Internet Engineering Task Force (IETF)
Request for Comments: 6627
Category: Informational
ISSN: 2070-1721

G. Karagiannis
University of Twente
K. Chan
Consultant
T. Moncaster
University of Cambridge
M. Menth
University of Tuebingen
P. Eardley
B. Briscoe
BT

July 2012

Overview of Pre-Congestion Notification Encoding

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 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 that 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 those approaches along with an explanation of the constraints that apply to any solution.

Status of This Memo

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 (AC), to decide whether to admit or block a new flow request, and flow termination (FT), 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 the decision points that take decisions about the flow admission and termination on this basis [RFC6661] [RFC6662].

This PCN information has to be encoded into the IP header. This requires at least three different codepoints: one for PCN-traffic that has not been marked, one for traffic that has been marked by the threshold meter, and one for traffic that has been marked by the excess-traffic-meter.

Since unused codepoints are not available for that purpose in the IP header (versions 4 and 6), already used codepoints must be reused, which imposes additional constraints on the design and applicability of PCN-based AC and FT. This document summarizes these issues as a record of the PCN working group discussions and for the benefit of the wider IETF community.

In Section 2, we briefly point out the PCN encoding requirement imposed by metering and marking algorithms, and by special packet drop strategies. The Differentiated Services field (6 bits -- see [RFC3260] updating [RFC2474] in this respect) and the Explicit Congestion Notification (ECN) field (2 bits) [RFC3168] have been selected to be reused 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 considered 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

The choice of metering and marking algorithms and the way they are applied to PCN-based AC and FT 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 that are termed PCN-excess-rate and PCN-threshold-rate, respectively. When traffic for PCN-flows enters a PCN-domain, the PCN-ingress-node sets a codepoint in the IP header indicating that the packet is subject to PCN-metering and PCN-marking and that it is not-marked (NM). The two metering and marking algorithms possibly re-mark PCN-packets as excess-traffic-marked (ETM) or threshold-marked (ThM).

Excess-traffic-marking ETM-marks all not-ETM-marked PCN-traffic that is in excess of the PCN-excess-rate. To that end, the algorithm needs to know whether a PCN-packet has already been marked with ETM 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 such knowledge is needed 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 rates are configured with the PCN-admissible-rate (AR) and the PCN-supportable-rate (SR), respectively. Threshold-marking meters all PCN-traffic, but re-marks only NM-traffic to ThM. Excess-traffic-marking meters only NM- and ThM-traffic and re-marks it 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.

```
          NM -----------> ThM
            \              /
             \            /
              \          /
                > ETM <
```

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 [RFC6661]. 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 defined threshold, called 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-sent-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.

```
NM --------> ETM
```

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 [RFC6662]. 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. Then it requests the PCN-sent-rate from the corresponding PCN-ingress-node and 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 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 [Menth12] [Menth10].

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 (AR) and the PCN-supportable-rate (SR), 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 are used for PCN-traffic that is not yet re-marked (like NM with single and dual marking), and their specific use is determined 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.

```
          not-ThM       not-ETM
              \             /
               \           /
                \         /
                  > PM <
```

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

An edge behavior for PSDM has been presented in [Menth09] and [PCN-MS-AC]. We call it PSDM-PCN. In contrast to CL-PCN and SM-PCN, AC is realized by reusing initial signaling messages for probing purposes. 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 message traverses the PCN-domain and is re-marked, then the corresponding admission request is blocked. This is a lightweight probing mechanism that does not generate extra traffic and does not introduce probing delay. 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.

PSDM has three major disadvantages. First, signalling traffic needs to be marked with a PCN-enabled DSCP so that it either shares the same queue as data traffic, which may not be desired by some operators, or multiple PCN-enabled DSCPs are needed, which is not a pragmatic solution. Second, reservations for PCN-flows need to be triggered by a path-coupled end-to-end signalling protocol, which restricts the choice of the signalling protocol. And third, the selected signalling protocols must be adapted to take advantage of PCN-marked signalling messages for admission decisions, which incurs some extra effort before PSDM can be used.

The advantages are that the AC algorithm is more accurate than the one of CL-PCN and SM-PCN [Menth12], that often only a single DSCP is needed, and that the new tunneling rules in [RFC6040] are not needed for deployment (Section 3.3.3).

2.2.4. Preferential Packet Dropping

The termination algorithms described in [RFC6661] and [RFC6662] require the preferential dropping of ETM-marked packets to avoid over-termination in the case of packet loss. An analysis explaining this phenomenon can be found in Section 4 of [Menth10].

Thus, [RFC5670] recommends that ETM-marked packets "SHOULD be preferentially dropped". As a consequence, droppers must have access to the exact marking information of PCN-packets.

3. Encoding Constraints

The PCN working group decided to use a combination of the 6-bit Differentiated Services (DS) field and the ECN field for the encoding of the PCN-marks (see [RFC6660]). This section describes the criteria that are used to compare the resulting encoding options described in Section 4.

