PCN Working Group
Internet-Draft
Intended status: Informational
Expires: March 14, 2010

D. Satoh
H. Ueno
Y. Maeda
O. Phanachet
NTT-AT

September 10, 2009

Single PCN Threshold Marking by using PCN baseline encoding for both admission and termination controls draft-satoh-pcn-st-marking-02

Status of this Memo

Copyright Notice

Abstract

Pre-congestion notification (PCN) gives early warning of congestion by metering and marking packets in order to protect the quality of service of inelastic flows. PCN traffic load is divide into three pre-congestion states by two rates that [I-D.ietf.pcn.architecture] defines per link: PCN-admissible- and PCN-supportable-rates. PCN admission control and flow termination mechanisms operate in accordance with these three states. [I-D.ietf.pcn.baseline.encoding] defines two PCN encoding states. This document proposes an algorithm for marking and metering by using PCN baseline encoding for both flow admission and flow termination. The ratio of marked packets determines the three link states: no packets marked, some packets marked, and all packets marked. To achieve this marking behaviour, we use two token buckets. One is not used for marking but for a marking switch; the other is used for marking. The token bucket for marking has two thresholds. One is TBthreshold.threshold, already defined in [I-D.ietf-pcn-marking-behaviour], and the other is a new threshold, which is set to be the number of bits of a metered-packet smaller than the token bucket size. Therefore, the new threshold is larger than TBthreshold.threshold. If the amount of tokens is less than TBthreshold.threshold, all the packets are marked as defined in [I-D.ietf-pcn-marking-behaviour]. If the amount of tokens is less than the new threshold and greater than TBthreshold.threshold, one-Nth packets are marked. We evaluated the performance of admission control and flow termination using a simulation. For admission control, the results show that the performance of the algorithm was almost the same as, but slightly inferior to, that of CL [draft-briscoe-tsvwg-cl-phb-03]. For flow termination, the performance of the algorithm was almost the same as CL when the load was 1.2 times the supportable rate, but it was superior to CL when the load was high (two times the supportable rate). Furthermore, in the algorithm, over termination percentages of all the bottleneck links are almost the same in the case of multi-bottleneck. In CL, the over termination percentages of all the bottleneck links are different and those at upstream bottleneck links are higher than those at downstream bottleneck links because of accumulation of marked packets.

Table of Contents

1. Introduction

Pre-congestion notification (PCN) provides information to support admission control and flow termination in order to protect the quality of service (QoS) of inelastic flows. Although several algorithms (e.g., [I-D.briscoe-tsvwg-cl-phb], [I-D.charny-pcn-single-marking], and [I-D.babiarz-pcn-3sm]) have been proposed to achieve PCN, only the single marking algorithm (SM) [I-D.charny-pcnsingle-marking] meets the requirement of baseline encoding.

This document proposes an algorithm for marking and metering by using PCN baseline encoding for both flow admission and flow termination. Our algorithm uses PCN-threshold marking while SM uses PCN-excess-traffic marking. Although our algorithm uses an elaborate mechanism, it chooses PCN-admissible- and PCN-supportable-rates independently, and it explicitly detects whether PCN traffic is greater than the PCN-supportable-rate. Furthermore, our algorithm is little affected by synchronization. Therefore, it can avoid degradation of admission accuracy caused by synchronization.

2. Terminology

The terminology used in this document conforms to the terminology of [ID.ietf-pcn-architecture] and [I-D.ietf-pcn-marking-behaviour].

3. Single Threshold-Marking

We describe an algorithm of marking and metering by using PCN baseline encoding for both flow admission and flow termination. This algorithm uses only the PCN-threshold-rate as the PCN-supportable-rate and does not use the PCN-excess-rate. We show a schematic of how the PCN admission control and flow termination mechanisms operate as the rate of PCN-traffic increases for a PCN-domain with three types of ratios of PCN-threshold-marked packets and three states of marking ratio in Fig. 1. Only two encoding states:un-marked and marked can be used for marking when PCN baseline encoding is used. We use two encoding state marking for the marking ratio to distinguish the three states. As shown in Fig. 1, no packets are PCN-marked at a rate less than the PCN-admissible-rate, some packets are PCN-marked at rates between the PCN-admissible- and PCN-supportable-rates, and all the packets are PCN-marked at rates greater than the PCN-supportable-rate. Admission control using this marking stops traffic from being admitted when the fraction of marked traffic for an ingress-egress aggregate (IEA) exceeds a configured threshold: the congestion level estimate (CLE). When the egress receives a certain sequential marked packets, it sends a message of

the receiving rate to an ingress. Then the ingress terminates flows based on the information of the sending and receiving rates.

The remainder of this section describes the possible operation of the system.

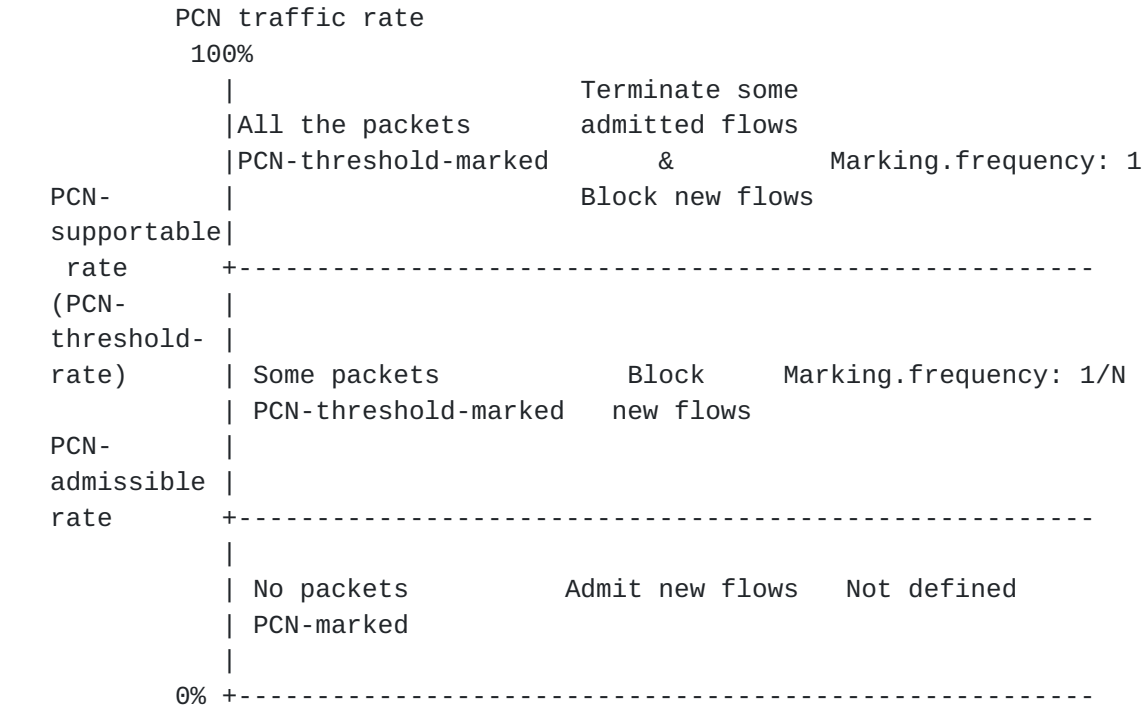

Figure 1: Ratio of PCN-threshold-marked packets, control operations, and Marking.frequency

3.1. Operation at PCN-interior-node

We explain here how to distinguish the three link states. The single threshold-marking algorithm (STM) uses two token buckets. One is not used for marking but as a marking switch, which we call the switch token bucket (SwTB). The other is used for marking, which we call the marking token bucket (MkTB). We also explain how to mark by using two token buckets in cooperation.

