Deliverable DA2.2.23
Policy and Charging Control Functionality with WLAN and PBRM

Janne Tervonen (NSN)

ICT SHOK Future Internet Programme
(ICT SHOK FI)

Phase 2: 1.6.2009 – 31.12.2010

Tivit, Yritysten tutkimus- ja kehittämisrahoitus, Päätös 516/09, 29.5.2009, Dnro 560/31/09

TKK, Tutkimusrahoituspäätös 40212/09, 29.5.2009, Dnro 925/31/09

www.futureinternet.fi

www.tivit.fi

This work was supported by TEKES as part of the Future Internet programme of TIVIT (Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT).

Executive summary / Internal release

Title: Policy and Charging Control Functionality with WLAN and PBRM

This document shortly describes the QoS mechanisms on both LTE / EPS and WLAN and how Policy-Based Resource Management could be used together with existing QoS mechanisms.

Content: Policy and Charging Control (PCC) functionality is an essential part of QoS mechanism of EPS. This document concentrates on PCC and how it can be used with practical WLAN deployments as well as together with PBRM.

Contact info: Janne Tervonen, janne.tervonen@nsn.com

Link: http://www.futureinternet.fi/publications.htm

Table of Contents

Abbreviations and Terminology

3GPP	3rd Generation Partnership Project
AAA	Authentication, Authorization & Accounting
AIFS	Arbitration Inter Frame Space
ANDSF	Access Network Discovery and Selection Function
AP	Access Point
API	Application Programming Interface
APN	Access Point Name
ARP	Allocation and Retention Priority
BBERF	Bearer Binding and Event Reporting Function
BBF	Bearer Binding Function
CDMA	Code Division Multiple Access
CW	Contention Window
DL	Downlink
DPI	Deep Packet Inspection
EDCA	Enhanced Distributed Channel Access
eNB	E-UTRAN Node B
EPC	Evolved Packet Core
EPS	Evolved Packet System
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
GBR	Guaranteed Bit Rate
GPRS	General Packet Radio Service
GTP	GPRS Tunneling Protocol
HCCA	HCF Controlled Channel Access
HCF	Hybrid Coordination Function
HSS	Home Subscriber Server
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
LTE	Long-Term Evolution
MAC	Medium Access Control
MBR	Max Bit Rate
MME	Mobility Management Entity
PDN	Packet Data Network
PDN GW	Packet Data Network Gateway
QCI	QoS Class Identifier
QoS	Quality of Service
P-CSCF	Proxy Call Session Control Function
PBRM	Policy-Based Resource Management
PCC	Policy and Charging Control
PCEF	Policy and Charging Enforcement Function
PCRF	Policy and Charging Rules Function
PMIP	Proxy Mobile IP
RTSP	Real Time Streaming Protocol
SDF	Service Data Flow
S-GW	Serving Gateway
SIP	Session Initiation Protocol
SPR	Subscription Repository Profile
TC	Traffic Category
TFT	Traffic Flow Template

TXOP	Transmission Opportunity
UE	User Equipment
UL	Uplink
WFA	Wi-Fi Alliance
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless LAN
WMM	Wi-Fi Multimedia

1 Introduction

This document is deliverable DA2.2.23 for activity 2.2 Task 7 of Future Internet program of TIVIT. This document covers the continuation of work for Policy-Based Resource Management (PBRM) within Task 7 during 2H2010. Earlier work related to PBRM can be found from other Future Internet deliverables [1], [2], [3], [4], [5] and [6].

This document concentrates on Policy and Charging Control (PCC) and how it can be used with practical WLAN deployments as well as together with PBRM. PCC framework is specified by 3GPP as a core network functionality. As was discussed in [5], PBRM can be realized with Access Network Discovery and Selection Function (ANDSF) in 3GPP environment. Thus, it is assumed in this document that ANDSF is used when discussing PBRM functionality.

PCC was already very shortly described in PBRM deliverable [5]. That introductory discussion is elaborated here for PCC framework as well as for ANDSF role.

Before going to the details of PCC, this document describes the existing quality of service concepts both in 3GPP LTE and WLAN networks. On chapter 3, PCC itself explored. On chapters 4 and 5, PCC usage with WLAN and PBRM is discussed, respectively. Finally, chapter 6 concludes this paper.

2 Quality of Service Concepts

In this chapter, the quality of service (QoS) concepts of two different systems are shortly described in general level: the first system to be considered is 3GPP-defined Evolved Packet System (EPS) and the second system is WLAN. The intention is not to compare or analyze the different Qos mechanisms, just briefly introduce what are the practical possibilities.

In general, QoS mechanisms are used to ensure certain level of service under varying conditions on radio access and core network. QoS mechanisms should allow the operator to enable service and subscriber differentiation and to control the performance experienced by a user. For example, with the adequate QoS mechanisms in place, the operator may provide differentiated treatment of the IP traffic for the same service depending on the type of the subscription the user has: e.g. gold user gets higher level of service than silver user, etc.

Depending on the system, certain level of service may be provided with various approaches: if the cost factors can be excluded, the most straightforward and technically simplest solution is to over-dimension all the critical resources on the network. However, this is very rarely a feasible approach in real life. Instead, to make a QoS mechanism successful – i.e. to make it really deployed in live networks – a delicate balance between the complexity and usefulness of the QoS mechanism need to be found.

2.1 QoS in 3GPP Evolved Packet System

3GPP Evolved Packet System (EPS) consists of both radio access network and core network. In the case of EPS, those are called E-UTRAN (Evolved Universal Terrestrial Radio Access Network) and Evolved Packet Core (EPC), respectively. The radio technology for E-UTRAN is called Long-Term Evolution (LTE), and often E-UTRAN and LTE are used as synonyms in common language to describe the radio access of EPS.

In Figure 6, a simplified general architecture of EPS is shown. Only some of the network elements are visible in the figure, for example AAA and charging related entities have been omitted. The architecture of EPS has been briefly described e.g. in earlier PBRM deliverable [5], and it is not further elaborated here. Good explanation of EPC and the role of the network elements can also be found from [7].

2.1.1 EPS and Bearer Concept

In EPS, the QoS is based on bearer concept: a bearer provides a logical transport channel between two network elements, e.g. the UE and eNodeB (eNB; the LTE base station). Each bearer is associated with a set of QoS parameters that describe the properties of the transport channel, for example bit rates, delay, etc. In EPS, the bearers are organized hierarchically: an EPS bearer represents the logical connection between the UE and PDN GW, i.e. the gateway that is used to access external IP networks. The EPS bearer and the QoS associated to it is realized in practice with another set of bearers between different network elements: radio bearer represents the connection between the UE and eNB, S1 and S5/S8 interface bearers are between eNB and Serving Gateway (S-GW), and S-GW and PDN GW, respectively. This EPS bearer concept is illustrated on Figure 1.

