REFERENCE ARCHITECTURE FOR END-TO-END QOS IN HETEROGENEOUS WIRELESS NETWORK ENVIRONMENTS

Sandra Frei[1, 2], Woldemar Fuhrmann[3], Andreas Rinkel[2] and Bogdan Ghita[1]

[1]Centre for Information Security and Network Research, University of Plymouth, Plymouth, United Kingdom
sandra.frei@plymouth.ac.uk
bogdan.ghita@plymouth.ac.uk
[2]University of Applied Sciences Rapperswil, Rapperswil, Switzerland
a.rinkel@hsr.ch
[3]University of Applied Sciences Darmstadt, Darmstadt, Germany
w.fuhrmann@fbi.h-da.de

ABSTRACT

The first step in providing end-to-end QoS in a wireless network is to model the current environment and identify the state of the art in QoS methods. The paper provides a reference architecture for end-to-end QoS in heterogeneous wireless network environments by giving an overview of the reference architecture and showing a detailed end-to-end session establishment including the necessary communication between service and transport layer. To estimate and compare the complexity of QoS signalling, an appraisal of the QoS signalling effort is given.

KEYWORDS

Reference Architecture, End-to-End QoS, Evolved Packet System, IMS, Policy Charging and Control

1. INTRODUCTION

One fundamental feature for next generation mobile networks (NGMN) should be their ability to integrate different wireless technologies, provide QoS and enable the integration in general application landscapes. This will demand significant flexibility from the next generation networks, a fact which will also lead to a correspondingly higher complexity. The control plane is responsible for establishing the end-to-end connection and therefore all underlying bearers have to be established to facilitate the sending of data through the user plane. Because the signalling is spanned over multiple layers, additional cooperation and processing between them is necessary to establish an end-to-end session. To support the cooperation between these layers and to reduce the signalling expense, a policy framework is needed. For evaluation purpose of complexity and signalling expense of upcoming end-to-end QoS solutions from further studies, a reference architecture for end-to-end QoS in heterogeneous wireless network environments is described in this paper on the basis of detailed sequences.

The paper is organized as follows: the next section provides an overview of the reference architecture by introducing the main components involved in the end-to-end QoS procedure. Section 3 introduces the mechanisms and components necessary to set up an end-to-end multimedia session within the reference architecture. The following section provides the establishment of a dedicated bearer with appropriate QoS. Section 5 gives a functional overview of the Policy Charging and Control (PCC) framework containing an estimation of the QoS signalling effort followed by section 6, presenting the conclusions of the paper and further work.

2. OVERVIEW OF THE REFERENCE ARCHITECTURE

It is expedient to build the reference architecture for end-to-end QoS upon 3GPP standards because these are approved and with the release 8, a basis is created which takes into account one fundamental feature: the ability to integrate different wireless technologies into the QoS concept. Such an integration of a wireless technology into the QoS concept is given with the integration of WiMAX [1]. A general description of the applied QoS control within the transport layer from the release 8 is provided in [2]. The used policy and charging framework evolved significantly from the previous release 7 to the new release 8 by supporting multiple access technologies [3]. An analysis of the necessary QoS signalling over the integrated Radio Access Network (RAN) and IP Multimedia Subsystem (IMS) is provided in [4] by using a service example. The signalling analysis is based on previous releases and only takes into account one of the policy and charging framework components, the Policy and Charging Rules Function (PCRF). As a basis for further studies, the reference architecture is analysed with detailed sequences of several QoS related scenarios within this paper. Furthermore an appraisal of the QoS signalling effort within the policy framework is provided.

The reference architecture describes services and therefore end to end network connectivity to heterogeneous wireless access networks is needed. This description allows full independence from the technology that is used in the access networks.

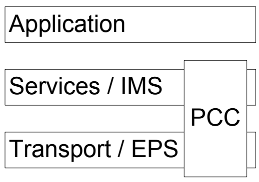

Figure 1 Overall view of the reference architecture

Figure 1 shows the general view of the reference architecture containing the three functional layers: application, services, which may contain the IP Multimedia Subsystem (IMS) (defined in [5]) and transport, which consists of the Evolved Packet System (EPS). The top layer is the application layer which provides the applications from the Public Land Mobile Network (PLMN) operators. The service layer enables PLMNs to offer IP Multimedia (IM) services and therefore defines the necessary elements and mechanisms.

