PCN
Internet-Draft
Intended status: Informational
Expires: April 22, 2011

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

October 22, 2010

Pre-Congestion Notification Encoding Comparison
draft-ietf-pcn-encoding-comparison-03

Abstract

Pre-congestion notification (PCN) is a link-specific and load-dependent packet marking mechanism for Differentiated Services networks. The packet markings are evaluated by egress nodes of PCN domains and the result is used for admission control and flow termination decisions. Two different types of markings have been defined. This document gives a summary of the marking requirements, the constraints to encode them in the current IP header (version 4 and above), and it explains why the PCN WG currently supports different encoding schemes for PCN marking.

http://www.ietf.org/shadow.html.

This Internet-Draft will expire on April 22, 2011.

Copyright Notice

Table of Contents

1. Introduction

Pre-congestion notification (PCN) is a link-specific and load-dependent packet marking mechanism for Differentiated Services networks which are then called PCN domains [RFC5559]. Links in such a network are configured with rate thresholds which are significantly lower than their bandwidth, and if the rate of specially marked PCN traffic exceeds such a threshold, PCN traffic is re-marked. These packet markings are used to support PCN-based admission control (AC) and flow termination (FT) in Differentiated Services networks in a simple way. High-priority traffic receives preferential treatment in a Differentiated Services network and AC limits the amount of high-priority traffic to avoid congestion on the links of the domain by explicitly admitting or blocking new flows that want to be carried by the high-priority per-hop behavior (PHB). In case of link or node failures, high-priority traffic may be rerouted so that links can be overloaded with admitted high-priority traffic. In such and other exceptional cases, FT can terminate some admitted flows to reduce the rate on affected links to a non-critical level.

A straightforward implementation of PCN-based AC and FT requires two rate thresholds per link: when an PCN-admissible-rate (AR) threshold is exceeded, new flows should be blocked and when a PCN-supportable-rate (SR) threshold is exceeded, some already admitted flows should be terminated (see Fig. 3 in [RFC5559]). This requires one metering and marking algorithms configured with the AR and another configured with the SR. The challenge is the encoding of the required PCN marks which requires at least three different codepoints for not-marked PCN traffic, PCN traffic re-marked by the marker configured with the AR, and PCN traffic re-marked by the marker configured with the SR. 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 AC and FT which have different pros and cons.

2. General PCN Encoding Requirements

Metering and marking algorithms, the way they are applied for PCN-based AC and FT, as well as packet drop strategies impose special 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 a reference rate which is 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 access to the information whether a PCN packet is already ETM marked or not. Threshold-marking re-marks all PCN traffic to ThM when the rate of PCN traffic exceeds the PCN-threshold-rate. Therefore, it does not need access to the exact marking information of a PCN packet.

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.

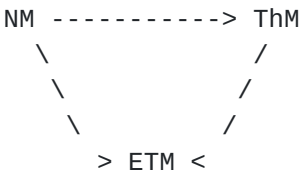


Figure 1: PCN Codepoint Re-Marking Diagram for Dual Marking (DM)

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 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 marked (ThM- and ETM-) PCN traffic is larger than a PCN-admission-threshold, 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 value chosen for the PCN-admission-threshold. 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 PCN-admission-threshold. Also FT works similarly to CL-PCN. The PCN-supportable-rate (SR) is only indirectly configured on any link by

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, and the amount of PCN traffic to be terminated is calculated by

termination-rate = PCN-ingress-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 threshold-marking is not available off the shelf. However, it works well only 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] [Menth08-Sub-9].

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). 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.

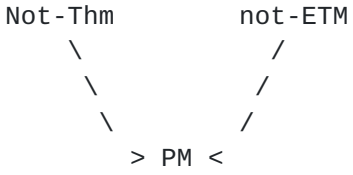


Figure 3: PCN Codepoint Re-Marking Diagram for Packet Specific Dual Marking (PSDM)

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 a packet probing mechanism. Only a single probe packet is needed for the admission decision of a new flow, and even that probe packet may

be part of an end-to-end signaling protocol so that no extra traffic is generated (implicit probing). When a flow requests admission, a probe packet is sent and the admission decision depends on whether the probe packet was re-marked or not. More specifically, the ingress node sets the ECN field of probe-packets to not-ThM and threshold marking configured with the PCN-admissible-rate possibly re-marks them to PM. In contrast, the ECN field of normal data packets is initially set to not-ETM and excess-traffic-marking configured with the PCN-supportable-rate possibly re-marks them to PM. For FT, the same algorithm may be used as for CL-PCN.

Disadvantages of this approach are that the implementation of the packet probing mechanism seems to be complex. Advantages are that the AC algorithm is more accurate than the one of CL-PCN and SM-PCN [Menth08-Sub-8] and that packet re-marking comes with fewer constraints. Section 4.4 will shows that this is an important feature.

2.2.4. Preferential Packet Dropping

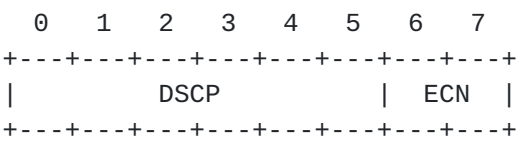
The termination algorithms proposed in the standards require preferential dropping of ETM-marked packets to avoid over-termination in case of packet loss [I-D.ietf-pcn-cl-edge-behaviour], [I-D.ietf-pcn-sm-edge-behaviour]. An analysis explaining this phenomenon can be found in section 4 of [Menth08-Sub-9]. Thus, preferential dropping of ETM-marked packets is also 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 and ECN fields. [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]).