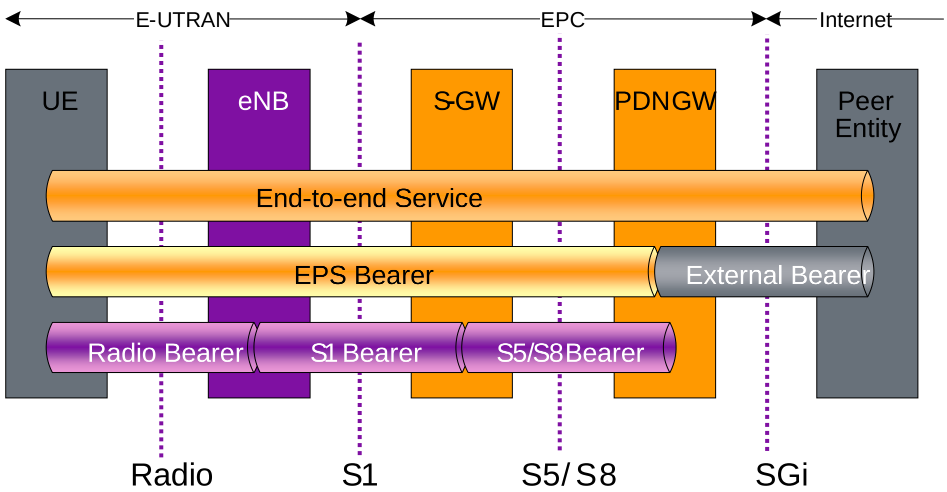

Figure 1. EPS bearer concept.

From the user point of view, an end-to-end connection to its peer entity residing in the Internet is realized with EPS bearer and so called external bearer that is just a logical name for the connectivity between EPC and the peer entity through the Internet. EPS QoS mechanisms can only affect on the transport connections within EPS: other QoS mechanisms may need to be applied for the "external bearer" in order to provide consistent quality of experience for the user. However, the QoS mechanisms outside EPS are not considered here.

In EPS, there are two types of bearers: guaranteed bit rate (GBR) and non-guaranteed bit rate (non-GBR) bearers. In general, GBR EPS bearers are used for services having strict real-time requirements, like voice, live video broadcast or gaming. For other types of services, non-GBR bearers are established.

2.1.2 EPS Bearer QoS Parameters

Each EPS bearer is associated with certain number of QoS parameters. In EPS, non-GBR bearers do only have two QoS parameters: QoS Class Identifier (QCI) and Allocation and Retention Priority (ARP). For GBR bearers, there are also two bit rate QoS parameters defined: guaranteed bit rate and maximum bit rate. Compared to the earlier QoS mechanism defined for 3G systems, a bunch of parameters have been removed: in 3G, there were too many QoS parameters creating overly complex QoS system that was never fully used in practice. In Figure 2, the new EPS QoS parameters are shown next to 3G's parameters.

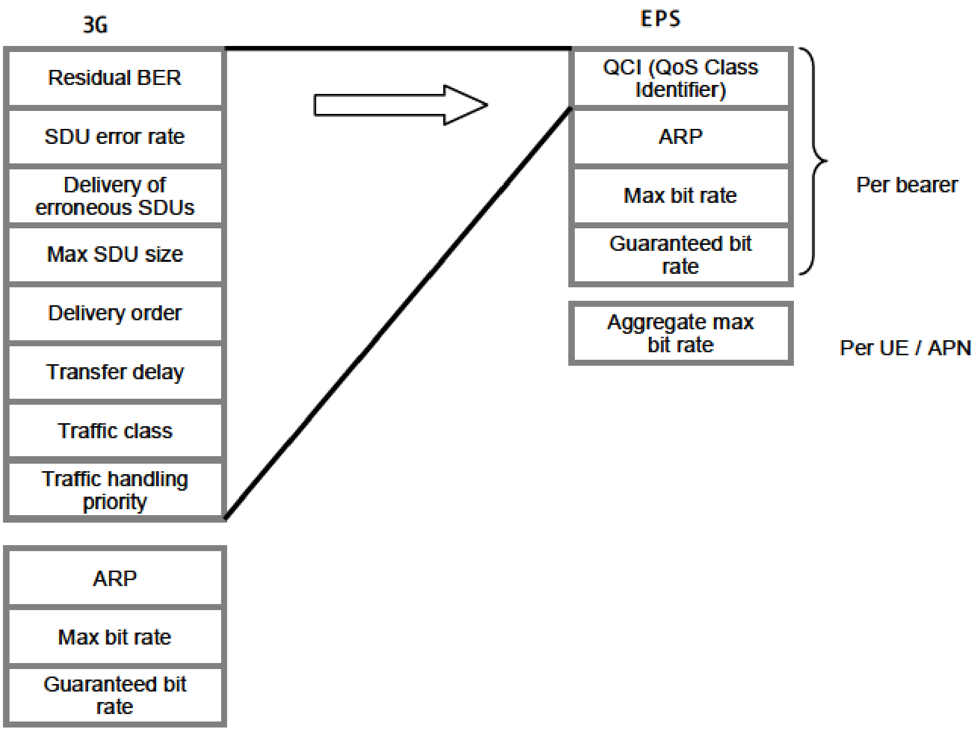

Figure 2. Reduction of QoS parameters from 3G to EPS.

QoS class identifier (QCI): QCI is maybe the most important QoS parameter. QCI itself is just an integer that points to a number of pre-defined, other QoS parameters. Certain QCI values have been standardized to reference specific QoS characteristics. These QoS characteristics describe what packet forwarding treatment the data traffic for that bearer receives edge-to-edge between the UE and PDN GW in terms of certain performance characteristics, such as priority, packet delay budget and packet error loss rate. The standardized characteristics are not signalled on any interface; instead each network element should understand what a certain QCI value means on interfaces of its own. The goal of standardizing QCI values with corresponding characteristics is to ensure that an application or a service receives the same minimum level of QoS in multi-vendor environment and in case of roaming. The standardized QCI values are shown in Figure 3. Each operator may also define new QCI values of its own. However, these operator-specific QCI values are valid only within that operator network.

QCI	Resource Type	Priority	Packet Delay Budget	Packet Error Loss Rate	Application
1	GBR	2	100 ms	10^{-2}	Conversational Voice
2		4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3		3	50 ms	10^{-3}	Real Time Gaming
4		5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100 ms	10^{-6}	IMS Signalling
6		6	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100 ms	10^{-3}	Voice, Video (Live Streaming) Interactive Gaming
8		8	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		9			

Figure 3. Standardized QCI values.

Allocation and retention priority (ARP): In situations where resources are scarce, the network can use the ARP parameter to prioritize establishment and modification of bearers with a high ARP value over bearers with a low ARP value. For example, emergency VoIP call with the highest ARP value should always get established in the network, at the expense of lower ARP value bearers.

Max bit rate (MBR) and guaranteed bit rate (GBR), only for GBR bearers: GBR is the bit rate that EPS will (try to) provide for the bearer, no matter what the situation in the network is. MBR is the upper bound of the bit rate that the bearer may temporarily get. For example, rate adaptive codecs for voice or video services may benefit for having bandwidth at least what GBR defines, and occasionally consume more bandwidth up to MBR.

Aggregate max bit rate: All the QoS parameters described above are specific for only one bearer. If the UE has several bearers active, e.g. due to launching several different types of applications simultaneously, every bearer has its own QoS parameters. In EPS, it is also possible to define what the combined bit rate is for the UE over all the active bearers. This QoS parameter can be used to ensure that single user does not consume too much network resources, e.g. by just establishing several bearers. It is possible to define aggregate max bit rate per UE or per APN (Access Point Name) for that subscriber. In practice in EPS, APN refers to a specific PDN GW, i.e. if the UE indicates it wants to use certain APN, all the traffic from/to the UE is routed through the same PDN GW associated to that APN. Aggregate max bit rate defined for that APN is then the upper limit for all the traffic from / to the UE via the specific PDN GW.

2.1.3 Network-controlled QoS

In EPS, the network makes all the decisions related to QoS and bearer management. UE may still request resources, i.e. initiate bearer establishment or modification procedures, but it is always the network that controls the EPS bearer state and mapping of different traffic to different bearers. The benefit of this mechanism is that the UE does not need to worry about the specifics of the QoS model of the access network. The application in the UE can instead rely on the network to ensure that the access-specific QoS procedures are executed as needed.