A detailed view of the reference architecture is presented in Figure 2. To support IM services the transport layer has to establish suitable bearers with appropriate QoS to transport multimedia data traffic. The bearer establishment is done through additional signalisation between the service and transport layer by using the Policy and Charging Control (PCC) framework (defined in [6]) as a mediator between the two layers. Three main components are involved in communication between the two layers: the Application Function (AF), located at the Proxy-Call Session Control Function (P-CSCF) in the IMS, the Policy and Charging Rules Function (PCRF) and the Policy and Charging Enforcement Function (PCEF) which is located in the EPS network. This relationship is also shown in Figure 2. The PCRF coordinates the establishment of IP-Connectivity Access Network (IP-CAN) bearers by processing incoming bearer establishment requests from the IMS related P-CSCF. The PCC contains the two main functions:

- Charging the Service Data Flows (SDF) through Online and Offline Charging System
- Policy Control which includes QoS control and signalling, gating control etc.

Since the focus of the paper is QoS the charging of the SDFs is out of scope.

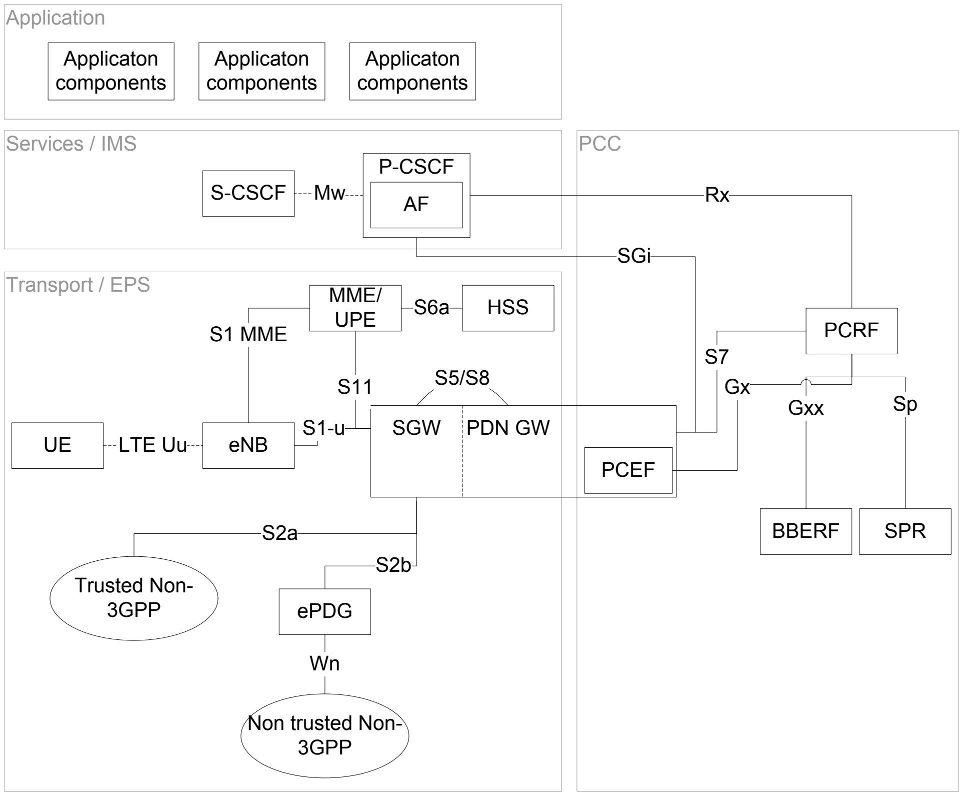

Figure 2 Detailed view of the reference architecture

3. End-to-end session

To show the process of an end-to end session establishment among the service and transport layers the following example is used: the user establishes a multimedia session using the Session Initiation Protocol (SIP). Multimedia sessions are controlled by the session layer e.g. IMS. To establish an end-to-end session the data transport bearers have to be established by the transport layer. Therefore the session layer has to inform the transport layer about the required resources.

