Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: December 15, 2011

G. Karagiannis
University of Twente
T. Taylor
K. Chan
Huawei Technologies
M. Menth
University of Tuebingen
P. Eardley
BT

June 15, 2011

Requirements for Signaling of (Pre-) Congestion Information in a DiffServ Domain
draft-ietf-pcn-signaling-requirements-05

Abstract

Precongestion notification (PCN) is a means for protecting quality of service for inelastic traffic admitted to a Diffserv domain. The overall PCN architecture is described in RFC 5559. This memo describes the requirements for the signaling applied within the PCN domain: (1) PCN-feedback-information is carried from the PCN-egress-node to the decision point;(2) the decision point may ask the PCN-ingress-node to measure, and report back, the rate of sent PCN-traffic between that PCN-ingress-node and PCN-egress-node. The decision point may be either collocated with the PCN-ingress-node or a centralized node (in the latter case, (2) is not required). The signaling requirements pertain in particular to two edge behaviours, "controlled load (CL)" and "single marking (SM)" [draft-ietf-pcn-cl-edge- behaviour-08], [draft-ietf-pcn-sm-edge-behaviour-05].

Status of this Memo

Copyright Notice

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Table of Contents

1. Introduction

The main objective of Pre-Congestion Notification (PCN) is to support the quality of service (QoS) of inelastic flows within a Diffserv domain in a simple, scalable, and robust fashion. Two mechanisms are used: admission control and flow termination. Admission control is used to decide whether to admit or block a new flow request while flow termination is used in abnormal circumstances to decide whether to terminate some of the existing flows. To support these two features, the overall rate of PCN-traffic is metered on every link in the domain, and PCN-packets are appropriately marked when certain configured rates are exceeded. These configured rates are below the rate of the link thus providing notification to boundary nodes about overloads before any congestion occurs (hence "pre-congestion" notification). The PCN-egress-nodes measure the rates of differently marked PCN traffic in periodic intervals and report these rates to the decision points for admission control and flow termination, based on which they take their decisions. The decision points may be collocated with the PCN-ingress-nodes or their function may be implemented in a centralized node.
For more details see [RFC5559],
[draft-ietf-pcn-cl-edge-behaviour-08],
[draft-ietf-pcn-sm-edge-behaviour-05].

This memo specifies the requirements on signaling protocols:
o to carry reports from a PCN-egress-node to the decision point,
o to carry requests, from the decision point to a PCN-ingress-node,

that trigger the PCN-ingress-node to measure the PCN-sent-rate,
- to carry reports, from a PCN-ingress-node to the decision point.

The latter two messages are only needed if the decision point and PCN-ingress-node are not collocated.

2. Signaling Requirements for Messages from the PCN-Egress-Nodes to Decision Point(s)

The PCN-egress-node measures- per ingress-egress-aggregate the rates of differently marked PCN-traffic in regular intervals. The measurement intervals are recommended to take a fixed value between 100 ms and 500 ms, see [draft-ietf-pcn-cl-edge-behaviour-08], [draft-ietf-pcn-sm-edge-behaviour-05]. At the end of each measurement interval, the PCN-egress-node calculates the congestion-level-estimate (CLE) based on these quantities. The PCN-egress-node MAY be configured to record a set of identifiers of PCN-flows for which it received excess-traffic-marked packets during the last measurement interval. The latter may be useful to perform flow termination in networks with multipath routing.

At the end of each measurement interval, or less frequently if "optional report suppression" is activated, see

[draft-ietf-pcn-cl-edge-behaviour-08], [draft-ietf-pcn-sm-edge-behaviour-05], the PCN-egress-node sends a report to the decision point.

For the SM edge behaviour, the report MUST contain:
- the identifier of the PCN-ingress-node and the identifier of the PCN-egress-node (typically their IP addresses); together they specify the ingress-egress-aggregate to which the report refers,
- the rate of not-marked PCN-traffic (NM-rate) in octets/second,
- rate of PCN-marked traffic in octets/second,
- the congestion-level-estimate, which is a number between zero and one.

For the CL edge behaviour, the report MUST contain:
- the identifier of the PCN-ingress-node and the identifier of the PCN-egress-node (typically their IP addresses); together they specify the ingress-egress-aggregate to which the report refers,

o the rate of threshold-marked PCN traffic (ThM-rate) in octets/second,
o rate of excess-traffic-marked traffic (ETM-rate) in octets/second,
o the congestion-level-estimate, which is a number between zero and one.

The number format and the rate units used by the signalling protocol will limit the maximum rate that PCN can use. If signalling space is tight it might be reasonable to impose a limit, but any such limit may impose unnecessary constraints in future.

