PCP Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 04, 2014

D. Wing
R. Penno
T. Reddy
Cisco
July 03, 2013

PCP Flowdata Option
draft-wing-pcp-flowdata-00

Abstract

This document defines a mechanism for a host to signal flow characteristics to the network, and the network to signal its ability to accommodate that flow back to the host. The mechanism defines a new PCP option for the existing MAP and PEER opcodes.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 04, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 2
2. Terminology 3
3. PCP FLOWDATA Option 3
3.1. Usage and Processing 4
3.2. Generating a PCP Request with FLOWDATA Option 5
3.3. Processing a Request with FLOWDATA Option 6
3.4. Processing a Response with FLOWDATA Option 7
3.5. Link or State Changes on PCP Server 7
3.6. Conflict Resolution 8
4. PCP FLOWDATA Option Data Fields 9
5. FLOWDATA Interaction with PCP Proxy 14
6. Network Authorization 15
7. Scaling Considerations 15
8. Security Considerations 15
9. IANA Considerations 15
10. Acknowledgements 15
11. References 15
11.1. Normative References 15
11.2. Informative References 16
Authors' Addresses 16

1. Introduction

Access networks often have insufficient bandwidth or other characteristics that prevent some applications from functioning as well as desired. Although the quality of wireless and wired access networks continue to improve, those access networks are often constrained for various reasons. This document provides a mechanism to signal the application's network requirements to the access network, so that certain network flows can receive service that is differentiated from other network flows. With this mechanism, a host can request the network provide certain characteristics for a flow in both the upstream and downstream directions. The network authorizes the request and signals back to the host that it can (fully or partially) accommodate the flow. This sort of signaling is useful for long-lived flows such as interactive audio/video, streaming video, and network control traffic (call signaling, routing protocols).

In order to obtain such differentiated service from a network, many previous mechanisms have been created for hosts to convey flow information to the network. The mechanism described in this document has several useful properties:

- Usable at the application level, without needing operating system support;

- o Abstracts layer 2 specifics, so host and applications can avoid layer 2-specific signaling;

- o Robust metadata support, to convey sufficient information to the network about the flow;

- o Differentiates service on the local network and the immediately adjacent access network, which is typically bandwidth constrained;

- o Deployable on a local network and its adjacent access link, without needing support of the remote host's network or support of the remote host;

- o Provides differentiated service for both directions of a flow, including flows that cross administrative boundaries (such as the Internet).

The mechanism described in this specification defines an extension to Port Control Protocol (PCP [RFC6887]). This may be surprising at first because PCP is considered as a protocol for managing mappings in NATs and firewalls. However, PCP does not require the network implement a NAT or to implement a firewall. This is an important point: this specification does not require the network operate a NAT, and does not require the network operate a firewall. At a high level, PCP provides bi-directional communication a flow to the network. PCP can recursively communicate flow information to a number of on-path devices using PCP itself ([I-D.cheshire-recursive-pcp], [I-D.ietf-pcp-proxy]) or using an SDN protocol. Such recursion provides the flow information to more devices on the path, allowing each of them to optimize the flow over their respective links.

2. Terminology

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 [RFC2119].

3. PCP FLOWDATA Option

The FLOWDATA option described in this document allows a host to signal the bi-directional characteristics of a flow to its PCP server. After signaling, the PCP server determines if it can accommodate that flow, making configuration changes if necessary to accommodate the flow, and returns information in the FLOWDATA option indicating its ability to accommodate the described flow.

3.1. Usage and Processing

A host may want to indicate to the network the priority of a flow after the flow has been established (typical if the host is operating as a client) or before the flow has been established (typical if the host is operating as a server). Both of these are supported and depicted in the following diagrams.

The following diagram shows how a connection is first established and then the flow is prioritized. This allows for the fastest connection setup time with the server.