This model works very well when the operator controls the service and has full knowledge of the characteristics of the service. For the services not known to the operator, e.g. a service downloaded from the Internet, UE needs to provide the required QoS information to the network. In practice, this can be done either with UE-initiated bearer modification procedures, or by using application-level signalling, e.g. based on SIP or RTSP.

2.2 QoS in WLAN

Standardization of WLAN is conducted in two standardization forums, in IEEE and in Wi-Fi Alliance (WFA). IEEE is taking forward the technical specification of WLAN by adding numerous new features and also improving the WLAN radio interface. WFA – the owner of the trademark Wi-Fi – has secured its place in WLAN ecosystem as a certification organization: in practice, all the WLAN products sold on market today have been certified by at least the mandatory WFA certification programs. After being certified, the WLAN product can carry the Wi-Fi logo(s), and the consumer should be sure that the product is compatible with other similarly certified WLAN products.

Figure 4. Wi-Fi logo for radio specifications 802.11a, b, g and n certified products.

In general, standardization of IEEE for WLAN proceed on its own, new features are added whenever someone thinks it could be useful. This has led into a situation where the IEEE 802.11 series of specifications contain a lot of useful features, but also a lot of not that useful features. This is where WFA steps in: from a bunch of features defined in IEEE 802.11 specifications, WFA picks those that really are required on consumer devices and defines them to be as a required part of a certification program. The general principle of WFA has been that it leaves the actual technical specification for the IEEE, and only concentrates on defining what parts of IEEE 802.11 specification need to be implemented (to get a WFA certificate).

The usage model of WLAN network varies from single AP home network installations to the networks consisting of thousands of APs. Depending also on the services offered with the WLAN network, the QoS requirements for different WLAN network deployments vary considerably. Basically, any WLAN vendors can define their own mechanisms how QoS is handled between the WLAN AP and rest of their network. But on WLAN radio interface, IEEE and WFA specifications have to be followed.

2.2.1 Available QoS Mechanisms on Radio Interface

IEEE 802.11e specification [8] defines improved WLAN MAC (Medium Access Control) mechanism for QoS on top of the original 802.11 features. Basically, 802.11e specification provides two separate mechanisms to realize QoS on WLAN radio interface: so called Enhanced Distributed Channel Access (EDCA) and more sophisticated HCF (Hybrid Coordination Function; a specific MAC access mechanism) Controlled Channel Access (HCCA).

Enhanced Distributed Channel Access (EDCA): EDCA is based on classifying traffic into four different Traffic Categories (TC): voice, video, best effort and background. The basic idea is that the higher priority traffic has higher chance to get so called Transmission Opportunity (TXOP) than lower priority traffic. In practice, transmission opportunity means the time the channel medium is reserved for that station (i.e. UE or access point) to transmit traffic. After the transmission opportunity has expired, the station needs to stop sending any data. All the stations following EDCA channel access are equally contending for getting the TXOP. The basic idea of EDCA for providing QoS is essentially the same as IETF-defined differentiated services (DiffServ).

For each EDCA traffic category, there are four QoS parameters defined: CWmin, CWmax, AIFS and Max TXOP. CWmin and CWmax parameters define the lower and upper bound for the Contention Window (CW) for specific traffic category. Contention window is the time that the station can use to acquire TXOP for the data packets belonging to a traffic category. Higher priority traffic has smaller CWmin and CWmax values than lower priority traffic. AIFS (Arbitration Inter Frame Space) stands for the time after the previous detected transmission the station needs to listen if the channel is idle, i.e. not used. Again, higher priority traffic gets smaller AIFS value than lower priority traffic. Max TXOP defines the maximum duration the station can send traffic without new contention: for lower priority traffic categories default Max TXOP is 0, only voice and video may send more than one packet per one contention.

In Figure 5, the basic principle of EDCA is illustrated. On that figure, there are three different TCs – i.e. traffic categories – contending for channel access. It should be noted that for each TC there is a separate channel access function, i.e. those three TCs shown on the figure may be contending for the channel access in the same device, or e.g. in three different devices. If there is a packet to be sent for a traffic category, the station waits after the previous heard transmission for the duration of AIFS; AIFS is shorter for the higher priority traffic. After AIFS period, the station starts so called backoff timer: initially, backoff timer is set to the value of CWmin for that traffic category. The station decrements backoff timer on every slot when the channel is idle. If the station notices the channel is busy – i.e. some other station or traffic category acquired the TXOP – the station needs to wait again for AIFS after the last heard transmission before continuing to decrement the backoff timer. In the Figure 5, this is denoted with "backoff" for the medium and low priority TC after the high priority traffic acquired TXOP. RTS stands for Request To Send control frame, and CTS for Clear To Send control frame, received from the peer station.

If the backoff timer gets to zero, and the channel is idle, the station acquires the TXOP for that traffic category and it may send the packet. If the backoff timer reaches zero but the channel happens to be busy, the station will not get the TXOP. Instead, a new backoff timer needs to be set: for the next round, backoff timer value gets doubled. However, CWmax value sets the upper bound for the new backoff timer.

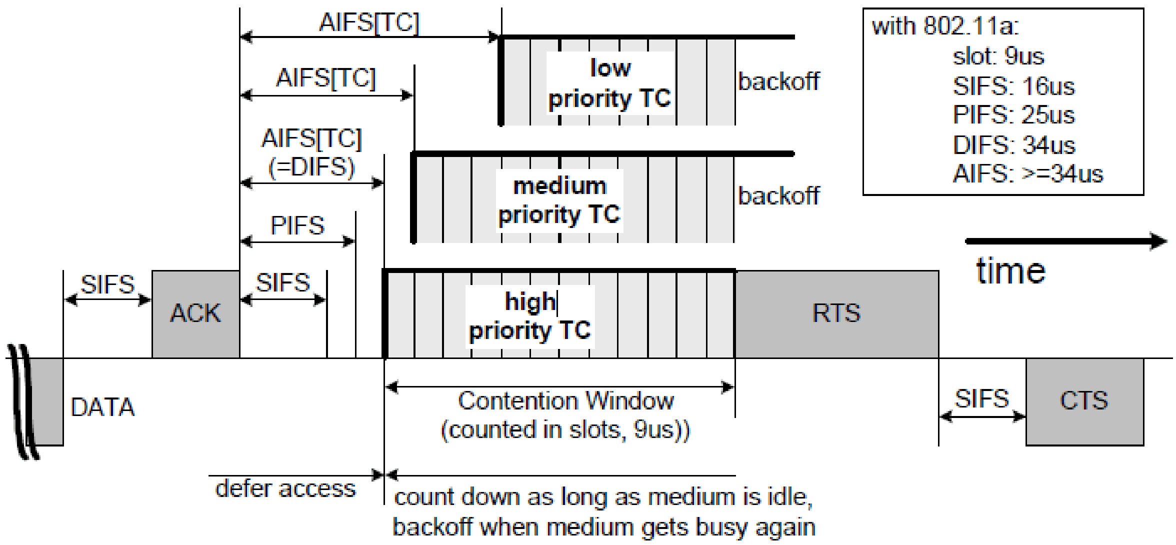

Figure 5. The basic principle of EDCA illustrated for three different traffic categories.

