Pre-Congestion Notification Using Packet-Specific Dual Marking

Michael Menth
Univ. of Würzburg, Germany

Jozef Babiarz
Nortel, Canada

Philip Eardley
BT Group, UK

***Abstract*—Pre-congestion notification (PCN) uses packet metering and marking within a PCN domain to notify PCN egress nodes about high load regimes in the network. One question is how to encode the PCN markings in packet headers. The problem is that the IPv4 packet header is short of available codepoints and that tunnelling mechanisms constrain solutions. This paper proposes packet-specific dual marking (PSDM) as a new encoding scheme that avoids these problems and also explains how to apply it to achieve PCN-based admission control and flow termination. Therefore, our proposal may improve the deployability of PCN in spite of the limited extensibility of the current Internet architecture.**

I. Introduction

Pre-congestion notification (PCN) is a new mechanism currently standardized by the IETF to facilitate PCN-based admission control (AC) and flow termination (FT) primarily for wired networks and inelastic realtime flows [1]. Traffic belonging to the PCN service class is prioritized over non-PCN traffic, which is essentially the DiffServ principle, and hence PCN traffic does not suffer from packet loss or delay when overload occurs in a network. In addition, the rate of PCN traffic is admission controlled so that overload cannot evolve within the PCN traffic class under normal operation. If the rate of PCN traffic becomes too large in case of a failure with subsequent rerouting, FT can remove some of the admitted traffic to restore a controlled load condition [2] on the overloaded link. The idea of PCN is that routers mark PCN packets on outgoing links when their PCN traffic rates exceed their configured admissible or supportable rates. Currently, PCN-based AC and FT is a per-domain concept. That means egress nodes evaluate the PCN packet markings and communicate the information about marked packets to ingress nodes which block admission requests for new PCN flows or terminate already admitted flows if required. An overview of existing techniques is provided in [3].

As mentioned above, PCN requires two different marking schemes to indicate whether current PCN traffic rate exceeds the admissible or supportable rate. When treating all packets in the same way, three different marking states are required: not marked (NM), admission-stop (AS) marked, and excess-traffic (ET) marked. The problem is that the IP header does not have available bits anymore. Therefore, the two bit explicit congestion notification (ECN) field of the VOICE-ADMIT Differentiated Services codepoint (DSCP) are proposed to be reused for PCN signaling. At first sight the resulting four encoding states would seem to be enough to signal three different states. However, one of them is needed for non-PCN traffic, and due to current encapsulation rules only re-marking to one out of the remaining three codepoints can survive tunneling within a PCN domain [4]. Extending the encoding to two re-marking options is possible [5] but consumes another DSCP which is too expensive given the shortage of DSCPs (A). One solution is to redefine the encapsulation rules, but this requires a lot of standardization effort which takes long time and is not clear whether this change will ever come [6], [7] (B). Another solution is to use the same, single metering and marking scheme for both AC and FT, but this constrains their accuracy and applicability [8], [9], [10] (C). The contribution of this paper is to propose a new solution: packet-specific dual marking (PSDM) [11], [12] (D). It uses feedback from probe packets for AC and feedback from data packets for FT. To that end, probe packets are subject only to exhaustive marking and data packets only to excess marking. Marked probe packets are interpreted as AS-marked, and marked PCN data packets are interpreted as ET-marked. This paper presents PSDM encoding, describes how admission control can be designed within these restrictions, and argues that existing FT methods [3] can be reused. The four potential solutions show that the current Internet architecture is packed so that extensions are difficult because compromises are needed to accommodate new features and mechanisms. However, we believe that the solution presented in this paper is good enough and fits well into today's architecture.

The paper is structured as follows. Sect. II explains basics of PCN. Sect. III reviews ECN, the restrictions imposed by encapsulation rules on PCN encoding, and the existing solutions (A-C) that have significant drawbacks. Sect. IV explains PSDM and the required edge behavior (D). Finally, Sect. V summarizes this work and draws conclusions.