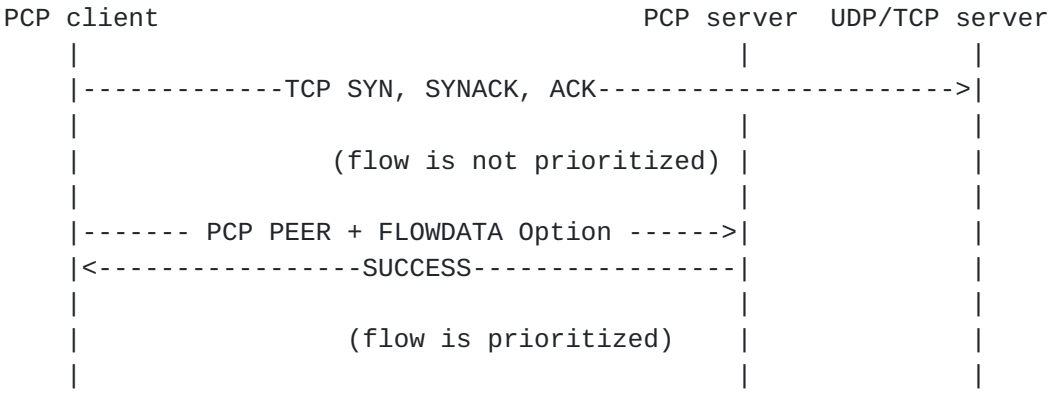

Figure 1: Message diagram, client connects first

The following diagram shows first asking the network to prioritize a flow, then establishing a flow. This is useful if the priority of the flow is more important than establishing the flow quickly.

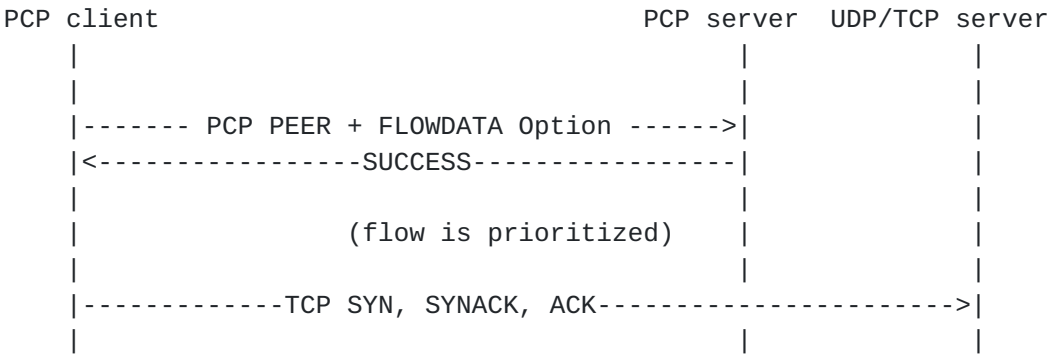

Figure 2: Message diagram, client sets priority first

The following diagram shows a PCP client getting a PCP MAP mapping for incoming flows with priority. This ensures that the PCP client has a mapping and all packets associated with the incoming TCP connections matching that mapping are prioritized. The PCP Client in this case could be a video server in a data center.

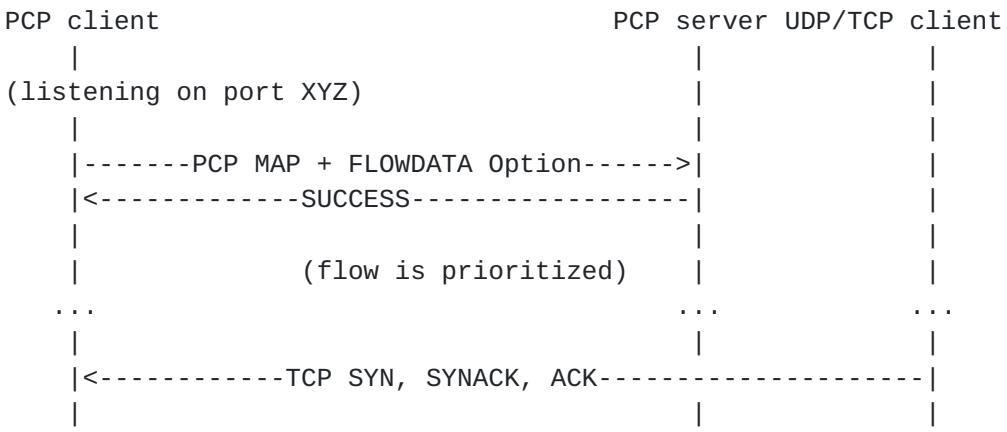

Figure 3: Message diagram, operating a server

The following diagram shows how two separate connections, where only one is active at a time, use the same instance identifier.

