Internet Engineering Task Force   F. Huang, Ed.
Internet-Draft   T. Taylor
Intended status: Standards Track   Huawei Technologies
Expires: January 14, 2011   G. Zorn
                                                           Network Zen
                                                         H. Tschofenig
                                                 Nokia Siemens Networks
                                                         July 13, 2010

        **Diameter Application To Transfer PCN Data From Edge Nodes To a
                      Centralized Decision Point
                  draft-huang-dime-pcn-collection-03**

Abstract

   Pre-congestion notification (PCN) is a technique for maintaining QoS
   for inelastic flows in a DIFFServ domain. The PCN architecture
   requires that egress nodes send regular reports of PCN-defined
   measurements to a decision point. It requires further that the
   decision point occasionally be able to request certain measurements
   from ingress nodes. The decision point can be located in different
   places in the network. This memo defines a Diameter application to
   support communications between the ingress and egress nodes and a
   Diameter server acting as a PCN decision point.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

   The objective of Pre-Congestion Notification (PCN) is to protect the
   quality of service (QoS) of inelastic flows within a Diffserv domain
   [RFC2475] in a simple, scalable and robust fashion. Admission
   control allows decisions on whether to admit or reject a new flow
   request, and (in abnormal circumstances, such as router failure) to
   provide flow termination of already admitted flows. These two
   mechanisms together aim to protect the QoS properties of previously
   admitted flows. To achieve this, the overall rate of the PCN traffic
   is metered on every link in the PCN 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 before any congestion occurs ("pre-congestion
   notification"). The level of marking allows decisions to be made
   about whether to admit or terminate. A more detailed description of
   the architecture can be found in [RFC5559].

   Marking statistics are gathered by egress nodes on a per-ingress-
   egress aggregate basis. If multipath routing is in use, egress nodes
   may also send a list of individual flows along with the marking
   statistics, if these flows have experienced an elevated level of pre-
   congestion.

   The reported statistics are processed to determine whether new flows
   can be admitted to the aggregate over the next measurement interval
   and whether some flows should be terminated to protect QoS for the
   remainder. (Flow termination is expected to be relatively
   infrequent, typically a result of network failure.) The admission
   state is based on a congestion level estimate (CLE), which the
   decision node derives from the statistics that the egress node passes
   to the decision point. The decision to terminate flows is made on
   the basis of a different criterion, also derived from those
   statistics. For further details see [I-D.SM-edge-behaviour] and
   [I-D.CL-edge-behaviour].

2. 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].

3. Procedures

   The following subsections discuss the processing requirements placed
   upon the various participating Diameter nodes by the PCN Data

Collection application.

3.1. Egress Node Behaviour

The egress node measures the traffic from a particular ingress node,
and calculates either two or three flow rates (octets per second) at
the ingress-egress-aggregate level, based on the PCN markings of the
packets it receives and the edge node behaviour that has been
deployed. The three rates are the PCN-unmarked rate (NM-Rate), the
PCN-threshold-marked rate (ThM-Rate), and the PCN-excess-traffic-
marked rate (ETM-Rate). The ThM-Rate is collected for the CL edge
node behaviour [I-D.CL-edge-behaviour], but not for the SM edge node
behaviour [I-D.SM-edge-behaviour]. The egress node calculates these
rates and reports them to the decision point every 100 to 500 ms.

The CL and SM edge node behaviour specifications also provide an
option for the egress node to calculate a congestion level estimate
(CLE), equal to the ratio of PCN-marked to total PCN traffic received
in the interval. The egress node reports the calculated CLE to the
decision point along with the basic traffic rates.

The CL and SM edge node behaviour specifications provide a further
option for report suppression when the calculated CLE is below a
reporting threshold for two or more successive reporting periods. It
is RECOMMENDED that the operator configure egress nodes to activate
the CLE reporting and report suppression options, to minimize the
reporting traffic volume in the network and the the processing load
at the decision point.

For the CL mode only, when multipath routing is in effect, the egress
node may be configured to collect identifiers of flows that
experienced PCN-excess-traffic-marking during the measurement
interval along with the calculated traffic rates.