The sections 3.1 and 3.2 describe the sequences needed to establish an end-to-end session. First the IP-CAN session establishment is described. The IP-CAN session consists of one or more bearers, set up between the UE and the Packet Data Network Gateway (PDN GW). The IP-CAN session establishment depends on the used access technology and therefore the sequence is described up to the Serving Gateway (SGW) because the sequence up to the SGW is independent for each access technology. For an example of an eUTRAN IP-CAN session see section 4. Section 3.2 contains the description of an end-to-end SIP session initiating including the QoS resource and reservation. The interworking to authorize and reserve the required resources between the service layer and the transport layer through the PCC is shown.

3.1. IP-CAN session establishment

In order to send data traffic from an originating to a terminating end point, an IP-CAN bearer has to be established. The IP-CAN session bearer establishment is taking place for the first time during the initial attachment of the UE. Such an IP-CAN bearer is also known under the term default IP-CAN bearer, because the applicable QoS is a result of the definitions from the operator in combination with the subscriber information provided by the Subscription Profile Repository (SPR). The QoS applied to such a default bearer is intended for applications that do not need a high QoS such as e.g. VoIP or video streaming would need it. The sequence of an IP-CAN session establishment is given in Figure 3.

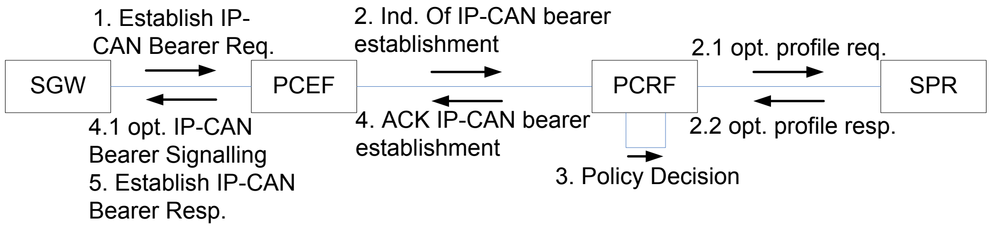

Figure 3 IP-CAN session establishment

If a gateway control session is required, it has to be established prior to the IP-CAN session establishment. The PCEF receives (1) an establishment IP-CAN Bearer request, assigns an IP address to the user and sends an indication of the IP-CAN bearer establishment (2) to the PCRF. If the PCRF has not stored the subscriber related information yet, it sends a query to the SPR (2.1) to get the required information (2.2). Afterwards the authorization and policy decision is performed by the PCRF (3) and the results are sent towards the PCEF (4). An IP-CAN bearer signalling (4.1) indicates the need for a new dedicated bearer establishment (see section 4). This is performed in case the bearer has to support a certain QoS. Step (5) finishes the IP-CAN session establishment by acknowledging the request.

3.2 End-to-End SIP session initiating

After setting up the default IP-CAN bearer, an end-to-end multimedia session can be established. Multimedia sessions are signalled with SIP over the IMS and require a high QoS because these types of sessions are prone to delay, jitter and packet loss and need a constant bit rate depending on the type of multimedia session. To ensure the required QoS towards the service layer, the transport layer authorises and reserves the QoS resources required using PCC as a mediator between the two layers.

The SIP invite message contains an initial Session Description Protocol (SDP) offer, containing information about the session. The invite message is sent from the UE to the originating P-CSCF (1) where it is forwarded to the terminating P-CSCF (2) until it reaches the end point (3) at the terminating network. The UE in the terminated network sends the same or modified offer as a response back to the P-CSCF (4) located in the terminating network.

Up to this point, the initiating of the session has neither established any bearer necessary for the data plane nor communicated the required QoS of the session to the underlying transport layer. To realise this interworking of the service and the transport layer the PCC is used as a mediator.