SwTB can be used when the traffic is below the admissible rate so that no packets are marked. Metering and marking using SwTB is similar to threshold metering and marking, although there is a difference between actual marking and having the marking switch ON. SwTB has one threshold, which is termed the SwTB threshold. Tokens of SwTB are added at the PCN-admissible-rate. Tokens are removed equal to the bit-size of the metered packet. These additions and

removals are independent of MkTB. If the metered traffic is sustained at a level greater than the PCN-admissible-rate, the tokens are less than the SwTB threshold. When the tokens are less than the SwTB threshold, the marking switch of SwTB is set to ON and the metered packet is threshold-marked in accordance with the meter based on MkTB. Otherwise, the marking switch of SwTB is OFF. When the marking switch is OFF, no packet is marked regardless of the meter of MkTB. Thus, the state of admitting new flows (seen in Fig. 1) is achieved.

MkTB can be used to mark packets in accordance with traffic rates. Metering and marking using MkTB is modified threshold metering and marking. MkTB has two thresholds. One, termed TBthreshold.shallow.threshold, is set to be smaller than the token bucket size by the bit-size of a metered packet. The other is TBthreshold.threshold in [I-D. pcn-marking-behaviour]. Regardless of the marking switch being ON or OFF, tokens of MkTB are added at the PCN-supportable-rate and tokens are removed equal to the bit-size of the metered packet. These additions and removals are independent of SwTB.

If the metered traffic is sustained at a level greater than the PCN-supportable-rate, the tokens are less than the TBthreshold.threshold. When the tokens are less than TBthreshold.threshold, all the metered packets are threshold-marked. The marking.frequency in this state is 1 because all the metered packets are marked. Relations among marking.frequency, marking behaviour, and PCN mechanisms are shown in Fig. 1. Thus, the state of terminating certain admitted flows and blocking new flows (seen in Fig. 1) is achieved.

If the metered traffic is between the PCN-admissible-rate and PCN-supportable-rate, the tokens are greater than the TBthreshold.threshold. If the metered traffic has rate fluctuations, the tokens are often less than TBthreshold.shallow.threshold. Note that the probability that the tokens are less than TBthreshold.shallow.threshold is approximately the ratio between the rate of the metered traffic and the PCN-supportable-rate [see Appendix D]. If the tokens are less than TBthreshold.shallow.threshold (the tokens are less than MkTB.size before the tokens of the arrived packet.size are removed), a metered packet is PCN-threshold-marked for every one-Nth. Note that one-Nth arrived packets are not marked. The marking.frequency in this state is 1/N. This marking with marking.frequency 1/N avoids accommodating marking through multi-bottleneck links to distinguish partial marking from marking of all packets.

This marking is expressed by the following pseudo code:

```
Input: pcn packet
//add tokens to the two token buckets
SwTB.fill = min(SwTB.size, SwTB.fill + SwTB.rate*(now - lastUpdate));
MkTB.fill = min(MkTB.size, MkTB.fill + MkTB.rate*(now - lastUpdate));
//remove tokens from the two token buckets
SwTB.fill = max( 0 , SwTB.fill - packet.size);
MkTB.fill = max( 0 , MkTB.fill - packet.size);
IF (SwTB.fill <= SwTB.threshold) THEN  //marking switch is ON
    IF  (MkTB.fill < MkTBthreshold.shallow.threshold) THEN
        IF (MkTB.fill > MkTBthreshold.threshold) THEN
            markCnt++;
            IF mod(markCnt, N) == 0 THEN //Marking.frequency = 1/N
               markCnt = 0;
               packet.mark = TM;
            ENDIF
        ELSE
            markCnt = 0;
            packet.mark = TM;
        ENDIF
    ENDIF
ELSE //marking switch is OFF
ENDIF
output: void
```

It can be rewritten as follows.

```
Input: pcn packet
//add tokens to the two token buckets
SwTB.fill = min(SwTB.size, SwTB.fill + SwTB.rate*(now - lastUpdate));
MkTB.fill = min(MkTB.size, MkTB.fill + MkTB.rate*(now - lastUpdate));
lastUpdate = now;
markOpe = ((SwTB.fill - packet.size) >= SwTB.Threshold) ? false:true;
IF (markOpe) THEN
   IF (MkTB.fill < MkTB.size) THEN
      markCnt ++;
      IF (MkTB.fill < MkTB.Threshold) THEN
         packet.mark = TM;
         markCnt = 0;
      ELSE IF (markCnt == N) THEN
         packet.mark = TM;
         markCnt = 0;
      ENDIF
   ENDIF
ENDIF
//remove tokens from each token bucket
SwTB.fill = max( 0 , SwTB.fill - packet.size);
```

```
MkTB.fill = max( 0 , MkTB.fill - packet.size);
Output: void
```

Here, SwTB.fill and MkTB.fill (TBthreshold.fill in [I-D.pcn-marking-behaviour]) represent the amount of tokens in the SwTB and MkTB, SwTB.size and MkTB.size (TBthreshold.max in [I-D.pcn-marking-behaviour]) represent the maximum values of SwTB.fill and MkTB.fill, and SwTB.threshold and MkTB.threshold represent the SwTB threshold and MkTB threshold (TBthreshold.threshold), respectively. Tokens of SwTB are added at the SwTB.rate whose value is the PCN-admissible-rate. Tokens of MkTB are added at the MkTB.rate whose value is the PCN-supportable-rate (PCN-threshold-rate). The time variables now and lastUpdate are the current time and the time when the fill state of TB was last updated, respectively. The byte of an arrived packet is packet.size. TM represents the state of the threshold-marked packet, packet.mark represents the state of the mark of a packet, and markCnt represents the number of marked packets.

The relation between PCN traffic rate and marking ratio is shown Fig. 2, where AR and SR represent PCN-admissible- and PCN-supportable-rates.

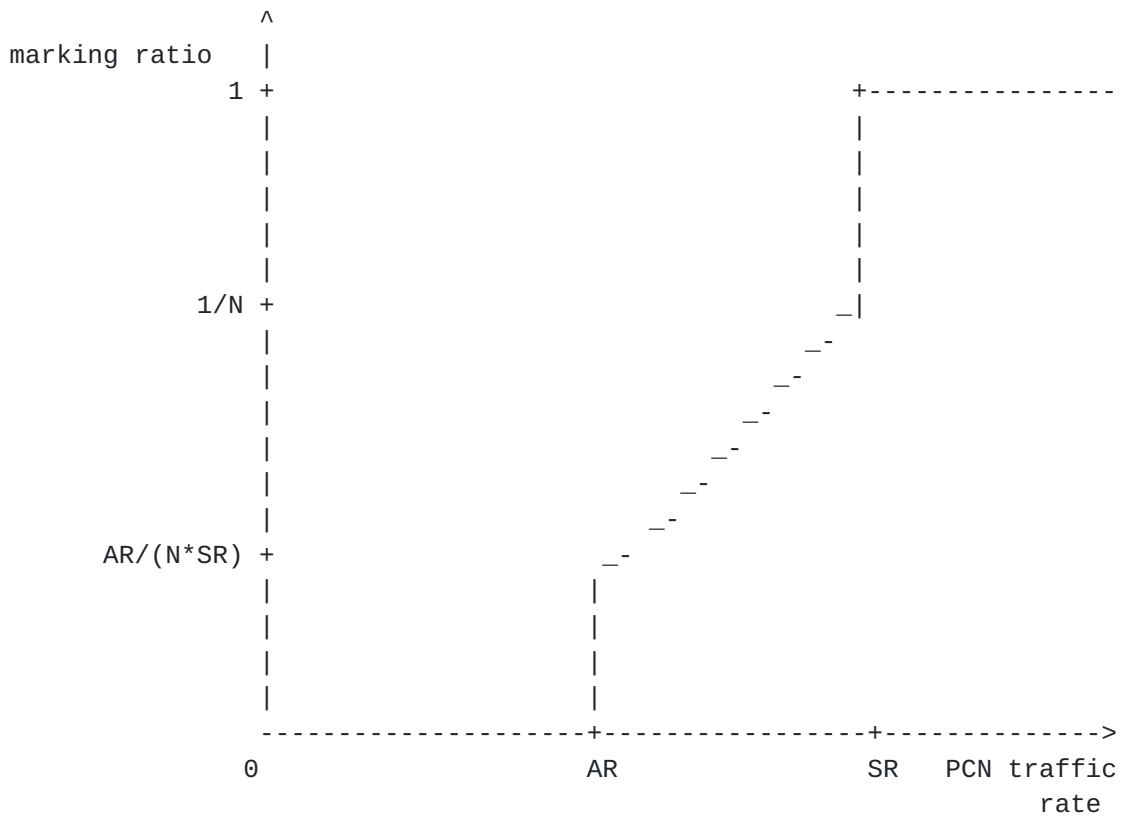

Figure 2: Relation between PCN traffic rate and marking ratio

3.2. Operation at PCN-egress-node

3.2.1. Operation for Admission Decision

A PCN-egress-node measures the ratio of PCN-threshold-marked packets on a per-ingress basis and reports to the PCN-ingress-node the congestion level estimate (CLE), which is the fraction of the marked traffic received from the ingress node and is exactly the same as that of CL, SM, and three state PCN marking algorithms (3sM).