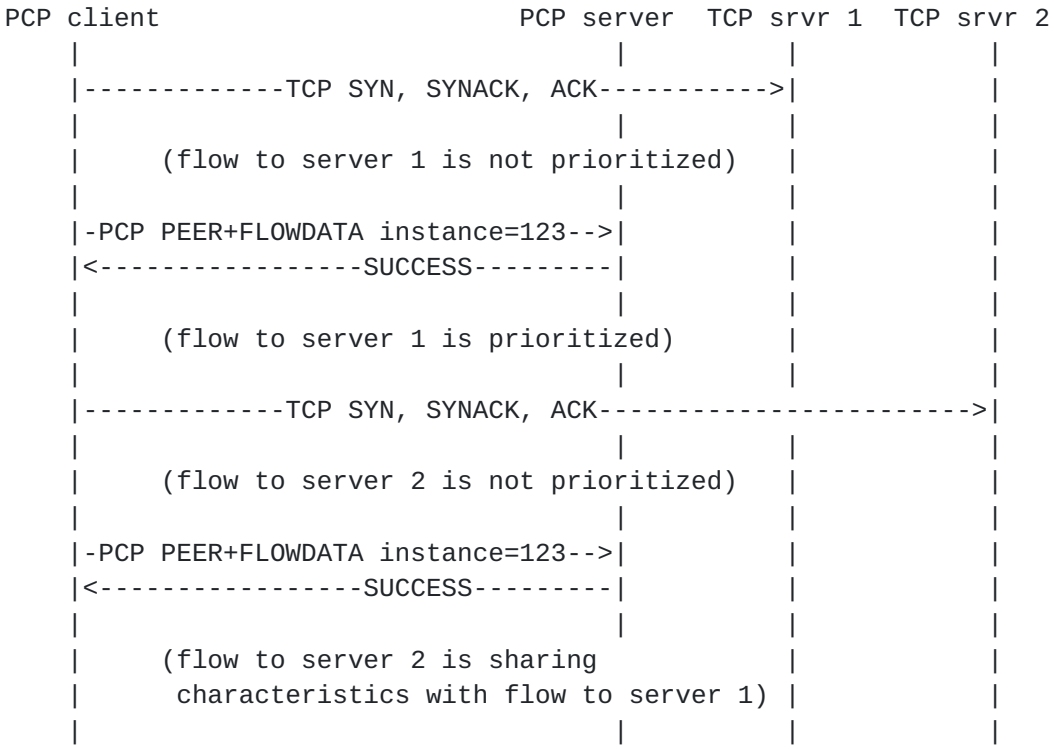

Figure 4: Message diagram with Instance Identifier

3.2. Generating a PCP Request with FLOWDATA Option

The PCP client first does all the processing described in Sections 8.1, 11.2, and 12.3 of [RFC6887] as appropriate for generating a MAP or PEER opcode request. Included in that request is a FLOWDATA option formatted as described in this document. For flows

established by the PCP client, the MAP or PEER request with FLOWDATA option can be sent before or after the PCP client has established any flows. For flows terminated by the PCP client (that is, when operating a server), the FLOWDATA option can be received and processed by the PCP server together with a MAP request or later during a MAP refresh request as shown in Figure 3.

3.3. Processing a Request with FLOWDATA Option

The PCP server performs processing in the order of the paragraphs below.

Upon receiving a PCP Request with FLOWDATA option first does the processing described in Section 8.2, 11.3, and 12.2 of [RFC6887], as appropriate for processing a MAP or PEER opcode request. If the MAP or PEER request contains the FLOWDATA option, the PCP server determines if the flow characteristics described in the FLOWDATA option can be accommodated by the network element controlled by the PCP server (that is, the router, NAT, or firewall controlled by the PCP server). To determine this, the PCP server might examine its static configuration and do bandwidth counting, or it might reconfigure the underlying network so that additional bandwidth is made available for this particular flow, or might perform other actions. If the PCP server determines the flow can only be partially accommodated, it returns values in the FLOWDATA fields that it can accommodate or returns 0 in those FLOWDATA fields where it has no information. In other words if the request indicated a low tolerance for delay but the PCP server and its controlled device determine that only high delay is available, the FLOWDATA response indicates high delay is available. The same sort of processing occurs on all of the FLOWDATA fields of the response (upstream and downstream delay tolerance, loss tolerance, jitter tolerance, minimum bandwidth, maximum bandwidth).

