PCP
Internet-Draft
Intended status: Standards Track
Expires: January 30, 2014

R. Penno
T. Reddy
D. Wing
B. VerSteeg
Cisco
M. Boucadair
France Telecom
July 29, 2013

PCP Usage for Quality of Service (QoS) in Mobile Networks
draft-penno-pcp-mobile-qos-00

Abstract

There are challenges to request quality of service for an application or network flow that is not part of a mobile network's Evolved Packet Core (EPC). This document addresses this issue by defining a mechanism to signal the desired characteristics of a flow to the Mobile Network from a User Equipment (UE) using Port Control Protocol (PCP). The signaled characteristics allow the Mobile Network to enforce appropriate policies such as prioritize that flow accordingly and trigger dedicated bearer activation or bearer modification procedure.

Table of Contents

1. Introduction

The use of Mobile Network for accessing the Internet and other data services via smartphones, tablets, and notebook/netbook computers has increased rapidly as a result of high-speed packet data networks such as HSPA and HSPA+; and now Long-Term Evolution (LTE) is being deployed. Mobile devices are becoming similar in capability to their desktop counterparts. From that perspective, it is feasible to run WebRTC, HTTP Adaptive Streaming (HAS), P2P applications on mobile devices. Mobile network needs to have a mechanism to prioritize such packet flows in both directions.

The Web Real-Time communication (WebRTC) framework [I-D.ietf-rtcweb-overview] provides the protocol building blocks to support direct, interactive, real-time communication using audio, video, collaboration, games, etc., between peer web-browsers. WebRTC application use Interactive Connectivity Establishment (ICE) protocol [RFC5245] for gathering candidates, prioritizing them, choosing default ones, exchanging them with the remote party, pairing them and ordering them into check lists. Once all of the above steps have

been completed the participating ICE agents can begin a phase of connectivity checks and eventually select a pair of candidates that will be used for real-time communication. The P2P streams (audio, video, data-channel) are dynamic, time-bound, encrypted and have different priorities. When WebRTC server is deployed in a 3rd party network trusted by the Mobile Network and the media session need to be prioritized, a mechanism is required to signal the flow characteristics (i.e., traffic performance requirements) of the media streams to the Mobile Network. However, the Mobile Network may not trust the host (UE) to signal the correct flow characteristics permitted by the WebRTC server.

PCP [RFC6887] provides a mechanism to describe a given flow to the network prior to actual session establishment. The primary driver for PCP has been creating port mappings on NAT and firewall devices. When doing this, PCP pushes flow information from the host into the network (specifically to the network's NAT or firewall device), and receives information back from the network (from the NAT or firewall device). This document uses PCP FLOWDATA option defined in [I-D.wing-pcp-flowdata] to convey the flow characteristics from the host to the Mobile Network, and allow the Mobile Network to prioritize that flow accordingly and trigger dedicated bearer activation or bearer modification procedure. This document also explains how the PCP Server in the Evolved Packet Core (EPC) maps the fields in PCP FLOWDATA option to 3GPP QCI, GBR values.

The mechanism described in this document has several useful properties :

a. Differentiated QoS services can be offered to third party applications. For third party applications differentiated QoS services can be installed even if the UE is behind NAT provided by the Mobile Network. In contrast, other mechanisms struggle to install differentiated QOS if the UE is behind NAT.

b. Mobile Network can authorize the differentiated service request from third party application because the proposed mechanism is compliant with the 3GPP's network-triggered QoS policy enforcement model.

c. This mechanism does not rely on DPI.

d. A UE can use single protocol no matter of the access technology; Abstracts layer 2 specifics, so host and applications can avoid layer 2-specific signaling even if their Internet connection is via 3G/4G or DOCSIS.

e. Usable at the application level, without needing operating system support

f. Robust metadata support, to convey sufficient information to the network about the flow;

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

h. Both high-priority and low-priority flows can be signalled, so that in overload situations operators can make low-priority flows yield to other flows through policing.

Note :

1. It is out of scope of this document to discuss the trade-offs between the proposed approach vs. deploying local WebRTC-IMS Gateways within the Mobile Network.

2. The mechanism described in this document provides QoS and network feedback for a variety of applications including interactive audio/video application such as WebRTC, streaming video, and network backup. The value is provided for the applications that are orchestrated through EPC and for applications that are delivered over the top.