3.2.2. Operation for Flow Termination Decision

The PCN-egress-node always measures the receiving rate per PCN-ingress-node regardless of whether receiving marked packets or not. When the PCN-egress-node receives L-sequential marked packets from a PCN-ingress-node, it sends a packet of information about the receiving rate to the PCN-ingress-node at an interval that is shorter than the measuring interval of the PCN-ingress-node. If the PCN-egress-node receives M-sequential unmarked packets, it sends a packet for canceling termination to the PCN-ingress-node.

3.3. Operation at the PCN-ingress-node

3.3.1. Admission Decision

Just as in CL and SM, the admission decision is based on the CLE. The ingress node stops admitting new flows if the CLE is greater than a predefined threshold (CLE threshold). The CLE threshold is chosen under the following maximum value. The maximum of the CLE threshold MUST be

$$\text{CLE threshold} = (1/N)*\text{rho} \quad , \qquad (1)$$

where rho represents the minimum ratio between the PCN-admissible-rate and PCN-supportable-rate via all the links between the PCN-ingress and egress nodes and N is the denominator of marking.frequency 1/N. Note that the marking ratio approaches the maximum CLE threshold described above when the load is the PCN-admissible-rate at a bottleneck link via all the links between the PCN-ingress and egress nodes as the number of superposition of flows approaches infinity, as described in Appendix D.

3.3.2. Flow Termination Decision

As soon as a PCN-ingress-node receives a packet with information about a receiving rate from a PCN-egress-node, it starts measuring the sending rate to the PCN-egress-node. It receives the packet multiple times during the measuring interval and maintains the minimum value of the receiving rate. At the end of the measurement interval, it terminates flows whose bandwidths are the same as the sending rate - (1-y)* min(receiving rate, sending rate). Because flow termination affects the receiving rate after transmission delay, the receiving rate can be larger than the sending rate. Therefore, min(receiving rate, sending rate) is used. The value of y is small enough to tolerate. It stops terminating flows or measuring the sending rate as soon as it receives the packet for canceling termination from the PCN-egress-node. This termination takes more time when the PCN-supportable-rate is much lower than the physical rate because the sending rate is almost the same as the receiving rate even under a congested situation.

4. Admission Issues

4.1. Effect of Synchronization

CL and SM suffer from synchronization [I-D.zhang-pcn-performance-evaluation, I-D.charny-pcn-single-marking]. Synchronization is defined in [I-D.zhang-pcn-performance-evaluation] as the phenomenon of some flows having all their packets marked, while other flows have none of their packets marked. This phenomenon easily occurs in the case of CBR flows. If the duration time of flows is infinity, a superposition of homogeneous CBR flows shows periodic behaviour for every interval between two sequential packets of a CBR flow. Even if the duration time is limited, the superposition of homogeneous CBR flows shows periodic behaviour when the number of flows is not changed. When excess-traffic-marking is applied, the periodic behaviour of the packets leads to periodic marking behaviour because the periodic behaviour of the packets leads to the periodic behaviour of tokens, which causes periodic marking behaviour. The period of marking behaviour is the same as that of the behaviour of the packets. The periodic marking behaviour lasts as long as the periodic behaviour of packets. Synchronization makes the admission and termination of CL and SM less accurate [I-D.zhang-pcn-performance-evaluation, I-D.charny-pcn-single-marking].

Although STM uses partial marking for admission control, the marking is not excess-traffic-marking but a kind of threshold-marking. It is true that the periodic behaviour of packets leads to the periodic behaviour of tokens, but the period of the marking behaviour is

different from that of behaviour of packets when the marking.frequency is not 1. The period of marking behaviour is longer than the periodic behaviour of packets. Thus, no flows have all their packets marked. Because the marking of STM is threshold marking, more marking occurs for metered packets when 1 is chosen as N at the marking of STM than when excess-traffic-marking is applied. The ratio of marked packets to all the packets is approximately the same as that of traffic load to the PCN-supportable-rate as shown in Appendix D. Thus, the ratio of flows that have no packets marked in STM is smaller than that in the excess-traffic-marking although some flows have no packets marked at all. Furthermore, when PCN traffic is close to the admissible rate, no packets are marked in a short interval because of traffic fluctuations. A metered packet is not marked when the rate of the PCN traffic is less than the admissible rate in the marking of STM. This breaks the periodic marking behaviour. Almost the same performance results were achieved with and without randomized traffic, which supports the description above. Therefore, STM is little affected by synchronization.

5. Termination Issues

5.1. Effect of dropping packets to Termination

When the PCN traffic is greater than the PCN-supportable-rate, all the packets are threshold-marked in STM. Dropping packets themselves does not influence the performance of STM termination. Therefore, a node can drop any packets. Preferential dropping packets is not necessary.

5.2. Effect of multi-bottleneck

In the case of multi-bottleneck links, over termination percentages of STM are almost the same among all the bottleneck links although CL makes over termination percentages of upstream bottleneck links worse because CL is affected by accumulation of marking.

5.3. Speed and Accuracy of Termination

Because a PCN-ingress-node terminates flows every measurement interval, the amount of termination can be roughly evaluated. Therefore, by using the ratio between the physical link speed and the PCN-supportable-rate, the value of y is estimated roughly in order to finish termination within a certain interval. However, this termination has a trade-off between termination speed and accuracy.

6. Performance evaluation

We compared the performance of STM with that of CL. For admission, an over admission percentage was chosen, and for flow termination, over termination and termination time were chosen as performance metrics.

For admission control, the performance of STM was almost the same as that of CL, although STM was slightly inferior to CL.

For flow termination, when the load was slightly larger than the PCN-supportable-rate and the traffic type was CBR, the over termination percentages of STM were inferior to those of CL. However, when the traffic type was VBR or SVD, the over termination percentages of STM were almost the same as those of CL. When the load was much larger than the PCN-supportable-rate, our simulation showed that the over termination percentages of STM were superior to those of CL. STM took more termination time than CL because STM terminates flows little by little. In the multi-bottleneck case, over termination percentages of STM are almost the same among all the bottleneck liks because STM is not affected by accumulation of marking.

7. Impact on PCN marking behaviour

The goal of this section is to propose two minor changes to the PCN-marking-behaviour framework as currently described in [I-D.ietf-pcn-marking-behaviour] in order to enable the single threshold-marking approach. We propose additions of a threshold meter function with two thresholds and a combination of the meter functions. A new threshold of the threshold meter function is larger than TBthreshold.threshold. If the amount of tokens is less than the new threshold of the threshold meter function, the metered packets are partially marked and not all marked. This marking function includes the ramp-marking in [I-D.briscoe-tsvwg-cl-phb]. The combination of SwTB and MkTB is an example of the combination of the meter functions.

8. Changes from -01 version

- Added simulation results of redifined termination time (Appendix C)
- Added admission control using fractional marking (Appendix E)
- Other minor edits

9. Acknowledgements

This research was partially supported by the National Institute of Information and Communications Technology (NICT), Tokyo, Japan.