II. Pre-Congestion Notification (PCN)

In this section we review the general idea of PCN-based admission control (AC) and flow termination (FT) and illustrate their application in a domain context in the Internet. We explain exhaustive and excess marking and give examples how PCN edge nodes turn the obtained PCN information into AC and FT decisions. An overview can be found in [3].

This work was funded by Deutsche Forschungsgemeinschaft (DFG) under grant TR257/18-2. The authors alone are responsible for the content of the paper.

A. Pre-Congestion Notification (PCN)

PCN defines a new traffic class that receives preferred treatment by nodes within a PCN domain. It provides information to support AC and FT for this traffic type. PCN introduces an admissible and a supportable rate threshold ($AR(l)$, $SR(l)$) for each link l of the PCN domain. This implies three different load regimes as illustrated in Fig. 1. If the PCN traffic rate $r(l)$ is below $AR(l)$, there is no pre-congestion and further flows may be admitted. If the PCN traffic rate $r(l)$ is above $AR(l)$, the link is *AR*-pre-congested and the rate above $AR(l)$ is *AR*-overload. In this state, no further flows should be admitted. If the PCN traffic rate $r(l)$ is above $SR(l)$, the link is *SR*-pre-congested and the rate above $SR(l)$ is *SR*-overload. In this state, some already admitted flows should be terminated to reduce the PCN rate $r(l)$ below $SR(l)$.

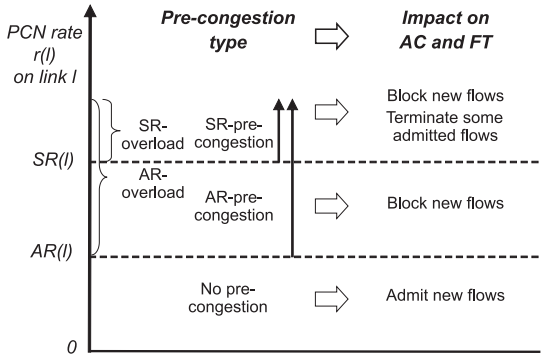

Fig. 1. The admissible and the supportable rate ($AR(l), SR(l)$) define three types of pre-congestion.

B. Edge-to-Edge PCN

Edge-to-edge PCN assumes that some end-to-end signalling protocol (e.g. SIP or RSVP) or a similar mechanism requests admission for a new flow that crosses a so-called PCN domain. This is similar to the IntServ-over-DiffServ concept [13]. Fig. 2 shows that edge-to-edge PCN is a per-domain QoS mechanism for the Internet and presents an alternative to RSVP clouds or extreme capacity overprovisioning. Traffic enters a PCN domain only through PCN ingress nodes and leaves it only through PCN egress nodes. Ingress nodes set a special header codepoint to make the packets distinguishable from other traffic and the egress nodes clear the codepoint. The nodes within a PCN domain are PCN nodes. They monitor the PCN traffic rate on their links and possibly re-mark the traffic in case of *AR*- or *SR*-pre-congestion. PCN egress nodes evaluate the markings of the traffic and send a digest to the AC and FT entities of the PCN domain. The overview in [3] presents different algorithms for these purposes.

C. PCN Metering and Marking

There are two basic marking strategies: excess and exhaustive marking. A token bucket based meter tracks whether a certain reference rate is exceeded. Exhaustive marking marks all packets when the PCN traffic rate exceeds the reference rate. When its reference rate is set to the admissible rate, exhaustive marking marks all packets in case of *AR*-pre-congestion and yields a very clear signal indicating that no more flows should be admitted. Excess marking marks only those packets that exceed the reference rate. When its reference rate is set to the supportable rate, the rate of marked packets corresponds to *SR*-overload. Egress nodes measure this rate to quantify the rate of PCN traffic to be terminated. For the description of AC and FT methods we assume the configuration presented in [14], [15]. Excess marking based on supportable rates meters all non-ET marked packets and re-marks some of them to "excess-traffic" (ET). Exhaustive marking based on admissible rates meters all PCN packets and re-marks all non-ET-marked packets to "admission-stop" (AS). Thus, in case of *AR*-pre-congestion, all packets are AS-marked and in case of *SR*-pre-congestion, some of the packets are ET-marked and the others are AS-marked.