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

Figure 4: The Struct of the DS Field

3.2. Constraints from the DSCP

The 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 DSCPs 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 differently marked PCN traffic.

3.2.2. Tunneling Rules

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 when PCN encoding uses more than a single DSCP.

3.2.3. Restoration of Original DSCPs at the Egress Node

PCN-based AC and FT may be requested for traffic with different DSCPs. Then, ingress nodes must mark the traffic of these flows as non-marked PCN traffic (DSCP nNM). Such traffic may be re-marked to threshold-marked or excess-traffic-marked PCN traffic (DSCP nThM, nETM) within the PCN domain. nNM, nThM, and nETM designate the DSCPs needed for encoding whereby they may all be the same DSCP n if the differentiation of the PCN codepoints is achieved in the ECN field. 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. A different PCN encoding ((mNM, mThM, mETM), (nNM, nThM, nETM), (oNM, oThM, oETM)) may be provided for different original DSCPs (i, j, k). The ingress re-marks incoming PCN traffic as not-marked, i.e., it maps DSCPs (i, j, k) to DSCPs (mNM, nNM, oNM). Within the PCN domain, PCN traffic may be re-marked. The egress node can restore the original DSCP by re-mapping ((mNM, mThM, mETM), (nNM, nThM, nETM), (oNM, oThM, oETM)) to (i, j, k). If PCN encoding uses N different DSCPs, this restoration technique requires N*M DSCPs where M is the number of original DSCPs that need to be differentiated. 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.

2. The original DSCP for the Packets of a flow may be signaled to the egress node. The egress node restores the original DSCP in the packets of this flow. This option assumes that all packets of a flow have the same DSCP and were not re-marked by previous network elements. A suitable signaling protocol is still missing and there are voice claiming that this solution cannot work for backbone networks.

3. PCN traffic may be tunneled from the ingress node to the egress node using the pipe model and PCN marking is applied only to the outer header. 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.

As option (3) impacts the user payload and option (2) is neither implemented nor able to preserve the real DSCP of the packets, option (1) is attractive for IP networks. However, it requires N*M different DSCPs for PCN encoding. To keep this option realistic,

only a single (N=1) DSCP should be used for PCN encoding.

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

TCP recognizes congestion in the Internet by experienced packet drops. The idea of Explicit Congestion Notification (ECN) [RFC3168] is that routers indicate incipient congestion to TCP receivers without dropping packets. To that end, the router re-marks packets appropriately. Figure 5 summarizes the codepoints defined in the ECN field.

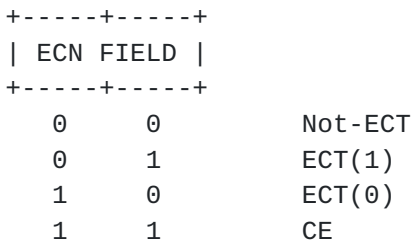


Figure 5: ECN Codepoints within the ECN field

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

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. As 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 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 this 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 as the PCN marking information is possibly not visible in the outer header of a packet.

3.3.2.3. Tunneling with IPSec

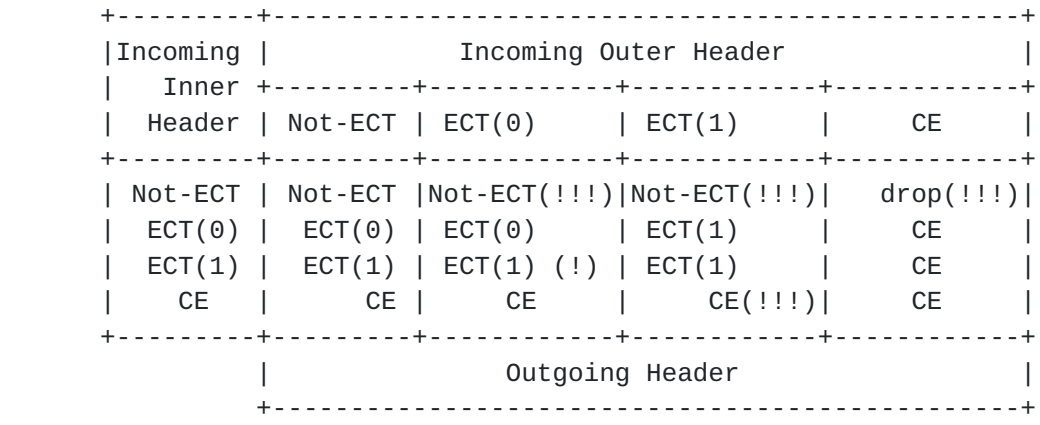
Tunneling has been defined in Sect. 5.1.2.1 of [RFC4301]. Upon encapsulation, the ECN field of the inner header is copied to the ECN field of the outer header. Decapsulation works as for the full-functionality option in Sect. 3.3.2. Tunneling with IPsec also requires that PCN re-marks the ECN field only to CE because any other

information is not copied to the inner header upon decapsulation and lost. In contrast to 3.3.2, 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 Option

[I-D.ietf-tsvwg-ecn-tunnel] proposes a new tunneling for ECN information that is intended to update the presented options. Upon encapsulation, with the compatibility mode, the ECN field of the outer header is reset to not-ECT while with the normal mode, the ECN field of the inner header is copied to the ECN field of the outer header (like the limited functionality option in 3.3.1).