3.1. Structure of the DS Field

Figure 4 shows the structure of the DS and ECN fields. [RFC0793] defined the 8-bit TOS octet and [RFC2474] redefined it as the DS field, including the two least significant bits as currently unused (CU). [RFC3168] assigned the two CU bits to ECN and [RFC3260] redefined the DS field as only the most significant 6-bits of the (former) IPv4 TOS octet, thus separating the two-bit ECN field from the DS field.

```
  0   1   2   3   4   5   6   7
+---+---+---+---+---+---+---+---+
|         DS            |  ECN  |
+---+---+---+---+---+---+---+---+
```

DS: Differentiated Services field [RFC2474], [RFC3260]
ECN: ECN field [RFC3168]

Figure 4: The Structure of the DS and ECN Fields

3.2. Constraints from the DS Field

The Differentiated Services Codepoint (DSCP) set in the DS field 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, which uses a DSCP (or DSCPs) that indicates 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 one additional DSCP for each DSCP originally used to indicate the PHB and in any case should not use more than two. Therefore, the DSCP should be used to indicate that traffic is subject to PCN-metering and PCN-marking, but not to differentiate various PCN-markings.

3.2.2. Handling of the DSCP in 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, if PCN encoding uses more than a single DSCP.

3.2.3. Restoration of Original DSCPs at the Egress Node

If PCN-marking does not alter the original DSCP, the traffic leaves the PCN-domain with its original DSCP. However, if the PCN-marking alters the DSCP, then some additional technique is needed to restore the original DSCP. A few possibilities are discussed:

1. 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.

2. 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.

3. PCN-traffic is tunneled across the PCN-domain. The pipe-tunneling model is applied, 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 pseudowires are potential solutions that scale well in backbone networks.

The most appropriate option depends on the specific circumstances an operator faces.

- o Option 1 is most suitable unless there is a shortage of available DSCPs.

- o Option 3 is suitable where the reduction of MTU is not liable to cause issues.

3.3. Constraints from the ECN Field

This section briefly reviews the structure and use of the ECN field. The ECN field may be redefined, but certain constraints apply [RFC4774]. The impact on PCN deployment is discussed, 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].

```
          +-----+-----+
          | ECN FIELD |
          +-----+-----+
          0     0          Not-ECT
          0     1          ECT(1)
          1     0          ECT(0)
          1     1          CE
```

Figure 5: ECN Codepoints within the ECN Field

ECT stands for "ECN-capable transport" and indicates that the senders 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. 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 reuse 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 content 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.

3.3.3. Handling of the ECN Field in 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; 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. [RFC6040] updates both of these RFCs to rationalize them into one consistent approach.

3.3.3.1. 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.

3.3.3.2. 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 PCN-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.

3.3.3.3. 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 described in Section 3.3.3.2. Tunneling with IPsec also requires that PCN re-mark the ECN field only to CE because any other information is not copied to the inner header upon decapsulation and is lost. In contrast to Section 3.3.3.2, with IPsec tunnels, CE marks of tunneled PCN-traffic remain visible for routers along the tunnel and to their meters, markers, and droppers.

3.3.3.4. ECN Tunneling

New tunneling rules for ECN are specified in [RFC6040], which 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. 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. Certain combinations of inner and outer ECN fields cannot result from any transition in any current or previous ECN tunneling specification. These currently unused (CU) combinations are indicated in Figure 6 by '(!!!)' or '(!)'; where '(!!!)' means the combination is CU and always potentially dangerous, while '(!)' means it is CU and possibly dangerous.

Arriving Inner Header	Arriving Outer Header			
	Not-ECT	ECT(0)	ECT(1)	CE
Not-ECT	Not-ECT	Not-ECT(!!!)	Not-ECT(!!!)	<drop>(!!!)
ECT(0)	ECT(0)	ECT(0)	ECT(1)	CE
ECT(1)	ECT(1)	ECT(1) (!)	ECT(1)	CE
CE	CE	CE	CE(!!!)	CE

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 '(!)'. ([RFC6040]; '(!!!)' means the combination is CU and always potentially dangerous, while '(!)' means it is CU and possibly dangerous.)

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

3.3.4. 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 restore the ECN field that 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 3.3.3.1). 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.

4. Comparison of Encoding Options

The PCN working group has studied four different PCN encodings, which redefine the ECN field. Figure 7 summarizes these PCN encodings. One, or at most two, different DSCPs are used to indicate PCN-traffic, and, only for these DSCPs, the semantics of the ECN field are redefined within the PCN-domain.

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 use the PCN-specific DSCP by setting the Not-PCN codepoint. Special per-hop behavior, 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 DSCPs, which may be chosen by the network operator. Not-PCN means that packets are not PCN-enabled. NM means not-marked. 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. So, baseline encoding supports SM-PCN.

The 01-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).

4.2. Encoding with 1 DSCP Providing 3 States

PCN 3-state encoding uses a single DSCP (3-in-1 encoding, [RFC6660]), 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 Sections 3.3.3.1 - 3.3.3.3.

Therefore, the full 3-in-1 encoding may only be used for PCN-domains implementing the new rules for ECN tunnelling [RFC6040] or for PCN-domains without tunnels. Currently, it is not clear how fast the new tunnelling rules will be deployed and this affects the applicability of the full 3-in-1 encoding. Where PCN-domains do contain legacy tunnel endpoints, a restricted subset of the full 3-in-1 encoding can be used that omits the '01' codepoint.