Step (5) contains the bridge from the IMS to the PCC framework containing the determination of the authorization and reservation of the required resources. The detailed procedure of authorization and reservation is shown in Figure 5. After the authorization is granted, the offer response is forwarded to the originated P-CSCF (6), where also an authorisation and reservation of the QoS resources is performed (7). Afterwards the offer response is forwarded to the UE (8), where a response confirmation (9) is sent to the P-CSCF.

Whenever new or modified media is defined, an authorization of the QoS resources is required.

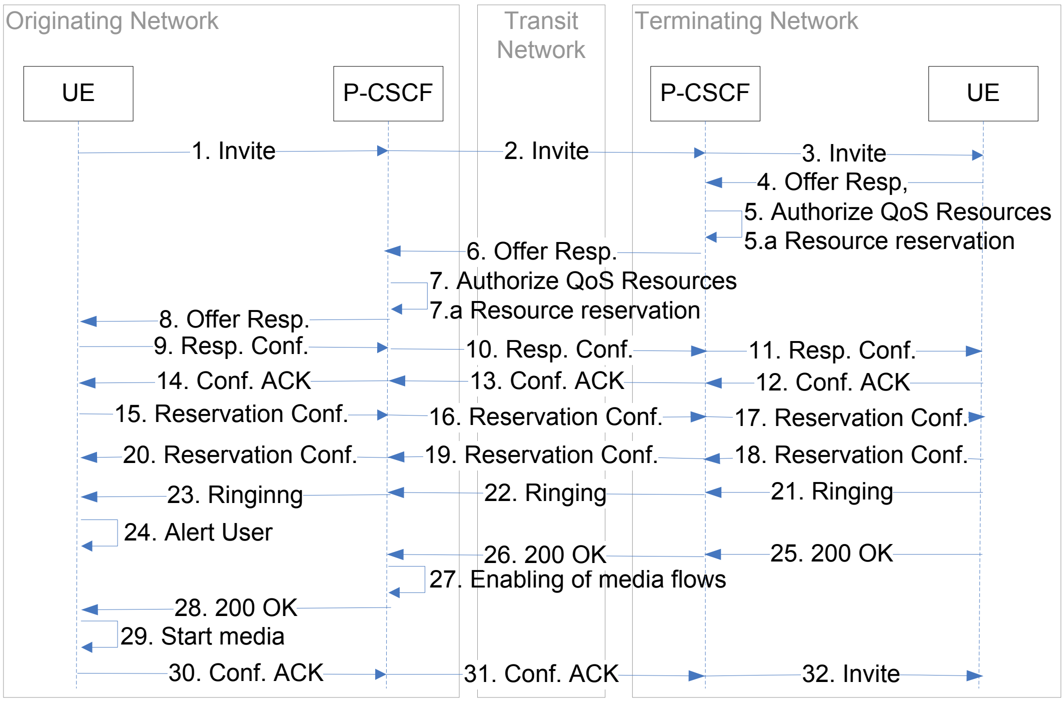

Figure 4 End-to-end SIP session initiating

The SIP signalling of the session is continued (step 10 – 26) until in step (27) the media flows are passed through the admission control within the PCC and are enabled. As a result, the user is allowed to use the allocated QoS for this session by the PCRF and is provided as an indication to the PCEF, unless this has already been done during the resource reservation procedure (based on PCC). The disabling of the media flows works the same way with an indication from PCRF to PCEF with a revocation of the allocated QoS resources for this session. Afterwards the SIP session signalling is continued the usual way.

3.2.1 Authorize QoS Resources

The PCC is used to authorize the QoS resources, as shown in Figure 5, so that the multimedia session is provided with the required QoS. The P-CSCF extracts (2) the relevant session information from the SDP (1) and forwards them to the PCRF (3). The PCRF uses the received information to calculate the appropriate authorisation (4) that enables the PCRF to authorize the required QoS resources.