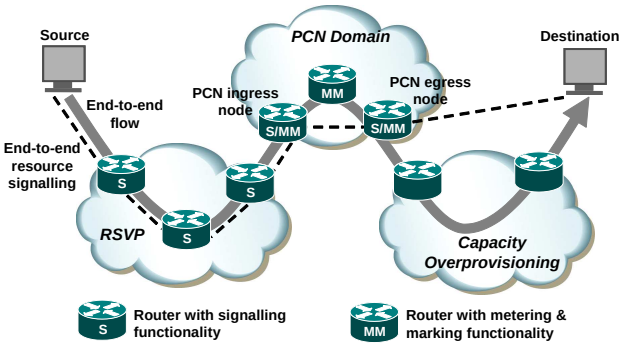

Fig. 2. Edge-to-edge PCN is triggered by admission requests from external signalling protocols and guarantees QoS within a single PCN domain.

D. Methods for Admission Control and Flow Termination

For a better understanding of PCN and to appreciate the advances of the new edge behaviors presented in Sect. IV, we review simple PCN-based AC- and FT-methods [3]. PCN ingress and egress nodes maintain information per ingress-egress aggregate (IEA). In particular, PCN egress nodes measure the rate of not marked, AS-marked, and ET-marked PCN traffic per IEA.

1) Admission Control: Ingress nodes keep per IEA an admission state K that indicates whether further flows for a particular IEA can be accepted or must be rejected. If egress nodes detect AS- or ET-marked packets for a particular IEA, they signal admission-stop to the corresponding ingress node. If AS- or ET-marked packets vanish, they signal admission-continue. Upon receipt of an admission-stop or admission-continue message the ingress node sets the admission state K to "block" or to "accept". Various implementations are possible, e.g. CLEBAC or OBAC [16].

2) Flow Termination: We introduce two different flow termination methods: measured rate termination (MRT) and marked flow termination (MFT). They can be applied as they are under PSDM.

With MRT egress nodes measure the traffic rate of received ET-marked PCN packets and communicate this rate to the ingress nodes. Upon receipt of such a message, the ingress

nodes terminate an appropriate set of PCN flows of the respective IEA. Further details and caveats can be found in [10]. With MFT the egress node maintains a credit counter for each admitted flow which is reduced by the amount of marked bytes received for that flow. When the counter becomes negative, the flow is terminated. This and other MFT methods have been proposed in [17], their performance has been evaluated, and recommendations have been given for their configuration.

The advantage of MFT compared to MRT is that only marked flows are terminated. In case of multipath routing, flows of a single IEA can be carried over different paths. If only one of them is congested, its important to remove the flows of the overloaded path. MFT achieves that while MRT does not achieve that goal.

III. Restrictions through ECN Encoding and Existing Solutions

In the previous section, we just talked about not marked, AS-marked, and ET-marked packets. In this section we show that it is difficult to encode these markings in the IP header. The explicit congestion notification (ECN) field is planned to be reused for PCN encoding. Therefore, we give a short overview of ECN and derive restrictions for PCN encoding due to encapsulation rules for ECN information. Existing solutions to this problem are to use more than a single DSCP for PCN encoding (A), to remove these restrictions through redefinition of the encapsulation rules for ECN information (B), or to use only a restricted set of PCN-based AC and FT algorithms that work with only a single marking scheme (C).