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, [PCN-3-in-2]) uses two different DSCPs to accommodate the three required codepoints NM, ThM, and ETM. It leaves some codepoints currently unused (CU), and also proposes a way to reuse them to store some information about the content of the ECN field before the packet enters the PCN-domain. 3-in-2 encoding works well with IPsec tunnels (see Section 3.3.3.3). 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. Further, if PCN is applied to more than one Diffserv traffic class, then two DSCPs are needed for each. 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 [PCN-PSDM] in the presence of IPsec tunnels (see Section 3.3.3.3). PSDM encoding also supports SM-PCN.

4.5. Standardized Encodings

The baseline encoding described in Section 4.1 is defined in [RFC5696]. The intention was to allow for experimental encodings to build upon this baseline. However, following the publication of [RFC6040], the working group decided to change its approach and instead standardize only one encoding (the 3-in-1 encoding [RFC6660] described in Section 4.2). Rather than defining the 3-in-1 encoding as a Standards Track extension to the existing baseline encoding [RFC5696], it was agreed that it is best to define a new Standards Track document that obsoletes [RFC5696].

5. Conclusion

This document summarizes the PCN working group's exploration of a number of approaches for encoding pre-congestion information into the IP header. It is presented as an informational archive. It provides details of those approaches along with an explanation of the constraints that apply. The working group has concluded that the "3-in-1" encoding should be published as a Standards Track RFC that obsoletes the encoding specified in [RFC5696].

The reasoning is as follows. During the early life of the working group, the working group decided on an approach of a standardized "baseline" encoding [RFC5696], plus a series of experimental encodings that would all build on the baseline encoding, each of which would be useful in specific circumstances. However, after the tunneling of ECN was standardized in [RFC6040], the PCN working group decided on a different approach -- to recommend just one encoding, the "3-in-1 encoding".

Although in theory "3-in-1" could be specified as a Standards Track extension to the "baseline" encoding, the working group decided that it would be cleaner to obsolete [RFC5696] and specify "3-in-1" encoding in a new, stand-alone RFC.

6. Security Implications

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

7. 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.

8. References

8.1. Normative References

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

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

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

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

8.2. Informative References

[PCN-MS-AC] Menth, M. and R. Geib, "Admission Control Using PCN-Marked Signaling", Work in Progress, February 2011.

[PCN-3-in-2] Briscoe, B., Moncaster, T., and M. Menth, "A PCN Encoding Using 2 DSCPs to Provide 3 or More States", Work in Progress, March 2012.

[PCN-PSDM] Menth, M., Babiarz, J., Moncaster, T., and B. Briscoe, "PCN Encoding for Packet-Specific Dual Marking (PSDM Encoding)", Work in Progress, March 2012.

[RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC2983] Black, D., "Differentiated Services and Tunnels", RFC 2983, October 2000.

[RFC3260] Grossman, D., "New Terminology and Clarifications for Diffserv", RFC 3260, April 2002.

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

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

[RFC5559] Eardley, P., Ed., "Pre-Congestion Notification (PCN) Architecture", RFC 5559, June 2009.

[RFC5670] Eardley, P., Ed., "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.

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

[RFC6660] Briscoe, B., Moncaster, T., and M. Menth, "Encoding Three Pre-Congestion Notification (PCN) States in the IP Header Using a Single Diffserv Codepoint (DSCP)", RFC 6660, July 2012.

[RFC6661] Charny, A., Huang, F., Karagiannis, G., Menth, M., and T. Taylor, Ed., "Pre-Congestion Notification (PCN) Boundary-Node Behavior for the Controlled Load (CL) Mode of Operation", RFC 6661, July 2012.

[RFC6662] Charny, A., Zhang, J., Karagiannis, G., Menth, M., and T. Taylor, "Pre-Congestion Notification (PCN) Boundary-Node Behavior for the Single Marking (SM) Mode of Operation", RFC 6662, July 2012.

[Menth09] 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), Dresden/Germany, June 2009.

[Menth12] Menth, M. and F. Lehrieder, "Performance of PCN-Based Admission Control under Challenging Conditions", IEEE/ACM Transactions on Networking, vol. 20, no. 2, April 2012.

[Menth10] Menth, M. and F. Lehrieder, "PCN-Based Measured Rate Termination", Computer Networks Journal, vol. 54, no. 3, Sept. 2010

Authors' Addresses

Georgios Karagiannis
University of Twente
P.O. Box 217
7500 AE Enschede,
The Netherlands
EMail: g.karagiannis@utwente.nl

Kwok Ho Chan
Consultant
EMail: khchan.work@gmail.com

Toby Moncaster
University of Cambridge Computer Laboratory
William Gates Building, J J Thomson Avenue
Cambridge, CB3 0FD
United Kingdom
EMail: Toby.Moncaster@cl.cam.ac.uk

Michael Menth
University of Tuebingen
Sand 13
72076 Tuebingen
Germany

Phone: +49-7071-2970505
EMail: menth@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