We thank Mr. Takayuki Uchida at U-software Corporation for helping with the simulations and Professor Shigeo Shioda at Chiba University for fruitful discussions. We also thank Assistant Professor Michael Menth at the University of Wuerzburg for his careful reading of this draft and fruitful discussions.

10. Appendix A: Simulation Setup and Environment

10.1. Network and Signaling Models

We used the three types of network topologies shown in the figures below for simulations. They are the same as those in [I-D.charny-pcn-single-marking] and [I-D.zhang-pcn-performance-evaluation], and the first two figures are the same as those in [I-D. briscoe-tsvwg-cl-phb]. The first type of network topology is single link (Fig. A.1). The second type is a multi-link network with a single bottleneck (termed "RTT") (Fig. A.2). The third type is a range of multi-bottleneck topologies (termed "Parking Lot") (Fig. A.3).

A single link between an ingress and an egress node is shown in Fig. A.1, where all flows enter at node A and depart from node B. This topology is used to basically verify the behaviour of the algorithms with respect to a single ingress-egress aggregate (IEA) in isolation.

Figure A.1: Simulated single-link network.

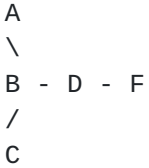

Figure A.2: Simulated multi-link network.

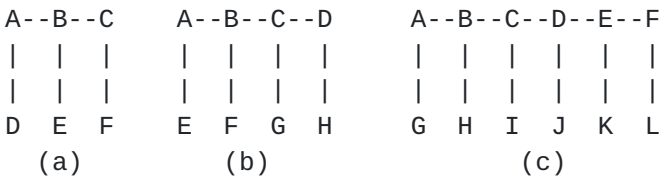

Figure A.3: Simulated multi-link network.

As shown in Fig. A.2, a set of ingresses (A, B, and C) are connected to an interior node in the network (D). This topology is used to study the behaviour of the algorithm where many IEAs share a single bottleneck link. The number of ingresses varied in different simulation experiments from 2 - 1800. All links generally have different propagation delays. Thus, these propagation delays were chosen randomly in the range of 1 - 100 ms. This node D in turn is connected to the egress (F). In this topology, different sets of flows between each ingress and the egress converge on a single link D-F, where the PCN algorithm is enabled. The capacities of the ingress links are not limiting, and hence no PCN is enabled on them. The bottleneck link D-F was modeled with a 10-ms propagation delay in all simulations. Therefore, the range of round-trip delays in the experiments was from 22 - 220 ms.

Another type of network of interest is that with a parking lot (PLT) topology, which has multi-bottleneck links. The simplest PLT with two bottlenecks is illustrated in Fig. A.3(a). A traffic matrix with this network on this topology is as follows.

- an aggregate of "2-hop" flows entering the network at A and leaving at C (via the two links A-B-C)

- an aggregate of "1-hop" flows entering the network at D and leaving at E (via A-B)

- an aggregate of "1-hop" flows entering the network at E and leaving at F (via B-C)

In the 2-hop PLT shown in Fig. A.3(a), the points of congestion are links A-B and B-C. The capacities of all the other links are not limited. We also experimented with larger PLT topologies with three bottlenecks (Fig. A.3(b)) and five bottlenecks (Fig. A.3(c)). In all cases, we simulated one ingress-egress pair that carried the aggregate of "long" flows traversing all N bottlenecks (where N is the number of bottleneck links in the PLT topology) and N ingress-egress pairs that carried flows traversing a single bottleneck link and exited at the next "hop". In all cases, only the "horizontal" links in Fig. A.3 were the bottlenecks, with non-limited capacities

of all "vertical" links. We named the bottleneck IDs in the order of lowest (upstream) to highest (downstream), e.g., bottleneck ID 1 is the upstream link between nodes A and B and bottleneck ID 2 is the downstream link between nodes B and C in Fig. A.3 (a). The propagation delays for all links in all PLT topologies were set to 1 ms.

Our simulations concentrated primarily on the range of capacities of "bottleneck" links with sufficient aggregation - above 128 Mbps for voice and 2.4 Gbps for SVD. Higher link speeds will generally result in higher levels of aggregation and hence generally better performance of measurement-based algorithms. Therefore, it seems reasonable to believe that the studied link speeds do provide meaningful evaluation targets.

In the simulation model, a call request arrives at the ingress, which immediately sends a message to the egress. The message arrives at the egress after the propagation time plus link processing time (but no queuing delay). When the egress receives this message, it immediately responds to the ingress with the current CLE. If the CLE is below the specified CLE threshold, the call is admitted. If otherwise, it is rejected. An admitted call sends packets in accordance with one of the chosen traffic models for the duration of the call (see next section). The propagation delay from the source to the ingress and from the destination to the egress is assumed to be negligible and is not modeled.

In the admission control simulation, the PCN-admissible- and PCN-supportable-rates were half the link speed and 90% of the link speed, respectively, and in the flow termination simulation, the PCN-admission-rate and PCN-supportable-rate were 20% and 40% of the link speed in all the links. The actual queue size was 80,000 packets, which was not the byte size because queue size was set by only the number of packets in the NS2 simulator.

10.2. Traffic Models

We use the same types of traffic as those of [I-D.briscoe-tsvwg-cl-phb], [I-D.charny-pcn-single-marking], and [I-D.zhang-pcn-performance-evaluation]. These are CBR voice, on-off traffic approximating voice with silence compression (termed "VBR"), and on-off traffic with higher peak and mean rates (we termed the latter "synthetic video" (SVD)). On-off traffic (VBR and SVD) is described with a two-state Markov chain, and on and off periods were exponentially distributed with the specified mean.

Each flow arrives according to the Poisson process. The distribution of flow duration was chosen to be exponentially distributed with a

mean of 1 min, regardless of the traffic type, which is the same as that in [I-D.briscoe-tsvwg-cl-phb] and [I-D.charny-pcn-single-marking] in admission simulation. The flow duration was infinity in termination simulation in order to observe the effect by only flow termination.

Traffic parameters for each type are summarized below.

CBR voice

- Packet length: 160 bytes
- Packet inter-arrival time: 20ms ((160*8)/(64*1000) s)
- Average rate: 64 Kbps

On-off traffic approximating voice with silence compression

- Packet length: 160 bytes
- Packet inter-arrival time during on period: 20 ms
- Long-term average rate: 21.76 Kbps
- On period mean duration: 340ms; during the on period traffic is sent with the CBR voice parameters described above
- Off period mean duration: 660 ms; no traffic is sent for the duration of the off period

SVD

- Packet length: 1500 bytes
- Packet inter-arrival time during on period: 1 ms
- Long term average rate: 4 Mbps
- On period mean duration: 340 ms
- Off period mean duration: 660 ms

10.3. Performance Metrics

In our experiments, we used the percent deviation of the mean rate of the expected load level as a performance metric. We term these "over-admission" and "over-termination" percentages, depending on the type of experiment.

In our experiments of admission control, we compared the actual achieved average throughput to the desired traffic load (PCN-admissible-rate). The desired traffic load is not the exact admissible rate because a signal packet for a new flow is sent for admission. Therefore, the desired traffic load is the admissible rate minus the amount of signal packets for non-admitted flows.

In our experiments of flow termination, we compared the actual achieved average throughput to the desired traffic load (PCN-supportable-rate) and measured termination time. As termination time, we measured time between the first termination and the termination by which the PCN rate is less than the PCN-supportable-rate. The actual termination time takes more time to measure rates at an ingress and egress nodes (200 ms for CL and 100 ms for STM) and latency one-way time from the PCN-egress-node to PCN-ingress-node.

10.4. Parameter Settings for STM

All the simulations were run with the following values.