A. Explicit Congestion Notification

Random early detection (RED) was originally presented in [18], and in [19] it was recommended for deployment in the Internet. It was intended to detect incipient link congestion and to throttle only some TCP flows early in order to avoid severe congestion and to improve the TCP throughput. RED measures the average buffer occupation *avg* in routers and packets are dropped or marked with a probability that increases linearly with the average queue length *avg*. Explicit congestion notification (ECN) is built on the idea of RED to signal incipient congestion to TCP senders in order to reduce their sending window [20]. Packets of non-ECN-capable flows can be differentiated by a "not-ECN-capable transport" (not-ECT, '00') codepoint from packets of a ECN-capable flow which have an "ECN-capable transport" (ECT) codepoint. In case of incipient congestion, RED gateways possibly drop not-ECT packets while they just switch the codepoint of ECT packets to "congestion experienced" (CE, '11') instead of discarding them. This improves the TCP throughput since packet retransmission is no longer needed in this case. Both the ECN encoding in the packet header and the behavior of ECN-capable senders and receivers after the reception of a marked packet is defined in [20]. ECN comes with two different codepoints for ECT: ECT(0) ('10') and ECT(1) ('01'). They serve as nonces to detect cheating network equipment or receivers [21] that do not conform to the ECN semantics. The four codepoints are encoded in the two CU (currently unused) bits of the DS field in the IP header which is a redefinition of the type of service octet [22]. The ECN bits can be redefined by other protocols and [23] provides guidelines for that. They are likely to be reused for encoding of PCN marks.

B. Encapsulation Rules for ECN Information and its Impact on PCN Encoding

PCN traffic will possibly be indicated by the Differentiated Services codepoint (DSCP) VOICE-ADMIT [24]. To allow usage of this DSCP also for non-PCN traffic, the ECN field is set to not-ECT in that case. Thus, PCN traffic can use the ECT(0), ECT(1), and CE codepoints for PCN marking. The encoding scheme must cope with tunnelling within PCN domains. However, various tunnelling schemes limit the persistence of re-marked ECN codepoints in an outer IP header of an encapsulated packet to a different degree. Two IP-in-IP tunnelling modes are defined in [20] and a third one in [25] for IP-in-IPsec tunnels.

The limited-functionality option in [20] requires that the ECN codepoint in the outer header is set to not-ECT such that ECN is disabled for all tunnel routers, i.e., RED gateways drop packets instead of mark them in case of congestion. The tunnel egress just decapsulates the packet and leaves the ECN codepoints of the inner packet header unchanged. This mode protects the inner IP header from being PCN-marked upon decapsulation. It can be used to tunnel ECN marks across PCN domains such that PCN marking is applied to the outer header and used within the PCN domain without affecting the ECN field of the inner header which is intended to be used by end systems.

The full-functionality option in [20] requires that the tunnel ingress router copies the ECN codepoint of the inner header to the outer header unless the inner header codepoint is CE. In this case, the outer header codepoint is set to ECT(0). This choice has been made for security reasons to disable the ECN fields of the outer header as a covert channel. Upon decapsulation, the ECN codepoint of the inner header remains unchanged unless the outer header ECN codepoint is CE. In this case, the inner header codepoint is also set to CE. This preserves outer header information if it is CE. However, the fact that CE marks of the inner header are not visible in the outer header may be a problem for excess marking as it takes already marked traffic into account and also for some flow termination methods that require preferential dropping of CE-marked packets [3].

Tunnelling with IPSec copies the inner header ECN bits to the outer header ECN bits [25, Sect. 5.1.2.1] upon encapsulation. Upon decapsulation, CE-marks of the outer header are copied into the inner header and the other marks are ignored. With this tunnelling mode, CE marks of the inner header become visible to all meters, markers, and droppers for tunnelled traffic. In addition, limited information from the outer header is propagated into the inner header. While the tunnelling modes proposed in [20] cannot support PCN marking over tunnels, IPSec tunnels are able to preserve at

least re-marked CE codepoints. However, re-marking packets to ECT(0) or ECT(1) in the outer header does not survive the decapsulation action.

Due to these tunnelling rules, baseline encoding [4] requires only the ECN field of the VOICE-ADMIT DSCP but provides only two encoding states. This limits PCN functionality.

C. Solution A: Using Two DSCPs for PCN Encoding

Three-state encoding [5] provides three different states but requires the ECN fields of two different DSCPs. As the IP header bits are scarce, it is not likely that such an encoding scheme will prevail. A PCN encoding scheme providing three encoding states using a single DSCP is still missing.

