# A Network Controlled QoS Model over the 3GPP System Architecture Evolution

Marius Iulian Corici, Fabricio Carvalho de Gouveia, Thomas Magedanz
*Fraunhofer Institute FOKUS*
*Berlin - Germany*
*{marius-iulian.corici, fabricio.gouveia, thomas.magedanz}@fokus.fraunhofer.de*

## Abstract

*Present network devices are able to access simultaneously services over various technologies e.g. UMTS, WLAN, WiMAX etc. each of these having different quality of service characteristics. Also for a specific service, different levels of quality could be offered depending on the client profile and on the momentary network capacities, giving the possibility of creating tiered access to resources. In order to enhance the network usability to these new conditions we propose an optimized, network driven QoS management and provisioning model. It is also considering a localization mechanism, doubled by a QoS control of the access networks. This paper focuses on a single operator scenario, which deploys multiple access technologies using an IMS signaled 3GPP System Architecture Evolution.*

## 1. Motivation

Multiple end-points have the possibility to access different networks of one or more technologies e.g. WiFi, UMTS or WiMAX and to receive and send data simultaneously over these networks, as depicted in Figure 1. In order to optimize the service continuity in these heterogeneous access networks various handover mechanisms are currently being studied. They concentrate especially on the seamless quality of the sessions, considering that the network where the user end-point is relocated after the handover has the necessary resources to support the service continuity.

However, the user end-points are the ones that decide to which access network to connect, having a general tendency of over-saturating the access where the cost is minimal. This leaves the network provider without the capacity to offer the required Quality of Service for the connected users, even though it could have provided services over other access networks.

Being agnostic to the momentary load of the networks, a multi-card end-point having the possibility of connecting to multiple access technologies will choose to which one to connect based on the user's preferences and on the knowledge gathered from prior connections. Therefore the network binding is decided without taking into consideration the momentary parameters of the network. In a worse case scenario, the user end-point connects to one network only to find out that its minimal requirements are not supported. Thus, for receiving the required service, it has to connect to another network, where this problem could repeat.

At this moment, the provisioning of resources is done by the user end-point. When a service is started, the end-point has to send some messages in order to ensure the quality of the data transmission; it tries to reserve some resources on a request-reply model. If the service is provided with different classes of resources requirements, the reservation has to be repeated until one of the classes is satisfied, which might be extremely time consuming over some of the access network technologies.

To remedy these shortcomings of the user end-point oriented QoS architecture, we propose a new model in which the signaled services trigger the reservation of resources without a direct involvement of the terminal. The decision is taken by crossing a set of filters, in a policy oriented manner. We concentrate on a scenario involving a single operator, which offers services on multiple access technologies.

The legacy compatibility with end-points that trigger the resource reservation was also taken into account, in order to offer a smooth transition between the present model and the one proposed by this paper.

The 2nd International Conference on Wireless
Broadband and Ultra Wideband Communications (AusWireless 2007)
0-7695-2842-2/07 $25.00 © 2007 IEEE

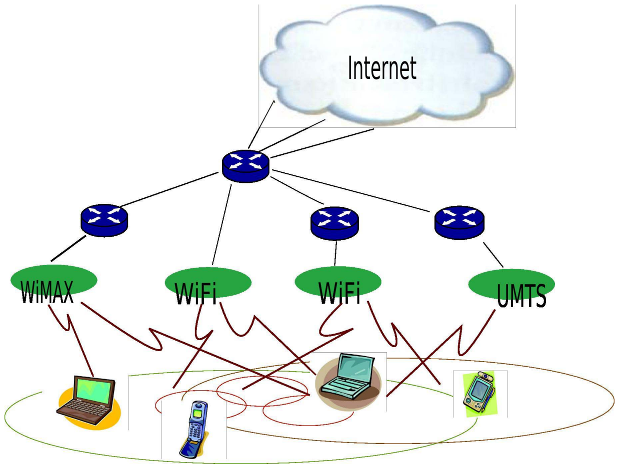

Figure 1. Typical Mobile Network Connectivity

## 2. Background

The System Architecture Evolution (SAE) ([1], [9]), offered by 3GPP, is designed to provision network convergence between different types of access networks, as depicted in Figure 2. An "Evolved Packet Core" was introduced in order to transparently unify the parameters of different technologies, like the UMTS, the 3GPP WLAN, non-3GPP access technologies and a future Evolved Radio Access Network. Each of these technologies comes with its own specific access functions. The core itself manages and stores the user end-point context in a Mobility Management Entity (MME) and the user end-point services and network information in a User Plane Entity (UPE) which are coalesce in one entity.

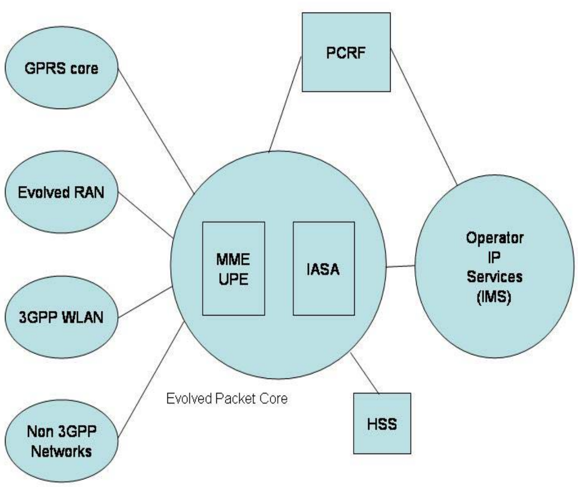

Figure 2. System Architecture Evolution

The traffic from all the networks is gathered in an Inter Access System Anchor (IASA), making the access technology transparent to other parties involved in the service provisioning.