- SwTB.size = 20 ms at PCN-admissible-rate
- SwTB.threshold = 10 ms at PCN-admissible-rate
- MkTB.size = 10 ms at PCN-supportable-rate
- MkTB.threshold = 8 ms at PCN-supportable-rate
- Marking.frequency = 1/3 and 1
- CLE threshold is varied in Table B.1
- EWMA weight is varied in Table B.1
- Number of sequential marked packets for termination = 100
- Number of sequential unmarked packets for stopping termination = 20
- Extra percentage of receiving or sending rate more than the difference between sending and receiving rates for termination at one time = 5.0
- Interval between sending the receiving rate = 33.3 ms
- Interval for measuring sending rate = 100 ms

10.5. Parameter Settings for CL

All the simulations were run with 5 ms at the PCN-supportable-rate as the token bucket size for termination and with the following virtual queue thresholds.

- Min-marking-threshold: 5 ms at the PCN-admissible-rate
- Max-marking-threshold: 15 ms at the PCN-admissible-rate
- Virtual-queue-upper-limit: 20 ms at the PCN-admissible-rate

The virtual-queue-upper-limit puts an upper bound on how much the virtual queue can grow. In the admission control simulation, the CLE threshold and EWMA weight were set as 0.05 and 0.3, respectively. All of the parameters of the virtual queue were the same as those in [I-D.briscoe-tsvwg-cl-phb], [I-D.charny-pcn-single-marking], and [I-D.zhang-pcn-performance-evaluation]. The value of 5 ms at the PCN-supportable-rate as the token bucket size corresponds to 500 packets in the case of CBR, 170 packets in the case of VBR, and 666 packets in the cases of SVD. The token bucket depth was set to 64 packets for CBR, and for on-off traffic, to 128 or higher (in [I-D.briscoe-tsvwg-cl-phb]) and to 256 packets (in [I-D.zhang-pcn-performance-evaluation]).

10.6. Simulation Environment

We used NS2 for our simulation experiments. Simulations were run on a Dell Power Edge 1950, Intel Quad-core Xeon 2.66GHz, 32GB RAM computer running Red Hat Enterprise Linux (64bit).

11. Appendix B: Admission Control Simulation

11.1. Parameter Settings for Admission Control

We evaluated over-admission percentages when the load of the bottleneck was five times the admissible rate, i.e., 2.5 times the link speed. The simulation time was 300 s, and simulation results during the time interval between 200 and 300 s were used for performance metrics to obtain the results in the steady state.

Egress measurement parameters for STM: In our simulations, CLE threshold was varied in Table B.1. The CLE was computed as an exponential weighted moving average (EWMA). The weight was varied in Table B.1. The CLE was computed on a per-packet basis.

Egress measurement parameters for CL: In our simulations, the chosen

CLE threshold was 0.05. The CLE was computed as an exponential weighted moving average (EWMA) with a weight of 0.3. The CLE was computed on a per-packet basis. These values were the same as those in Sect. 8.2.3 in [I-D.charny-pcn-single-marking].

11.2. Sensitivity to EWMA Weight and CLE Threshold

We simulated admission control of STM when the CLE threshold and EWMA weight were the six cases shown in Table B.1. The results of case 1 were shown in the previous draft. We simulated the admission control five times for each case. The smallest and largest over-admission percentage values of the averages of the five simulations are shown in Table B.2. The smallest and largest values of all five links are shown when the topology was PLT(c). The smallest and largest values can be over-admission percentages in a different link. Case 1, whose results were shown in the previous draft, was the worst case of the six in all of the simulations from the viewpoint of absolute deviation from no over- and under-admission. The results of case 6 are shown in the following sections because that case was the best.

Case	CLE threshold	EWMA weight
1	0.05	0.0005
2	0.05	0.01
3	0.05	0.03
4	0.05	0.3
5	0.00001	0.3
6	0.000001	0.3

Table B.1: CLE thresholds and EWMA weights in our simulation

Traffic	Topo.	Smallest	Largest
CBR	S.Link	0.028	0.285
	RTT10	0.585	0.844
	RTT70	-1.656	0.942
	RTT600	-5.218	0.234
	RTT1000	-6.030	-0.179
	PLT(c)	0.005	0.103
VBR	S.Link	0.979	1.300
	RTT10	-0.753	1.562
	RTT70	-3.249	1.462
	RTT600	-5.427	0.699
	RTT1800	-5.870	-0.077
	PLT(c)	0.408	0.752
SVD	S.Link	4.308	4.741
	RTT10	1.593	4.982
	RTT35	-1.981	5.999
	RTT140	-5.502	4.679
	RTT300	-7.192	1.302
	PLT(c)	0.993	2.382

Table B.2:Smallest and largest values of over-admission percentages of all pairs of CLE threshold and EWMA weight

11.3. Basic Evaluation

Over-admission percentage statistics were evaluated using the single link topology. Table B.3 gives results of an admission control

simulation when the load was five times the admissible rate. When the traffic type is CBR, the link speed is 128 Mbps and the load is the rate of 5000 flows. The call interval per ingress-egress aggregate (IEA) is 0.012 sec. When the traffic type is VBR, the link speed is 78.3 Mbps and the load is the rate of 9000 flows. The call interval per IEA is 0.0067 sec. When the traffic type is SVD, the link speed is 2.45 Gbps and the load is the rate of 1500 flows. The call interval per IEA is 0.04 sec. We show the average of the results of five simulations with different random seeds for each traffic type. The performance of STM was very good although it was inferior to that of CL.

Type	Over Admission %	
	STM	CL
CBR	0.065	0.028
VBR	1.249	0.979
SVD	4.678	4.476

Table B.3: Over-admission percentage statistics obtained with single link

11.4. Effect of Ingress-Egress Aggregation

We evaluated the effect of ingress-egress aggregation (IEA) using RTT topology. As with the simulations in [I-D.charny-pcn-single-marking], the aggregate load on the bottleneck was the same across each traffic type with the aggregate load being evenly divided among all ingresses.

11.4.1. CBR

Simulation results with traffic load conditions when the traffic type was CBR are shown in Table B.4. The link speed of the bottleneck was 128 Mbps for all the cases with varying numbers of ingresses. It corresponded to 2000 CBR flows. Thus, the PCN-admissible-rate corresponded to 1000 CBR flows. The load corresponded to 5000 flows. The second column in Table B.4 and the following tables is the expected admitted number of connections per IEA. STM(r) in Table B.4 represents the results of STM with randomized CBR traffic. Each packet of the randomized CBR traffic has an added delay distributed

uniformly from 0 to 50 ms. The results of both CBR traffic and randomized CBR traffic were almost the same, as shown in Table B.4. These simulation results show that the accuracy of STM was little affected by synchronization because synchronization effects are largest in the case of CBR with low IEA [I-D.charny-pcn-single-marking].

No. of ingresses	Expected No. of connections per IE pair	Over Admission (%)		
		STM	STM (r)	CL
10	100	0.750	0.786	0.761
70	14.3	0.890	0.867	0.727
600	1.6	0.122	-0.085	0.038
1000	1	-0.241	-0.433	-0.179

Table B.4: Over admission percentage statistics with CBR, RTT

11.4.2. VBR

Simulation results with traffic load conditions when the traffic type was VBR are shown in Table B.5. The link speed of the bottleneck was 78 Mbps for all the cases with varying numbers of ingresses. This link speed was 1800 times higher than the average rate of a VBR flow. Thus, the PCN-admissible-rate was 900 times higher than the average rate of a VBR flow. The load was 2.5 times higher than the link speed. The result of STM is almost the same as that of CL.

No. of ingresses	Expected No. of connections per IE pair	Over Admission (%) STM	CL
10	180	1.562	1.345
70	25.7	1.334	1.462
600	3	-0.483	0.699
1800	1	-1.822	-0.077

Table B.5: Over-admission percentage statistics with VBR and RTT

11.4.3. SVD

Simulation results with traffic load conditions when the traffic type was SVD are shown in Table B.6. The link speed of the bottleneck was 2448 Mbps for all the cases with varying numbers of ingresses. This link speed was 600 times higher than the average rate of a SVD flow. Thus, the PCN-admissible-rate was 300 times higher than the average rate of a SVD flow. The load was 2.5 times higher than the link speed. Although the result of STM is almost the same as that of CL, CL is slightly superior to STM.