D. Solution B: Redefinition of Tunnelling Rules

Recently, an attempt is made to redefine rules for tunnelling ECN information [6]. The draft points out that the limited ECN support was due to security reasons and that these concerns are not so severe that they justify the weak tunnelling support for the ECN field. It proposes to copy the complete ECN field from the inner header to the outer header upon encapsulation and from the outer header to the inner header upon decapsulation. The redefinition of the tunnelling rules for the ECN field assures that ECN information is propagated across protocol layers without loss in case of encapsulation and decapsulation. As a consequence, re-marking packets to any of the states ECT(0), ECT(1), or CE is persistent. This facilitates a new encoding scheme where ECT(0) corresponds to unmarked packets, ECT(1) to AS-marked packets, and CE to ET-marked packets [7]. This path is most sensible, but it will take long time until existing standards will be changed, and it is not sure whether this change will ever come. In addition, vendors need to change their products, at least those supporting PCN.

E. Solution C: PCN Using Only a Single Marking Scheme

Another option is to implement PCN with only a single marking scheme. That means that marking based on either the admissible or supportable rate can be implemented. In the first case only AC can be supported, in the second case only FT can be supported. In contrast, the method in [8], a supports both AC and FT when only excess marking based on the admissible rate is used. It requires that the supportable rate is a fixed multiple of the admissible rate on all links, i.e. $SR = AR \cdot u$. Admission is stopped for a specific IEA when its egress node observes a small amount of AS-marked packets. Flows are terminated when the measured PCN traffic rate at the egress rate is larger than u times the rate of the measured AS-marked PCN traffic. The traffic rate to be terminated is essentially the difference between these two rate values. This clever implementation works well in case of large aggregates and for single path routing. Under other conditions problems occur [9], [10] and the limited configuration flexibility through the coupling of *AR* and *SR* values can lead to bandwidth inefficiencies in resilient networks [26].

IV. Admission Control and Flow Termination Methods for PSDM

In this section, we explain packet-specific dual marking (PSDM) and how it may be used in a PCN context. We present various new PCN edge behaviors to support AC using PSDM. Some of them require the notion of ingress-egress aggregates (IEAs), others do not and can, therefore, easily cope with multipath routing. We do not further elaborate on FT methods as any method presented in [3] can be applied.

A. Packet-Specific Dual Marking

PSDM assumes two different types of PCN traffic: data traffic and probe traffic and we assume that they can be differentiated somehow. Data traffic is only subject to excess marking based on the supportable rate. The objective is to get quantitative feedback about how much PCN traffic must be terminated in case of *SR*-overload. Probe traffic is only subject to exhaustive marking based on the admissible rate. The objective is to get clear feedback whether additional flows can still be admitted. This concept does not require that AS-marked traffic is possibly re-marked to ET. It is a dual marking that is packet-specific, therefore, we call it packet-specific dual marking (PSDM). To hide specifics about packet formats from routers, PSDM encoding indicates which metering and marking scheme packets are subject to.

TABLE I
Interpretation of the ECN field under PSDM encoding.

Codepoint	ECN	PSDM	Interpretation for VOICE-ADMIT
'00'	not-ECT	not-PCN	not PCN
'01'	ECT(1)	not-EcM	not excess-marked PCN
'10'	ECT(0)	not-EhM	not exhaustive-marked PCN
'11'	CE	M	marked PCN

Table I shows how ECN codepoints are reused by PSDM encoding. The assumption is that the VOICE-ADMIT DSCP is used for PCN traffic. The VOICE-ADMIT DSCP is currently under standardization and is also used by non-PCN traffic. Such packets should use the not-PCN (not-ECT) codepoint while the other codepoints indicate PCN traffic. Not-EcM-marked PCN traffic is subject to excess marking and not-EhM-marked PCN traffic is subject to exhaustive marking. Excess marking meters only not-EcM-marked packets and possibly re-marks them to M. Exhaustive marking meters all PCN packets, but re-marks only not-EhM-marked packets to M. As packets are re-marked only to the M (CE) codepoint, this encoding survives IPSec tunnels. PCN ingress nodes mark PCN data packets with not-EcM and PCN probe packets with not-EhM and they are possibly re-marked to the same codepoint M. Therefore, PCN egress nodes must be able to differentiate PCN data and probe packets. They interpret marked probe packets as AS-marked and marked data packets as ET-marked.