The egress node MUST use a Congestion Report Request (CRR) command to
send the calculated rates and the flow identifiers, when applicable,
to the decision node. The Event-Timestamp AVP MUST be present within
the CRR, and MUST indicate the ending time of the latest measurement
period represented by the information within the message. At least
one instance of the Aggregate-PCN-Egress-Data AVP MUST be present.
Multiple instances of this AVP MAY be present, but each instance MUST
report on a different ingress-egress-aggregate.

3.2. Decision Point Behaviour

If the decision point receives an Congestion-Report-Request (CRR)
identified as belonging to the PCN Data Collection application, it
MUST acknowledge the message with a Congestion-Report-Answer (CRA).

The decision point usage of the information provided by PCN-Congestion-Info and PCN-Excess- Flow-Info AVPs is described in the applicable edge behavior specification ([I-D.SM-edge-behaviour] or [I-D.CL-edge-behaviour].

The decision point MAY send a Measurement-Poll-Request (MPR) to the ingress node for a specific ingress-egress-aggregate for which flow termination appears to be required. The decision point will make the decision about sending the MPR depending on the content of the CRR relating to the ingress-egress- aggregate concerned. The decision point SHOULD copy the Event-Timestamp AVP it received in the CRR to the MPR.

If the decision point receives a successful Measurement-Poll-Answer message, it uses the information contained in the PCN-Sent-Info AVP to determine how much traffic to terminate, as described in [I-D.SM-edge-behaviour] or [I-D.CL-edge-behaviour].

3.3. Ingress Node Behaviour

When an ingress node receives an MPR, it MUST generate a Measurement-Poll-Answer message containing an instance of the PCN-Sent-Info AVP. The I-E-Aggregate-Id within the PCN-Sent-Info AVP MUST be the same as received in the MPR, and the I-E-Aggregate-Sent-Rate MUST be a rate measured for that aggregate. If Event-Timestamp is present in the MPR, the measurement upon which I-E-Aggregate-Sent-Rate is based SHOULD be that for a period after the time given by Event-Timestamp. In any case, Event- Timestamp MUST be present in the MPA, and if it is, MUST give the end-time of the measurement period upon which I-E-Aggregate-Sent-Rate is based.

4. Diameter PCN Data Collection Application

4.1. Advertising Application Support

Clients, servers, and proxies supporting the PCN Data Collection application MUST advertise support by including the value <AID> in the Auth-Application-Id of Congestion-Report-Request (CRR), Congestion-Report-Answer (CRA), Measurement-Poll-Request (MPR), and Measurement-Poll-Answer (MPA) messages.

4.2. Session Management

Diameter sessions for this application are implicitly terminated. An implicitly terminated session is one for which the server does not maintain session state information. The client does not need to send any re- authorization or session termination requests to the server.

The Diameter base protocol includes the Auth-Session-State AVP as the
mechanism for the implementation of implicitly terminated sessions.

The client (server) SHALL include in its requests (responses) the
Auth-Session-State AVP set to the value NO_STATE_MAINTAINED (1), as
described in RFC 3588 [RFC3588]. As a consequence, the server does
not maintain any state information about this session and the client
does not need to send any session termination request. Neither the
Authorization-Lifetime AVP nor the Session-Timeout AVP SHALL be
present in requests or responses.

4.3. Commands

The PCN Data Collection application defines four new commands,
Congestion-Report-Request(CRR), Congestion-Report-Answer(CRA),
Measurement-Poll-Request (MPR) and Measurement-Poll-Answer (MPA).

4.3.1. Congestion-Report-Request (CRR) Command

The egress node sends the Congestion-Report-Request (CRR) command,
indicated by the Command-Code field set to <CC1> and the Command
Flags' 'R' bit set, to report PCN marking statistics and possibly
individual flow identifiers. Multiple reports for different ingress-
egress-aggregates MAY be included in the same message, as described
in Section 3.1.

