ECN is an end-to-end service, it is desirable that the egress node of a PCN domain restores the ECN field a PCN packet had at the ingress node. There are basically two options. PCN traffic may be tunneled between ingress and egress node using limited functionality tunnels (see Section 3.3.2.1). Then, PCN marking is applied only to the outer header, and the original ECN field is restored after decapsulation. However, this reduces the MTU from the perspective of the user. Another option is to use some intelligent encoding that preserves the ECN codepoints. However, a viable solution is not known.

3.3.4. Redefinition of the ECN Field

The ECN field may be redefined for other purposes and [RFC4774] gives guidelines for that. Essentially, not-ECT-marked packets must never be re-marked to ECT or CE because not-ECT-capable end systems do not reduce their transmission rate when receiving CE-marked packets. This is a threat to the stability of the Internet. Moreover, CE-marked packets must not be re-marked to not-ECT or ECT, because then ECN-capable end systems cannot reduce their transmission rate.

The re-use of the ECN field for PCN encoding has some impact on the deployment of PCN. First, routers within a PCN domain must not apply ECN re-marking when the ECN field has PCN semantics. Second, before a PCN packet leaves the PCN domain, the egress nodes must either (A) reset the ECN field of the packet to the contents it had when entering the PCN domain or (B) reset its ECN field to not-ECT. According to Section 3.3.3, tunneling ECN traffic through a PCN domain may help to implement (A). When (B) applies, CE-marked packets must never become PCN packets within a PCN domain as the egress node resets their ECN field to not-ECT. The ingress node may drop such traffic instead.

4. Comparison of Encoding Options

The PCN WG has studied four different PCN encodings, which redefine the ECN field as summarized in Figure 7. PCN semantics apply only to one or at most two specific DSCPs, and therefore ECN semantics do not apply to them. When a PCN-ingress-node classifies a packet as a PCN-packet it sets its PCN-codepoint to not-marked (NM). Non-PCN traffic can also to be sent with the PCN-specific DSCP, by setting the Not-PCN codepoint. Special per hop behaviour, defined in [RFC5670], applies to PCN-traffic.

ECN Bits	00	10	01	11	DSCP
RFC 3168	Not-ECT	ECT(0)	ECT(1)	CE	Any
Baseline	Not-PCN	NM	EXP	PM	PCN-n
3-In-1	Not-PCN	NM	ThM	ETM	PCN-n
3-In-2	Not-PCN	NM	CU	ThM	PCN-n
	Not-PCN	CU	CU	ETM	PCN-m
PSDM	Not-PCN	Not-ETM	Not-ThM	PM	PCN-n

Notes: PCN-n, PCN-m under the DSCP column denotes PCN Compatible DiffServ codepoints. Not-PCN means that packets are not PCN enabled. NM means Not-Marked to signal a not-pre-congested path. CU means Currently Unused.

Figure 7: Semantics of the ECN field for various encoding types

4.1. Baseline Encoding

With baseline encoding [RFC5696], the NM codepoint can be re-marked only to PCN-marked (PM). Excess-traffic-marking uses PM as ETM, threshold-marking uses PM as ThM, and only one of the two marking schemes can be used.

The 10-codepoint is reserved for experimental purposes (EXP) and the

other defined PCN encoding schemes can be seen as extensions of baseline encoding by appropriate redefinition of EXP. Baseline encoding [RFC5696] works well with IPsec tunnels (see Section 3.3.2.3).

4.2. Encoding with 1 DSCP Providing 3 States

PCN 3-state encoding extension in a single DSCP (3-in-1 encoding, [I-D.ietf-pcn-3-in-1-encoding] extends the baseline encoding and supports the simultaneous use of both excess-traffic-marking and threshold-marking. 3-in-1 encoding well supports the preferred CL-PCN and also SM-PCN.

The problem with 3-in-1 encoding is that the 10-codepoint does not survive decapsulation with the tunneling options in Section 3.3.2.1 - 3.3.2.3. Therefore, 3-in-1 encoding may be used only for PCN domains implementing the new rules for ECN tunneling [RFC6040], see Section 3.3.2.4), or where it is known that there are no tunnels in the PCN domain. Currently it is not clear how fast the new tunneling rules will be deployed, but the applicability of 3-in-1-encoding depends on that.

