Performance evaluation of termination in CL-algorithm
draft-satoh-pcn-performance-termination-00

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on January 4, 2010.

Copyright Notice

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents in effect on the date of publication of this document (http://trustee.ietf.org/license-info). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Abstract

Pre-congestion notification (PCN) gives information to support
admission control and flow termination in order to protect the quality of service (QoS) of inelastic flows. [I-D.taylor-cl-edge-behaviour] describes one boundary node behaviours for three-state measurement-based load control, known informally as CL [I-D.briscoe- tsvwg-cl-phb]. In [I-D.taylor-cl-edge-behaviour], flow termination is required if excess-traffic-marked packets were observed and the end of one measurement period MUST be the beginning of the next one, independently of current flow conditions. According to this termination, PCN-flows in some ingress-egress (IE) pairs may be terminated during measurement period of other IE pairs unless round-trip times (RTT) of all the IE pairs are the same. We illustrate that this can lead to over-termination. Our simulation confirms that accuracy of termination is improved when no PCN-flows in some IE pairs are terminated during measurement period of other IE pairs.

Table of Contents

1. Introduction .................................................. 3
2. Terminology ................................................... 3
3. Unintended termination ....................................... 3
   3.1. Example 1 ................................................. 4
   3.2. Example 2 ................................................. 7
   3.3. Example 3 ................................................. 7
4. Acknowledgements ............................................... 7
5. Appendix: Simulation evaluation .............................. 7
   5.1. Network topology and rerouting ......................... 8
   5.2. Traffic .................................................. 8
   5.3. Parameter setting for CL ................................ 9
   5.4. Simulation results ...................................... 9
6. IANA Considerations ........................................... 10
7. Security Considerations ....................................... 11
8. Informative References ....................................... 11
Authors’ Addresses .............................................. 11
1. Introduction

Pre-congestion notification (PCN) gives information to support admission control and flow termination in order to protect the quality of service (QoS) of inelastic flows. Flow termination is a new control whose function is terminating already admitted PCN-flows. Termination is necessary even if admission control is provided because rerouting by failures makes PCN traffic rate higher than PCN-supportable-rate.

Menth and Lehrieder evaluate the performance of measured rate termination [I-D.menth-pcn-performance]. In [I-D.menth-pcn-performance], they pointed out that indirect measured rate termination can lead to substantial over-termination when packets are metered and marked before dropping. Furthermore, they also pointed out that over-termination can occur without dropping of packets when the measurement periods of the ingress and the egress are not in synchronization, that is, both measurement intervals do not cover the same data.

In this memo, we illustrate that over-termination also occurs when during a measurement period of an IE the load of another IE using the same bottleneck link is reduced due to termination. This situation occurs when IEs sharing a common bottleneck have significantly different round trip times (RTT) within the PCN domain.

2. Terminology

The terminology used in this document conforms to the terminology of [RFC5559], [I-D.ietf-pcn-marking-behaviour], [I-D.taylor-cl-edge-behaviour], [I-D.briscoe-tsvwg-cl-phb], and [I-D.menth-pcn-performance].

3. Unintended termination

[I-D.taylor-cl-edge-behaviour] describes flow termination is required if excess-traffic-marked packets were observed. It also describes the end of one measurement period MUST be the beginning of the next one, independently of current flow conditions.

If a bottleneck link that contains IE pairs whose RTTs are different, unintended termination can occur under the condition described above. We illustrate it by using a simple network model.

We assume a network of Figure 1 after rerouting due to failure. Three ingress nodes: A, B, and C are connected to interior node D.
with links of different propagation delay and interior node D is connected to egress node F. We assume propagation delay between nodes B and D is the same as that between nodes C and D, and they are different from that between nodes A and D to simplify the model. We consider the case of bottleneck link between nodes D and F. We call IE pair A-F as IE#1, IE pair B-F as IE#2, and IE pair C-F as IE#3. We illustrate that unintended termination can occur by using the three examples in Table 1. RTT of IE#1 is smaller than those of IE#2 and IE#3 in Example 1. RTT of IE#1 is larger than those of IE#2 and IE#3 in Example 2. In Example 3, rates of IE pairs in Example 3 are twice higher than those in Examples 2 and 3 and RTTs in Example 3 are the same as those in Example 1. The PCN-supportable rate (SR) is 160Mbps in all the examples. Thus, termination is necessary. The link speed of the bottleneck is much more than 480 Mbps. Thus no dropping occurs. Furthermore, we assume the ratio of unmarked rate to total rate of an IE pair is that of SR to the total rate of the link. The duration of the measurement period is 100 ms, but it does not influence the unintended termination directly.

A
\ 
B - D - F
/ 
C

Figure 1: A network in which IE pairs have different RTTs.

<table>
<thead>
<tr>
<th>IE pair</th>
<th>Ingress</th>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate</td>
<td>RTT</td>
<td>Rate</td>
<td>RTT</td>
</tr>
<tr>
<td></td>
<td>[Mbps]</td>
<td>[ms]</td>
<td>[Mbps]</td>
<td>[ms]</td>
</tr>
<tr>
<td>IE#1</td>
<td>A</td>
<td>360</td>
<td>100</td>
<td>360</td>
</tr>
<tr>
<td>IE#2</td>
<td>B</td>
<td>60</td>
<td>150</td>
<td>60</td>
</tr>
<tr>
<td>IE#3</td>
<td>C</td>
<td>60</td>
<td>150</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 1: Rates and RTTs of IE pairs.