B. A Short Note on Probing

We call all PCN traffic probe traffic that is not PCN data traffic and whose PCN feedback is possibly used for AC decisions. The notion of probe traffic is sometimes seen in a narrower sense, i.e. possibly several probe packets are

generated at the arrival of an admission request to test the pre-congestion state of the new flow's prospective path across the PCN domain. This entails significant management overhead and admission delay especially when multiple probe packets are sent per flow. These drawbacks do not apply for probing in general so that probing cannot be viewed per se as evil. The following AC methods use probing, but they do not have these drawbacks.

C. Admission Control Methods for IEAs

We first describe two AC methods that are similar to the one presented in Sect. II-D1. They also keep an admission state K per IEA. Both methods require that for all IEAs probe packets are regularly sent from the PCN ingress node to the PCN egress node. The size of the probe packets can be chosen arbitrarily small as exhaustive marking is not sensitive to packet sizes. The PCN egress node detects potential *AR*-pre-congestion and informs the PCN ingress node with admission-stop and admission-continue messages to update the corresponding admission state K. In the following we propose two different approaches for PCN egress nodes to detect *AR*-pre-congestion.

1) Observation-Based AC Using Probe Packets: When the PCN egress node receives an M-marked probe packet or detects a missing probe packet, it sends an admission-stop message to the corresponding PCN ingress node and sets a timer for the minimum block interval to a configurable value T_{block}. The timer is reset by consecutive arrivals of M-marked probe packets. When the timer expires, an admission-continue message is sent to the PCN ingress node.

2) Congestion Level Estimate (CLE) Based AC Using Probe Packets: The PCN egress node proceeds in measurement intervals. It tracks the number of missing probe packets or probe packets received with an M-mark during a measurement interval and at its end it calculates a congestion level estimate (CLE) as the fraction of this number and the number of overall received and missing probe packets. If the CLE is smaller than a configurable value T_{CLE}^{ACont}, an admission-continue message is sent to the PCN ingress node. If the CLE is larger than a configurable value T_{CLE}^{AStop}, an admission-stop message is sent to the PCN ingress node.

D. Admission Control Based on Implicit per Flow Probing

We briefly review RSVP and explain how its signalling messages can be reused for implicit per-flow probing.

1) A Brief Summary of RSVP: Realtime flows are usually accompanied by end-to-end signalling. A popular protocol example is RSVP [27]. With RSVP, the data source issues a PATH message which is carried hop-by-hop over the same path future data packets will go. To that end, the PATH message uses the same source and destination address as future data packets and also all other header fields that are possible input for routing and load balancing decisions need to be the same. When a PATH message arrives at an RSVP-capable node, a PATH state is established pointing to the previous hop before the PATH message is forwarded further downstream. When the PATH message arrives at the destination, the destination triggers the end-to-end reservation for the flow by sending a RESV message upstream along the nodes that set up a PATH state. In these nodes, the RESV message is processed. In particular, resource AC is performed for the new flow request and if it succeeds, the node forwards the RESV message to the previous hop recorded by the PATH state. This two pass signalling approach guarantees that the reservation is done on the downstream path of the future data flow. In contrast to PATH messages, RESV messages have the source address of the sending node and the destination address of the hop pointed to by the PATH state. That way, the information about the downstream next hop of the future data stream is conveyed to the previous hop and the flow-related information is stored in a RESV state. RSVP is a soft-state protocol, i.e., the PATH and RESV control messages are periodically sent to keep the PATH and RESV states alive and, thereby, the flow reservations. AC needs to be performed for a flow only once when no RESV state is set up, yet.