Upon decapsulation, the scheme 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.



Currently unused combinations are indicated by '(!!!)' or '(!)'

Figure 6: New IP in IP Decapsulation Behaviour

3.3.3. Restoration of the Original ECN Field at the EgressNode

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 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 [RFC2474] 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 packet 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 produces four different PCN encodings which redefine the ECN field which are summarized in Figure 7. While ECN semantics apply to all DSCPs, PCN semantics apply only to one or at most two special DSCPs n and m which need further specification. These DSCPs imply a special PHB. All PCN encodings allow the simultaneous use of the taken DSCP n or m also for non-PCN traffic by marking such traffic with a Not-PCN codepoint. Generally, the ECN field of PCN Packets entering a PCN domain is set to not-marked (NM).

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 code points. CU means Currently Unused. NM means Not-Marked to represent Not Pre-Congested. Not-PCN means that packets are not PCN enabled.

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 3.3.3.3).

As baseline encoding supports only two PCN-codepoints, only a single metering and marking scheme can be used. It supports SM-PCN or only AC according to CL-PCN so that only threshold-marking is needed. As mentioned before, SM-PCN may be inaccurate and a missing FT function is also a severe disadvantage.

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] extents 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 3.3.1 - 3.3.3. Therefore, 3-in-1 encoding may be used only for PCN domains implementing the new rules for ECN tunneling (draft-..., see Section 3.3.3.4). 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 3.3.3). It well supports the preferred CL-PCN and also SM-PCN.

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 than 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 2.3. 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 3.3.3). PSDM encoding also supports SM-PCN.

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 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 SM-PCN which is important for the deployment of PCN-based AC and FT if threshold-marking is not offered by vendors.

6. Security Implications

Packets from normal precedence and higher precedence sessions [ITU-MLPP] aren't distinguishable by PCN Interior Nodes. This prevents an attacker specifically targeting, in the data plane, higher precedence packets (perhaps for DoS or for eavesdropping). However, PCN End Nodes can access this information to help decide whether to admit or terminate a flow. The separation of network information provided by the Interior Nodes and the precedence information at the PCN End Nodes allows simpler, easier and better focused security enforcement.

PCN End Nodes police packets to ensure a flow sticks within its agreed limit. This is similar to the existing IntServ behaviour. Between them the PCN End Nodes must fully encircle the PCN-Region, otherwise packets could enter the PCN-Region without being subject to admission control, which would potentially destroy the QoS of existing flows.

It is assumed that all the Interior Nodes and PCN End Nodes run PCN and trust each other (ie the PCN-enabled Internet Region is a controlled environment). For instance a non-PCN router wouldn't be able to alert that it's suffering pre-congestion, which potentially would lead to too many calls being admitted (or too few being terminated). Worse, a rogue router could perform attacks such as marking all packets so that no flows were admitted.

So security requirements are focussed at specific parts of the PCN-Region:

The PCN End Nodes become the trust points. The degree of trust required depends on the kinds of decisions it has to make and the kinds of information it needs to make them. For example when the PCN End Node needs to know the contents of the sessions for making the decisions, when the contents are highly classified, the security requirements for the PCN End Nodes involved will also need to be high.

PCN-marking by the Interior Nodes along the packet forwarding path needs to be trusted, because the PCN End Nodes rely on this information.

7. IANA Considerations

This memo includes no request to IANA.

8. Acknowledgements

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

Appendix A. Considerations for Selection of PCN Encoding

This document provides an historical account on the examination of the different ways to encode pre-congestion notification (PCN) [RFC5559] information in IP packets for transporting the information from the PCN ingress nodes, through the PCN interior nodes, to the PCN egress nodes. Documenting the examination results to indicate the reasoning behind the approach in selection of PCN encoding in IP packets. The appendix provides additional information and keeps an account of the reasoning and lessons learned from the consideration of the different encoding choices.

There are a number of criteria that affect the choice of encoding to be used. The key ones are:

1. The support of the required encoding states to satisfy the functional requirement of PCN. These required encoding states may need two, or three, or four encoding code points to represent.

2. Compliance with RFC 4774 [RFC4774] if the ECN field is to be re-used for PCN encoding.

3. Compliance with the requirements for specifying DSCPs and DSCP per-hop-behaviour groups [RFC2474].

4. Any PCN marking has to carry the '11' codepoint in the ECN field since this is the only codepoint that is guaranteed to be copied down into the tunneling inner header upon decapsulation. This criterion is related to the constraints that any PCN encoding needs to survive being tunnelled through either an IP in IP tunnel or an IPsec Tunnel.

5. Co-existence of PCN and not-PCN traffic: It is important to note that the scarcity of pool 1 DSCPs coupled with the fact that PCN is envisaged as a marking behaviour that could be applied to a number of different DSCPs makes it essential that we provide a not-PCN state. Because PCN re-defines the meaning of the ECN field for such DSCPs it is important to allow an operator to still use the DSCP for traffic that isn't PCN-enabled. This is achieved by providing a Not-PCN state within the encoding scheme.