3. Administrative-related considerations between the administrative entity managing the third party application server and the Mobile Network are out of scope of this document.

2. Notational Conventions

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].

This note uses terminology defined in [RFC5245], [RFC6459].

WebRTC Server : Web Server that supports WebRTC.

High-Speed Packet Access : The High-Speed Packet Access (HSPA) and HSPA+ are enhanced versions of the Wideband Code Division Multiple Access (WCDMA) and UTRAN, thus providing more data throughput and lower latencies.

3. QoS in Cellular Networks

3GPP has standardized QoS for EPC (Enhanced Packet Core) from Release 8 [TS23.107]. 3GPP QoS policy configuration defines access agnostic QoS parameters that can be used to provide service differentiation in multi vendor and operator deployments. The concept of a bearer is used as the basic construct for which QoS treatment is applied for uplink and downlink packet flows between the Mobile Node (MN) and gateway [TS23.401]. A bearer may have more than one packet filter associated and this is called a Traffic Flow Template (TFT). IP source address, source port, IP destination, destination port, L4 protocol, Type of service/Traffic class type, Security parameter index etc identify a packet filter. Each UE can have one or multiple bearers associated with its registration, each supporting different QoS characteristics. An UpLink Traffic Flow Template (UL TFT) is the set of uplink packet filters in a TFT. A DownLink Traffic Flow Template (DL TFT) is the set of downlink packet filters in a TFT.

The access agnostic QoS parameters associated with each bearer are QCI (QoS Class Identifier), ARP (Allocation and Retention Priority), MBR (Maximum Bit Rate) and optionally GBR (Guaranteed Bit Rate) explained in [TS23.203]. QCI is a scalar that defines packet forwarding criteria in the network. Mapping of QCI values to DSCP is well understood and GSMA has defined standard means of mapping between these scalars [GSMA-IR34]. Primarily LTE offers two types of bearer: Guaranteed Bit rate bearer for real time communication, e.g., Voice calls etc and Non-Guaranteed bit rate bearer, e.g., best effort traffic for web access etc. Packets mapped to the same EPS bearer receive the same bearer level packet forwarding treatment. For example QCI value 1 is typically used for Conversational Voice and the standardized flow characteristics for QCI value 1 are Packet delay of 100 ms and Packet error loss Rate of 10 to the power -2.

3G and LTE networks also provide extensive support for accounting and charging already, for example using the Policy Charging Control (PCC) architecture. In the EPS, per-user information is normally part of the user profile (stored in the Home Subscriber Server) that would be accessed by PCC entities such as the PCRF for dynamic updates, enforcement etc.

4. Solution Overview

In the below topology, The main involved functional elements are:

- UE (User Equipment) is a mobile node.

- The evolved NodeB (eNB) is a base station entity that supports the Long-Term Evolution (LTE) air interface. It is part of the access network that provides radio resource management, header compression, security and connectivity to the core network through

the S1 interface. In an LTE network, the control plane signaling traffic and the data traffic are handled separately. The eNBs transmit the control traffic and data traffic separately via two logically separate interfaces.

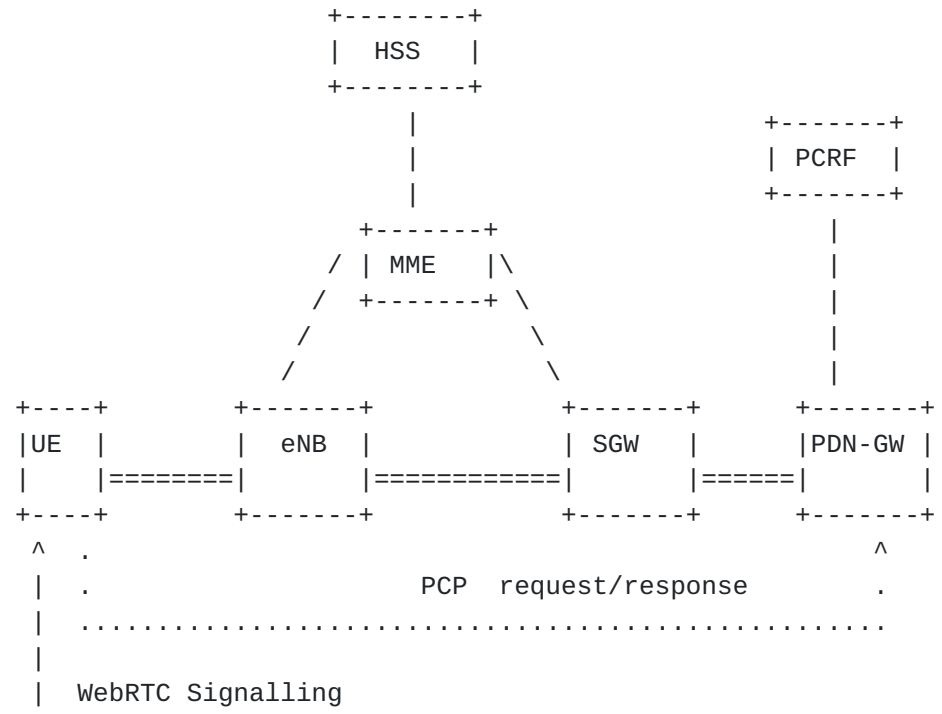
- The Serving gateway, SGW, is the mobility anchor and manages the user plane data tunnels during the inter-eNB handovers. It tunnels all user data packets and buffers downlink IP packets destined for UEs that happen to be in idle mode.