No. of ingresses	Expected No. of connections per IE pair	Over Admission (%) STM	CL
10	30	4.982	4.870
35	8.4	5.261	4.898
140	2	3.595	2.597
300	1	-0.033	0.221

Table B.6: Over-admission percentage statistics with SVD and RTT

11.5. Effect of Multi-bottleneck

We evaluated the effect of a multi-bottleneck with PLT topology. The over- admission-percentage of each bottleneck ID is shown in Table B.7. The bottleneck IDs are named in the order of lowest (upstream) to highest (downstream). The alphabets in Table B.7 shows nodes in Fig. A.3(c). When CBR was used as the traffic type, the link speed of all the bottlenecks was 128 Mbps and the call-interval per IEA was 0.024 s. When VBR was used as the traffic type, the link speed of all the bottlenecks was 78 Mbps and the call-interval per IEA was 0.01 s. When SVD was used as the traffic type, the link speed of all the bottlenecks was 2448 Mbps and the call-interval per IEA was 0.08 s. Both show good performance. In the cases of CBR and VBR, CL is slightly superior to STM. In the case of SVD, STM is slightly superior to CL.

Traffic Type	Algorithm	Bottleneck ID 1(A-B)	2(B-C)	3(C-D)	4(D-E)	5(E-F)
CBR	STM	0.060	0.043	0.038	0.062	0.056
	CL	0.019	0.025	0.015	0.005	0.009
VBR	STM	0.690	0.684	0.708	0.749	0.752
	CL	0.560	0.540	0.585	0.614	0.620
SVD	STM	1.441	1.910	1.348	1.602	1.551
	CL	1.887	1.908	1.952	2.123	2.382

Table B.7: Over admission percentage with multi-bottleneck

11.6. Fairness among Different Ingress-Egress Pairs

In [I-D.charny-pcn-single-marking], fairness is illustrated using the ratio between the bandwidth of the long-haul aggregates and the short-haul aggregates. We used CBR traffic with the load of each bottleneck being five times the PCN-admissible-rate. The fairness results for different topologies are shown in Table B.9. We measured the ratios of the average throughput of short-haul aggregates to that of long-haul aggregates during the simulation time at the first bottleneck. The average of five simulations is displayed. The link speed and call interval conditions in Table B.7 were used in this experiment. Both show good performance.

	Topo	Simulation Time (s)							
		10	20	30	40	50	60	70	80
STM	PLT2	0.98	1.00	1.05	1.11	1.18	1.25	1.32	1.39
CL		1.02	1.04	1.08	1.15	1.22	1.30	1.38	1.46
STM	PLT3	0.98	1.04	1.13	1.24	1.35	1.46	1.57	1.67
CL		0.98	1.00	1.05	1.12	1.21	1.30	1.41	1.52
STM	PLT5	1.02	1.06	1.17	1.31	1.46	1.62	1.79	1.97
CL		1.02	1.05	1.14	1.28	1.43	1.59	1.75	1.93

Table B.8: Fairness performance

12. Appendix C: Flow termination

We evaluated over-termination percentages and the time to terminate the necessary amount of traffic (termination time) when the load of the bottleneck was 1.25 and 2.0 times the PCN-supportable-rate. Over-termination percentages are defined as (PCN-supportable-rate - the rate after termination)/PCN-supportable-rate) expressed as a percentage. The PCN-admissible-rate and PCN-supportable-rate were 20% and 40% of the link speed, respectively, in all the links. The load was lower than the PCN-admissible-rate at the beginning of the simulation. The simulation time was 100 s. The duration of all the flows was infinity. One IE-pair was generated at half the simulation time (50 s). Each flow of the IE-pair arrived in accordance with uniform distribution within the average of the packet interval in a flow. This simulated the change of a route when there was a failure. Over-termination percentages were calculated using the average rate during the time interval between 80 and 100 s. In all the tables, termination time is shown as the time between the first termination and the termination by which the PCN traffic rate is less than PCN-supportable-rate. We detected that the PCN traffic rate was less than PCN-supportable-rate by comparing the result of multiplication of the number of micro flows and the average rate of a micro flow with the PCN-supportable-rate. The CL needs time for measuring SAR and the sending rate and latency one-way time from the PCN-egress-node to PCN-ingress-node. The STM needs time for measuring the sending rate and the same latency as the CL.

When the traffic type was CBR, the link speed was 320 Mbps and the load was the rate of 2500 and 4000 flows. When the traffic type was VBR, the link speed was 109 Mbps and the load was the rate of 2500

and 4000 flows. When the traffic type was SVD, the link speed was 4 Gbps and the load was the rate of 500 and 800 flows. We used randomized CBR, VBR, and SVD in termination simulations. Each packet of the randomized CBR, VBR, and SVD traffic has an added delay distributed uniformly from 0 to 50 ms. We show the average of the results of five simulations with different random seeds for each traffic type.

12.1. Basic evaluation

The results of a termination control simulation with the single link topology are shown in Table C.1. When the load was 1.25 times the PCN-supportable-rate (SR in Table), the accuracy of the proposed control was almost the same as that of the CL control in the cases of VBR and SVD, although the proposed control was inferior to the CL control in the case of CBR. However, when the load was 2.0 times the PCN-supportable-rate, the proposed control was more accurate than the CL control. We show the average of results of five simulations with different random seeds for each traffic type. Termination times in Table C.1 are include measuring time for SAR and the sending rate in CL and measuring time for the sending rate in STM. The acutual termination time is the sum of the value of Table C.1 and latency one-way time from the PCN-egress-node to PCN-ingress-node. Termination time in the following tables also include measuring time for SAR and the sending rate in CL and measuring time for the sending rate in STM.

Traffic Type	Load (x SR)	Over termination(%)		Termination time (s)	
		STM	CL	STM	CL
CBR	1.25	3.260	0.531	0.400	0.201
VBR	1.25	13.201	13.785	0.381	0.250
SVD	1.25	12.367	14.552	0.400	0.707
CBR	2.0	3.461	12.441	1.300	0.224
VBR	2.0	4.910	25.833	1.202	0.243
SVD	2.0	12.527	24.503	1.220	0.659

Table C.1: Over-termination percentage statistics and termination time obtained with single link

12.2. Effect of Ingress-Egress Aggregation

We evaluated the effect of ingress-egress aggregation (IEA) with RTT topology. As with the simulations in [I-D.charny-pcn-single-marking], the aggregate load on the bottleneck was the same across each traffic type, with the aggregate load being evenly divided among all ingresses.

12.2.1. CBR

Simulation results obtained with traffic load conditions when the traffic type was CBR are shown in Table C.2. SR in the second column represents the PCN-supportable-rate. The link speed of the bottleneck was 320 Mbps and the propagation delay between the ingress and egress nodes was 1 ms for all the cases of varying numbers of ingresses. When the load is 1.25 times of PCN-supportable-rate, over termination percentages of STM are larger than those of CL and termination times of STM are slightly larger than those of CL. When the load is 2.0 times of PCN-supportable-rate, over termination percentages of STM are smaller than those of CL and termination times of STM are over one second larger than those of CL.

No. of Ingress	Load (x SR)	Over termination (%)		Termination time (s)	
		STM	CL	STM	CL
2	1.25	6.310	2.810	0.514	0.277
10		6.070	3.270	0.486	0.345
35		7.451	4.170	0.500	0.405
70		6.341	4.780	0.541	0.431
2	2.0	7.211	12.960	1.461	0.298
10		6.560	12.560	1.463	0.392
35		6.171	13.370	1.491	0.451
70		5.851	14.300	1.582	0.400

Table C.2: Over-termination percentage statistics and termination time obtained with CBR and RTT

12.2.2. VBR

Simulation results obtained with traffic load conditions when the traffic type was VBR are shown in Table C.3. The link speed of the bottleneck was 108.8 Mbps and the propagation delay between the ingress and egress nodes was 1 ms for all the cases of varying

numbers of ingresses. When the load is 1.25 times of PCN-supportable-rate, over termination percentages of STM are smaller than those of CL except the case of the number of ingress 10and termination times of STM are slightly larger than those of CL. When the load is 2.0 times of PCN-supportable-rate, over termination percentages of STM are much smaller than those of CL and termination times of STM are over one second larger than those of CL.