Message format:

```
  <CRR> ::= < Diameter Header: CC1, REQ, PXY >
            < Session-Id >
            { Auth-Application-Id }
            { Auth-Session-State }
            { Origin-Host }
            { Origin-Realm }
            { Destination-Realm }
          * ( Aggregate-PCN-Egress-Data )
            ( Event-Timestamp ]
            [ Destination-Host ]
          * [ Proxy-Info ]
          * [ Route-Record ]
          * [ AVP ]
```

At least one instance of the Aggregate-PCN-Egress-Data AVP MUST be
present. The value of the Session-Id AVP MUST be unique and SHOULD
be set according to the recommendations in Section 8.8 of RFC 3588
[RFC3588].

4.3.2. Congestion-Report-Answer (CRA) Command

   The decision point uses the Congestion-Report-Answer (CRA) command,
   indicated by the Command-Code field set to <CC1> and the Command
   Flags' 'R' bit cleared, to acknowledge a Congestion-Report-Request
   command sent by an egress node. The Congestion-Report-Answer command
   contains the same Session-Id as the corresponding request. The
   Event-Timestamp AVP MUST be present and MUST contain the value
   received in the CRR that is being acknowledged.

   Message format:

```
     <CRA> ::= < Diameter Header: CC1, PXY >
               < Session-Id >
               { Auth-Application-Id }
               { Auth-Session-State }
               { Result-Code }
               { Origin-Host }
               { Origin-Realm }
               ( Event-Timestamp )
               [ Error-Message ]
               [ Error-Reporting-Host ]
               [ Failed-AVP ]
             * [ Proxy-Info ]
             * [ AVP ]
```

4.3.3. Measurement-Poll-Request (MPR) Command

   The decision point sends the Measurement-Poll-Request (MPR) command,
   indicated by the Command-Code field set to <CC2> and the Command
   Flags' 'R' bit set, to request that an ingress node report the rate
   at which PCN- marked traffic has been forwarded to a given ingress-
   egress aggregate, measured over a time period constrained as
   described in Section 3.3. The value of the Session-Id AVP MUST be
   unique and SHOULD be set according to the recommendations in Section
   8.8 of RFC 3588 [RFC3588]. The I-E-Aggregate-Id AVP MUST be present.
   The Event-Timestamp AVP SHOULD be present.

   Message format:

```
   <MPR> ::= < Diameter Header: CC2, REQ, PXY >
             < Session-Id >
             { Auth-Application-Id }
             { Auth-Session-State }
             { Origin-Host }
             { Origin-Realm }
             { Destination-Realm }
             { I-E-Aggregate-Id }
             [ Destination-Host ]
             [ Event-Timestamp ]
           * [ Proxy-Info ]
           * [ Route-Record ]
           * [ AVP ]
```

4.3.4. Measurement-Poll-Answer (MPA) Command

The ingress node sends the Measurement-Poll-Answer (MPA) command,
indicated by the Command-Code field set to <CC2> and the Command
Flags' 'R' bit cleared, in response to an MPR sent by the decision
point. The Session-Id MUST be copied from the MPR to which the MPA
is responding. The PCN-Sent-Info and Event-Timestamp AVPs MUST be
present.

Message format:

```
   <MPA> ::= < Diameter Header: CC2, PXY >
             < Session-Id >
             { Auth-Application-Id }
             { Auth-Session-State }
             { Result-Code }
             { Origin-Host }
             { Origin-Realm }
             { PCN-Sent-Info }
             ( Event-Timestamp )
             [ Error-Message ]
             [ Error-Reporting-Host ]
             [ Failed-AVP ]
           * [ Proxy-Info ]
           * [ AVP ]
```

4.4. Attribute Value Pairs (AVPs)

This section describes the AVPs specific to the PCN Data Collection
application. The 'M' bit MUST be set and the 'V' bit MUST NOT be set
for all of these AVPs when used in the PCN Data Collection
application.

4.4.1. I-E-Aggregate-Id AVP

The I-E-Aggregate-Id AVP (AVP code <AVP1>) is of type Grouped and
identifies a specific aggregate.