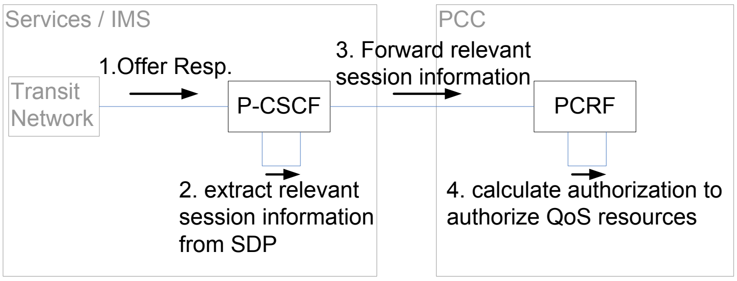

Figure 5 Authorize QoS resources

The resource reservation is initiated either by the IP-CAN itself shown in Figure 6 A (Push mode) or by the UE shown in Figure 6 B (Pull mode). The choice of the resource reservation procedure depends on the selected Bearer Control Mode (BCM), either UE only or UE/Network (NW). The push mode is applied in Figure 4 step (5a, 7a).

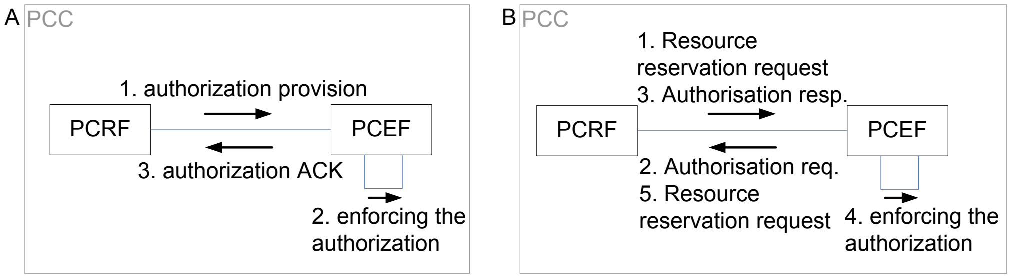

Figure 6 A Push mode B Pull mode

In Figure 6 A the BCM UE / Network (NW) is selected and therefore the authorization is pushed from PCRF towards PCEF. Figure 6 B shows the case where the BCM UE only mode is chosen. The authorization is pulled from the PCRF by the PCEF. The enforcing of the authorization (Figure 6 A step 2 and Figure 6 B step 4) either contains the modification of an existing IP-CAN bearer or the establishment of a new one.

4. ESTABLISHMENT OF A DEDICATED BEARER IN THE TRANSPORT NET

The transport layer provides appropriate bearer service on request of the service layer using the PCC as a mediator. The dedicated bearer supports the required QoS level and is network initiated. Figure 7 presents the establishment of a dedicated bearer in the eUTRAN network where dynamic PCC is deployed.

