Internet Engineering Task Force   G. Karagiannis
Internet-Draft   University of Twente
Intended status: Informational   T. Taylor
Expires: January 5, 2011   K. Chan
                                                  Huawei Technologies
                                                           M. Menth
                                             University of Wurzburg
                                                       July 5, 2010

**Requirements for Signaling of (Pre-) Congestion Information in a
DiffServ Domain
draft-ietf-pcn-signaling-requirements-00**

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: PCN feedback is carried from the PCN-egress-node to the
   decision point and the decision point may demand for the measurement
   and delivery of the PCN rate sent at the PCN-ingress-node. The
   decision point may be either collocated with the PCN-ingress-node or
   a centralized node.

Status of this Memo

This Internet-Draft will expire on January 5, 2011.

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 as so-called PCN feedback 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-06],
   [draft-ietf-pcn-sm-edge-behaviour-03].

   Thus, signaling is needed to transport PCN feedback from PCN-egress-
   nodes towards the decision point. Moreover, signaling is needed that
   the decision point can trigger the PCN-ingress-node to measure the
   PCN traffic rate and send these measurement results to the decision
   point.

   This memo briefly describes the signaled content and specifies the
   requirements that have to be satisfied by the signaling protocols.

1.1. Terminology

   In addition to the terms defined in [RFC5559], this document uses the
   following terms:

   Decision Point:

      The node that makes the decision about which flows to admit and to
      terminate. In a given network deployment, this may be the ingress
      node or a centralized control node. Of course, regardless of the
      location of the decision point, the ingress node is the point
      where the decisions are enforced.

   PCN egress feedback:

      Content used by the PCN-egress-node to report and inform the
      decision point about measurements required during flow
      admission and flow termination decisions.

   PCN ingress feedback:

      Content used by the PCN-ingress-node to report and inform the
      decision point about measurements required during flow termination
      decisions.

   ingress rate request:

      A message sent by the decision point towards the PCN-ingress-node
      to request the PCN-ingress-node to measure and report the value of
      the rate of admitted PCN traffic for a given
      ingress-egress-aggregate.

   Congestion level estimate (CLE)
      A value derived from the measurement of PCN packets calculated at a
      PCN-egress-node for a given ingress-egress-aggregate, representing
      the ratio of marked to total PCN traffic (measured in octets) over
      a short period. For further details see
      [draft-ietf-pcn-cl-edge-behaviour-06] and [draft-ietf-pcn-sm-edge-
      behaviour-03].

2. Signaling requirements between PCN-egress-nodes and Decision Point

   The PCN-egress-node measures the rates of differently marked PCN
   traffic in regular intervals and signals them as PCN egress feedback
   to the decision point.
   This section describes the PCN egress feedback and the requirements
   that apply to signaling protocols used for the transport of PCN
   feedback from PCN-egress-nodes to decision points.
   Note that if the decision point and the PCN-ingress-node are
   collocated, then the signaling requirements described in this section
   apply to the signaling between PCN-egress-nodes and PCN-ingress-
   nodes.

2.1 PCN Reporting Frequency

   The specification of PCN-based admission control and flow termination
   in [draft-ietf-pcn-cl-edge-behaviour-06], [draft-ietf-pcn-sm-edge-
   behaviour-03] suggest measurement and reporting intervals at the PCN-
   egress-nodes of 100 to 500 ms. The PCN reporting frequency can provide

some level of reliability. Therefore, it is considered that for regularly reported information, additional reliability mechanisms are not needed, see Section 2.3.6. The following PCN contents are sent regularly: rate of not-marked PCN traffic, rate of threshold-marked PCN traffic, rate of excess-traffic-marked PCN traffic, CLE.

2.2 Signaled PCN egress Feedback

The PCN-egress-node measures per ingress-egress-aggregate the
following rates
   o rate of not-marked PCN traffic;
   o rate of threshold-marked PCN traffic, which applies to CL edge
     behaviour only;
   o rate of excess-traffic-marked PCN traffic.
   o Congestion level estimate (CLE)

The rate values are reported in octets/second to the decision point each
time that the PCN-egress-node calculates them and when this is supported
via configuration. CLE is only reported to the decision point when this
is supported via configuration.
For more details see [draft-ietf-pcn-cl-edge-behaviour-06], [draft-ietf-
pcn-sm-edge-behaviour-03].

If multipath routing is enabled, the PCN-egress-node tracks a list of
flows for which it has recently received excess-traffic-marked
packets. The list of these flow IDs is included in the PCN feedback
because these flows are candidates for termination.
The representation of a flow ID depends on the surrounding
environment, e.g., "pure IP", MPLS, GMPLS, etc. Examples of such flow ID
representations can be found in [RFC2205], [RFC3175] [RFC3209],
[RFC3473]. The list SHOULD be a concatenation of flow IDs associated with
the flows that are candidates for termination. The format of a list
containing flow ID_1 to flow ID_n SHOULD be:
list flow IDs = <flow ID_1> <flow ID_2> ... <flow_ID_n>.