3.1. Example 1

As seen in Figure 2, marked and unmarked rates of IE#1 just after failure are 240Mbps and 120Mbps, those of IE#2 are 40Mbps and 20Mbps, and those of IE#3 are 40Mbps and 20Mbps. Sustainable aggregate rate
(SAR) of IE#1 is 120Mbps, that of IE#2 is 20Mbps, and that of IE#3 is 20Mbps. The sum of the three SAR is 160Mbps, which is equal to SR.

PCN traffic rate

```
+-----------------------------------------------+                                               |
|             IE#3(mark):   40Mbps              |                                               |
|-----------------------------------------------|                                               |
|             IE#3(unmark): 20Mbps               |                                               |
+-----------------------------------------------+                                               |
|             IE#2(mark):   40Mbps              |                                               |
|-----------------------------------------------|                                               |
|             IE#2(unmark): 20Mbps               |                                               |
+-----------------------------------------------+                                               |
|                                               |                                               |
|             IE#1(mark):   240Mbps             |                                               |
+-----------------------------------------------+                                               |
|                                               |                                               |
|             IE#1(unmark): 120Mbps             |                                               |
+-----------------------------------------------+                                               |
```

Figure 2: Marked and unmarked rate of IE pairs just after failure.

We assume that the next measurement period at the PCN egress node begins after the termination of IE#1 but before the termination of IE#2 and IE#3. This is likely since these IEs have significantly larger RTTs. Figure 2 shows the measured rates of marked and unmarked traffic in this measurement period. The total rate is 240Mbps, marked and unmarked rates of IE#1 are 40Mbps and 80Mbps,
those of IE#2 are 20Mbps and 40Mbps, and those of IE#3 are 20Mbps and 40Mbps in the first half of the measurement period. The total rate is 160Mbps, unmarked rate of IE#1 is 120Mbps, that of IE#2 is 20Mbps, and that of IE#3 is 20Mbps in the second half of the measurement period. The total rate of the three IE pairs is 160Mbps, which equals SR. If the measurement period begin with the half of the period, no termination is necessary. However, egresses of IE#1, IE#2, and IE#3 measure SAR as 100Mbps (= (80+120)/2), 30Mbps (= (40+20)/2), 30Mbps (= (40+20)/2). The sum of the SAR equals SR.

Figure 3: Marked and unmarked rate of IE pairs after termination of IE#1 and before termination IE#2 and IE#3 at the beginning of the measurement period.

Figure 3 illustrates that the second half of the measurement period shows that sufficiently many flows are already terminated. However, SAR is calculated over the entire measurement period and, therefore, it is underestimated. Thus, the PCN ingress node of IE#1 again terminates flows because the reported SAR is 100Mbps which is smaller than the existing 120 Mbps PCN traffic rate in IE#1. The ingresses of IE#2 and IE#3 receive SAR=30 Mbps which is 10 Mbps more than their existing PCN traffic rates of 20 Mbps. Figure 4 shows the situation in the next measurement interval. Then, over-termination of 20Mbps is visible because IE#1 has terminated another 20 Mbps by then.
3.2. Example 2

Flows of IE#1, which has the highest rate, are terminated at first and flows of IE#2 and IE#3 are terminated in Example 1. By using example 2, we show that over-termination also occurs and the amount of unintended termination is changed when the flows of an IE pair whose rate is not highest rate are terminated first. According to the similar discussion in Example 1, the amount of unintended termination is 12Mbps although that in Example is 20Mbps.

3.3. Example 3

Example 3 has the twice higher load than Example 1. According to the similar discussion in Example 1, the amount of unintended termination is 33.35Mbps although that in Example 1 is 20Mbps.

4. Acknowledgements

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

5. Appendix: Simulation evaluation

The unintended termination illustrated in the previous section can be avoided if SAR is measured only after all IEs have finished their termination and the effect of their termination steps have become visible at the PCN egress node. We evaluated how much unintended termination can be avoidable by simulation.
5.1. Network topology and rerouting

We simulated flow termination with CL for the simple topology given in Figure 1 when the number of ingresses were 3, 11, 36, 71, and 101. One IE pair had the highest rate and others had the same rate. The total rate without the IE pair whose rate was the highest was the same as PCN-admissible-rate. The IE pair was rerouted by failure. Hence, the load was higher than SR at the bottleneck link. We simulated the load of 1.25SR and 2.0SR with CBR and VBR traffic.

All links had different propagation delays which were chosen randomly in the range of 1ms - 100 ms. The bottleneck link D-F was modeled with a 10ms propagation delay in all simulations. Therefore, the range of round-trip delays in the experiments ranged from 22ms to 220ms.