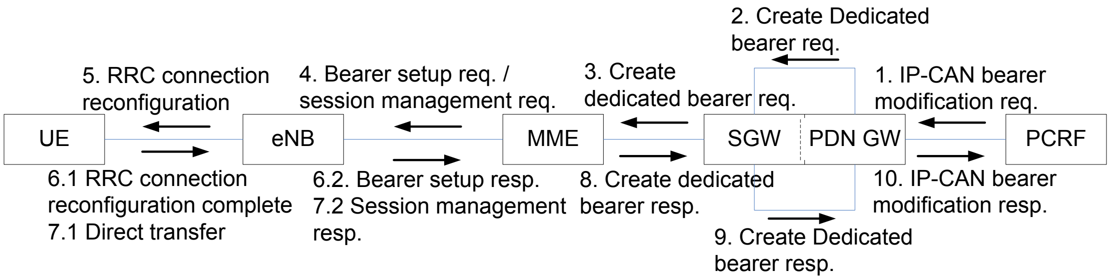

Figure 7 Dedicated bearer establishment

The PCRF sends a PCC decision provision message (1) containing the QoS policy to the PDN GW where the QoS policy values are assigned to bearer level QoS parameters of the EPS bearer. The PDN GW sends a dedicated bearer request containing the QoS information to the SGW (2) where it is forwarded to the MME (3). If the UE is in the idle mode, paging is performed either by the SGW or the MME depending on whether UTRAN/GERAN or eUTRAN is used. The MME sends a bearer setup request containing the session management request, EPS bearer identity, EPS bearer QoS parameters and other configuration parameters to the eNB (4). The eNB does the mapping of the EPS QoS to the dependent Radio bearer QoS and provides it with the session management request and the EPS radio bearer identity in an RRC connection reconfiguration message to the UE (5). The bearer activation is acknowledged towards the eNB (6.1) where the acknowledgement of the bearer activation is forwarded to the MME (6.2). This indicates if the EPS bearer QoS could be allocated or not. The UE Non Access Stratum (NAS) layer creates a session management response that is sent in a direct transfer message (7.1) to the eNB where the session management response is sent to the MME (7.2). The MME acknowledges the bearer activation to the SGW (8) once both responses (6.2, 7.2) have been received by the MME. The SGW forwards the created dedicated bearer response message to the PDN GW (9) where the PDN GW informs the PCRF whether or not the QoS policy could be enforced (10).

Within the described sequence above the GPRS Tunnelling protocol (GTP) is applied to the S5/S8 reference point. It is a tunnelling protocol used in the GERAN and UTRAN networks. The EPS enables the use of the Proxy Mobile IP (PMIP), another tunnelling protocol, on the interface S5/S8 between the SGW and the PDN GW instead of the GTP. If the PMIP is applied, the steps (1, 2, 9, 10) would be different from what is being presented in [7].

5. POLICY CHARGING AND CONTROL

To enable the cooperation between the service and transport layer, the PCC is used as a mediator between them. While the PCC is responsible for decision making on the allocation and authorization of the required bearer resources on request of the service layer, it also provides session admission control in cooperation with the transport layer. The focus of this paper is QoS provision and therefore the charging system of the PCC is not part of this paper.

The policy control provides the following functions:

- Binding, i.e. the generation of an association between a Service Data Flow (SDF) and the IP-CAN bearer (section 5.1).
- Admission control of SDFs (see Figure 4, step 27)
- Event reporting, e.g. notification to trigger new behaviour in the user plane. The event reporting is not explained further in this paper because it is out of scope.
- QoS control, i.e. the authorisation and enforcement of the maximum QoS that is authorised for an SDF or an IP-CAN bearer.
- IP-CAN bearer establishment

Section 5.1 introduces the binding mechanism the PCC is using. This is relevant to associate the application to the transport layer session, where the QoS parameters are used within the applied rules. Furthermore an analysis of the QoS signalling traffic is given in section 5.2 to appraise the signalling effort within the PCC. This appraisal is necessary to be able to compare the QoS signalling effort of other solutions with those of the reference architecture.

5.1. Binding Mechanism

To introduce the binding mechanism the example from section 3, the end-to-end multimedia session is used. Data traffic from a multimedia session is sent from the originating access network through the transit network e.g. IMS to the terminating access network until the end point is reached. One challenge of the PCC is to associate the session from the service layer to the corresponding IP-CAN bearer of the transport layer regarding the correlated QoS. This is done with the binding mechanism containing a three step procedure:

1. Session binding
2. PCC and QoS rule authorization
3. Bearer binding

The **session binding** is needed for the association of the service layer session, the Application Function (AF), to the corresponding transport layer session, the IP-CAN session. This assignment is done by the PCRF by using e.g. the following IP-CAN parameters:

UE IP address, UE identity, information about the accessed Packet Data Network (PDN).

The **PCC and QoS rule authorization** is used to fix the QoS level by selecting QoS parameters for PCC and QoS rules. The authorization of the PCC and QoS rules leads to determination if the user can get access to the required service or not and defines the constraints. This decision is performed by the PCRF.

The **bearer binding** contains the association of the PCC and QoS rules to an IP-CAN bearer within an IP-CAN session. The PCC rules are installed within the already known IP-CAN session. The decision if an existing IP-CAN bearer can be used or a new IP-CAN bearer has to be established, is made within the bearer binding.