The values of EDCA QoS parameters – Cwmin, CWmax, AIFS and Max TXOP – are not fixed in the specifications, only default values are given. In practice, this means that each vendor or WLAN network operator may set those QoS parameters as he wishes. For a WLAN network operator, this may be beneficial: the QoS parameters can be set so that the WLAN network supports the services the operator is providing, e.g. the parameters are tuned so that VoIP works well in practice in most situations. However, from user point of view, there is a risk that differently configured networks provide completely different user experience: the same application with the same traffic category gets better QoS on a WLAN network, while on another network QoS is considerable worse.

With EDCA, it is possible to define that certain traffic gets better chance for the requested transmission capacity. However, it is not possible to guarantee e.g. VoIP service will work in any network load conditions. With careful network planning and putting an admission control in place (i.e. not letting too many users in), it seems to be possible to build EDCA-based WLAN network that provides adequate QoS also for real time services.

HCF Controlled Channel Access (HCCA): With HCCA, the channel time is divided to contention-free and contention periods. During contention periods, EDCA access mechanism is applied, as described above. For the contention-free period, WLAN AP controls the usage of channel. In practice, WLAN AP can decide when contention-free period is initiated. This can be done for example by sending so called QoS CF-Poll message to a station (UE in this case) that WLAN AP wants to give TXOP. The starting time and the maximum duration of each TXOP is specified by the WLAN AP as signalled in the QoS CF-Poll frames. During the contention-free period, only the WLAN AP can grant TXOPs by sending QoS CF-Poll frames.

Non-AP stations – as UEs are called in IEEE 802.11 specifications – can also indicate to WLAN AP how much traffic they have in their buffers per traffic category. Based on this information, WLAN AP may grant TXOPs to the UEs so that higher priority traffic gets more channel time than lower priority traffic. Basically, it is implementation specific how WLAN AP distributes the TXOPs among the associated UEs.

Since the control of granting TXOPs in HCCA is in a single point (i.e. in WLAN AP), it is possible to build QoS mechanisms for WLAN that correspond the EPS guaranteed bit rate bearer concept. However, HCCA is fairly complex and it requires support from both WLAN AP and the UEs.

2.2.2 WFA and Wi-Fi Multimedia

Wi-Fi Multimedia (WMM) is the certification program of WFA for WLAN QoS mechanisms. Only EDCA is defined as mandatory QoS mechanism for WMM. What this means in practice is that there are no products that would support HCCA. Further, WMM itself is optional certification program for Wi-Fi devices. Currently, most of the new smart phones do support WMM, but for example the latest iPhone is the only iPhone having WMM support.

It is not known to the author what the ratio is between WMM-capable and non-WMM capable deployed WLAN networks in the world. However, it is probably safe to assume that there are a large number of installed WLAN networks that do not have any QoS mechanisms available. Thus, even if the consumer's UE supported WMM, the QoS mechanisms may or may not be available in a WLAN network that UE connects to.

3 3GPP Policy and Charging Control Overview

In the previous chapter, the available QoS mechanisms for 3GPP EPS and WLAN were described. This chapter will concentrate how EPS can realize the QoS mechanisms for 3GPP networks: for that, Policy and Charging Control (PCC) functionality is introduced.

Initially, PCC was brought into 3GPP core network in Release-5 in 2002. The objective was to provide for the operators service-based end-to-end QoS mechanisms for IP Multimedia Subsystem (IMS) services run on top of GPRS network. In practice, IMS can be used to provide e.g. voice services and related supplementary services over packet networks, although before LTE IMS is not that widely used in practice.

EPS and LTE were initially defined in 3GPP Release-8 that was officially finalized at the end of 2008. As was described in chapter 2.1.3, EPS only supports network-controlled QoS paradigm, where the network – i.e. EPC – makes the decisions to establish or modify a bearer based on the provided QoS parameters. In EPC, PCC is the functionality that enables centralized control on QoS.

3.1 PCC Architecture

PCC provides operators advanced tools for service-aware QoS and charging control. The main function of the PCC is to manage QoS of each connection established through the EPC, and also to ensure that the agreed QoS level is maintained in spite of changing conditions (e.g. handovers). PCC is also used for various charging-related tasks, e.g. to check in real-time that there are enough credits to establish a new connection, etc. However, charging aspects of PCC are not further considered in this document.

The main functional entities of PCC are shown in Figure 6. The figure is a bit simplified, for example the interfaces to charging functions have been omitted. In EPC, it is possible to use different mobility protocols in EPC internal interface: for example, the S5/S8 interface between S-GW and PDN GW can either be based on GTP or PMIP. If GTP is used, the EPS bearers are terminated in the PDN GW, and thus PDN GW can use the bearer procedures to control the EPS bearers. In this case from PCC point of view, it is enough to have PCEF functionality in PDN GW. But if PMIP is used on S5/S8, PDN GW loses the visibility for single bearers: only one PMIP tunnel is established between the S-GW and PDN GW, and this tunnel carries all the traffic. With PMIP variant, EPS bearers effectively terminate at S-GW (instead of PDN GW), and for PCC operations a specific functionality in S-GW is needed: this is called Bearer Binding and Event Reporting Function (BBERF). However, BBERF is not visible in the figure below. For the sake of simplicity, the remainder of this chapter concentrates only on the scenario where BBERF is not present. Thus, it is assumed that GTP is used on S5/S8 interface between S-GW and PDN GW.

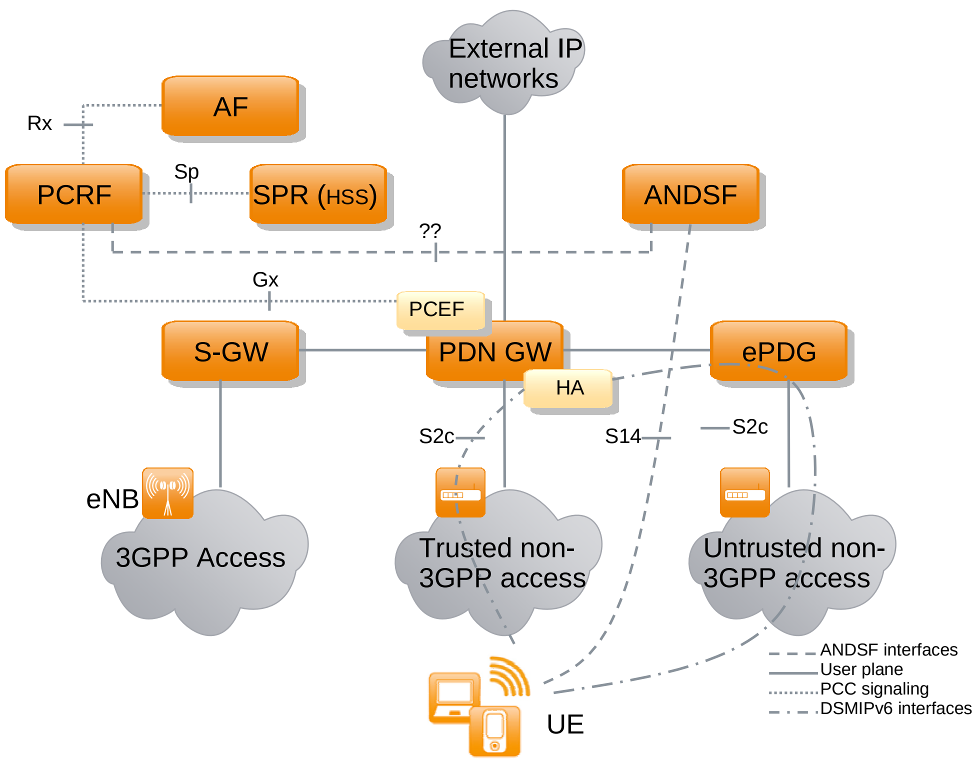

Figure 6. Simplified EPS architecture together with general PCC architecture.