No. of Ingress	Load (x SR)	Over termination (%)		Termination time (s)	
		STM	CL	STM	CL
2	1.25	9.696	11.320	0.591	0.316
10		11.226	10.564	0.487	0.368
35		9.971	12.857	0.511	0.407
100		8.112	15.422	0.679	0.379
2	2.0	5.377	20.768	1.480	0.322
10		5.536	22.791	1.461	0.364
35		5.271	22.670	1.418	0.409
100		4.534	25.033	1.644	0.418

Table C.3: Over-termination percentage statistics and termination time obtained with VBR and RTT

12.2.3. SVD

Simulation results obtained with traffic load conditions when the traffic type was SVD are shown in Table C.4. The link speed of the bottleneck was 4.00 Gbps and the propagation delay between the ingress and egress nodes was 1 ms for all the cases of varying numbers of ingresses. When the load is 1.25 times of PCN-supportable-rate, over termination percentages of STM are smaller than those of CL and termination times of STM are over one second larger than those of CL. When the load is 2.0 times of PCN-supportable-rate, over termination percentages of STM are much smaller than those of CL and termination times of STM are over 1.5 second larger than those of CL.

No. of Ingress	Load (x SR)	Over termination (%)		Termination time (s)	
		STM	CL	STM	CL
2	1.25	12.078	15.748	0.469	0.452
10		12.906	16.016	0.340	0.367
35		12.616	20.396	0.274	0.339
2	2.0	12.214	24.970	1.369	0.584
10		11.592	26.763	1.181	0.409
35		13.390	31.583	1.072	0.398

Table C.4: Over-termination percentage statistics and termination time obtained with SVD and RTT

12.3. Effect of Multi-bottleneck

We evaluated the effect of multi-bottlenecks using PLT(c) topology. The over-termination-percentage of each bottleneck ID and termination time of the long-haul IEA are shown in Table C.5. These bottleneck IDs are the same as those in Table B.7. The link speed of each traffic type was the same as that in simulations of the effect of IEA. The propagation delay between the ingress and egress nodes was 1 ms. Over termination percentages of STM are almost the same among all the bottleneck links although those of CL worsen as the bottleneck IDs decrease (upstream). CL is affected by accumulation of marking. STM is not affected by it because all the packets are marked in termination.

The results of termination time show in Table C.6. When the load is 1.25 times of PCN-supportable-rate, termination times of STM are almost the same as those of CL. When the load is 2.0 times of PCN-supportable-rate, termination times of STM are over one second larger than those of CL.

Load (x SR)	Traffic Type	Alg.	Over termination (%) Bottleneck ID 1(A-B)	2(B-C)	3(C-D)	4(D-E)	5(E-F)
1.25	CBR	STM	4.491	4.551	4.321	4.311	4.370
		CL	6.351	2.742	1.341	0.882	0.672
	VBR	STM	15.443	15.893	15.729	15.134	15.741
		CL	17.321	13.343	12.459	11.201	11.127
	SVD	STM	15.941	15.489	15.219	15.877	15.229
		CL	18.467	15.535	14.487	13.950	14.536
2.0	CBR	STM	4.371	4.552	4.301	4.241	4.741
		CL	20.291	12.601	9.571	8.050	7.550
	VBR	STM	8.941	6.447	6.619	6.486	6.332
		CL	27.015	14.756	14.408	14.510	14.242
	SVD	STM	18.019	15.547	16.355	16.334	15.983
		CL	27.115	15.151	14.097	14.263	13.472

Table C.5: Over-termination percentage statistics and termination time obtained with multi-bottleneck

Load (x SR)	Traffic Type	Algorithm	Termination time (s)
1.25	CBR	STM	0.402
		CL	0.212
	VBR	STM	0.402
		CL	0.226
	SVD	STM	0.408
		CL	0.430
2.0	CBR	STM	1.302
		CL	0.209
	VBR	STM	1.221
		CL	0.581
	SVD	STM	1.325
		CL	0.459

Table C.6: Termination time obtained with multi-bottleneck

13. Appendix D: Why is TBthreshold.shallow.threshold set to be smaller than the token bucket size by the bit-size of a metered packet?

The origin of the threshold marking is a virtual queue [I-D.briscoe-tsvwg-cl-phb]. Therefore queueing theory is applied to marking. Little's formulas are one of the most powerful relationship for G/G/1 in queueing theory where G/G/1 is described using Kendall's notation. Both arrival process and service time distribution of G/G/1 is described as a general distribution and 1 represents the number of servers. Furthermore, Little's formulas are valid for any queue discipline such as (first-in, first-out) FIFO, (Last-in, first-out) LIFO, and (service in random order) SIRO. TBthreshold.shallow.threshold is introduced to apply the Little's formulas to marking.

At first, we explain Little's formulas in queueing theory [see Gross and Harris]. Little's formulas are

$$L = \lambda W \qquad (D.1)$$

and

$$L_q = \lambda W_q, \qquad (D.2)$$

where L represents the mean number of customers in the system and L_q represents the expected number of customers in queue, lambda represents the average rate of customers entering the queueing system, W_q represents the expected time a customer spends waiting in the queue prior to entering service, and W represents the expected total time a customer spends in the queueing system. In queueing theory, the term customer is used in a general sense and does not imply necessarily a human customer.

From eqs. (D.1) and (D.2), the following equation is derived,

$$L - L_q = \lambda(W-W_q) = \lambda/\mu = \rho \qquad (D.3)$$

where mu represents the mean service rate, that is, 1/mu is the mean service time and rho is traffic intensity. The left-hand side of eq.(D.3) is the expected number of customers in service in the steady state. Furthermore,

$$L = 1*p_1 + 2*p_2 + 3*p_3 + ... \qquad (D.4)$$

and

$$L_q = 0*p_1 + 1*p_2 + 2*p_3 + ... \qquad (D.5)$$

where p_n represents the probability that the number of customers in the system is n in the steady state. From eqs. (D.4) and (D.5),

$$L - L_q = p_1 + p_2 + p_3 + ... = 1 - p_0 = p_b \qquad (D.5)$$

where p_b is the probability that the server is busy in the steady state.

If we get p_b, we get the traffic intensity. Marking by using TBthreshold.shallow.threshold is a way to get p_b. The marking makes each packet marked when the previous packet is in service. Each packet is not marked only when no packet is in a virtual queue at the packet's arrived time. The marking ratio gives an approximation of p_b. Both are different generally because the time average and the customer average are different, and p_b is the time average when the server is busy and the marking ratio is the customer average. However, both are the same when the arrival proces is a Poisson

process. This property is called PASTA, which is abbreviation that Poisson arrivals see time averages. Furthermore, if an arrival process is described as a renewal process, an infinite superposition of them is the Poisson process {Cox]. The renewal process is a large class of stochastic processes. The inter-arrival time is independent and identically distributed (IID). Therefore the marking ratio by using TBthreshold.shallow.threshold is expected to give a good approximation of the traffic intensity.

14. Appendix E: Admission control using fractional marking

We show another option of operation at PCN-interior-node. The relation between PCN traffic rate and marking ratio is shown in Fig. E.1, where AR and SR represent PCN-admissible- and PCN-supportable-rates. If the PCN traffic rate is less than AR, no traffic is marked. If the PCN traffic rate is between AR and SR, 1/N of the traffic is marked. If the PCN traffic rate is greater than SR, all the traffic is marked.