2.3 Signaling requirements

This section describes the requirements for signaling protocols that
are used to carry the PCN egress feedback from PCN-egress-nodes to
the decision point.

2.3.1 Priority of signaling messages

Signaling messages SHOULD have a higher priority than data packets.
This is needed to avoid as much as possible the situations that
during severe overload cases the signaling messages are dropped
within the PCN domain.

2.3.2 Local information exchange

Signaling messages MUST be able to carry the PCN egress feedback from
the PCN-egress-node to the decision point.

[2.3.3](#) Carry identification of PCN edge nodes

The signaling protocol MUST be able to carry identification
(address information) of the PCN edge nodes. This is required due to the
fact that the decision point needs to be able to associate the received
signaling message with the PCN edge node that sent this message.

However, the identification of the PCN edge nodes
MUST NOT be visible to non-PCN nodes outside the PCN domain.

2.3.4 Carry identification of ingress-egress-aggregates

The signaling protocol MUST be able to carry identification
(address information) of the ingress-egress-aggregates. It is
proposed to identify them using the addresses of the PCN-ingress-node
and PCN-egress-node between which they pass. If each of the edge
nodes do not have unique addresses, then other identifiers could be
used.

2.3.5 Signaling load

The load generated by the signaling protocol to carry the PCN egress
Feedback from the PCN-egress-nodes to the decision point SHOULD be
minimized as much as possible.

2.3.6 Reliability

There are situations that messages need to be received in a
reliable way. There are different ways of achieving reliability. The
solution of achieving this reliability is out of the scope of this
document. However, it is considered that when information is received on
a regular fashion, additional reliability measures are not
required. The list with flow IDs associated with the excess-traffic-
marked flows is not sent regularly, hence SHOULD be sent reliably.

2.3.7 Security

The signaling support may need security protection against replay
attacks. The security services to be supported are:
 o) Message authentication and integrity: an attacker could cause denial
    of service using impersonation. Moreover, an attacker could cause a
    denial of service by modifying message contents. Therefore, message
    authentication and integrity SHOULD be supported.
 o) Message confidentiality: There could be situations where the PCN
    signaling messages should not be visible to non authorised nodes. In
    such cases, PCN message confidentiality MAY be supported.

2.4. Filter specifications

In PCN the ingress and egress nodes should be able to identify the
ingress-egress-aggregate to which each flow belongs. Moreover, the
egress node also needs to associate an aggregate with the address of
the ingress node for receiving reports, if the ingress node is the

decision point. The filter specification at the PCN-egress-nodes
depends on the surrounding environment, e.g., pure IP, MPLS, GMPLS.

In this document, a possible IP filter spec for pure IP is given as an example. In this case the filter spec should be able to identify a flow using (all or a subset of the) following information:

o source IP address;

o destination IP address;

o protocol identifier and higher layer (port) addressing;

o flow label (typical for IPv6);

o SPI field for IPsec encapsulated data;

o DSCP/TOS field.

o IP address of PCN-ingress-node

o IP address of PCN-egress-node

3. **Signaling Requirements between Decision Point and PCN-ingress-nodes**

The decision point monitors and uses the PCN egress feedback sent by the PCN-egress-node. There are situations that the decision point must obtain an estimate of the rate at which PCN-traffic is being admitted to the aggregate from the PCN-ingress-node.
In order to receive this information the decision point has to request from the PCN-ingress-node to send the value of the PCN traffic admitted to a certain aggregate.
Note that if the decision point and the PCN-ingress-node are collocated, then the information exchanges between the decision point and PCN-ingress-node are internal operations.

3.1 **Signaled PCN ingress Feedback**

The PCN-ingress-node measures per ingress-egress-aggregate the following rate
   o rate of admitted PCN traffic

This value is reported in octets/second to the decision point as soon as possible after receiving the request from the decision point. .

3.2 **Signaled decision point trigger**

The decision point uses the "ingress rate request" to request from the PCN-ingress-node to send for a certain ingress-egress-aggregate, the value of the admitted PCN traffic rate. The "ingress rate

request" message identifies the ingress-egress-aggregate for which
the admitted PCN traffic rate is required.

3.3 Signaling requirements

The same signaling requirements described in Section 2.3 apply for
this situation.

The only difference is the fact that these signaling
requirements apply for the signaling messages that have to be sent
between the decision point and PCN-ingress-nodes. Moreover, since the
"ingress rate request" message sent by the decision point towards
the PCN-ingress-node and the admitted PCN traffic rate sent by the
PCN-ingress-node towards the decision point are not sent regularly,
they SHOULD be delivered reliably.

4. Security Considerations

[RFC5559] provides a general description of the security
considerations for PCN. This memo introduces the additional security
considerations described in Section 2.3.7.

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-06] 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)", June 2010.

[draft-ietf-pcn-sm-edge-behaviour-03] 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)", June 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.

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 Wurzburg
   Institute of Computer Science
   Room B206
   Am Hubland, Wuerzburg D-97074
   Germany
   Email: menth@informatik.uni-wuerzburg.de