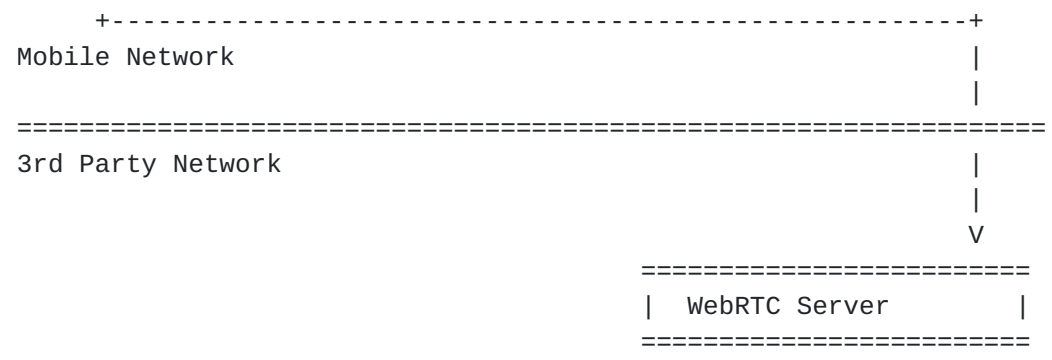
- Policy and Charging Rule Function (PCRF) which is responsible for determining which policy and charging control rules are to be applied [TS23.203].

- Policy and Charging Enforcement Function (PCEF) which performs policy enforcement (e.g., Quality of Service (QoS)) and flow-based charging [TS23.203]. PCEF is co-located with PDN-GW. PDN-GW is also responsible for IP address allocation to the UE, packet filtering, and policy-based control of flows.

- Application Function (AF) is an element offering applications that require dynamic policy and/or charging control [TS23.203].

- The Home Subscriber Server, HSS, is a database that contains user subscriptions and QoS profiles. The Mobility Management Entity, MME, is responsible for user authentication, bearer establishment and modification and maintenance of the UE context.





PCP interdomain - WebRTC

4.1. Network-triggered QoS

This section describes the existing steps applicable to any other network that requires authorization from third party application to permit differentiated QOS service request from UE which has been discussed in [I-D.wing-pcp-third-party-authz].

1. PCP client determines the PCP server to use by using the mechanisms explained in section 8.1 of [RFC6887]. In case of the GTP-based S5/S8 interface, the PDN-GW is the first-hop router for the UE, and in the case of PMIPv6-based S5/S8, the SGW is the first-hop router. PCP server could be co-located with the PDN-GW. For instance PCP client can also learn the PCP server address using DHCP [I-D.ietf-pcp-dhcp] and behavior to be followed by the PCP client to contact its PCP server(s) is explained in [I-D.ietf-pcp-server-selection]. The other benefits of using PCP are explained in [I-D.penno-rtcweb-pcp].

2. Once ICE [RFC5245] processing has completed, an updated offer/answer exchange takes place. WebRTC server is aware of the active media path after the controlling ICE endpoint follows the procedures in Section 11.1 of [RFC5245], specifically to send updated offer if the candidates in the m and c lines for the media stream (called the DEFAULT CANDIDATES) do not match ICE's SELECTED CANDIDATES (also see Appendix B.9 of [RFC5245]).