A PCP server that processes the FLOWDATA option is likely to create state for that flow (e.g., for bandwidth counting so that the bandwidth is returned to the bandwidth pool when the flow lifetime expires). Because Memory and other resources limit how much state can be created, the PCP server MUST implement a policy limit so that all state is not consumed by one host. It MAY also implement other limits, such as rate limits. The PCP server can implement its own policy to remove flows from its memory, such as FIFO. If a host has exceeded its quota, the existing error USER_EX_QUOTA SHOULD be returned.

If the PCP server can accommodate the flow as described in the FLOWDATA option, and can create the mapping as described in the MAP or PEER opcode, it sends a PCP response with the SUCCESS response

code, and includes the FLOWDATA option filled in according to Section 4.

After performing the above steps, the router creates state (if necessary for its implementation) and sends SUCCESS response code to the client with the data fields in the FLOWDATA option properly filled out.

3.4. Processing a Response with FLOWDATA Option

The PCP client performs processing in the order of the paragraphs below.

Upon receiving a PCP response, the PCP client performs the normal processing described in Section 8.3 of [RFC6887].

If the PCP response was SUCCESS (0), the PCP server has created a mapping. If the PCP response contains the FLOWDATA option, the FLOWDATA fields indicate if the network could accommodate the requested flow characteristics. The PCP client can use that information to influence the traffic it sends and receives on the network. For example, if the FLOWDATA response indicates the network can accommodate a flow of a certain downstream bandwidth, the PCP client will likely achieve the best result if it does not initiate a flow that exceeds that bandwidth.

> Note to implementers: PCP allows the server to send multiple responses to a single request. This means that after sending a request and receiving a (positive) response, a subsequent response might be sent updating the information about the flow, should the network conditions change. The response could carry a FLOWDATA option where the data fields contain different values from the first response. This might occur, for example, if a competing high-bandwidth flow has finished, more bandwidth is available for this host; the DSL line rate might have improved (or degraded); the link speed may have been dynamically increased (or decreased). Thus, a PCP client should expect these subsequent responses and react accordingly.

3.5. Link or State Changes on PCP Server

After the PCP server has sent a SUCCESS response code including the FLOWDATA option, link characteristics might change causing a flow to no longer be accommodated by the network (e.g., link speed degrades) or for the PCP server to flush a flow from its list of prioritized flows (e.g., due to memory constraints). Whenever the network can no longer accommodate a flow, the PCP server MUST inform the PCP client by sending a mapping update response including an updated FLOWDATA

option, following the same procedure as a Mapping Update (Section 14.2 of [RFC6887]). As with PCP without FLOWDATA, if the PCP server loses all its state it will alert the PCP clients using rapid recovery (Section 14 of [RFC6887]) which also indicates loss of FLOWDATA state in the network.

Note: it is also possible that originally-requested flowdata could be accommodated (e.g., link speed improved). We might want to signal to endpoints that they should ask again for their originally-requested flowdata. This is for future study.

3.6. Conflict Resolution

It is possible that two hosts send requests with different thresholds for delay or jitter or different values for bandwidth in each direction, and their requests arrive at the same PCP server. An example is a media streamer and a television within the same home where one indicates its sending bandwidth is higher than the other indicates its receiving bandwidth. As another example, the indicated tolerance for delay might be different.

If this occurs, it is RECOMMENDED that the PCP server use the smaller bandwidth and stricter delay/loss tolerance (that is, the lower tolerance to delay or jitter), and issue a FLOWDATA update so both PCP clients receive the same information. The diagram below depicts a conflict message flow.