There are several different usage models for PCC. Depending on the use case, not all the functional elements are required. For example, if application-level signalling is used between the UE and EPC to initiate EPS bearer setup, Application Function (AF) can be used to provide QoS information to PCRF about the application to be set up. Such application level signalling could e.g. be SIP signalling when using IMS services. For the IMS, AF corresponds to the P-CSCF function. If the operator does not have need for the AF (i.e. there is no need for forwarding application-level information to PCC), it can be omitted from the PCC deployment.

Policy and Charging Rules Function (PCRF) is the central PCC architecture entity, and it can be considered as the control function for the whole PCC. PCRF may receive session information from AF over Rx interface as well as information from the access network via Gx interface from Policy and Charging Enforcement Function (PCEF). Also, PCRF may receive subscription information from the Subscription Repository Profile (SPR). The PCRF takes the available information, as well as configured operator policies, into account and creates service-session level policy decisions. These are called "PCC decisions". The decisions are then provided to the PCEF (and the BBERF, if that is used). Another task of PCRF is to forward event reports between PCEF and AF, e.g. to notify AF that PCEF has detected a subscriber to exceed monthly quota, etc. In practice, PCRF is a standalone server that is marketed to the operators as part of the complete PCC offering.

Policy and Charging Enforcement Function (PCEF) resides in PDG GW. The PCEF enforces PCC decisions (e.g. maximum bit rate policing) received from the PCRF. Also, the PCEF may perform measurements of user plane traffic (e.g. traffic volume per bearer or per user, duration of the sessions, etc.) and report them to the charging functionality or to PCRF.

Subscription Repository Profile (SPR) contains subscription information, such as user specific policies and data. SPR can be an independent repository as part of PCRF, e.g. for small scale deployments, or SPR can be realized as Home Subscriber Server (HSS) for regular mobile operators.

Not officially part of PCC framework, but in practice used also with PCC is so called DPI (Deep Packet Inspection). DPI refers to inspection of data packet non-header part e.g. for detecting certain traffic type. In practice, DPI is part of some gateway functionality, e.g. integrated into PDN GW. When residing in the same network element as PCEF, it is possible to use PCC signalling from PCRF to control also the usage of DPI functionality.

3.2 PCC Decisions and Rules

The PCRF is in charge of making PCC decisions. These decisions define how the PCC framework treats certain service data flow or e.g. certain user. The decisions can be based on input from a number of different sources, e.g.:

- Operator static configuration in the PCRF that defines the policies applied to given services
- Subscription information for a given user, received from the SPR (e.g. HSS)
- Information about the services received from the AF

- Information from network, e.g. from PCEF, that indicates a change in the existing PCC decision is needed, e.g. transfer volume quota exceeded, etc.

Based on the available information, the PCRF provides its decision in the form of so called "PCC Rules". A PCC rule contains a set of information that is used by the PCEF and the charging system. In the PCC rule, there are three main components: service data flow detection, charging information and policy control. In this document, we concentrate on QoS aspects of PCC, so charging information is not further considered.

Service Data Flow (SDF) detection allows the PCEF to identify the IP packets that belong to a specific service session. SDF template is used to identify the service data flows. SDF template contains a description of the IP flow and typically consists of source and destination IP addresses, source and destination port numbers and the protocol type used in the data portion of the IP packet. These five parameters are often referred to as IP 5-tuple. It is also possible to define SDF with only some parameters of the 5-tuple.

Policy control information of the PCC rule consists of so called gate status and a number of QoS parameters: QCI, MBR, GBR and ARP. The gate status indicates whether the virtual gate for that service data flow is open or closed: if the PCRF decides that the "gate" for a service data flow is closed, it updates the PCC rules in PCEF (i.e. in PDN GW). When PCEF receives a packet matching to a SDF template – i.e. a 5-tuple – whose gate status is set to closed, the PCEF will drop the packet. The virtual gate for a SDF could be set to "closed" e.g. if pre-paid quota has been already consumed, or type of traffic is not allowed in the operator network, etc.

The general purpose and usage of QoS parameters was discussed in chapter 2.1.2. However, there is one important difference between EPS bearer and PCC rule QoS parameters: the QoS parameters of a PCC rule are applied only to a specific service data flow (identified with SDF template, i.e. 5-tuple or a number of 5-tuples), but EPS bearer QoS parameters are applicable for an EPS bearer. A single EPS bearer may be used to carry traffic of several service data flows. In practice with PCC rule QoS parameters, it is possible to have finer-grained QoS mechanism: unique QoS characteristics can be applied for each specific service data flow, e.g. application. On the following table, the PCC rule QoS parameters are shown.

Table 1. PCC rule QoS parameters [10].

PCC Element	Description
Rule identifier	Uniquely identifies the PCC rule. It is used between PCRF and PCEF for referencing PCC rules.
Service data flow detection	*This clause defines the method for detecting packets belonging to a service data flow.*
Precedence	Determines the order, in which the service data flow templates are applied at service data flow detection.
Service data flow template	A list of service data flow filters for the detection of the service data flow.
Policy control	*This clause defines how the PCEF shall apply policy control for the service data flow.*
Gate status	The gate status indicates whether the service data flow, detected by the service data flow template, may pass (Gate is open) or shall be discarded (Gate is closed) at the PCEF.
QoS class identifier (QCI)	Identifier for the authorized QoS parameters for the service data flow.
UL and DL maximum bitrates	Maximum bit rates authorized for the service data flow in UL and DL.
UL and DL guaranteed bitrates	Guaranteed bit rates authorized for the service data flow in UL and DL.
ARP	The Allocation and Retention Priority for the service data flow consisting of the priority level, the pre-emption capability and the pre-emption vulnerability

It is possible to deploy PCC as static or dynamic. Static PCC means that relevant network elements are configured with some pre-defined QoS parameters, i.e. there is no dynamic information exchange between different PCC functional elements. In practice, some set or all of the above described PCC rule QoS parameters are configured to PCEF for certain type traffic. PCEF then treats traffic matching to the SDF template according to the pre-defined parameters. In this document, static PCC is not further considered, i.e. this document concentrates on dynamic PCC.

3.3 PCC Usage

So how is the PCC used in practice? In this chapter, some PCC usage is considered from few practical use scenarios' point of view.

3.3.1 Session Establishment with PCC

When a user decides to launch an application (or responses to service request from a peer entity, e.g. accepts incoming VoIP call), the UE has to indicate to the network that a new service is being launched. With PCC, there are two ways to do it: either the application-level signaling is used via AF (refer to chapter 3.1 PCC entities), or normal LTE bearer signaling is used. Both of these options are shortly described below.

Using of Application Function (AF) to signal the QoS requirements to PCRF is illustrated in Figure 7 below. Each step is shortly explained in the following.