Appendix A.1. Encoding Options

There are couple of methods to carry the PCN marks. The method used affects the possible encoding options. Hence when we describe the different encoding options in this appendix, we group them based on how the encoding states are carried.

The encoding transport methods considered are:

1. using the combination of the ECN and DSCP bits of a data packet header

2. using only the DSCP bits of a data packet header

3. using only the ECN bits of a data packet header

We discuss the encoding options for each of the encoding transport methods separately in their own subsections.

The main required encoding states for PCN capable packets are listed below:

- Not-Marked (NM), for indication of No Pre-Congestion Indication.

- Admission Marked (AM), for indication of Flow Admission Information.

- Termination Marked (TM), for indication of Flow Termination Information.

- Affected Marked (AfM), for indication of ECMP Information.

- Not-PCN, for indication of packets that are not PCN-enabled.

A total of five main required encoding states for PCN capable packets.

Appendix B. Encoding Using ECN and DSCP Fields

The use of both DSCP and ECN fields is following an approach that allows a clean traffic treatment separation of PCN Capable traffic and Non PCN Capable traffic. This natural use of the DSCP field, to provide treatment differentiation of packets using different DSCP encoding, is one way of providing the "PCN Capable Packet" encoding state. The using of this approach allows us to focus on encoding the four required PCN Encoding States using the two ECN bits.

ECN Bits	00	10	01	11	DSCP
RFC 3168	Not-ECT	ECT(0)	ECT(1)	CE	NA
Option 1	AM	NM	NM	TM	PCN-1
Option 2	AfM	NM	NM	AM/TM	PCN-1
Option 3	NM	NA	NA	AM/TM	PCN-1
Option 4	Not-PCN	NM	EXP	AM/TM	PCN-1
Option 5	NM	NA	NA	AM	PCN-1
				TM	PCN-2
Option 6	AM	NM	TM	NA	PCN-1

Notes: NA means Not Applicable. PCN-1, PCN-2 under the DSCP column denotes specific DSCPs used to indicate PCN capable packets. AM/TM means the two encoding states are sharing the same encoding bit pattern. NM means Not-Marked to represent Not Pre-Congested. Not-PCN means that packets are not PCN enabled.

Figure 8: Encoding of PCN Information Using DSCP and ECN Fields

In Figure 8, we listed the fundamental options when both DSCP and ECN fields are used. In Option 4 the ECN codepoints '01' and '10' could both be used for NM encoding or one of them could be used for NM encoding and the other for experimental encoding. There are couple of variations of the theme provided by these options. One way of comparing these options is by examining the pros and cons of the different ways the four code points provided by the two ECN bits are used. We group these discussions in the following way:

1. The '01' and '10' code points.

2. The '11' code point.

3. The '00' code point.

We discuss each of them in the following sub-sections.

Appendix B.1. The Use of '01' and '10' Encoding for PCN

There can be different degrees of usage of the '01' and '10' code points by PCN:

1. PCN Does NOT use the '01' and '10' code points, see Option 3 and Option 5. This will be the safest choice. But this choice will leave us with only two usable code points, unless we want to deploy more than one PCN DSCPs. Even when the PCN domain does not use these code points, the PCN domain still have to handle the receiving of '01' and '10' packets at ingress. The notion of safe comes in two flavors, first if there is any packets in the PCN domain having the '01' or '10' encoding, it is immediately known that these are packets in error, either they are leaked into the PCN domain in error or are set to '01' or '10' in error inside the PCN domain. In both cases, action can be taken. The second flavor of safe is if a legitimate PCN packet leaks out of the PCN domain, it will not have the '01' or '10' encoding and should not cause an ECN router to mistaken the PCN packets to be ECN packets.

2. PCN uses '01' and '10' code points in an ECN friendly manner, see Options 1, 2, 4, and 6. One ECN friendly manner is to have both '01' and '10' to mean "PCN Capable Packet". The determination of ECN friendliness depends on the use of code points beside '01' and '10'. Furthermore, the use of '01' and '10' codepoints allow the transport of the Not-PCN encoding, see Option 4.

Appendix B.2. The Use of '11' Encoding for PCN

Not using the '11' code point for PCN will be a safe choice from the ECN semantic point of view, see Option 6. However, this will reduce the possible numger of encoding codepoints to three. And the '11' code point is the only code point that survive tunneling. The encoding codepoint '11' is used in Options 1, 2, 3, 4, 5.

Appendix B.3. The Use of '00' Encoding for PCN

The '00' codepoint are used by ECN to indicate Not ECN enabled. A safe use of '00' codepoint by PCN will be to indicat Not PCN enabled, as in Option 4. The other usage may have problem in some of the environments.

Appendix B.4. Benefits of Using DSCP and ECN Fields

A major feature of using both DSCP and ECN fields is the ability to use the inherent nature of DiffServ for traffic class separation to allow PCN treatment be applied to PCN traffic, without concerns of applying PCN treatment to none PCN traffic and vise versa. This feature frees this approach for PCN encoding from some of the concerns raised by RFC 4774 [RFC4774]. This feature will also keep none PCN Capable traffic out of the PCN treatment mechanisms, allowing the PCN treatment mechanisms focus on their respective PCN

tasks.