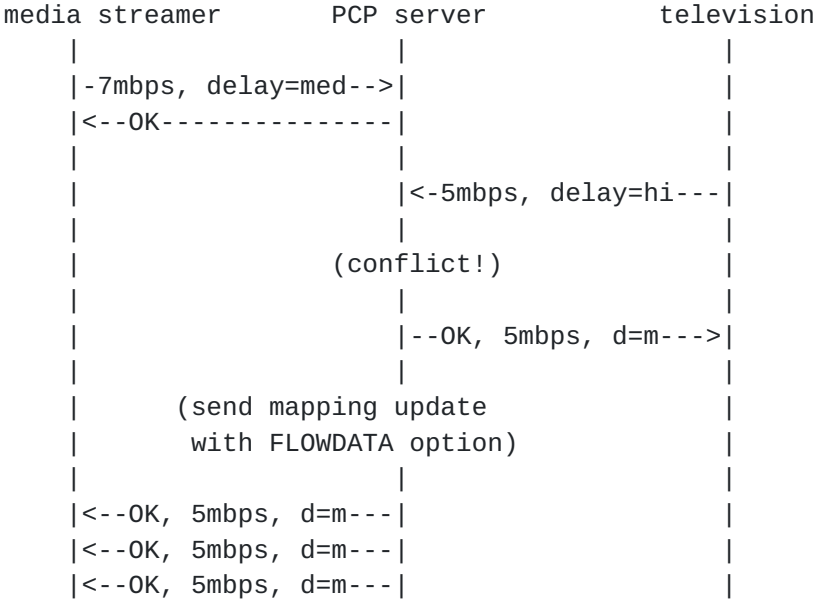

Figure 5: Message diagram, resolving conflict

It is also possible for one PCP client to think two flows should use the same instance identifier but the other PCP client to use different instance identifiers for those two flows. In this case, the operation of the PCP server (and the device it controls) is implementation specific.

4. PCP FLOWDATA Option Data Fields

The FLOWDATA option has the following characteristics:

Option Name: FLOWDATA
Number: (to be assigned by IANA)
Purpose: Describe flow characteristics to the network
Valid for Opcodes: MAP, PEER
Length: 24 octets
May appear in: request. May appear in response only if it appeared in the associated request.
Maximum occurrences: 1

The FLOWDATA option request has the following format.

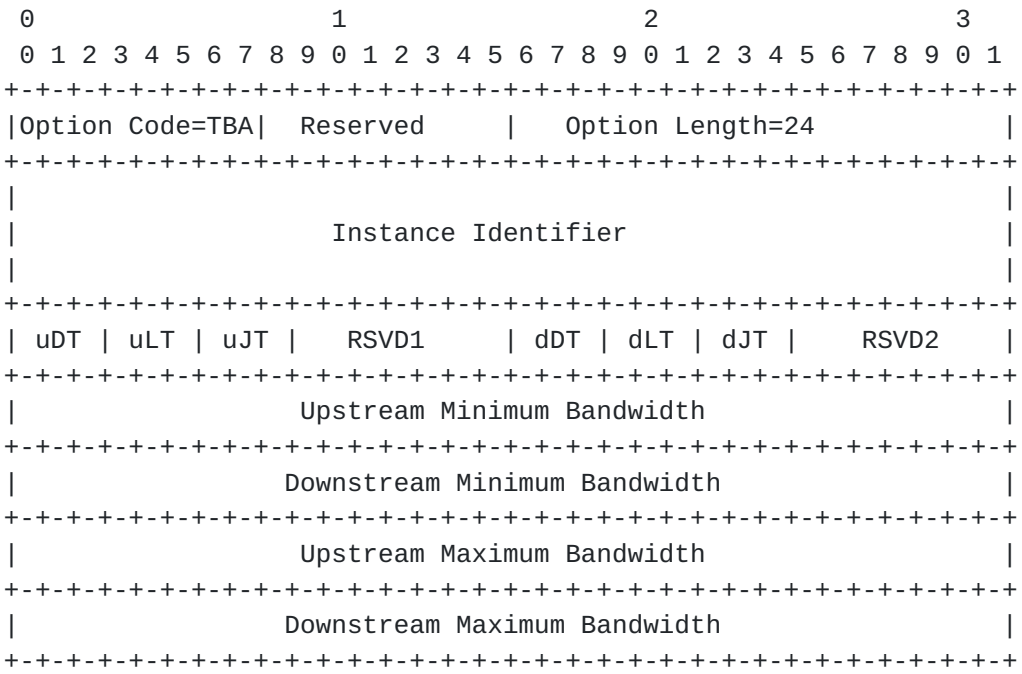

Figure 6: FLOWDATA Option

Description of the fields:

