Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 7, 2010

F. Huang
Huawei Technologies
G. Zorn, Ed.
Network Zen
July 6, 2009

The Diameter Precongestion Notification (PCN) Data Collection Application
draft-huang-dime-pcn-collection-00

Status of this Memo

Copyright Notice

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 reports of congestion-related events (flow admission state change, excess flow) reliably to a policy decision point. The ITU-T is working on a variant of this architecture which places the policy decision point in a central node rather than ingress or egress nodes of the network. In this case the policy decision point must request and obtain certain data from an ingress node when it receives an excess flow report affecting that ingress node. This memo defines a Diameter application to support egress node reporting and data collection from the ingress node. The nature of the data flows requires the policy decision point to act both as server and as client. Hence this memo draws upon the precedent established by the Rw application (RFC 5431 and ITU-T Recommendation Q.3303.3).

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. Two mechanisms are used: admission control, to decide whether to admit or block a new flow request, and (in abnormal circumstances) flow termination to decide whether to terminate some of the existing flows. Together they protect the QoS 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. For a full description of the PCN architecture, see [RFC5559].

Marking statistics are gathered by egress nodes on a per-ingress-egress aggregate basis. They 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 egress node reports to a decision point whenever the CLE value passes a set threshold (upward or downward). The decision to terminate flows is made on the basis of a different criterion. When the egress node detects that this criterion has been satisfied, it sends a report to the decision node providing measurement values that are used to

determine the total volume of traffic that must be terminated. If equal cost multipath (ECMP) routing is in use, it also sends a list of individual flows that were marked at the termination level.

2. Related Work

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 (PDP) [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 PDP. Egress node reports are sent by the egress node acting as client to the PDP acting as server. Data generated at the ingress node are needed only when flow termination is required. They are requested by the PDP acting as client and sent in responses by the ingress node acting as server. The PDP 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:

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

- 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.

3. 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].

4. The 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. Diameter Session Usage

Diameter sessions are implicitly terminated. An implicitly terminated session is one for which the server does not maintain 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 when the congestion level estimate (CLE) moves above or drops below the pre-congestion reporting threshold, or when an excess flow condition is detected. Multiple reports MAY be included in the same message, as described in Section 4.5.2.

Message format:

```
  <CRR> ::= < Diameter Header: CC1, REQ, PXY >
            < Session-Id >
            { Auth Application Id }
            { Origin-Host }
            { Origin-Realm }
            { Destination-Realm }
            [ Destination-Host ]
            [ Event-Timestamp ]
          * [ PCN-Congestion-Info ]
          * [ PCN-Excess-Flow-Info ]
          * [ Proxy-Info ]
          * [ Route-Record ]
          * [ AVP ]
```

At least one instance of the PCN-Congestion-Info or the PCN-Excess-Flow-Info AVP MUST be present.

4.3.2. Congestion-Report-Answer (CRA) Command

The PDP 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 an Congestion-Report-Request command sent by an egress node. The Congestion-Report-Answer command contains the same Session-Id as the corresponding request.

Message format:

```
  <CRA> ::= < Diameter Header: CC1, PXY >
            < Session-Id >
            { Auth Application Id }
            { Result-Code }
            { Origin-Host }
            { Origin-Realm }
```

```
                [ Error-Message ]
                [ Error-Reporting-Host ]
                [ Failed-AVP ]
                [ Event-Timestamp ]
              * [ Proxy-Info ]
              * [ AVP ]
```

4.3.3. Measurement-Poll-Request (MPR) Command

The PDP 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 given measurement period as described in Section 4.5.4.

Message format:

```
  <MPR> ::= < Diameter Header: CC2, REQ, PXY >
            < Session-Id >
            { Auth Application Id }
            { 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 PDP.

Message format:


```
  <MPA> ::= < Diameter Header: CC2, PXY >
```

```
                < Session-Id >
                { Auth Application Id }
                { Result-Code }
                { Origin-Host }
                { Origin-Realm }
                { PCN-Sent-Info }
                [ Error-Message ]
                [ Error-Reporting-Host ]
                [ Failed-AVP ]
                [ Event-Timestamp ]
              * [ 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 UTF8String. It identifies a specific ingress-egress aggregate flow. The internal structure of the I-E-Aggregate-Id value is network dependent.