4.3. Encoding with 2 DSCPs Providing 3 or More States

PCN encoding using 2 DSCPs to provide 3 or more states (3-in-2 encoding, [I-D.ietf-pcn-3-state-encoding] uses two different DSCPs to accommodate the three required codepoints NM, ThM, and ETM. It leaves some codepoints currently unused (CU) and proposes also one way how to reuse them to store some information about the content of the ECN field before the packet entered the PCN domain. 3-in-2 encoding works well with IPsec tunnels (see Section 3.3.2.3). This type of encoding can support both CL-PCN and SM-PCN schemes.

The disadvantage of 3-in-2 encoding is that it consumes two DSCPs. Moreover, the direct application of this encoding scheme to other technologies like MPLS, where even fewer bits are available for the encoding of DSCPs is more difficult.

4.4. Encoding for Packet Specific Dual Marking (PSDM)

PCN encoding for packet-specific dual marking (PSDM) is designed to support PSDM-PCN outlined in Section 2.2.3. It is the only proposal that supports PCN-based AC and FT with only a single DSCP [I-D.ietf-pcn-psdm-encoding] in the presence of IPsec tunnels (see Section 3.3.2.3). PSDM encoding also supports SM-PCN.

4.5. Standardized encodings

The working group decided to take forward two encodings to RFC status: baseline and 3-in-1 (Sections 4.1, 4.2). The 3-in-1 encoding is on the

standards track; RFC 5696 only envisaged experimental use of the 10-codepoint and therefore is updated by the 3-in-1 encoding.

5. Conclusion

In this document, we have summarized various requirements for PCN encodings and the constraints imposed by the redefinition of the DS field for PCN encodings. We presented an overview of the currently supported PCN encodings and explained their pros and cons. As the accuracy of controlled load PCN (CL-PCN) is good enough and its complexity is acceptable, the redefinition of ECN tunneling rules and their deployment is desirable so that 3-in-1 encoding can support CL-PCN using only a single DSCP. Moreover, it also supports single marking PCN (SM-PCN) that is important for the deployment of PCN-based admission control and flow termination, as threshold-marking may not be supported by all vendors.

6. Security Implications

[RFC5559] provides a general description of the security considerations for PCN. This memo does not introduce additional security considerations.

7. IANA Considerations

This memo includes no request to IANA.

8. Acknowledgements

We would like to acknowledge the members of the PCN working group and Gorry Fairhust for the discussions that generated and improved the contents of this memo.

9. References

9.1. Normative References

9.2. Informative References

[I-D.ietf-pcn-cl-edge-behaviour]
Charny, A., Huang, F., Karagiannis, G., Menth, M., and T. Taylor, "PCN Boundary Node Behaviour for the Controlled

Load (CL) Mode of Operation",
draft-ietf-pcn-cl-edge-behaviour-08 (work in progress),
December 2010.

[I-D.ietf-pcn-sm-edge-behaviour]
Charny, A., Karagiannis, G., Menth, M., and T. Taylor,
"PCN Boundary Node Behaviour for the Single Marking (SM)
Mode of Operation", draft-ietf-pcn-sm-edge-behaviour-05
(work in progress), December 2010.

[I-D.ietf-pcn-3-in-1-encoding]
Briscoe, B. and T. Moncaster, "PCN 3-State Encoding
Extension in a single DSCP",
draft-ietf-pcn-3-in-1-encoding-04 (work in progress),
January 2011.

[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.

[I-D.ietf-pcn-psdm-encoding]
Menth, M., Babiarz, J., Moncaster, T., and B. Briscoe,
"PCN Encoding for Packet-Specific Dual Marking (PSDM
Encoding)", draft-ietf-pcn-psdm-encoding-01 (work in
progress), March 2010.

[I-D.babiarz-pcn-3sm]
Babiarz, J., Liu, X., Chan, K., and M. Menth, "Three State
PCN Marking", draft-babiarz-pcn-3sm-01 (work in progress),
November 2007.

[I-D.charny-pcn-single-marking]
Charny, A., Zhang, X., Faucheur, F., and V. Liatsos, "Pre-
Congestion Notification Using Single Marking for Admission
and Termination", draft-charny-pcn-single-marking-03
(work in progress), November 2007.