Instance Identifier: 96 bit identifier, unique to each simultaneously-active flow. This is a pseudo random number that MUST be generated following the procedures described in [RFC4086].

uDT: Upstream Delay Tolerance, 0=no information available, 1=very low, 2=low, 3=medium, 4=high.

uLT: Upstream Loss Tolerance, 0=no information available, 1=very low, 2=low, 3=medium, 4=high.

uJT: Upstream Jitter Tolerance, 0=no information available, 1=very low, 2=low, 3=medium, 4=high.

RSVD1: Reserved (7 bits), MUST be ignored on reception and MUST be 0 on transmission.

dDT: Downstream Delay Tolerance, 0=no information available, 1=very low, 2=low, 3=medium, 4=high.

dLT: Downstream Loss Tolerance, 0=no information available, 1=very low, 2=low, 3=medium, 4=high.

dJT: Downstream Jitter Tolerance, 0=no information available, 1=very low, 2=low, 3=medium, 4=high.

RSVD2: Reserved (7 bits), MUST be ignored on reception and MUST be 0 on transmission.

Upstream Minimum Bandwidth Measures bandwidth sent by the PCP client. Value is in octets per second. The value 0 means no information is available.

Downstream Minimum Bandwidth Measures bandwidth sent to the PCP client. Value is in octets per second. The value 0 means no information is available.

Upstream Maximum Bandwidth: Measures bandwidth sent by the PCP client. Value is in octets per second. The value 0 means no information is available.

Downstream Maximum Bandwidth Measures bandwidth sent to the PCP client. Value is in octets per second. The value 0 means no information is available.

The instance identifier accommodates network traffic where multiple 5-tuples exist for a particular data flow, but the bandwidth flows only over the aggregate of the multiple 5-tuples. A use-case for this identifier is TCP video streaming which retrieves short pieces

of the movie, often over separate TCP connections for load balancing, which would use the same Instance Identifier for each TCP connection. An instance is considered unique if the combination of the PCP client's IP address and the instance identifier are unique.

> Discussion point: Minimum and maximum value of bandwidth is 1 byte per second to 4 gigaBYTES per second. We probably need to express higher bandwidth, and maybe also lower bandwidth?

Different applications have different needs for their flows. The following table is derived from [RFC4594] to serve as a guideline for tolerance to loss, delay and jitter for some sample applications.

Service Class Name	Traffic Characteristics	Tolerance to Loss	Tolerance to Delay	Tolerance to Jitter
Network Control	Variable size packets, mostly inelastic short messages, but traffic can also burst (e.g., OSPF)	Low	Low	High
Telephony	Fixed-size small packets, constant emission rate, inelastic and low-rate flows (e.g., G.711, G.729)	Very Low	Very Low	Very Low
Signaling	Variable size packets, some what bursty short-lived flows	Low	Low	High
Multimedia Conferencing	Variable size packets, constant transmit interval, rate adaptive, reacts to loss	Low - Medium	Very Low	Low
Real-Time Interactive	RTP/UDP streams, inelastic, mostly variable rate	Low	Very Low	Low
Multimedia Streaming	Variable size packets, elastic with variable rate	Low - Medium	Medium	High
Broadcast Video	Constant and variable rate, inelastic, non-bursty flows	Very Low	Medium	Low
Low-Latency Data	Variable rate, bursty short-lived elastic flows	Low	Low - Medium	High
OAM	Variable size packets, elastic & inelastic flows	Low	Medium	High
High-Throughput Data	Variable rate, bursty long-lived elastic flows	Low	Medium - High	High
Standard	A bit of everything	0	0	0
Low-Priority Data	Non-real-time and elastic (e.g., network backup)	High	High	High

The FLOWDATA Option response has the following format. The fields indicate what the network can accommodate of the request.