3. To provide differentiated QOS, the WebRTC server generates cryptographic token and metadata for prioritizing the media streams which is passed to the WebRTC endpoint. In this scenario PCP client on the UE is the third-party application obtaining limited access to an PCP server (resource server) on behalf of the WebRTC server (resource owner). The PCP TOKEN_ACCESS option defined in [I-D.wing-pcp-third-party-authz] must be included in the PCP request sent to the PCP server. This TOKEN_ACCESS option

is created by the PCP client using the access token, key id etc received from the authorization server using OAuth 2.0 [RFC6749]. The PCP client populates the fields in FLOWDATA option using the metadata provided by the authorization server. The PCP client sends the PCP request with MAP or PEER opcodes with the above PCP options to the PCP server. This mechanism is required so that the PCP server in the Evolved Packet (EPC) can validate that the PCP request for specific flow characteristics is initiated by the UE because of using a trusted 3rd party WebRTC Server.

4. The PCP server identifies the authorization server using the Domain Name in the PCP ACCESS_TOKEN option. The PCP server validates the fields in TOKEN_ACCESS option using the mechanism explained in section 5.2 of [I-D.wing-pcp-third-party-authz]. If the token is successfully validated then the authorization server returns the token bound authorization data in response. The token bound authorization data would be flow characteristics like upstream and downstream minimum bandwidth, delay, loss etc. The PCP server then matches this token bound authorization data with what is requested in the PCP FLOWDATA option. If the authorization sets match, the PCP server honors the PCP request made by the PCP client.

4.2. PCP to 3GPP

This section describes steps involved with processing PCP FLOWDATA option to initiate bearer activation for each media stream.

1. The PCP FLOWDATA option has all the required fields to trigger dedicated bearer activation or modification with relevant QCI, GBR values. UpLink Traffic Flow Template (UL TFT) and DownLink Traffic Flow Template (DL TFT) would be installed in both directions for the media stream. For example IP source address, source port, IP destination address, destination port, L4 protocol will be used from the PCP request (PEER opcode) to create packet filter which is associated with UL TFT. The advantage of this technique is no changes are required to TFT definition. PCP success response would be sent without waiting for network-initiated bearer activation or modification to be complete: i.e., PCP success response would be sent based on the resource availability to setup or modify bearers.

2. Using the fields in PCP FLOWDATA option listed in the below table, relevant QCI value will be determined to initiate bearer activation or modification procedure. Upstream and Downstream Bandwidth Minimum values will be set to zero in PCP FLOWDATA option to indicate QCI values in the range 5-8. Non-zero Bandwidth Minimum value in FLOWDATA option will be mapped to GBR

to determine if the requested bitrate can be provided or not. GBR is provided only for QCI values 1 to 4.

(Fields in PCP FLOWDATA option - uDT, uLT, dDT, dLT)

QCI	Delay	Loss	Example Services
1	Low	Medium	Conversational Voice
2	Medium	Low	Conversational Video
3	Very Low	Low	Real Time Gaming
4	Medium	Very Low	Non-conversational Video, buffered streaming
5	Low	Very Low	IMS Signalling
6	Medium	Very Low	Video (Buffered Streaming)
7	Low	Low	Voice, Video (Live streaming)
8	Medium	Low	web access
9	High	Low	e-mail

PCP FLOWDATA to QCI Mapping

3. The PDN-GW will communicate with the PCRF to trigger the appropriate Policy charging and control (PCC) decision based on which PDN-GW will initiate bearer activation or modification procedure.

4. If PCP authentication [I-D.ietf-pcp-authentication] is used then the PCP server can also provide identity of the UE to PCRF.

5. After the call is terminated PCP client informs the PCP server to close the mapping. The Authorization Server also informs the PCP server to revoke the access token after the call is terminated which is discussed in section 5.2 of [I-D.wing-pcp-third-party-authz]. This step triggers bearer deactivation procedure discussed in section 5.4.4.1 of [TS23.401].

5. Security Considerations

Security considerations discussed in [RFC6887] and PCP authentication [I-D.ietf-pcp-authentication] are to be taken into account.

6. IANA Considerations

None.

7. Acknowledgements

Authors would like to thank Harold Lassers, Basavraj Patil, Thomas Anderson for their comments and review.

8. References

8.1. Normative References

[I-D.ietf-rtcweb-overview]
Alvestrand, H., "Overview: Real Time Protocols for Brower-based Applications", draft-ietf-rtcweb-overview-06 (work in progress), February 2013.