This approach also leaves the ECN field available totally for PCN encoding states purposes. Removing the need to carry the Not-PCN Encoding in the ECN field.

Appendix B.5. Drawbacks of Using DSCP and ECN Fields

The use of both DSCP and ECN fields will require the setting aside of one (or possibly two) DSCP for use by PCN. This may add complexity to the PCN encoding standardization effort.

Appendix B.6. Comparing DSCP and ECN Fields Encoding Options

Here we discuss the differences between the different encoding options when both DSCP and ECN fields are used. There are many encoding options, we have provided the ones we think are favorable in Figure 8.

When DSCP is used to differentiate between PCN capable and Not-PCN capable traffic, the encoding of "Not-PCN" in the ECN field is not required. This is the motivation for Option 1 in Figure 8, where the encoding "00" for "Not-ECT" is being used for "AM" (Admission Marking) encoding state. The encodings "01" and "10" for "ECT(1)" and "ECT(0)" supports the required encoding states for "Not Pre-Congested Marking" (PCN), and reserving them for any "Nonce Marking" if necessary. With the possible additional encoding of "PCN(A)" and "PCN(T)" in place of "ECT(1)" and "ECT(0)" for indicating percentage of Admission Marked traffic and percentage of Termination Marked traffic when the algorithm benefits from such additional information.

Option 2 in Figure 8 uses the "00" encoding for "AfM". With '01' and '10' encoding the same as for Option 1, requiring the use of "11" encoding for both "AM" (Admission Mark) and "TM" (Termination Mark) states or requiring the allocation of a DSCP for encoding the "TM" state.

Option 4 is the only option that can fulfill both criteria 4 and 5, listed in Appendix A.

Appendix B.7. Concerns on Alternate Semantics for the ECN Field

Section 2 of [RFC4774] raised couple of concerns for usage of alternate semantics for the ECN field. We try to address each of the concerns in this section.

1. Section 3.1 of [RFC4774] discusses Concern 1: "How routers know which ECN semantics to use with which packets." This use of DSCP

and ECN for encoding PCN states address this by following the recommendation of [RFC4774] on using a diffserv codepoint to identify the packets using the alternate ECN semantics. This diffserv codepoint may possibly be a new diffserv codepoint to minimize the possible confusion between using the old per hop behavior of the codepoint and the using of the alternate ECN semantics per hop behavior of the codepoint.

2. Section 4 of [RFC4774] discusses Concern 2: "How does the possible presence of old routers affect the performance of the alternate ECN connections." With the notion of old routers meaning routers that performs RFC 3168 ECN processing instead of PCN processing. An answer to this question is given by assuming that the environment using the alternate ECN semantics is envisioned to be within a single administrative domain, and it has the ability to ensure that all routers along the path understand and agree to the use of the alternate ECN semantics for the traffic identified by the use of a diffserv codepoint. This uses option 2 indicated in section 4.2 of [RFC4774]. But incase there is a mis-configuration, the choice of encoding may make a difference:

   * With encoding Option 1, the old routers will interprete:

      + '00' encoding as Not-ECT, and will drop AM marked packets. The PCN edge nodes should not admit traffic that it does not receive, hence the PCN admission functionality should be OK.

      + '01' encoding as ECT(1), which indicates ECN capable and can be remarked to '11' to indicate congestion experienced. The RFC 3168 ECN CE encoding have the same functionality as the PCN TM encoding, to reduce the offered traffic load. Hence the PCN termination functionality should be OK.

      + '10' encoding as ECT(0). The discussion for '01' above applies equally to this encoding.

      + '11' encoding as CE. The old router should use this encoding to reduce the offered traffic load and should not remark this to any other ECN encoding, the same functionality the PCN TM encoding requires, hence should be OK for PCN.

      The above discussion for Option 1 applies equally for PCN traffic leaked out of the PCN domain and interpreted by RFC 3168 ECN nodes.

* With encoding Option 2, the old routers will interpret:

  + '00' encoding as Not-ECT, and will drop AfM marked packets. This may possibly affect the efficiency of the Affected Marking functionality.

  + '01' encoding as ECT(1), which indicates ECN capable and can be remarked to '11' to indicate congestion experienced. The RFC 3168 ECN CE encoding have the same functionality as the PCN TM encoding, to reduce the offered traffic load. Depending on the PCN algorithm on how AM and TM share the same '11' encoding, this may or may not affect the functionality of PCN.

  + '10' encoding as ECT(0). The discussion for '01' above applies equally to this encoding.

  + '11' encoding as CE. The old router should use this encoding to reduce the offered traffic load and should not remark this to any other ECN encoding. Depending on the PCN algorithm on how AM and TM share the same '11' encoding, this may or may not affect the functionality of PCN.

  The above discussion for Option 2 applies equally for PCN traffic leaked out of the PCN domain and interpreted by RFC 3168 ECN nodes.

3. Concern 3: "How does the possible presence of old routers affect the coexistence of the alternate ECN traffic with competing traffic on the path." Within the PCN domain, the PCN (alternate ECN) traffic is separated from the other traffic using diffserv. If by mis-configuration, an old routers that does not understand PCN handles PCN traffic, the PCN traffic will get the per hop behavior as the other traffic, hence not receiving the benefits of PCN at the old router, but will not affect the coexistence of the PCN and the other traffic. If the old router uses RFC 3168 ECN congestion treatment, then the discussion for Concern 2 above applies.