2) Modification of Standard RSVP to Perform PCN-Based AC: We assume that interior nodes of a PCN domain are RSVP-disabled. That means, they just forward RSVP messages without processing them and PCN ingress and egress nodes are neighboring RSVP-capable nodes. As a consequence, PCN ingress nodes decide whether new flows can be admitted and carried through the PCN domain or not. When the initial PATH message travels downstream, it is marked with not-EhM by the ingress node to indicate to PCN nodes that this packet is subject to exhaustive marking. It is possibly re-marked to M and eventually received by the PCN egress node. If no PATH state can be found for this flow at the PCN egress node, this PATH message is the first one and not a refresh message. If the PATH message is the first of the flow and if it is marked with M, the RSVP engine sends back a PATHERR message to reject the flow. If the PATH message is still marked with not-EhM, the RSVP PATH state is established at the PCN egress node and the PATH message is forwarded further downstream. Refresh messages are just forwarded according to standard RSVP. When the PATH message arrives at the destination and a RESV message is sent back along the nodes with a PATH state. Eventually, the corresponding RESV message arrives at the PCN ingress node. When no RESV state is set up yet, this is the first RESV message and admission control must be performed. By the mere fact that the RESV message arrives, the PCN ingress node knows that the corresponding initial PATH message was not marked. Thus, it can admit any PCN flow for which a new RESV message arrives. Note that RSVP is only an example for a two-pass end-to-end signalling protocol and the principle can be adapted to others.

E. Comparison with Other Deployment Scenarios

The advantage of PSDM compared to other dual marking solutions is that it requires only a single DSCP for encoding. While it is not clear whether a redefinition of ECN tunneling rules will ever come, PSDM encoding can be immediately standardized. Single marking uses only feedback from excess

marking based on admissible link rates to support AC and FT. However, such a scheme leads to significant inaccuracies for when IEAs carry only little traffic and it is not extensible to multipath routing [9], [10]. Therefore, PCN-based AC and FT using PSDM outperforms existing solutions that are feasible in today's Internet architecture.

F. Comparison of PSDM Encoding with Baseline Encoding

Baseline encoding is similar to PSDM encoding as it remarks not-marked packets (NM, ECT(0)) to marked (M, CE). It supports either excess or exhaustive marking but does not reveal which of them applies so that routers must be explicitly configured. With PSDM encoding PCN ingress nodes can mark all packets with not-EhM when only AC is implemented using exhaustive marking. Single marking can be supported when PCN ingress nodes mark all PCN packets with not-EcM as it requires excess marking. And the presented AC and FT deployment can be supported when PCN ingress nodes mark PCN probe packets with not-EhM and PCN data packets with not-EcM.

G. Caveats and Open Issues

PSDM encoding is less extensible than baseline encoding in the sense that baseline encoding can be extended to 3-in-1 encoding [7] or 3-state encoding [5] while PSDM already defines the meaning of ECT(1). Probe-based AC requires the definition of a probe packet format that can be easily recognized by PCN egress nodes. Implicit per-flow probing needs an upgrade of RSVP operation in PCN edge nodes. Probing for IEAs generates additional traffic. Probe packets can be arbitrarily small, e.g. 100 bytes every 50 ms, but this already leads to a probing overhead of 16 Kbit/s per IEA. However, this is certainly only an issue when capacity is scarce and the number of flows per IEA is low, but then AC methods without probing fail anyway [9].

V. Conclusion

We have illustrated the principle ideas of PCN-based admission control (AC) and flow termination (FT). Two different packet metering and marking schemes are needed, but there are not enough available codepoints to encode their markings. Packet-specific dual marking (PSDM) was presented as a new encoding scheme for PCN. It concurrently supports both marking schemes, but only one of them per packet. This is a restriction of the general PCN idea so that existing AC methods cannot be applied. We have presented simple edge behaviors for PCN ingress and egress nodes to implement PCN-based AC and FT using PSDM. They require probing to support AC. They are highly accurate in the sense that over-admission or overtermination is unlikely to occur. In addition, some of them can be can be applied even in networks with multipath routing which is not possible with other existing proposals. The presented PSDM encoding scheme is very flexible so that other than the presented edge behaviors can be easily supported. We proposed both PSDM and the new AC methods in IETF for deployment of PCN-based AC and FT in the current Internet architecture without compromising its applicability [11], [12].

Acknowledgment

The authors would like to thank Bob Briscoe and Toby Moncaster for stimulating ideas and fruitful discussions.