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

[RFC5245] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", RFC 5245, April 2010.

[RFC5389] Rosenberg, J., Mahy, R., Matthews, P., and D. Wing, "Session Traversal Utilities for NAT (STUN)", RFC 5389, October 2008.

[RFC6407] Weis, B., Rowles, S., and T. Hardjono, "The Group Domain of Interpretation", RFC 6407, October 2011.

[RFC6459] Korhonen, J., Soininen, J., Patil, B., Savolainen, T., Bajko, G., and K. Iisakkila, "IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS)", RFC 6459, January 2012.

[RFC6749] Hardt, D., "The OAuth 2.0 Authorization Framework", RFC 6749, October 2012.

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

8.2. Informative References

[GSMA-IR34]
, "Inter-Service Provider Backbone Guidelines 5.0, 22 December 2010", September 2012.

[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-dhcp]
Boucadair, M., Penno, R., and D. Wing, "DHCP Options for the Port Control Protocol (PCP)", draft-ietf-pcp-dhcp-07 (work in progress), March 2013.

[I-D.ietf-pcp-server-selection]
Boucadair, M., Penno, R., Wing, D., Patil, P., and T. Reddy, "PCP Server Selection", draft-ietf-pcp-server-selection-01 (work in progress), May 2013.

[I-D.ietf-rtcweb-security-arch]
Rescorla, E., "WebRTC Security Architecture", draft-ietf-rtcweb-security-arch-07 (work in progress), July 2013.

[I-D.penno-rtcweb-pcp]
Penno, R., Reddy, T., Wing, D., and M. Boucadair, "PCP Considerations for WebRTC Usage", draft-penno-rtcweb-pcp-00 (work in progress), May 2013.

[I-D.reddy-rtcweb-mobile]
Reddy, T., Kaippallimalil, J., R, R., and R. Ejzak, "Considerations with WebRTC in Mobile Networks", draft-reddy-rtcweb-mobile-03 (work in progress), May 2013.

[I-D.wing-pcp-flowdata]
Wing, D., Penno, R., and T. Reddy, "PCP Flowdata Option", draft-wing-pcp-flowdata-00 (work in progress), July 2013.

[I-D.wing-pcp-third-party-authz]
Wing, D., Reddy, T., Patil, P., and R. Penno, "PCP Extension for Third Party Authorization", draft-wing-pcp-third-party-authz-00 (work in progress), May 2013.

[RFC6342] Koodli, R., "Mobile Networks Considerations for IPv6 Deployment", RFC 6342, August 2011.

[TS23.107] 3GPP, ., "End-to-End Quality of Service (QoS) Concept and Architecture, Release 10, 3GPP TS 23.207, V10.0.0 (2011-03)", September 2012.

[TS23.203] 3GPP, ., "3GPP, "Policy and charging control architecture", 3GPP TS 23.203 10.5.0, December 2011.", September 2012.

[TS23.401] 3GPP, ., "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 11), 3GPP TS 23.401, V11.2.0 (2012-06).", September 2012.

Appendix A.

A.1. Other techniques

- UE can also request bearer resource modification for an E-UTRAN as explained in Section 5.4.5 of [TS23.401]. The procedure allows the UE to request modification of bearer resources (e.g., allocation or release of resources) for one traffic flow aggregate with a specific QoS demand. Alternatively, the procedure allows the UE to request modification of the packet filters used for an active traffic flow aggregate, without changing QoS. If accepted by the network, the request invokes either the Dedicated Bearer Activation Procedure or the Bearer Modification Procedure. However this technique is not widely deployed and only network-controlled quality of service is widely used.

- After certain QoS parameters are established, the UE or the network may want to change those QoS parameters. This is supported in both 3GPP [TS23.401] and PCP FLOWDATA.

- Bearers modification, creation procedures when Application Server like WebRTC is deployed in 3GPP network is explained in section 4.3 of [I-D.reddy-rtcweb-mobile].

- TODO : OneAPI.

Authors' Addresses

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

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

Email: dwing@cisco.com

Bill VerSteeg
Cisco Systems, Inc.
5030 Sugarloaf Parkway
Lawrenceville 30044
USA

Email: billvs@cisco.com

Mohamed Boucadair
France Telecom
Rennes 35000
France

Email: mohamed.boucadair@orange.com