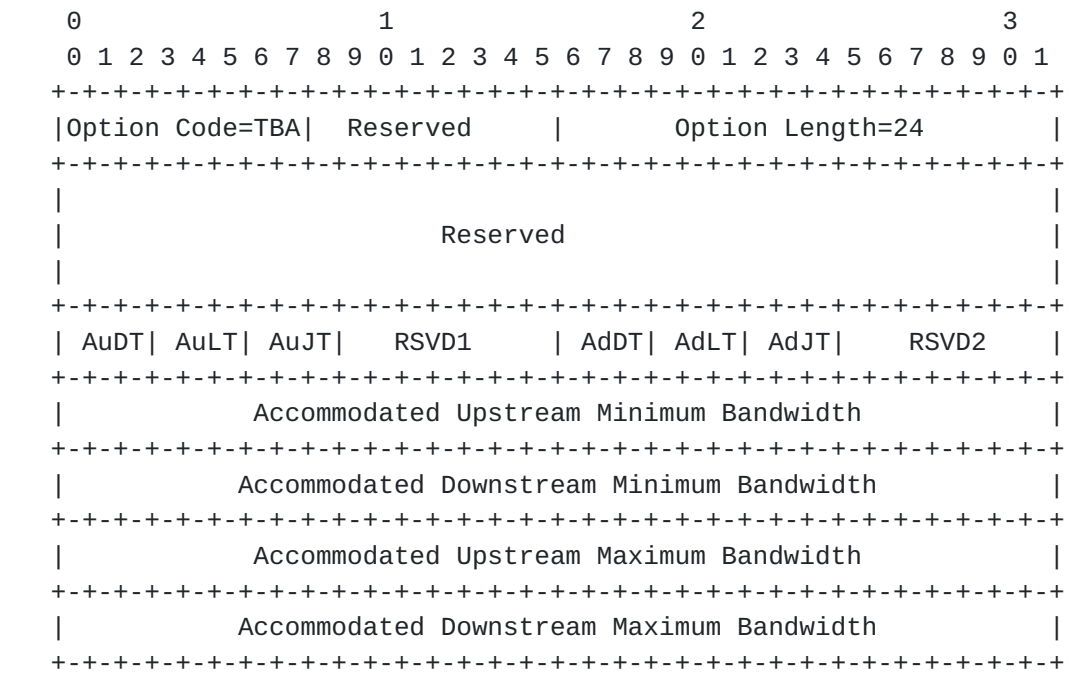

Figure 7: FLOWDATA Option

Description of the fields:

Reserved: 96 bits, MUST be ignored on reception and MUST be 0 on transmission.

AuDT: Accommodated Upstream Delay Tolerance, 0=no information available, 1=able to accommodate very low, 2=able to accommodate low, 3=able to accommodate medium, 4=able to accommodate high.

AuLT: Accommodated Upstream Loss Tolerance, 0=no information available, 1=able to accommodate very low, 2=able to accommodate low, 3=able to accommodate medium, 4=able to accommodate high.

AuJT: Accommodated Upstream Jitter Tolerance, 0=no information available, 1=able to accommodate very low, 2=able to accommodate low, 3=able to accommodate medium, 4=able to accommodate high.

RSVD1: Reserved (7 bits), MUST be ignored on reception and MUST be 0 on transmission.

AdDT: Accommodated Downstream Delay Tolerance, 0=no information available, 1=able to accommodate very low, 2=able to accommodate low, 3=able to accommodate medium, 4=able to accommodate high..

AdLT: Accommodated Downstream Loss Tolerance, 0=no information available, 1=able to accommodate very low, 2=able to accommodate low, 3=able to accommodate medium, 4=able to accommodate high.

AdJT: Accommodated Downstream Jitter Tolerance, 0=no information available, 1=able to accommodate very low, 2=able to accommodate low, 3=able to accommodate medium, 4=able to accommodate high.

RSVD2: Reserved (7 bits), MUST be ignored on reception and MUST be 0 on transmission.

Accommodated Upstream Minimum Bandwidth Bandwidth the network can accommodate for this flow, sent by the PCP client. Value in bytes per second. 0 means no information is available.