The PCC rules applied for the binding mechanism can either be dynamic or predefined. The dynamic PCC rules are generated in the PCRF and are provisioned towards the PCEF where they are installed. The predefined PCC rules are preconfigured in the PCEF. The PCRF is able to activate or deactivate a preconfigured PCC rule or even a set of PCC rules.

5.1.1 Bearer level QoS parameters

Bearer level QoS parameters are used to identify the QoS level of SDFs as well as of EPS bearers. Therefore an SDF as well as an EPS bearer is associated with a QoS class identifier (QCI), an Allocation and Retention Priority (ARP) and an authorized bit rate for up- and downlink: Guaranteed Bit Rate (GBR) and Maximum Bit Rate (MBR).

The QCI is a scalar which is a reference to access node specific forward behaviour for a SDF/EPS bearer. The QCIs are preconfigured in the access nodes. In case of resource limitations, like on a radio link, the ARP sets the priority to decide which request has to be rejected or accepted. The ARP prioritises bearer and is useful during resource limitations like e.g. congestion, handover or disaster scenario as a decision help on which bearer has to be dropped.

The Maximum Bit Rate (MBR) is used to set an upper bit rate limit of an SDF/EPS bearer. In case the bit rate of an SDF exceeds the MBR, traffic shaping could be applied to prevent this SDF/EPS bearer from flooding the network. The Guaranteed Bit Rate (GBR) is applied to services which require a guaranteed QoS. (All parameters [8])

5.2 Appraisal of QoS signalling effort within PCC

To appraise the complexity of the QoS signalling effort in the described reference architecture, a signalling traffic analysis from the PCC is given in the following. The diameter request / answer message pairs are given as the average in Bytes. That is apparent from the RFCs [9], [10], [11] as well as the reference point specifications [12], [13]. The SIP signalling as well as the signalling of the default and dedicated bearer are not considered. Since the SIP signalling has to be established either way, as well as the establishment of the default and dedicated bearer in order to transport data between UE and PDN GW.
Figure 8 shows the establishment of an IP-CAN session, the default bearer, which is needed to be established before a dedicated bearer can be established. The diameter message CCR/CCA (Credit Control Request/Answer) are used to request and provide PCC rules for a bearer.

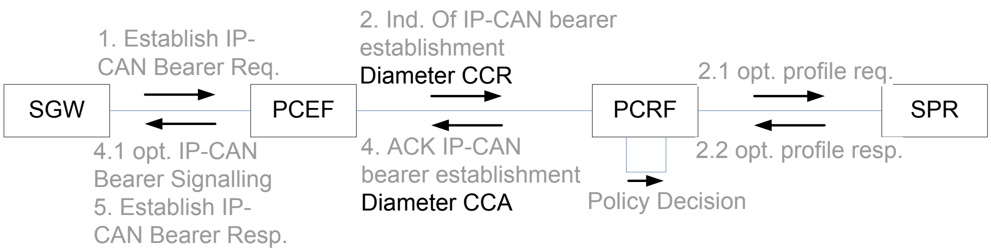

Figure 8 IP-CAN session establishment signalling effort

The IP-CAN session establishment is performed in the originated and terminated network respectively.

Table 1 Calculation of IP-CAN session establishment signalling effort

Network	**Diameter Messages**	**Bytes**	**Total Byes/ network**
Originated	2 CCR/CCA	500	1000
Terminated	2 CCR/CCA	500	1000
Total in Bytes			**2000**

Figure 9 shows the authorization of QoS resources which is necessary to establish a SIP session on the service layer. The diameter messages AAR/AAA (Request/Answer) are used to provide and acknowledge session information. To provide and acknowledge PCC and QoS rules the diameter messages RAR/RAA (Re-Auth-Request/Re-Auth-Answer) are used.