[I-D.westberg-pcn-load-control]
Westberg, L., Bhargava, A., Bader, A., Karagiannis, G.,
and H. Mekkes, "LC-PCN: The Load Control PCN Solution",
draft-westberg-pcn-load-control-05 (work in progress),
November 2008.

[I-D.briscoe-tsvwg-cl-phb]
Briscoe, B., "Pre-Congestion Notification marking", draft-briscoe-tsvwg-cl-phb-03 (work in progress), October 2006.

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

[RFC0793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, September 1981.

[RFC1633] Braden, B., Clark, D., and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.

[RFC2211] Wroclawski, J., "Specification of the Controlled-Load Network Element Service", RFC 2211, September 1997.

[RFC2309] Braden, B., Clark, D., Crowcroft, J., Davie, B., Deering, S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G., Partridge, C., Peterson, L., Ramakrishnan, K., Shenker, S., Wroclawski, J., and L. Zhang, "Recommendations on Queue Management and Congestion Avoidance in the Internet", RFC 2309, April 1998.

[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.

[RFC2597] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

[RFC2983] Black, D., "Differentiated Services and Tunnels",

RFC 2983, October 2000.

[RFC2998] Bernet, Y., Ford, P., Yavatkar, R., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J., and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks", RFC 2998, November 2000.

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

[RFC3246] Davie, B., Charny, A., Bennet, J., Benson, K., Le Boudec, J., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, March 2002.

[RFC3247] Charny, A., Bennet, J., Benson, K., Boudec, J., Chiu, A., Courtney, W., Davari, S., Firoiu, V., Kalmanek, C., and K. Ramakrishnan, "Supplemental Information for the New Definition of the EF PHB (Expedited Forwarding Per-Hop Behavior)", RFC 3247, March 2002.

[RFC3540] Spring, N., Wetherall, D., and D. Ely, "Robust Explicit Congestion Notification (ECN) Signaling with Nonces", RFC 3540, June 2003.

[RFC3955] Leinen, S., "Evaluation of Candidate Protocols for IP Flow Information Export (IPFIX)", RFC 3955, October 2004.

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

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

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

[RFC5129] Davie, B., Briscoe, B., and J. Tay, "Explicit Congestion Marking in MPLS", RFC 5129, January 2008.

[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.

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

[Menth08-Sub-8] Menth, M. and F. Lehrieder, "Applicability of PCN-Based Admission Control", https://atlas2.informatik.uni-tuebingen.de/menth/Menth08-Sub-8.pdf.

[Menth10q] Menth, M. and F. Lehrieder, "PCN-Based Measured Rate Termination", Computer Networks Journal, vol. 54, no. 3, Sept. 2010

.

[DClark] Clark, D., Shenker, S., and L. Zhang, "Supporting Real-Time Applications in an Integrated Services Packet Network: Architecture and Mechanisms", Proceedings of SIGCOMM '92 at Baltimore MD, August 1992.

[ITU-MLPP] "Multilevel Precedence and Pre-emption Service (MLPP)", ITU-T Recommendation I.255.3, 1990.

[Reid] Reid, A., "Economics and Scalability of QoS Solutions", BT Technology Journal Vol 23 No 2, April 2005.

Authors' Addresses

Georgios Karagiannis
University of Twente
P.O. Box 217
7500 AE Enschede,
The Netherlands

Email: g.karagiannis@ewi.utwente.nl

Kwok Ho Chan
Huawei Technologies
125 Nagog Park
Acton, MA 01720
USA

Email: khchan@huawei.com

Toby Moncaster
Moncaster Internet Consulting
Dukes
Layer Marney
Colchester
CO5 9UZ

Email: toby@moncaster.com

Michael Menth
Chair of Communication Networks
University of Tuebingen
Sand 13
72076 Tuebingen
Germany

Email: menth@informatik.uni-tuebingen.de

Philip Eardley
BT
B54/77, Sirius House Adastral Park Martlesham Heath
Ipswich, Suffolk IP5 3RE
United Kingdom

Email: philip.eardley@bt.com

Bob Briscoe
BT
B54/77, Sirius House Adastral Park Martlesham Heath
Ipswich, Suffolk IP5 3RE
United Kingdom

Email: bob.briscoe@bt.com