4. Concern 4: "How well does the alternate ECN traffic perform." The performance of the different proposed PCN (alternate ECN) metering and marking algorithms are currently under study with their simulation and study results described by their respective documents.

The environment using the alternate ECN semantics is envisioned to be within a single administrative domain. With the ability to ensure

that all routers along the path understand and agree to the use of the alternate ECN semantics for the traffic identified by the use of a Diffserv codepoint. This uses option 2 indicated in section 4.2 of [RFC4774].

Appendix C. Encoding Using DSCP Field

In this type of encoding and transport method the congestion and precongestion information is encoded into the 6 DSCP bits that are transported in the IP header of the data packets. Four possible alternatives can be distinguished, as can be seen in Figure 9, with details provided by [I-D.westberg-pcn-load-control]. Option 7 needs 2 additional DSCP values, Options 8 and 9 need three additional DSCP values and Option 10 needs four additional DSCP values. Note that all additional and experimental DSCP values are representing and are associated with the same PHB. The 1st, 2nd, 3rd, and 4th DSCP values are representing DSCP values that are assigned by IANA as DSCP experimental values, see [RFC2211]. Furthermore, all options listed in Figure 2 are able to support the Not-PCN encoding state.

DSCP Bits	Original	Add DSCP 1	Add DSCP 2	Add DSCP 3	Add DSCP 4
Option 7	Not-PCN	UM	AM/TM	NA	NA
Option 8	Not-PCN	UM	AM/TM	AfM	NA
Option 9	Not-PCN	UM	AM	TM	NA
Option 10	Not-PCN	UM	AM	TM	AfM

Notes: Not-PCN means the packet is not PCN capable. UM for Un-Marked meaning Not Pre-Congested

Figure 9: Encoding of PCN Information Using DSCP Field

Appendix C.1. Benefits of Using DSCP Field

The main benefits of using the DSCP field for PCN encoding are:

- it is not affecting the end-to-end ECN semantics and therefore the issues and concerns raised in [RFC4774] are not applicable for this encoding scheme.

- it is not affected by the PCN tunneling issues.

- all 4 DSCP encoding options depicted in Figure 9 can support the PCN capable not congested/UnMarked (UM) indication, the admission control (AM) and flow termination (TM) encoding states.

- the experimental DSCPs are lightly standardized and therefore, the rules on how to apply and use them are limited. This provides a high flexibility to network operators to apply and use them in different settings.

- simple packet classification, since a router needs only to read the DSCP field, instead of reading both DSCP and ECN fields.

- Option 8 and 10 support the Affected Marking (AfM) encoding, which according to [I-D.westberg-pcn-load-control], it has benefits if the PCN-domain operates ECMP routing and is not using DSCP for route selection.

- by using an additional DSCP to encode the not congested PCN state, all PCN-ingress-nodes can be configured to encode this state into all packets that are entering the PCN domain and are PCN aware. This will solve any PCN-egress-node misconfiguration problems, which can allow a AM/TM or SM encoded packet to exit a PCN-domain.

Appendix C.2. Drawbacks of Using DSCP Field

The main drawbacks of using the DSCP field for PCN encoding are the following:

this type of encoding needs to use per PHB, in addition to the original DSCP and depending on the encoding option used, one, two, three, or four DSCP values, respectively. These additional DSCP values can be taken from the DSCP values that are not defined by standards action, see [RFC2211]. Note that all the additional DSCP values are representing and are associated with one PHB. The value of this DSCP/PHB can either follow a standards action or use a value that is applied for experimental or local use. It is important to note that the number of the DSCP values used for local or experimental use is restricted and therefore the number of different PHBs supported in the PCN domain will also be restricted.

applying the DSCP field as PCN encoding transport within an PCN aware MPLS domain, see [RFC5129], can be problematic due to the scarce packet header real-estate.

when the PCN-domain is operating ECMP that uses DSCP to select the routes, a risk of mis-ordering of packets within a flow might occur. The impact of this drawback depends on the following:

1. the level of deployment of ECMP algorithms that use DSCP for route selection;

2. mis-ordering of packets within a flow when there is termination marking may be acceptable;

3. the possibility of configuring the ECMP algorithms that use DSCP for route selection in the PCN-domain that the used PCN aware DSCPs are belonging to the same PHB and therefore, all these DSCP values should be converted to one preconfigured DSCP value before applying it in the ECMP routing algorithm. Note that all the additional experimental DSCPs that are used within PCN are belonging to the same PHB.

Appendix D. Encoding Using ECN Field

This section takes the approach 3 option indicated in Appendix A.1. Which the DSCP field only indicates the packet forwarding behavior, for which both PCN Capable and Non PCN Capable traffic use/share the same DSCP. This approach requires the use of the Not PCN Capable Encoding State to be encoding using the ECN bits. Hence this section describes the encoding options that uses only the ECN field (without the DSCP field) available in the IP header of the data packets to encode the PCN states.

The use of the same DSCP for both PCN Capable and Non PCN Capable also opens the question of having PCN and RFC 3168 ECN traffic using the same DSCP. Which increases the importance of satisfying the concerns indicated in RFC 4774.