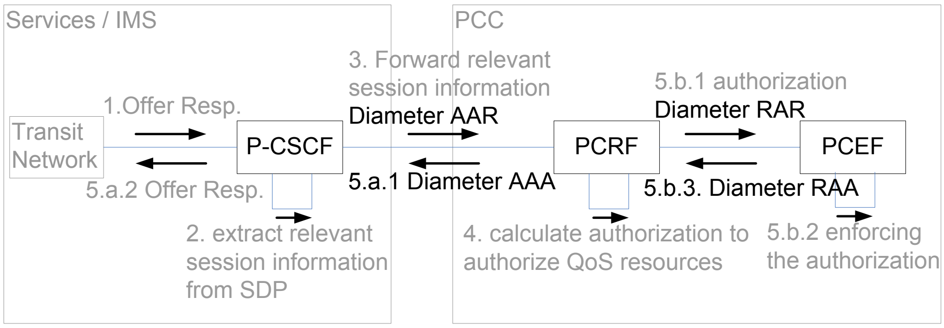

Figure 9 Authorize QoS Resource signalling effort

Note that step 5.a.x is executed in parallel to step 5.b.x.
The authorize QoS resources is also performed in the originated and terminated network respectively.

Table 2 Calculation of authorize QoS resources signalling effort

Network	**Diameter Messages**	**Bytes**	**Total Byes/ network**
Originated	4 AAR/AAA	240	960
Terminated	4 RAR/RAA	280	1120
Total in Bytes			**2080**

From Table 1 the sum of the QoS related signalling from the default bearer establishment is 2000 Bytes, which is not much. As well as the sum of the QoS related signalling to authorize QoS resources from Table 2, that is 2080 Bytes. Whenever the QoS of a session changes the authorization of the QoS resources has to be performed and therefore 2080 Bytes are required for QoS signalling. However, the complexity level of the QoS related signalling in the PCC is rather low.

6. CONCLUSIONS

The paper described a functional reference architecture for end-to-end QoS. An end-to-end session establishment model is shown in the paper, including the establishment of default and dedicated bearer, where the dedicated bearer provides an appropriate bearer service towards the service layer. Based on this end-to-end session establishment model, the cooperation of the service and transport layer via the PCC is shown. With an appraisal of the QoS signalling effort a basis for comparison has been created. The reference architecture provides the starting point for comparing end-to-end QoS solutions against this reference architecture in further studies. The next steps are the integration of mobility into policy and therefore this reference architecture for end-to-end QoS can be used as a basis to verify the impact of mobility to end-to-end QoS as well as the impact on the QoS signalling effort.

REFERENCES

[1] Cherian, S., Feder, P., Sadeghi, B., Wisenocker, R. (2008) “Integration of the IMS/PCC framework into the mobile WiMAX network”, Communications Magazine, IEEE, Volume 46, Issue 10, pp 66 – 73

[2] Ekstrom, H. (2009) “QoS control in the 3GPP evolved packet system”, Communications Magazine, IEEE, Volume 47, Issue 2, pp 76 – 83

[3] Pastor Balbas, J.-J., Rommer, S., Stenfelt, J. (2009) “Policy and charging control in the evolved packet system”, Communications Magazine, IEEE, Volume 47, Issue 2, pp 68 – 74

[4] Zaghloul, S., Jukan, A., Alanqar, W. (2007) “Extending QoS from Radio Access to an All-IP Core in 3G Networks: An Operator's Perspective”, Communications Magazine, IEEE, Volume 45, Issue 9, pp 124 – 132

[5] 3GPP TS 23.228 (2008) “IP Multimedia Subsystem (IMS)”, Stage 2, Release 8

[6] 3GPP TS 23.203 (2008) “Policy and charging control architecture”, Release 8

[7] 3GPP TS 23.402 (2008) “Architecture enhancements for non-3GPP accesses”, Release 8

[8] 3GPP TS 23.401 (2008) “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access”, Release 8

[9] IETF RFC 3588 (2003) “Diameter Base Protocol”, Request For Comments

[10] IETF RFC 4005 (2005) “Diameter Network Access Server Application”, Request For Comments

[11] IETF RFC 4006 (2005) “Diameter Credit-Control Application”, Request For Comments

[12] 3GPP TS 29.212 (2008) “Policy and Charging Control over Gx reference point”, Release 8

[13] 3GPP TS 29.214 (2008) “Policy and Charging Control over Rx reference point”, Release 8