The I-E-Aggregate-Id AVP has the following format:

EDITOR'S NOTE: may need to add something that works for MPLS.

```
   I-E-Aggregate-Id ::= < AVP Header: AVP1 >
                         [ PCN-Ingress-Node-Address ]
                         [ PCN-Egress-Node-Address ]
                         [ VLAN-ID-Range ]
                       * [ AVP ]
```

4.4.2. PCN-Ingress-Node-Address AVP

The PCN-Ingress-Node-Address AVP (AVP code <AVP2>) is of type Grouped
and contains the address of a PCN-Ingress-Node.

The PCN-Ingress-Node-Address AVP has the following format:

```
   PCN-Ingress-Node-Address ::= < AVP Header: AVP2 >
                                 [ Framed-IP-Address ]
                                 [ Framed-IPv6-Prefix ]
```

4.4.3. PCN-Egress-Node-Address AVP

The PCN-Egress-Node-Address AVP (AVP code <AVP3>) is of type Grouped
and contains the address of a PCN-Egress-Node.

The PCN-Egress-Node-Address AVP has the following format:

```
   PCN-Egress-Node-Address ::= < AVP Header: AVP3 >
                                [ Framed-IP-Address ]
                                [ Framed-IPv6-Prefix ]
```

4.4.4. Framed-IP-Address AVP

The Framed-IP-Address AVP is defined in the NASREQ application (RFC
4005 [RFC4005]).

4.4.5. Framed-IPv6-prefix AVP

The Framed-IPv6-prefix AVP is defined in the NASREQ application (RFC
4005 [RFC4005]).

[4.4.6](#). **VLAN-ID-Range AVP**

   The VLAN-ID-Range AVP, defined in ietf-dime-qos-attributes [I-D.ietf-
   dime-qos-attributes], is of type Grouped and specifies the VLAN range
   to match.

[4.4.7](#). **Aggregate-PCN-Egress-Data AVP**

   The PCN-Congestion-Info AVP (AVP code <AVP4>) is of type Grouped. It
   identifies an ingress-egress aggregate, reports the current value of
   the PCN-unmarked, PCN-excess-traffic-marked, and optionally, PCN-
   threshold- marked traffic rates and the CLE, and MAY identify zero or
   more flows experiencing excess-traffic-marking.

   The PCN-Congestion-Info AVP has the following format:

```
   Aggregate-PCN-Egress-Data ::= < AVP Header: AVP4 >
                                 { I-E-Aggregate-Id }
                                 { NM-Rate }
                                 { ETM-Rate }
                                 [ ThM-Rate ]
                                 [ CLE-Value ]
                               * [ Classifier ]
                               * [ AVP ]
```

[4.4.8](#). **NM-Rate AVP**

   The NM-Rate AVP (AVP code <AVP5>) is of type Unsigned32. It gives
   the calculated rate of receipt of PCN-unmarked traffic in octets per
   second for a given ingress-egress-aggregate.

[4.4.9](#). **ETM-Rate AVP**

   The ETM-Rate AVP (AVP code <AVP6>) is of type Unsigned32. It gives
   the calculated rate of receipt of excess-traffic-marked traffic in
   octets per second for a given ingress-egress-aggregate.

[4.4.10](#). **ThM-Rate AVP**

   The ThM-Rate AVP (AVP code <AVP7>) is of type Unsigned32. It gives
   the calculated rate of receipt of threshold-marked traffic in octets
   per second for a given ingress-egress-aggregate.

[4.4.11](#). **CLE-Value AVP**

   The CLE-Value AVP (AVP code <AVP8>) is of type Unsigned32. It gives
   the calculated ratio of traffic rates of PCN-marked traffic to total
   PCN traffic received at the egress node, multiplied by 1000 and

truncated, for a given ingress-egress-aggregate. By construction,
the value of the CLE-Value AVP ranges from 0 to 1000.

4.4.12. Classifier AVP

The Classifier AVP (AVP Code 511) is a grouped AVP that consists of a
set of attributes that specify how to match a packet. The Classifier
AVP is defined in [RFC5777]. In the present specification its
purpose is to identify a flow that is experiencing excess-traffic-
marking.

4.4.13. PCN-Sent-Info AVP

The PCN-Sent-Info AVP (AVP code <AVP9>) is of type Grouped. It
provides the estimated rate of flow of PCN traffic in octets per
second that the ingress node has admitted to a given ingress-egress-
aggregate.