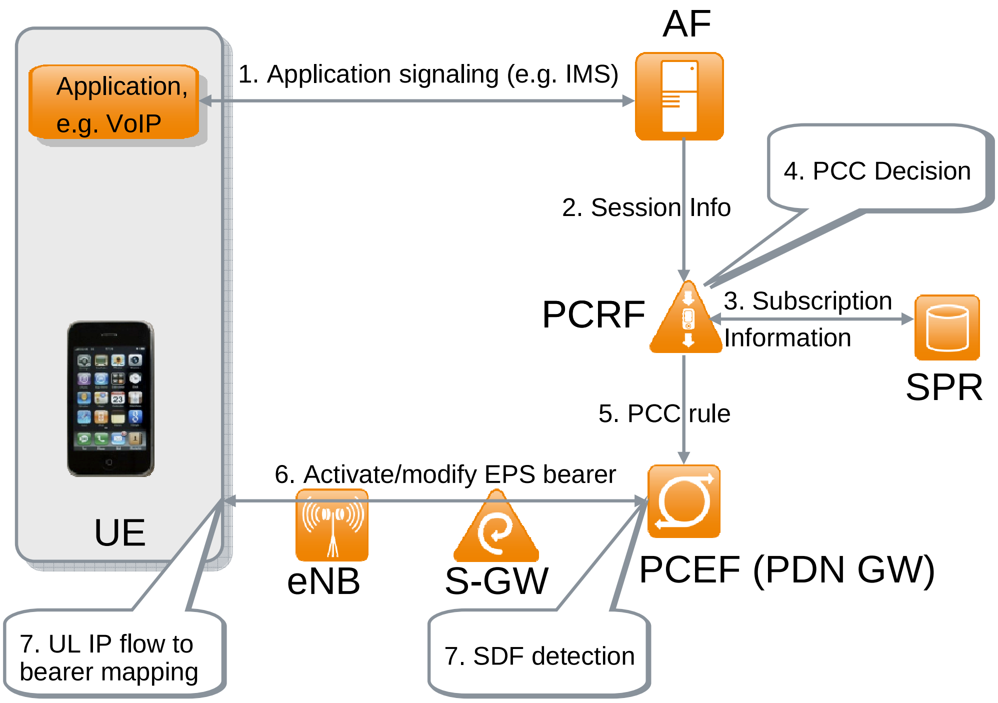

Figure 7. PCC signaling with Application Function.

1. The user initiates a VoIP call that is realized in this example with IMS. End-to-end VoIP call establishment signalling is performed. AF (P-CSCF in the case of IMS) functionality captures QoS related information from this signalling.

2. Based on the information gathered during VoIP call setup, the AF provides the PCRF with the service-related information over the Rx interface. This information typically includes QoS information (type of service, bit rate requirements) as well as traffic parameters (like the 5-tuple to identify the traffic).

3. The PCRF may request subscription-related information from the SPR (e.g. HSS), like what services are allowed for the subscriber, etc. PCRF may also have been performed this step earlier, independently of the application setup.

4. The PCRF takes the session information received from AF, operator-defined service-polices and subscription information from SPR into account when making the PCC decisions. The results of the decision are PCC rule(s).

5. The PCC rules are sent by the PCRF to PCEF (PDN GW) in Policy and Charging Rules Provision message. The PCEF will enforce the policy decisions according to the received PCC rule(s). All the user plane traffic for a user passes through PCEF (assuming traffic is routed via EPC, i.e. not offloaded from the core network). Thus, PCEF can monitor all the traffic sent between the UE and core network.

6. The PCEF (PDN GW) installs the PCC rule(s) and performs so called bearer binding to ensure that the traffic for this service receives appropriate QoS. Depending on the result of the bearer binding, a new bearer may be established, or an existing one is modified to accommodate the new traffic. Bearer management procedures for LTE are described in 3GPP TS 23.401 [11]. PCC bearer binding operation is shortly described on 3.3.2.

7. After the successful VoIP call setup and the associated bearer establishment or modification, the application data transfer can start. During a new bearer establishment or modification of an existing bearer, UE receives so called Traffic Flow Templates (TFT) that effectively contain the same information as SDF detection (i.e. 5-tuple). With TFT, UE is able to map in uplink direction certain service data flow to its associated LTE radio bearer. On the downlink direction, PCEF will perform SDF detection (with SDF template, including 5-tuple) to detect the IP flow for this service. Based on the SDF detection, the IP flow is transported over the appropriate bearer towards the UE.

Next for the scenario when AF is not used. There can be several cases when the AF is not available or it cannot be used: for example, the operator may have decided that AF is not needed, or the user is launching such a service that does not support the application-level signalling so that AF could be utilized (e.g. the user has downloaded an app from the Internet). For these cases, UE is required to initiate bearer resource modification, following the normal bearer management procedures, as defined 3GPP TS 23.401 [11]. This also means that either the UE or the application can provide the QoS information (e.g. via a specific API) for UE's bearer management module that can further provide this information to the network and PCC. In Figure 8, it is shown the PCC signalling for the scenario where UE-initiated bearer modification request triggers the PCC procedures (i.e. when no AF is used). The steps involved in this scenario are shortly explained below.

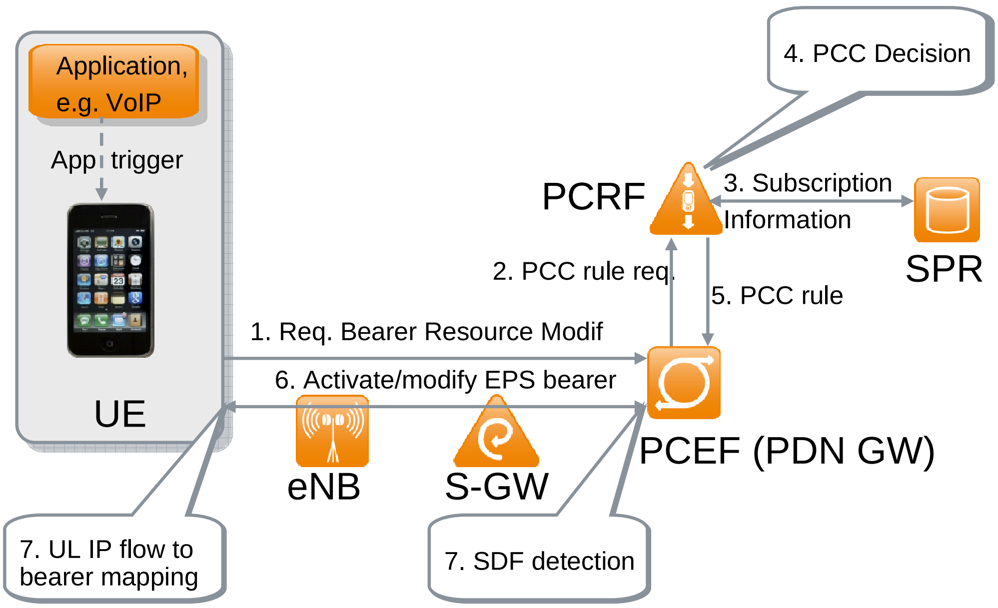

Figure 8. PCC signaling without Application Function.

1. The user launches an application that is e.g. downloaded from the Internet. The application does not support application-level signalling so that AF could be used. Thus, the application needs to be able to indicate its QoS requirements to the UE's entity that is responsible for bearer management. After having gathered the necessary information, the UE sends Request Bearer Resource Modification message to the network. Among other things, that message contains the requested QoS parameters. The MME (Mobility Management Entity) network element (not visible in the figure) receives the message from the UE, and forwards its contents to PDN GW / PCEF.

2. The PCEF requests for PCC rules from PCRF by sending Indication of IP CAN Session Establishment message with the associated information (e.g. QoS parameters).

After the step 2, the procedure continues as with the AF scenario, i.e. the steps from 3. to 7. are identical in both scenarios.

In the examples above, only the session establishments have been considered. Similar mechanisms have been defined for modifying an existing session as well as terminating the session. Those scenarios are not further elaborated here.