For both CL and SM edge behaviours, the report MAY also contain:
o a set of PCN-flow identifiers (see Section 2.1).

The signaling report can either be sent directly to the decision point or it can "piggy-back", i.e., be included within some other message that passes through the PCN-egress-node and then reaches the decision point.

Signaling messages SHOULD have a higher priority than data packets to deliver them quickly and to avoid that they are dropped in case of overload.

The load generated by the signaling protocol SHOULD be minimized. We give three examples that may help to achieve that goal:
o piggy-backing the reports by the PCN-egress-nodes to the decision point(s) onto other signaling messages that are already in place,
o reducing the amount of reports to be sent by optional report suppression,
o combining reports for different ingress-egress-aggregates in a single message (if they are for the same decision point).

As PCN reports are sent regularly, additional reliability mechanisms are not needed. This also holds in the presence of optional report suppression as reports are sent periodically if actions by the

decision point(s) are needed, see [draft-ietf-pcn-cl-edge-behaviour--08], [draft-ietf-pcn-sm-edge-behaviour-05].

2.1 Specification of PCN-Flow Identifiers

The representation of a PCN-flow identifier depends on the surrounding environment, e.g., pure IP, MPLS, GMPLS, etc. Examples of such PCN-flow identifier representations can be found in [RFC2205], [RFC3175] [RFC3209], [RFC3473], [RFC4804].

In pure IP networks, the identifier may consist of a subset of the following information:

- source IP address;
- destination IP address
- protocol identifier and higher layer (port) addressing
- flow label (typical for IPv6)
- SPI field for IPsec encapsulated data
- DSCP/TOS field

Note, where a PCN-flow consists of a collection of microflows, then the PCN-flow is identified by the PCN-ingress-node's and PCN-egress-node's identifiers (typically their IP addresses), which are already part of the report.

3. Signaling Requirements for Messages between Decision Point(s) and PCN-Ingress-Nodes

Through request-response signaling between the decision point and PCN-ingress-node, the decision point requests and in response the PCN-ingress-node measures and reports the PCN-sent-rate for a specific ingress-egress-aggregate. Signaling is needed only if the decision point and PCN-ingress-node are not collocated.

The request MUST contain:

- the identifier of the PCN-ingress-node and the identifier of the PCN-egress-node; together they determine the ingress-egress-aggregate for which the PCN-sent-rate is requested,
- the identifier of the decision point that requests the PCN-sent-rate.

The report MUST contain:

- the PCN-sent-rate in octets/second,
- the identifier of the PCN-ingress-node and the identifier of the PCN-egress-node.

The request MUST be addressed to the PCN-ingress-node, and the report MUST be addressed to the decision point that requested it.

The request and the report SHOULD be sent with high priority and reliably, because they are sent only when flow termination is needed, which is an urgent action.

4. Security Considerations

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

5. IANA Considerations

This memo includes no request to IANA.

6. Acknowledgements

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

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

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

[draft-ietf-pcn-cl-edge-behaviour-08] T. Taylor, A, Charny, F. Huang, G. Karagiannis, M. Menth, "PCN Boundary Node Behaviour for the Controlled Load (CL) Mode of Operation (Work in progress)", December 2010.

[draft-ietf-pcn-sm-edge-behaviour-05] A. Charny, J. Zhang, G. Karagiannis, M. Menth, T. Taylor, "PCN Boundary Node Behaviour for the Single Marking (SM) Mode of Operation (Work in progress)", December 2010.

7.2. Informative References

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

[RFC3175] Baker, F., Iturralde, C. Le Faucher, F., Davie, B., "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, 2001.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

[RFC4804] F. Le Faucheur, "Aggregation of Resource ReSerVation Protocol (RSVP) Reservations over MPLS TE/DS-TE Tunnels", RFC 4804, February 2007.

Authors' Addresses

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

Tom Taylor
Huawei Technologies
1852 Lorraine Ave.
Ottawa, Ontario K1H 6Z8
Canada
Phone: +1 613 680 2675
Email: tom111.taylor@bell.net

Kwok Ho Chan
Huawei Technologies

125 Nagog Park
Acton, MA 01720
USA
Email: khchan@huawei.com

Michael Menth
University of Tuebingen
Department of Computer Science
Chair of Communication Networks
Sand 13
72076 Tuebingen
Germany
Phone: +49 7071 29 70505
Email: menth@informatik.uni-tuebingen.de

Philip Eardley
BT
B54/77, Sirius House Adastral Park Martlesham Heath
Ipswich, Suffolk IP5 3RE
United Kingdom
EMail: philip.eardley@bt.com