Accommodated Downstream Minimum Bandwidth Bandwidth the network can accommodate for this flow, sent to the PCP client. Value in bytes per second. 0 means no information is available.

Accommodated Upstream Maximum Bandwidth: Bandwidth the network can accommodate for this flow, sent by the PCP client. Value in bytes per second. 0 means no information is available.

Accommodated Downstream Maximum Bandwidth Bandwidth the network can accommodate for this flow, sent to the PCP client. Maximum Downstream bandwidth in bytes per second, 0 means no information is available.

5. FLOWDATA Interaction with PCP Proxy

The FLOWDATA option is optional to process. A PCP Proxy performs the functions described in [I-D.ietf-pcp-proxy], and if the PCP request contains the FLOWDATA option it also performs the functions described in this section.

The PCP request containing the FLOWDATA option SHOULD be proxied normally, so that the upstream PCP server can be aware of the entire request. The PCP proxy MAY have its own policies specific to the FLOWDATA option which require it to modify the FLOWDATA values request (e.g., reduce bandwidth for a certain PCP client).

After proxying the message containing FLOWDATA, when the PCP proxy receives the associated PCP response, the PCP proxy MAY reduce the bandwidth values or use worse (higher) values for delay, loss, or jitter tolerance. It MUST NOT increase the bandwidth or use better (lower) values for the delay, loss, or jitter tolerance.

6. Network Authorization

Oftentimes the endpoints themselves are not authorized to request network resources, but instead authorization has to first be obtained from a network element such as a call controller or policy element. To accommodate such deployments, third party authorization can be used with FLOWDATA . At a high level, this authorization works by the PCP client first obtaining a cryptographic token from the authorizing network element (e.g., call controller) and includes that token in the PCP request. The PCP server in the network validates the token and grants access.

7. Scaling Considerations

The network elements need only act upon those flows explicitly signaled by a PCP client, instead of all possible flows that a host generates.

Short lived flows (e.g., HTTP/1.0) or best-effort flows would receive little to no benefit from the signaling described in this document. As explained in Section 3.3, the PCP server will limit excessive flowdata requests, so hosts are encouraged to be conservative in how many flows are signaled with flowdata.

8. Security Considerations

On some networks, only certain users or certain applications are authorized to signal the priority of a flow. This authorization can be achieved with PCP client authentication [I-D.ietf-pcp-authentication].

9. IANA Considerations

IANA is requested to assign a new PCP option called FLOWDATA from the optional to process range (128-255) in the [pcp-iana] registry.

10. Acknowledgements

Thanks to Anca Zamfir for review comments.

11. References

11.1. Normative References

[I-D.ietf-pcp-authentication]
Wasserman, M., Hartman, S., and D. Zhang, "Port Control Protocol (PCP) Authentication Mechanism", draft-ietf-pcp-authentication-01 (work in progress), October 2012.

[I-D.ietf-pcp-proxy]
Boucadair, M., Penno, R., and D. Wing, "Port Control Protocol (PCP) Proxy Function", draft-ietf-pcp-proxy-03 (work in progress), June 2013.

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

[RFC4086] Eastlake, D., Schiller, J., and S. Crocker, "Randomness Requirements for Security", BCP 106, RFC 4086, June 2005.

[RFC6887] Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, April 2013.

11.2. Informative References

[I-D.cheshire-recursive-pcp]
Cheshire, S., "Recursive PCP", draft-cheshire-recursive-pcp-02 (work in progress), March 2013.

[RFC4594] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, August 2006.

[pcp-iana]
IANA, "Port Control Protocol (PCP) Parameters", May 2013, <http://www.iana.org/assignments/pcp-parameters/pcp-parameters.xml#options>.

Authors' Addresses

Dan Wing
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, California 95134
USA

Email: dwing@cisco.com

Reinaldo Penno
Cisco Systems, Inc.
170 West Tasman Drive
San Jose 95134
USA

Email: repenno@cisco.com

Tirumaleswar Reddy
Cisco Systems, Inc.
Cessna Business Park, Varthur Hobli
Sarjapur Marathalli Outer Ring Road
Bangalore, Karnataka 560103
India

Email: tireddy@cisco.com