3.3.2 Bearer Binding

The PCC rules need to be mapped to a corresponding bearer in the access network to ensure that the packets receive appropriate QoS treatment. This mapping is one of the central components of PCC. The association between a PCC rule and a bearer is referred to as bearer binding. The bearer binding is done by the Bearer Binding Function (BBF) that is located either in the PCEF or in the BBERF (BBERF is used with LTE radio access only when PMIP is used on S5/S8 interface). When the PCEF receives new or modified PCC rules, the BBF evaluates whether or not it is possible to use the existing bearers. If one of the existing bearers can be used, e.g. if a bearer with corresponding QCI and ARP values already exists, the BBF may just initiate bearer modification procedures to adjust e.g. the bit rates of that bearer. If it is not possible to use any existing bearer, the BBF – i.e. PDN GW – initiates the establishment of a suitable new bearer. For PCC rules with GBR requirements, the BBF must also ensure that the authorized QoS of the PCC rule can be provided through the network.

When using LTE radio access, the BBF relies on the EPS bearer procedures when establishing or modifying a bearer for PCC rules activation. It is the task of the BBF to interact with the appropriate EPS QoS procedures in order to ensure the appropriate QoS all the way between the UE and PDN GW. As described in chapter 2.1.1, EPS bearers are in practice realized with interface-dedicated bearers: for example, radio bearer needs to be setup on radio interface, etc.

Since the bearer procedures are triggered from BBF (i.e. PCEF / PDN GW), there is no visibility of the available radio interface resources for the PDN GW; this information in LTE is only kept within eNB. Thus, when PDN GW initiates a bearer establishment or modification for the PCC rule activation/change, PDN GW can only request Serving GW and eNB to establish suitable bearers on corresponding interfaces. If it is not possible to establish e.g. the requested bearer on radio interface due to lack of resources, eNB will indicate back to PDN GW that the bearer establishment / modification was a failure. Then it is up to PDN GW to forward this information also to PCRF that decides how to proceed (e.g. reject the service establishment, etc.). In the signaling figures above, it was assumed that the step 6. was successful for all the required bearer operations.

3.3.3 PCC Rule Enforcement

After the application connection has been successfully setup as described in chapter 3.3.1, PCC should take care of ensuring QoS also for the lifetime of the application. This is the task of PCEF: since all the traffic from/to the UE is going through PCEF / PDN GW, PCEF is able to monitor that certain service data flow really gets QoS that it should. For example, if a dedicated bearer is setup for a service that requested guaranteed bit rate service, PCEF is able to detect if EPS bearer is not any more able to provide enough bandwidth. After such an event has been detected in PCEF, the PCEF may send an indication to PCRF for further actions.

As such, PCC framework cannot influence how eNB realizes radio interface QoS: during a service and corresponding bearer establishment, PCC signals the requested QoS parameters also for the corresponding radio bearer establishment, according to EPS bearer model. After that, it is up to the eNB to ensure that it fulfills the QoS requirements. If for some reason – e.g. due to congestion, or the UE entered tunnel – it is not anymore possible, PCEF can only detect the situation and inform PCRF about it. If the service is such that it cannot work with lower QoS parameters, in practice the only possibility for the PCRF is then to initiate release of the connection (assuming no other network element has already initiated the release).

However, for the opposite direction PCC can effectively be used to enforce QoS: if for example certain user consumes too much traffic, PCEF can detect that the aggregate max bit rate (refer to chapter 2.1.2) gets too high for that user and PCEF / PDN GW can start to limit the bit rate the user gets. In practice, this can be done by making the user’s packets to wait in PCEF / PDN GW queues longer than other users’ traffic.

Also, with PCC it is possible to implement mechanisms that allow the operator to define certain monthly user-specific quota e.g. for flat-rate subscribers. When the subscriber has consumed all his quota (e.g. 5 GB per month), PCEF starts applying considerably lower maximum bit rate for that user. When this happens, it is also possible to redirect the user to a web portal that user can use to buy extra quota and get his maximum bit rate QoS parameters back to the original level.

There are several use cases that PCC framework enables. Above, only few of them have been mentioned. Apart from providing the basic QoS signaling mechanisms for LTE-based networks, the use cases for PCC are mostly for controlling and monitoring the traffic going through the operator’s network.

As a summary, it could be said that the main use of PCC for the operator is to enable efficient tools for traffic engineering. In addition to that, PCC also provides the mechanisms for ensuring QoS in EPS networks.

4 Policy and Charging Control with WLAN

In 3GPP TS 23.402 [12], it is defined how PCC can be used together with non-3GPP accesses. In this context, non-3GPP access refers to everything else than 3GPP-defined accesses, i.e.

CDMA2000, WiMAX and WLAN. Both CDMA2000 and WiMAX define their own bearer models, and those can be integrated with PCC at least to some extent. However, for the WLAN the situation is a bit different.

In order to exploit all the QoS mechanisms of PCC with WLAN, the WLAN should support some kind of bearer model equivalent to EPS bearer model. As was discussed in chapter 2.2, only the EDCA channel access mechanism is the mandatory feature of the (optional) Wireless Multimedia (WMM) certification program of WFA. In practice, this means that no QoS reservation mechanism do exist in the current or near future WLAN network or terminal equipment. Thus, PCC cannot be used with WLAN to ensure certain level of QoS for services run over the WLAN access.

Although PCC QoS ensuring mechanisms cannot be used with WLAN access, there are a few aspects in PCC that are either independent of access or related to the interworking between EPS and the access-specific QoS mechanism. When the UE is accessing the services over WLAN access via EPC core (i.e. traffic is not offloaded directly to the Internet from WLAN access), an APN – i.e. PDN GW – needs to be assigned also for WLAN traffic. One QoS parameter that is common to all access (connected to EPC) is the APN-AMBR (APN Aggregate Maximum Bit rate, refer to 2.1.2). The APN-AMBR is enforced by the PDN GW and can be enforced independently of which access the UE may be using.

In addition to limiting the maximum bit rates for WLAN access, PCC framework could also be used for subscriber-specific policies (from SPR): e.g. the same monthly quotas could be applied to subscribers no matter what is the access they are using, PCEF "gate" is set to closed for certain type of traffic, etc. In general, the traffic engineering mechanisms that are realized only with the PCEF (i.e. PDN GW) can also be used with WLAN access, when the traffic is routed via EPC. Also, the PCC charging functions could be applied for WLAN in the similar way as they applied for 3GPP access (charging issues not further elaborated in this document).

5 Policy and Charging Control and PBRM

As was discussed in [5], PBRM concept can be realized in 3GPP domain with Access Network Discovery and Selection Function (ANDSF). Since this document has concentrated on 3GPP PCC framework, also PBRM functionality is considered from 3GPP and ANDSF point of view. Currently in 3GPP, ANDSF has been defined as an independent function from other core network elements, i.e. there are no standardized interfaces between ANDSF and EPC core network elements (excluding possibly ANDSF – HSS interface in future). In past, there have been some discussions whether there should be an interface between ANDSF and PCRF, but it was agreed that there is no need to standardize such an interface.

However, that does not prevent network equipment manufacturers to implement their own proprietary interfaces for ANDSF. The basic idea of ANDSF is to provide static network selection policies to the UEs. After receiving these policies, UEs may use it to facilitate the upcoming network selection decisions. Since the information ANDSF provides is static in nature, there is no need to provide information from ANDSF to PCC. However, the information PCC gathers could be used to modify the network selection policies ANDSF provides. In below, one possible scenario for combining PCC framework and PBRM (i.e. ANDSF) is shortly discussed.