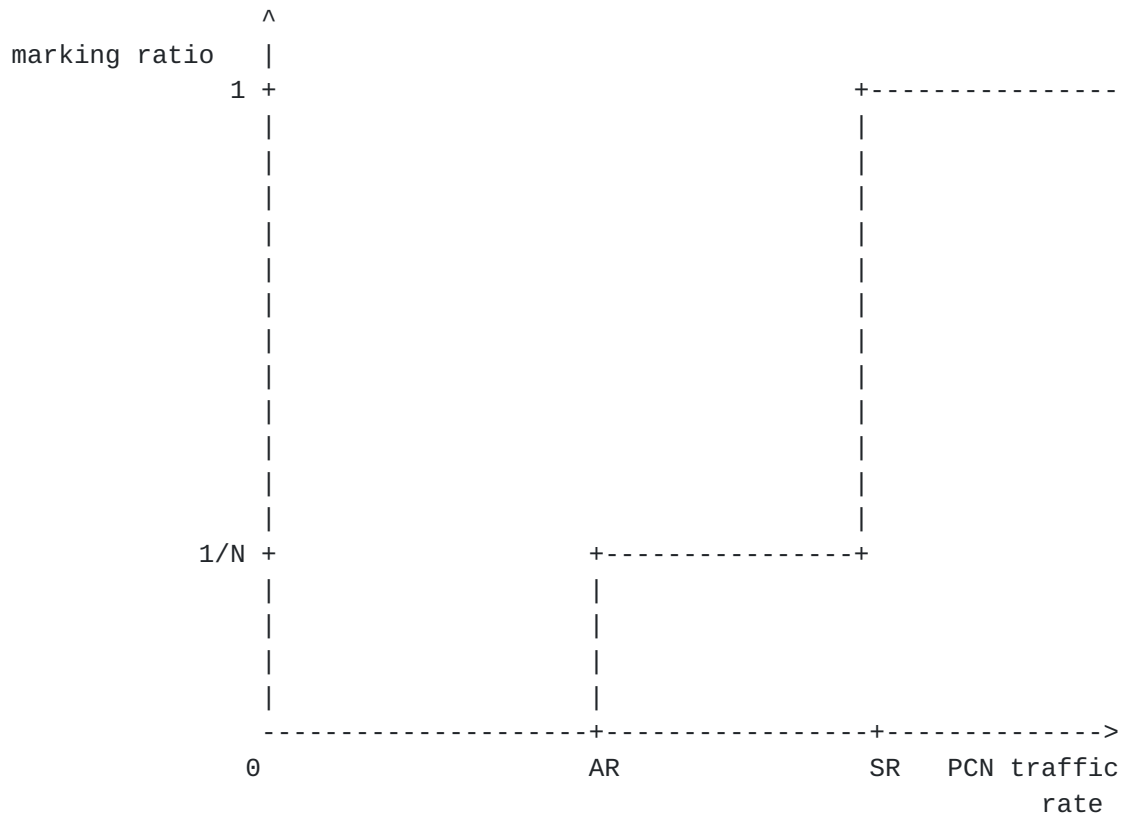

Figure E.1: Relation between PCN traffic rate and marking ratio

To achieve this marking behavior, we use two token buckets for a

modified threshold marking called fractional marking [MichaelPCNSurvey] and the threshold marking. The fractional marking marks 1/N of the PCN traffic when the PCN traffic rate is greater than a configured bit rate although the threshold marking marks all PCN packets if the PCN traffic rate is greater than a configured bit rate. The fractional metering and marking is applied to one token bucket whose tokens are added at the PCN-admissible-rate. The threshold metering and marking is applied to the other token bucket whose tokens are added at the PCN-supportable-rate. Tokens in both buckets are removed equal to the size in bits of the metered-packet. The additions and removals in one token bucket are independent of the other token bucket.

The marking explained above is expressed by the pseudo code of the following algorithm, where FrTB.fill and ThTB.fill represent the fill states of the token bucket for the fractional marking and threshold marking, respectively. FrTB.size and ThTB.size (TBthreshold.max in [I-D.pcn-marking-behaviour]) represent the token bucket sizes and packet.size represents the size of the measured packet. lastUpdate represents the time when both token buckets were last updated and now represents the current time. FrTB.threshold and ThTB.threshold represent configured thresholds. Cnt is a byte counter for marking 1/N of the measured traffic. M represents the marked state, and packet.mark represents the state of the mark of a packet. A two-step function, as shown in Fig. E.1 is achieved by this marking algorithm. This algorithm enables us to distinguish AR- and SR- pre-congestion states using two encoding states.

```
Input: pcn packet
//add tokens to the two token buckets
FrTB.fill = min(FrTB.size, FrTB.fill + FrTB.rate*(now - lastUpdate));
ThTB.fill = min(ThTB.size, ThTB.fill + ThTB.rate*(now - lastUpdate));
lastUpdate = now;
IF(FrTB.fill < FrTB.threshold) THEN
 IF(Cnt < 0) THEN
   packet.mark = M;
   Cnt=Cnt+N*packet.size;
 ENDIF
   Cnt=Cnt-packet.size;
ENDIF
IF(ThTB.fill < ThTB.threshold)
   packet.mark = M;
ENDIF
//remove tokens from each token bucket
FrTB.fill = max(0, FrTB.fill - packet.size);
ThTB.fill = max(0, ThTB.fill - packet.size);
Output: void
```

15. IANA Considerations

TBD

16. Security Considerations

TBD

17. Informative References

[Cox] Cox, D., "Renewal Theory", 1962.

[Gross and Harris]
Gross, D. and C. Harris, "Fundamentals of Queueing Theory", 1998.

[I-D.babiarz-pcn-3sm]
Babiarz, J., Liu, X-G., Chan, K., and M. Menth, "Three State PCN Marking", November 2007.

[I-D.briscoe-tsvwg-cl-phb]
Briscoe, B., Eardley, P., Songhurst, D., Le Faucheur, F., Charny, A., Liatsos, V., Babiarz, J., Chan, K., Dudley, S., Karagiannis, G., Bader, A., and L. Westberg, "Pre-Congestion Notification Marking", October 2006.

[I-D.charny-pcn-single-marking]
Charny, A., Zhang, J., Le Faucheur, F., and V. Liatsos, "Pre-Congestion Notification Using Single Marking for Admission and Termination", November 2007.

[I-D.ietf-pcn-architecture]
Eardley, P., "Pre-Congestion Notification Architecture", January 2009.

[I-D.ietf-pcn-baseline-encoding]
Moncaster, T., Briscoe, B., and M. Menth, "Baseline Encoding and Transport of Pre-Congestion Information", Feburary 2009.

[I-D.ietf-pcn-marking-behaviour]
Eardley, P., "Marking behaviour of PCN-nodes", October 2008.

[I-D.westberg-pcn-load-control]
Westberg, L., "LC-PCN: The Load Control PCN Solution",

November 2008.

[I-D.zhang-pcn-performance-evaluation]
Zhang, J., Charny, A., Liatsos, V., and F. Le Faucheur, "Performance Evaluation of CL-PHB Admission and Termination Algorithms", July 2007.

[MichaelPCNSurvey]
Menth, M., Lehrieder, F., Briscoe, B., Eardley, P., Moncaster, T., Babiarz, J., Charny, A., Zhang, J., Taylor, T., Chan, K., Satoh, D., Geib, R., and G. Karagiannis, "A Survey of PCN-Based Admission Control and Flow Termination", 2010.

Authors' Addresses

Daisuke Satoh
NTT Advanced Technology Corporation
1-19-18, Nakacho
Musashino-shi, Tokyo 180-0006
Japan

Email: daisuke.satoh@ntt-at.co.jp

Harutaka Ueno
NTT Advanced Technology Corporation

Email: harutaka.ueno@ntt-at.co.jp

Yukari Maeda
NTT Advanced Technology Corporation

Email: yukari.maeda@ntt-at.co.jp

Oratai Phanachet
NTT Advanced Technology Corporation

Email: oratai.phanachet@ntt-at.co.jp