4.4.2. PCN-Congestion-Info AVP

The PCN-Congestion-Info AVP (AVP code <AVP2>) is of type Grouped. It identifies an ingress-egress aggregate, reports the current value of the congestion level estimate (CLE), and indicates whether the report is generated because the CLE has risen above the reporting threshold or because it has fallen below the reporting threshold.

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

```
  PCN-Congestion-Info ::= < AVP Header: AVP2 >
                                { I-E-Aggregate-Id }
                                { CLE-Value }
                                { CLE-Report-Reason }
                              * [ AVP ]
```

4.4.3. CLE-Value AVP

The CLE-Value AVP (AVP code <AVP3>) is of type Float32. It gives the current (smoothed) congestion level estimate as a fraction between 0.0 and 1.0.

4.4.4. CLE-Report-Reason AVP

The CLE-Report-Reason AVP (AVP code <AVP4>) is of type Enumerated. The following values are defined in this document:

PRECONGESTION_ONSET (0)
The current CLE (reported in CLE-Value) is above the configured onset reporting threshold. The CLE derived in the previous measurement period was below that threshold.

PRECONGESTION_END (1) The current CLE (reported in CLE-Value) is below the configured end-of-precongestion reporting threshold, which may have the same value as the onset reporting threshold. The CLE derived in the previous measurement period was above that threshold.

4.4.5. PCN-Excess-Flow-Info AVP

The PCN-Excess-Flow-Info AVP (AVP code <AVP5>) is of type Grouped. It identifies an ingress-egress aggregate, reports a rate of excess traffic for that aggregate, and MAY identify a number of individual flows within that aggregate that experienced the markings that led to the generation of the PCN-Excess-Flow-Info AVP. Precise details of the conditions under which this AVP is generated and how the indivdual flows are selected are given in the specification for the PCN edge behaviour deployed in the domain.

The PCN-Excess-Flow-Info AVP has the following format:

```
PCN-Excess-Flow-Info ::= < AVP Header: AVP5 >
                           { I-E-Aggregate-Id }
                           { I-E-Aggregate-Excess-Rate }
                         * [ Flow Number ]
                         * [ Flow Description ]
                         * [ AVP ]
```

4.4.6. I-E-Aggregate-Excess-Rate AVP

The I-E-Aggregate-Excess-Rate AVP (AVP code <AVP6>) is of type Unsigned32. It gives the rate of flow of excess traffic in octets per second that the egress node derived for the identified ingress-egress aggregate for the measurement period ending at the time given by the Event-Timestamp AVP (if present).

4.4.7. Flow Number AVP

The Flow Number AVP (AVP code 509) is of type Unsigned32, and it contains the ordinal number of the IP flow(s). The rules for how the

ordinal number should be assigned are not described in this document.

4.4.8. Flow Description

The Flow Description AVP (AVP code 507) is of type IPFilterRule, and defines a packet filter for an IP flow with the following information:

- Direction (in or out).
- Source and destination IP address (possibly masked).
- Protocol.
- Source and destination port (list or ranges).

The b type shall be used with the following restrictions:

- Only the Action "permit" shall be used.
- No "options" shall be used.
- The invert modifier "!" for addresses shall not be used.
- The keyword "assigned" shall not be used.

The Flow description AVP shall be used to describe a single IP flow. The direction "in" refers to uplink IP flows, and the direction "out" refers to downlink IP flows.

4.4.9. PCN-Sent-Info AVP

The PCN-Sent-Info AVP (AVP code <AVP7>) is of type Grouped. It provides the rate of flow of PCN-marked traffic in octets per second that the ingress node derived for the identified ingress-egress aggregate for the measurement period ending at the time given by the Event-Timestamp AVP (if present).

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

```
PCN-Sent-Info ::= < AVP Header: AVP8 >
                  { I-E-Aggregate-Id }
                  { I-E-Aggregate-Sent-Rate }
                * [ AVP ]
```


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

The I-E-Aggregate-Sent-Rate AVP (AVP code <AVP8>) is of type Unsigned32. It gives the rate of flow of PCN-marked 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 (if present).

4.5. Procedures

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

4.5.1. Overall Procedures

The egress node measures the traffic from a particular ingress node, and calculates the congestion level estimate(CLE) at the ingress-egress aggregate level. The egress node may compare the CLE calculated at the current interval with the CLE calculated at the last interval, if the difference of the two CLEs exceeds a preset range, the egress node sends the feedback information, including at least the current CLE, to the PDP. After receiving the feedback information, the PDP saves the it for admission control and flow termination. After receiving a service flow request, the PDP can determine whether to admit the request or not based on the feedback information. Besides, the PDP also decides whether some of the admitted flows need to be terminated. The PDP needs to signal to the ingress node the decision about admission or termination.