ECN Bits	00	01	10	11	DSCP
RFC 3168	Not-ECT	ECT(1)	ECT(0)	CE	NA
Option 11	Not-PCN	AM	PCN	TM	NA
Option 12	Not-PCN	PCN	PCN	AM/TM	NA
Option 13	Not-PCN	AfM	PCN	AM/TM	NA

Figure 10: Encoding of PCN Information Using ECN Field

In Figure 10, we listed the fundamental options when only the ECN field is used. Like in Figure 8, there are variations of the theme provided by these options. For example, when both "01" and "10" encoding are used for NPM in Option 5, they can be interpreted as PCN(A) and PCN(T) instead of just PCN. Using the PCN(A) and PCN(T) variation provides the additional information of the ratio of packets AM marked to packets Not AM marked, and the ratio of packets TM marked to packets Not TM marked. Having these ratios being independent from one another.

For Option 11, the use of '01' for AM and '10' for PCN can be swapped and provide the same functionality. For Option 13, the use of '01' for AfM and '10' for PCN can also be swapped without change of functionality.

Appendix D.1. Benefits of Using ECN Field

The using of only the ECN field for encoding PCN encoding states allow more efficient use of the DSCP field, not requiring the allocation of PCN specific DSCP values.

This approach also opens the question of possibly having both PCN and ECN traffic using the same DSCP.

When the same treatment can be provided to both ECN and PCN traffic to achieve each of ECN and PCN purpose, then not having DiffServ as separation between ECN and PCN traffic may be a benefit. Under such circumstances, having the same encoding between ECN and PCN may be desireable. But this can only be true if the requirement set forth in [RFC4774] for alternate ECN semantics can be satisfied.

If the same treatment can be applied to both ECN and PCN traffic, then:

- The first issue of [RFC4774]: "How routers know which ECN semantics to use with which packets." may be solved because there are no difference in the treatments of ECN and PCN packets, hence they can use the same semanics.

- The second and third issues of [RFC4774]: "How does the possible presence of old routers affect the performance of the alternate ECN connections." and "How does the possible presence of old routers affect the coexistence of the alternate ECN traffic with competing traffic on the path." are also solved because there are no difference in the treatment of ECN and PCN packets.

- The forth issue of [RFC4774]: "How well does the alternate ECN traffic perform." are dependent on the algorithm used, and should be provided by the respective algorithm document, and not in the scope of this document.

Appendix D.2. Drawbacks of Using ECN Field

Notice this group of encoding options does not use DiffServ code points for PCN encoding. With this group of encoding options, the required states of "PCN Capable Transport"/"None PCN Capable Transport" must be encoded using the ECN field. Leaving less encoding real estate to carry the remaining required PCN encoding states. Another drawback is without the protection/separation capability provided by DiffServ, it is typically harder to satisfy the requirement set forth in [RFC4774] for alternate ECN semantics.

Appendix D.3. Concerns on Alternate Semantics for the ECN Field

Section 2 of [RFC4774] raised couple of concerns for usage of alternate semantics for the ECN field. We try to address each of the concerns in this section.

1. Section 3.1 of [RFC4774] discusses Concern 1: "How routers know which ECN semantics to use with which packets." When this group of PCN encodings are used without the use of DSCP, routers can not distinguished PCN encoded packets from RFC 3168 ECN encoded packets. Hence there needs to be some kind of differentiation between PCN and RFC 3168 ECN packets, may be using PCN for real-time traffic types (with specific DSCP) and ECN for elastic traffic (with specific DSCP). And only distinguishing PCN Capable and Non-PCN Capable packets in real-time traffic. Only distinguishing ECT and Not-ECT packets in elastic traffic. But not having PCN and ECN traffic together.

2. Section 4 of [RFC4774] discusses Concern 2: "How does the possible presence of old routers affect the performance of the alternate ECN connections." With the notion of old routers meaning routers that performs RFC 3168 ECN processing instead of PCN processing, or drop packets instead of encoding the congestion information. The easy answer is the environment using the alternate ECN semantics is envisioned to be within a single administrative domain. With the ability to ensure that all routers along the path understand and agree to the use of the alternate ECN semantics for the traffic identified to be PCN Capable. This uses option 2 indicated in section 4.2 of [RFC4774]. But incase there is mis-configuration, the choice of encoding may make a difference:

* With encoding Option 11, the old routers will interpret:

  + '00' encoding as Not-ECT, and will drop Not-PCN marked packets when congestion is detected. With '00' the encoding for Not-PCN, requiring the same functionality as Not-ECT, the presence of old routers will not affect the performance of PCN functionality.

  + '01' encoding as ECT(1), which indicates ECN capable and can be remarked to '11' to indicate congestion experienced. For Option 11, the old router can possibly remark AM to TM. This puts a burden on the metering and marking algorithms to treat TM encoded packets to indicate stop admission. This may or may not be acceptable, depending on the algorithm.

  + '10' encoding as ECT(0), which indicates ECN capable and can be remarked to '11' to indicate congestion experienced. The RFC 3168 ECN CE encoding have the same functionality as the PCN TM encoding, to reduce the offered traffic load. Hence the PCN termination functionality should be OK.