As shown in Figure 9 below, in this scenario PDN GW has a built-in deep packet inspection (DPI) function. The idea is to use DPI to detect the services or applications that are consuming most of the network resources: in this case, it is assumed that the LTE is the scarce resources that should be protected from overload. For example, if there is a lot of peer-to-peer traffic, DPI can detect that from the contents of peer-to-peer data packets and provide information about it e.g. to PCRF. Here it is assumed that such information is possible to deliver from PCEF to PCRF, either using standardized or proprietary interfaces. Also, if there is a lot of YouTube traffic on the network, it is possible to detect that e.g. by screening source addresses of the data packets going through PDN GW towards the UEs. The idea here is to detect only few of the most resource-consuming applications: it is not necessary to identify every possible application running through the EPC, only the most resource-heavy. This identification is denoted with step 1 on the figure.

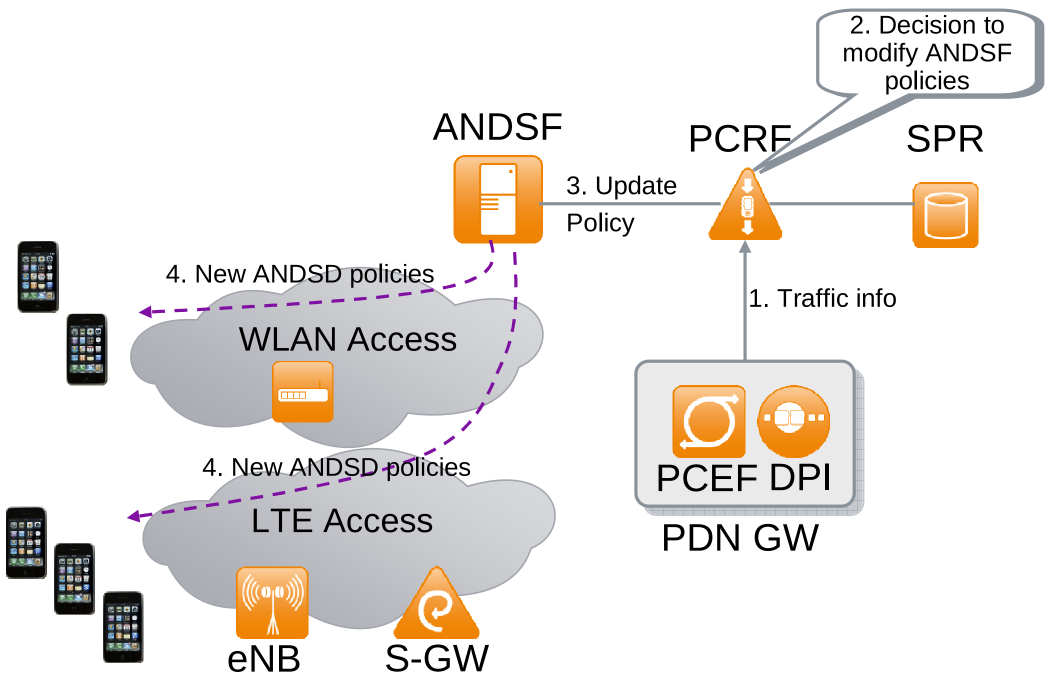

Figure 9. ANDSF interworking with PCC.

When the PCRF has received information from PCEF and/or DPI that certain applications are consuming excessive amount of network resources, PCRF can try to figure out how to solve the situation. It should be noted that this logic may reside also in another network element, e.g. in a stand-alone server, but here it is assumed that PCRF functionality is extended to have the needed logic. In this situation, PCRF could, for example, make new PCC decisions that limit the bandwidth for those couple of most consuming applications and send corresponding PCC rules to PCEF. However, here PCRF decides that another mean is more efficient: PCRF is aware of the existing general ANDSF network selection information (due to the proprietary interface between the ANDSF and PCRF), and PCRF notices that it could modify ANDSF policies so that those few resource-heavy applications are offloaded from LTE access to use WLAN. ANDSF Release10 specification allows this: it is possible to define application-specific policies that effectively guide the traffic of that application into a specific WLAN network

After making the decision illustrated on the step 2 on the figure, PCRF informs ANDSF about the modified ANDSF network selection policies (step 3 on the figure). It should be emphasized here that the interface between PCRF and ANDSF is not standardized, i.e. it is based on proprietary mechanisms. When ANDSF receives this information from PCRF, it will modify its network selection policies accordingly. After that, ANDSF may use push mechanism to update the policies immediately on the UEs, or wait that each UE contacts ANDSF for a new policy update. When the UEs get the updated policies, they can start using to offload e.g. all peer-to-peer and YouTube traffic over specific WLAN networks, when available.

With this kind of mechanism, it is possible to build network functionality that allows dynamic steering of traffic between different accesses, namely between 3GPP and WLAN accesses.

6 Conclusions

In this document, the 3GPP Policy and Charging Control framework has been discussed. Also, the underlying QoS mechanisms of two different radio access networks – LTE and WLAN – were described. EPC and LTE access can provide tools for defining both non-guaranteed and guaranteed bit rate bearers. Together with PCC, this provides an efficient mechanism for the operator to manage the QoS that the different applications require.

However, due to the characteristics of WLAN, there is no corresponding bearer and QoS ensuring mechanisms available in WLAN. Thus, the usage of WLAN together with PCC is much more limited than with EPC and LTE. Still, operators can benefit from PCC also with WLAN: it is possible to define e.g. bit rate limiting PCC rules also for WLAN traffic.

From operator point of view, there are two main benefits from PCC: it can be used to ensure QoS for a single user or even a single application, but it is also a useful tool for traffic engineering in EPC and LTE networks. PCC enables many use cases that the operators benefit in managing the traffic flowing through its networks.

For PBRM (and ANDSF), the PCC could provide a dynamic mechanism for managing different access network usage: with the information received from PCC framework, it is possible to dynamically influence what radio access network is used for what traffic.

7 References

[1] Policy-Based Resource Management, Future Internet Program deliverable 2.4.1, version 1.0, Janne Tervonen, 19.12.2008.

[2] Network Selection Simulations, Future Internet Program deliverable 2.4.3, version 1.0, Pasi Seittenranta, Sami Heinonen, Timo Ketonen, Janne Tervonen, 31.5.2009.

[3] Automated PBRM Server Configuration, Future Internet Program deliverable 2.4.2, version 1.0, Sami Heinonen, Janne Tervonen, 26.6.2009.

[4] Service Based Access Selection with PBRM, Future Internet Program deliverable 2.2.21, version 1.0, Pasi Seittenranta, Janne Tervonen, 4.6.2009.

[5] Realization of Policy-Based Resource Management Concept, Future Internet Program deliverable 2.2.20, version 1.0, Janne Tervonen, Jari Mustajärvi, 16.02.2009.

[6] Offloading Traffic from Cellular Networks with PBRM, Future Internet Program deliverable 2.2.22, version 1.0, Janne Tervonen, 1.2.2011.

[7] SAE and the Evolved Packet Core, Driving the Mobile Broadband Revolution, Magnus Olsson, Shabnam Sultana, Stefan Rommer, Lards Frid, Catherine Mulligan, Academic Press, 2009.

[8] IEEE Std 802.11e™-2005, IEEE Standard for Information technology — Telecommunications and information exchange between systems — Local and metropolitan area networks — Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements, 11 November 2005.

[9] IEEE 802.11e Wireless LAN for Quality of Service, Stefan Mangold, Sunghyun Choi, Peter May, Ole Klein, Guido Hiertz, Lothar Stibor,

[10] 3GPP TS 23.203, Policy and charging control architecture (Release 10), v10.1.0, September 2010.

[11] 3GPP TS 23.401, Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 10), v10.1.0, September 2010.

[12] 3GPP TS 23.402, Architecture enhancements for non-3GPP accesses (Release 10), v10.1.0, September 2010.