The PCN-Sent-Info AVP has the following format:

```
    PCN-Sent-Info ::= < AVP Header: AVP9 >
                        { I-E-Aggregate-Id }
                        { I-E-Aggregate-Sent-Rate }
                    * [ AVP ]
```

4.4.14. I-E-Aggregate-Sent-Rate AVP

The I-E-Aggregate-Sent-Rate AVP (AVP code <AVP10>) is of type
Unsigned32. It gives the estimated rate of flow of PCN traffic in
octets per second that the ingress node forwarded to the identified
ingress-egress aggregate, calculated for the measurement period
ending at the time given by the Event-Timestamp AVP.

4.5. AVP Occurrence Tables

The following table presents the AVPs defined in this document and
their occurrences in Diameter messages. Note that AVPs that can only
be present within a Grouped AVP are not represented in this table.

The table uses the following symbols:

0: The AVP MUST NOT be present in the message.

1: One instance of the AVP MUST be present in the message.

1+: One or more instances of the AVP MUST be present in the message.


```
                                +-----------------------+
                                |     Command-Code      |
                                |-----+-----+-----+-----+
      AVP Name                  | CRR | CRA | MPR | MPA |
      --------------------------|-----+-----+-----+-----+
      Aggregate-PCN-Egress-Data | 1+  |  0  |  0  |  0  |
      PCN-Sent-Info             |  0  |  0  |  0  |  1  |
      I-E-Aggregate-Id          | (*) |  0  |  1  | (*) |
                                +-----+-----+-----+-----+
```


(*): Note that the I-E-Aggregate-Id AVP appears alone in the MPR
command and is contained within the Aggregate-PCN-Egress-Data grouped
AVP (CRR command) and the PCN-Sent-Info grouped AVP (MPA command).

5. IANA Considerations

Upon publication of this memo as an RFC, IANA is requested to assign
values as described in the following sections.

5.1. Diameter Application Identifier

An application identifier for Diameter PCN Data Collection (<AID>,
Section 4.1) must be assigned according to the policy specified in
Section 11.3 of [RFC3588].

5.2. Diameter Command Codes

Codes must be assigned for the following commands according to the
policy specified in [RFC3588], Section 11.2.1:

o Congestion-Report-Request (CRR) and Congestion-Report-Answer (CRA)
   (<CC1>, Section 4.3.1 and Section 4.3.2).

o Measurement-Poll-Request (MPR) and Measurement-Poll-Answer (MPA)
   (<CC4>, Section 4.3.3 and Section 4.3.4).

5.3. Attribute-Value Pairs

Codes must be assigned for the following AVPs using the policy
specified in [RFC3588], Section 11.1.1:

o I-E-Aggregate-Id (<AVP1>, [Section 4.4.1](#))

o PCN-Ingress-Node-Address (<AVP2>, [Section 4.4.2](#))

o PCN-Egress-Node-Address (<AVP3>, [Section 4.4.3](#))

o Aggregate-PCN-Egress-Data (<AVP4>, [Section 4.4.7](#))

o NM-Rate (<AVP5>, [Section 4.4.8](#))

o ETM-Rate (<AVP6>, [Section 4.4.9](#))

o ThM-Rate (<AVP7>, [Section 4.4.10](#))

o CLE-Value (<AVP8>, [Section 4.4.11](#))

o PCN-Sent-Info (<AVP9>, [Section 4.4.13](#))

o I-E-Aggregate-Sent-Rate (<AVP10>, [Section 4.4.14](#)).

[6](#). Security Considerations

The following sections discuss the security threats against the
Diameter PCN Data Collection application and describe some
countermeasures.

[6.1](#). Traffic Security

Application traffic MUST be secured as specified in [RFC 3588](#)
[[RFC3588](#)] (i.e., through the use of (preferably) TLS or IPsec). In
the absence of appropriate protection, all manner (including man-in-
the-middle) of attacks are possible, potentially resulting in the
inappropriate termination and non-admittance of flows.