References

[1] P. Eardley (ed.), "Pre-Congestion Notification Architecture," http://tools.ietf.org/id/draft-ietf-pcn-architecture-09.txt, Jan. 2009.

[2] J. Wroclawski, "RFC2211: Specification of the Controlled-Load Network Element Service," Sep. 1997.

[3] M. Menth et al., "PCN-Based Admission Control and Flow Termination," in *currently under revision for IEEE Communications Surveys & Tutorials*, 2009.

[4] T. Moncaster, B. Briscoe, and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information," http://tools.ietf.org/id/draft-ietf-pcn-baseline-encoding-02.txt, Feb. 2009.

[5] ——, "A Three State Extended PCN Encoding Scheme," http://tools.ietf.org/id/draft-moncaster-pcn-3-state-encoding-00.txt, Jun. 2008.

[6] B. Briscoe, "Layered Encapsulation of Congestion Notification," http://tools.ietf.org/id/draft-briscoe-tsvwg-ecn-tunnel-01.txt, Oct. 2008.

[7] ——, "PCN 3-State Encoding Extension in a single DSCP," http://tools.ietf.org/id/draft-briscoe-pcn-3-in-1-encoding-00.txt, Oct. 2008.

[8] A. Charny et al., "Pre-Congestion Notification Using Single Marking for Admission and Pre-emption," http://tools.ietf.org/id/draft-charny-pcn-single-marking-03.txt, Nov. 2007.

[9] M. Menth and F. Lehrieder, "Applicability of PCN-Based Admission Control," in *currently under submission*, 2008.

[10] ——, "PCN-Based Measured Rate Termination," in *currently under submission*, 2008.

[11] M. Menth, J. Babiarz, T. Moncaster, and B. Briscoe, "PCN Encoding for Packet-Specific Dual Marking (PSDM)," http://tools.ietf.org/id/draft-menth-pcn-psdm-encoding-00.txt, Jul. 2008.

[12] M. Menth, "Deployment Models for PCN-Based Admission Control and Flow Termination Using Packet-Specific Dual Marking (PSDM)," http://tools.ietf.org/id/draft-menth-pcn-psdm-deployment-00.txt, Oct. 2008.

[13] Y. Bernet et al., "RFC2998: A Framework for Integrated Services Operation over Diffserv Networks," Nov. 2000.

[14] B. Briscoe et al., "An Edge-to-Edge Deployment Model for Pre-Congestion Notification: Admission Control over a DiffServ Region," http://tools.ietf.org/id/draft-briscoe-tsvwg-cl-architecture-04.txt, Oct. 2006.

[15] B. Briscoe et al., "Pre-Congestion Notification Marking," http://tools.ietf.org/id/draft-briscoe-tsvwg-cl-phb-03.txt, Oct. 2006.

[16] M. Menth and F. Lehrieder, "Performance Evaluation of PCN-Based Admission Control," in *IWQoS*, 2008.

[17] ——, "PCN-Based Marked Flow Termination," in *currently under submission*, 2008.

[18] S. Floyd and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance," *IEEE/ACM ToN*, vol. 1, no. 4, 1993.

[19] B. Braden et al., "RFC2309: Recommendations on Queue Management and Congestion Avoidance in the Internet," Apr. 1998.

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

[21] N. Spring, D. Wetherall, and D. Ely, "RFC3540: Robust Explicit Congestion Notification (ECN)," Jun. 2003.

[22] K. Nichols et al., "RFC2474: Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," Dec. 1998.

[23] S. Floyd, "RFC4774: Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) Field," Feb. 2007.

[24] F. Baker, J. Polk, and M. Dolly, "DSCPs for Capacity-Admitted Traffic," http://www.ietf.org/internet-drafts/draft-ietf-tsvwg-admitted-realtime-dscp-05.txt, Nov. 2008.

[25] S. Kent and K. Seo, "RFC4301: Security Architecture for the Internet Protocol," Dec. 2005.

[26] M. Menth and M. Hartmann, "Threshold Configuration and Routing Optimization for PCN-Based Resilient Admission Control," *accepted for Computer Networks*, 2009.

[27] B. Braden et al., "RFC2205: Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification," Sep. 1997.