  + '11' encoding as CE. The old router should use this encoding to reduce the offered traffic load and should not remark this to any other ECN encoding, the same functionality the PCN TM encoding requires, hence should be OK for PCN.

  The above discussion for Option 11 applies equally for PCN traffic leaked out of the PCN domain and interpreted by RFC 3168 ECN nodes.

* With encoding Option 12, the old routers will interprete:

  + '00' encoding as Not-ECT, and will drop Not-PCN marked packets when congestion is detected. With '00' the encoding for Not-PCN, requiring the same functionality as Not-ECT, the presence of old routers will not affect the performance of PCN functionality.

  + '01' encoding as ECT(1), which indicates ECN capable and can be remarked to '11' to indicate congestion experienced. The RFC 3168 ECN CE encoding have the same functionality as the PCN TM encoding, to reduce the offered traffic load. Depending on the PCN algorithm on how AM and TM share the same '11' encoding, this may or may not affect the functionality of PCN.

+ '10' encoding as ECT(0). The discussion for '01' above applies equally to this encoding.

+ '11' encoding as CE. The old router should use this encoding to reduce the offered traffic load and should not remark this to any other ECN encoding. Depending on the PCN algorithm on how AM and TM share the same '11' encoding, this may or may not affect the functionality of PCN.

The above discussion for Option 12 applies equally for PCN traffic leaked out of the PCN domain and interpreted by RFC 3168 ECN nodes.

* With encoding Option 13, the old routers will interprete:

+ '00' encoding as Not-ECT, and will drop Not-PCN marked packets when congestion is detected. With '00' the encoding for Not-PCN, requiring the same functionality as Not-ECT, the presence of old routers will not affect the performance of PCN functionality.

+ '01' encoding as ECT(1), which indicates ECN capable and can be remarked to '11' to indicate congestion experienced. For Option 13, the old router can possibly remark AfM to TM. This may or may not be acceptable, depending on the algorithm's Affected Marking functionality.

+ '10' encoding as ECT(1), which indicates ECN capable and can be remarked to '11' to indicate congestion experienced. The RFC 3168 ECN CE encoding have the same functionality as the PCN TM encoding, to reduce the offered traffic load. Depending on the PCN algorithm on how AM and TM share the same '11' encoding, this may or may not affect the functionality of PCN.

+ '11' encoding as CE. The old router should use this encoding to reduce the offered traffic load and should not remark this to any other ECN encoding. Depending on the PCN algorithm on how AM and TM share the same '11' encoding, this may or may not affect the functionality of PCN.

The above discussion for Option 13 applies equally for PCN traffic leaked out of the PCN domain and interpreted by RFC 3168 ECN nodes.

3. Concern 3: "How does the possible presence of old routers affect the coexistence of the alternate ECN traffic with competing traffic on the path." If RFC 3168 ECN and PCN traffic are to be treated within a single DiffServ PHB, because with these encoding there is no way to differentiate between the ECN packets from the PCN traffic, the metering and marking algorithm used must be totally friendly between ECN and PCN traffic, else they will affect each other in possibly non-acceptable ways. These encoding will work OK with traffic besides ECN because of the use of 'Not-PCN' encoding.

4. Concern 4: "How well does the alternate ECN traffic perform." The performance of the different proposed PCN (alternate ECN) metering and marking algorithms are currently under study with their simulation and study results described by their respective documents.

Appendix E. Encoding Choice Considerations

With the discussions and constraints indicated by this document, the use of both the DSCP [RFC2474] and the ECN [RFC3168] fields are necessary to fullfill the requirement of encoding and transporting the PCN marks.

9. 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-07 (work in progress), September 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-04 (work in progress), September 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-02 (work in progress), March 2010.

[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.ietf-tsvwg-ecn-tunnel]
Briscoe, B., "Tunnelling of Explicit Congestion Notification", draft-ietf-tsvwg-ecn-tunnel-10 (work in progress), August 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.

[I-D.ietf-tsvwg-admitted-realtime-dscp]
Baker, F., Polk, J., and M. Dolly, "DSCP for Capacity-Admitted Traffic", draft-ietf-tsvwg-admitted-realtime-dscp-07 (work in progress), March 2010.

[I-D.ietf-tsvwg-mlef-concerns]
Baker, F. and J. Polk, "MLEF Without Capacity Admission Does Not Satisfy MLPP Requirements", draft-ietf-tsvwg-mlef-concerns-00 (work in progress),

February 2005.

[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", http://www3.informatik.uni-wuerzburg.de/staff/menth/Publications/papers/Menth08-Sub-8.pdf.

[Menth08-Sub-9]
Menth, M. and F. Lehrieder, "PCN-Based Measured Rate Termination", Accepted for Computer Networks Journal, Elsevier preprint available at http://www3.informatik.uni-wuerzburg.de/staff/menth/Publications/papers/Menth08-Sub-9.pdf, February 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

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

Email: khchan@huawei.com

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

Email: g.karagiannis@ewi.utwente.nl

Toby Moncaster
BT Research
B54/70, Sirius House Adastral Park Martlesham Heath
Ipswich, Suffolk IP5 3RE
United Kingdom

Email: toby.moncaster@bt.com

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

Email: menth@informatik.uni-wuerzburg.de

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

Email: philip.eardley@bt.com

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

Email: bob.briscoe@bt.com