[6.2](#). Device Security

Compromise of an ingress node by an attacker could result in the
inappropriate refusal of admittance to valid flows, while the
compromise of an egress node could allow the termination of valid
flows.

Compromise of the decision point could result in both denial of
admission to new flows and termination of existing flows, enabling an
attacker to essentially control PCN traffic on the affected network.

[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.

   [RFC3588] Calhoun, P., Loughney, J., Guttman, E., Zorn, G., and J.
              Arkko, "Diameter Base Protocol", RFC 3588, September 2003.

   [RFC4005] Calhoun, P., Zorn, G., Spence, D., and D. Mitton,
              "Diameter Network Access Server Application", RFC 4005,
              August 2005.

   [RFC5777] Korhonen, J., Tschofenig, H., Arumaithurai, M., Jones, M.,
              and A. Lior, "Traffic Classification and Quality of
              Service (QoS) Attributes for Diameter", RFC 5777,
              February 2010.

7.2. Informative References

   [I-D.CL-edge-behaviour]
              Charny, A., Huang, F., Karagiannis, G., Menth, M., and T.
              Taylor, "PCN Boundary Node Behaviour for the Controlled
              Load (CL) Mode of Operation (Work In progress)",
              June 2010.

   [I-D.SM-edge-behaviour]
              Charny, A., Zhang, J., Karagiannis, G., Menth, M., and T.
              Taylor, "PCN Boundary Node Behaviour for the Single
              Marking (SM) Mode of Operation (Work in progress)",
              June 2010.

   [Q.3303.3]
              ITU-T, "Resource control protocol No. 3 -- Protocols at
              the Rw interface between a policy decision physical entity
              (PD-PE) and a policy enforcement physical entity (PE-PE):
              Diameter", ITU-T Recommendation Q.3303.3, May 2008.

              <http://www.itu.int/rec/T-REC-Q.3303.3>

   [RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z.,
              and W. Weiss, "An Architecture for Differentiated
              Services", RFC 2475, December 1998.

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

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

Appendix A. Appendix A. Related Work in ITU-T

   The ITU-T is doing work to exploit the PCN technology in an
   environment where the decisions are made by a central policy decision
   point (decision point) [Q.3303.3], which needs the information
   generated by PCN marking to support per-flow decisions on admission
   and termination. This memo defines a Diameter application to
   transfer the information from edge nodes to the decision point.
   Egress node reports are sent by the egress node acting as client to
   the decision point acting as server. Data generated at the ingress
   node are needed only when flow termination is required. They are
   requested by the decision point acting as client and sent in
   responses by the ingress node acting as server. The decision point
   thus acts both as client and as server in the same application. The
   Rw application [RFC5431] provides a precedent for such an
   application.

   The PCN Data Collection application is related to existing ITU-T
   applications as follows:

   o The Rs application allows application-level functions to request
      flow admission for individual application flows.

   o The Rw application provides the control linkage between a Policy
      Decision Point and an ingress router, to pass down decisions on
      flow admission following either the push or the pull model. The
      Rw application also passes flow termination decisions.

   As can be seen from this brief description, the PCN Data Collection
   application defined in this memo is complementary to the Rw
   application. Within the strict terms of the ITU-T architecture, it
   is a realization of a different interface, the Rc interface.
   However, the PCN Data Collection application is intended for use in
   any of a number of architectures based on a centralized policy
   decision element.

Authors' Addresses

   Fortune Huang (editor)
   Huawei Technologies
   Section F, Huawei
   Industrial Base
   Bantian Longgang, Shenzhen 518129
   P.R. China

   Email: fqhuang@huawei.com

Tom Taylor
Huawei Technologies
Ottawa, Ontario
Canada

Email: tom111.taylor@bell.net

Glen Zorn
Network Zen
1310 East Thomas Street
#306
Seattle,, Washington 98102
USA

Phone: +1 (206) 377-9035
Email: gwz@net-zen.net

Hannes Tschofenig
Nokia Siemens Networks
Linnoitustie 6
Espoo 02600
Finland

Phone: +358 (50) 4871445
Email: Hannes.Tschofenig@nsn.com
URI: http://www.tschofenig.priv.at