Internet Engineering Task Force (IETF)
Request for Comments: 6663
Category: Informational
ISSN: 2070-1721

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

July 2012

Requirements for Signaling of Pre-Congestion Information in a Diffserv Domain

Abstract

Pre-Congestion 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 first case, (2) is not required). The signaling requirements pertain in particular to two edge behaviors, Controlled Load (CL) and Single Marking (SM).

Table of Contents

1. Introduction 2
1.1. Requirements Language 3
2. Signaling Requirements for Messages from the PCN-Egress-Nodes to Decision Point(s) 3
3. Signaling Requirements for Messages between Decision Point(s) and PCN-Ingress-Nodes 5
4. Security Considerations 5
5. Acknowledgments 6
6. References 6
6.1. Normative References 6
6.2. Informative References 6

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; the Decision Points use these rates to make 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], [RFC6661], and [RFC6662].

This memo specifies the requirements on signaling protocols:
- to carry reports from a PCN-egress-node to the Decision Point,
- 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.

1.1. 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].

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 [RFC6661] and [RFC6662]. 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 [RFC6661] and [RFC6662]), the PCN-egress-node sends a report to the Decision Point.

For the SM edge behavior, 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, and
- the rate of PCN-marked traffic (PM-rate) in octets/second.

For the CL edge behavior, 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,
- the rate of threshold-marked PCN traffic (ThM-rate) in octets/second, and
- the rate of excess-traffic-marked traffic (ETM-rate) in octets/second.

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

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.

As described in [RFC6661], PCN reports from the PCN-egress-node to the Decision Point may contain flow identifiers for individual flows within an ingress-egress-aggregate that have recently experienced excess-marking. Hence, the PCN report messages used by the PCN CL edge behavior MUST be capable of carrying sequences of octet strings constituting such identifiers.

Signaling messages SHOULD have a higher priority and a lower drop precedence than PCN-packets (see [RFC5559]) in order to deliver them quickly and to prevent them from being dropped in case of overload.

The load generated by the signaling protocol SHOULD be minimized. We give three methods that may help to achieve that goal:

1. piggy-backing the reports by the PCN-egress-nodes to the Decision Point(s) onto other signaling messages that are already in place,
2. reducing the amount of reports to be sent by optional report suppression, or
3. 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 [RFC6661] and [RFC6662].

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, and
- the identifier of the Decision Point that requests the PCN-sent-rate.

The report MUST contain:
- the PCN-sent-rate in octets/second, and
- 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.

Because they are sent only when flow termination is needed (which is an urgent action), the request and the report SHOULD be sent with high priority, with a lower drop precedence than PCN-packets, and in a reliable manner.

Note that a complete system description for a PCN-domain with centralized Decision Point includes the signaling from Decision Point to the PCN-ingress-nodes to control flow admission and termination. However, this is a known problem (with solutions provided in [RFC3084] and [RFC5431], for example), and it lies outside the scope of the present document.

4. Security Considerations

[RFC5559] provides a general description of the security considerations for PCN. This memo relies on the security-related requirements of the PCN signaling, provided in [RFC5559]. In particular, the signaling between the PCN-boundary-nodes must be protected from attacks. For example, the recipient needs to validate that the message is indeed from the node that claims to have sent it. Possible measures include digest authentication and protection against replay and man-in-the-middle attacks.

Specifically for the generic aggregate RSVP protocol, additional protection methods against security attacks are described in [RFC4860].

5. Acknowledgments

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

6. References

6.1. Normative References

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

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

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

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

6.2. Informative References

[RFC3084] Chan, K., Seligson, J., Durham, D., Gai, S., McCloghrie, K., Herzog, S., Reichmeyer, F., Yavatkar, R., and A. Smith, "COPS Usage for Policy Provisioning (COPS-PR)", RFC 3084, March 2001.

[RFC4860] Le Faucheur, F., Davie, B., Bose, P., Christou, C., and M. Davenport, "Generic Aggregate Resource ReSerVation Protocol (RSVP) Reservations", RFC 4860, May 2007.

[RFC5431] Sun, D., "Diameter ITU-T Rw Policy Enforcement Interface Application", RFC 5431, March 2009.

Authors' Addresses

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

Tom Taylor
Huawei Technologies
Ottawa
Canada
EMail: tom.taylor.stds@gmail.com

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

Michael Menth
University of Tuebingen
Sand 13
72076 Tuebingen
Germany
Phone: +49-7071-2970505
EMail: menth@uni-tuebingen.de

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