5.2. Traffic

We used the same types of traffic as those of [I-D. briscoe-tsvwg-cl-phb], [I-D. charny-pcn-single-marking], [I-D. zhang-pcn-performance-evaluation]. These were CBR and on/off process (VBR) that was described with two state Markov chain and whose on and off periods were exponentially distributed with the specified mean. The distribution of flow duration of all the flows was infinity in order to eliminate other causes to decrease the number of flows than termination.

Traffic parameters for each type are summarized below:

CBR voice

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

On-off traffic approximating voice with silence compression

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

5.3. Parameter setting for CL

All the simulations were run with 5ms at PCN-supportable-rate as token bucket size for Termination. The value of 5ms at PCN-supportable-rate as the token bucket size corresponds to 500 packets in the case of CBR and VBR.

5.4. Simulation results

We evaluated over-termination percentages to terminate the necessary amount of traffic when the load of the bottleneck was 1.25 and 2.0 times SR. Over-termination percentages is defined as (SR - the rate after termination)/SR) expressed in percentage. SR was 40% of the link speed in the link D-F. The load was lower than PCN-admissible-rate at the beginning of the simulation. The simulation time was 100s. One IE-pair was generated at half the simulation time (50s). Each flow of the IE-pair arrived according to 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 100s. During this period, no termination occurred.

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. We used randomized CBR and VBR in termination simulations. Each packet of the randomized CBR and VBR traffic was added a delay distributed uniformly from 0 to 50ms. We show the average of the results of five simulations with different random seeds for each traffic type. In Tables 2 and 3, we show over-termination percentages in the case of no inter-measurement period time(IMPT) and the case of 350ms as IMPT. In this simulation, no termination of some IE pairs occurred during measurement period of other IE pairs because the maximum of RTT was 220ms and measuring time at an ingress was 100ms and the sum of them was less than IMPT 350ms.

In the case of no IMPT in Table 2, over-termination percentages in the case of 2.0SR load is higher than those in the case of 1.25SR load. The more bandwidth was terminated during measurement period, the more over-termination was observed as shown in the previous section. On the other hand, in the case of 350ms as IMPT, over-termination percentages were the same in both loads.
### Table 2: Over-termination percentage statistics with CBR.

The same phenomena as the case of CBR were observed in the case of VBR as seen in Table 3. However, the phenomena in the case of CBR were more clearly observed than those in the case of VBR.

<table>
<thead>
<tr>
<th>No. of Ingress</th>
<th>Load (x SR)</th>
<th>Over-termination %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IMPT=None</td>
<td>IMPT=350ms</td>
</tr>
<tr>
<td>3</td>
<td>3.611</td>
<td>0.121</td>
</tr>
<tr>
<td>11</td>
<td>4.070</td>
<td>0.400</td>
</tr>
<tr>
<td>36</td>
<td>5.740</td>
<td>1.250</td>
</tr>
<tr>
<td>71</td>
<td>7.570</td>
<td>2.380</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of Ingress</th>
<th>Load (x SR)</th>
<th>Over-termination %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IMPT=None</td>
<td>IMPT=350ms</td>
</tr>
<tr>
<td>3</td>
<td>8.910</td>
<td>0.119</td>
</tr>
<tr>
<td>11</td>
<td>9.670</td>
<td>0.371</td>
</tr>
<tr>
<td>36</td>
<td>11.080</td>
<td>1.150</td>
</tr>
<tr>
<td>71</td>
<td>12.570</td>
<td>2.271</td>
</tr>
</tbody>
</table>

### Table 3: Over-termination percentage statistics with VBR.

<table>
<thead>
<tr>
<th>No. of Ingress</th>
<th>Load (x SR)</th>
<th>Over-termination %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IMPT=None</td>
<td>IMPT=350ms</td>
</tr>
<tr>
<td>3</td>
<td>8.873</td>
<td>5.398</td>
</tr>
<tr>
<td>11</td>
<td>8.891</td>
<td>5.582</td>
</tr>
<tr>
<td>36</td>
<td>12.952</td>
<td>5.457</td>
</tr>
<tr>
<td>101</td>
<td>18.733</td>
<td>6.704</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of Ingress</th>
<th>Load (x SR)</th>
<th>Over-termination %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IMPT=None</td>
<td>IMPT=350ms</td>
</tr>
<tr>
<td>3</td>
<td>10.195</td>
<td>5.617</td>
</tr>
<tr>
<td>11</td>
<td>12.373</td>
<td>5.516</td>
</tr>
<tr>
<td>36</td>
<td>15.935</td>
<td>5.267</td>
</tr>
<tr>
<td>101</td>
<td>17.796</td>
<td>4.923</td>
</tr>
</tbody>
</table>

6. IANA Considerations

TBD
7.  Security Considerations

    TBD

8.  Informative References

[I-D.briscoe-tsvwg-cl-phb]

[I-D.charny-pcn-single-marking]

[I-D.ietf-pcn-marking-behaviour]

[I-D.menth-pcn-performance]

[I-D.zhang-pcn-performance-evaluation]


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
1-19-18, Nakacho
Musashino-shi, Tokyo  180-0006
Japan

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

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

Email: menth@informatik.uni-wuerzburg.de