4.5.2. Egress Node Behaviour

For each ingress-egress aggregate flow it serves, the egress node meters received traffic for PCN markings, recomputes its smoothed congestion level estimate, and determines whether there is excess flow in successive measurement periods in accordance with the PCN edge behaviour specification deployed in the domain. When a change in the smoothed congestion level estimate causes it to cross a reporting threshold, either upward or downward, the egress node MUST send an Accounting-Request message to the PDP. Similarly, the egress node MUST send an Accounting-Request message to the PDP when excess flow is detected for an ingress-egress aggregate served by that node. The Session-Id is irrelevant to the PCN Data Collection application and MAY have any value that conforms to [RFC3588]. The Account-Record-Type for the message MUST be set to EVENT_RECORD (1). The Accounting-Record-Number AVP SHOULD be set to 0.

The Event-Timestamp AVP SHOULD be present, and SHOULD provide the

ending time of the measurement period from which the data triggering the generation of the message were derived. At least one instance either of the PCN-Congestion-Info or the PCN-Excess-Flow-Info AVP MUST be present. Both AVPs MAY be present for the same ingress-egress aggregate, if both apply according to the edge behaviour specification. Multiple instances of either AVP MAY be present, but each instance MUST report on a different ingress-egress aggregate.

4.5.3. PDP Behaviour

If the PDP receives an Accounting-Request (ACR) identified as belonging to the PCN Data Collection application, it MUST acknowledge the message with an Accounting-Answer (ACA). The PDP usage of the information provided by PCN-Congestion-Info and PCN-Excess-Flow-Info AVPs is described in the applicable edge behaviour specification.

When the PDP receives an ACR containing an PCN-Excess-Flow-Info AVP, it MAY send a Measurement-Poll-Request (MPR) to the ingress node for the aggregate concerned. The Account-Record-Type for the message MUST be set to EVENT_RECORD (1). The I-E-Aggregate-Id MUST identify the ingress-egress aggregate flow for which information is being requested. The Event-Timestamp MUST be present if it was present in the ACR that contained the PCN-Excess-Flow-Info AVP, and MUST have the same value.

If the PDP receives a successful Measurement-Poll-Answer message, it uses the information contained in the PCN-Sent-Info AVP as described in the applicable edge behaviour specification.

4.5.4. 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 Account-Record-Type for the message MUST be set to EVENT_RECORD (1). The Accounting-Record-Number AVP SHOULD be set to 0. 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 the latest measurement period ending before or at the time given by Event-Timestamp, if available. In any case, Event-Timestamp SHOULD 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.

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 RFC 3588.

5.2. Diameter Command Codes

Command codes must be assigned for Congestion-Report-Request (CRR) (<CC1>, Section 4.3.1), Congestion-Report-Answer (MPA) (<CC1>, Section 4.3.2), Measurement-Poll-Request (MPR) (<CC2>, Section 4.3.3) and Measurement-Poll-Answer (MPA) (<CC2>, Section 4.3.4) commands according to the policy specified in RFC 3588, Section 11.2.1.

5.3. Attribute-Value Pairs

Codes must be assigned for the following AVPs using the policy specified in RFC 3588, Section 11.1.1:

I-E-Aggregate-Id (<AVP1>, Section 4.4.1)

PCN-Congestion-Info (<AVP2>, Section 4.4.2)

CLE-Value (<AVP3>, Section 4.4.3)

CLE-Report-Reason (<AVP4>, Section 4.4.4)

PCN-Excess-Flow-Info (<AVP5>, Section 4.4.5)

I-E-Aggregate-Excess-Rate (<AVP6>, Section 4.4.6)

PCN-Sent-Info (<AVP7>, Section 4.4.9)

I-E-Aggregate-Sent-Rate (<AVP8>, Section 4.4.10)

6. Security Considerations

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

6.1. Traffic Security

Application taffic MUST be secured as specified in RFC 3588 (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-adittance 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 PDP 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.

7.2. Informative References

[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", 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.

Authors' Addresses

Fortune Huang
Huawei Technologies

Section F
Huawei Industrial Base
Bantian Longgang, Shenzhen 518129
P.R. China

Email: fqhuang@huawei.com

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

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