For the authentication/authorization of the user access a Home Subscriber Server is used. An evolved Policy and Charging Resource Function (PCRF) [2] is connected to the core in order to provide the necessary information for controlling the data information received from the user end-points.

The IP Multimedia System (IMS) [6] represents today the global service delivery platform. The IMS is a complete signaling framework, able to integrate different types of services in a unified manner as seen from the user's perspective, using as signaling protocol the Session Initiation Protocol (SIP) [3]. The IMS structure also enables the connectivity of devices using different access networks in a unified manner [7], reducing the management cost of the operators that deploy multiple types of access technologies.

The Open IMS Core developed at FOKUS [7] serves as the signaling platform for the architecture here presented. It contains the structures necessary for user registration and user reachability. The user end-point access to the network is done in a network controlled manner. Also access to various services, both local to the home domain of the operator and remote to other parties is done on a network technology agnostic level, offering an interface easy to integrate in various user end-point devices.

By its ability to integrate multiple services as application servers, the IMS Core enables various services to be defined and deployed. The SAE architecture is expected to become the part of the IMS able to connect the signaling with the QoS on the access networks in the 3GPP IMS specifications, thus connecting the offered signaled services to the data transmission path in a standardized manner.

Even though the Evolved Packet Core offers a convergent model for accessing the required services, it does not consider user end-points with the capacity of parallel attachment to multiple access systems. Thus the reservation of resources and its connection to mobility still remain as open issues.

A multiple access could be resource consuming if no prior information is exchanged with the Packet Core. The decision of connecting to a specific access point is taken by the user end-point, based on some criteria like cost, signal strength and prior knowledge [4]. For a possible second connection, even though the first one has been established, the user end-point continues to consider only those local policies. Thus, it remains agnostic of the load information that is kept by the core, decreasing the possibility of finding a network able to offer the required surplus of resources.

Some of the access technologies, like WLAN [5], use frequencies from the public spectrum. Thus the

The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless 2007)
0-7695-2842-2/07 $25.00 © 2007 IEEE

user is left to choose from multiple networks, from which some pertain to the operator and can offer the required services and some pertain to other independent parties. In order to eliminate this shortcoming, knowledge of the candidate networks is required.

An alternative is to store all the information about all the networks of the operator in the user end-point. This will lead to memory consumption and to a possible duplication from malicious third-parties. The same problem would appear in the case a static criterion of determining the possible good networks is inserted in the terminal. Therefore the only acceptable alternative is to receive information about usable networks, using another communication channel, if established. This information should be restricted to the geographical position of the user end-point.

By having multiple interfaces, on a service provisioning, a malicious user might send multiple resource reservation requests on all the interfaces available. Using the actual QoS model, in which the user has to reserve resources on the data path, various types of attacks could affect the network.

Also in the public frequency networks the users can not be completely identified. Other devices could "take over" session provisioning of a user end-point and duplicate it, creating a denial of service attack. In order to limit this possibility a network provisioned QoS model was considered.

Presently the services could be offered on multiple levels of quality. This is due to the coalescence of multiple services in a new service e.g. voice and whiteboard and to the appearance of different transmission mechanisms that could adapt in real-time to different bandwidth changes e.g. the video codecs H.261, H.263. Therefore the QoS provisioning mechanism should adapt to this evolution in the service.

In the actual architectures, the reservation of a tiered service is done by multiple requests sent by the user end-point until the constraints of one class of reservation is satisfied. For example for an IPTV service multiple transmission rates could be considered. If the best one considered by the user end-point cannot be satisfied, a request for the subsequent one is required. This process can repeat until the resources for one class can be reserved.

This could lead to overly-extended session setup delays, which affect the user perception of the service as a whole. The model here proposed is able to solve the resource allocation using only one step request.

# 3. Concept Description

Figure 3 depicts the enhancements to the 3GPP SAE. A new function was introduced: the QoS Information Function (QIF) which keeps track of the resources that are available on the access networks and pre-reserves them on requests coming from the PCRF.

The information is gathered from the MME which communicates directly to the technology specific enforcement points. Triggered by the service signaling, the PCRF filters the requested profile of the user according to the information received from the QIF. After the policy class is decided for the required service, it is enforced to the MME/UPE and to the IASA. The MME/UPE is sending the QoS request to the technology specific enforcement points, while the IASA reserves resources on the anchor point.

The new interfaces introduced, have both the function of enforcement of different policies and of transmitting information about the momentary load of the resource on which the policies are enforced. Through interface I1, information about network resources availability is sent from the QIF to the PCRF. Using this information, the user profile and the cost information the PCRF is able to decide which the most appropriate access technology for the user is. The QIF receives information using interface I2 from the MME/UPE. When a specific QoS is enforced on the access network by the MME/UPE, it also sends a message to the QIF.

The enforcement of the QoS resources is done by the PCRF both on the MME/UPE using interface I3 and on the IASA using interface I4. It is necessary for the resources to be enforced on the anchor point (IASA) as to be considered for the connection between the operator network and the other networks. The QoS request has also to be enforced on the MME/UPE in order to be sent to all the technology specific entities in the path between the user end-point and the IASA.

I5 is a the typical IMS interface (Rx+) on which information about the IP filters necessary to identify the session and the QoS requirements of the user end-point are sent. Using this interface messages about policy control and charging are transferred from the user end-point to the Packet Core, in a protected way, by passing through the IMS signaling infrastructure.

If one of the connections of the user end-point is done using a large coverage network, like the 3GPP UMTS, a Tracking Area is defined for the user in the MME. Having this information, the MME is able to determine the networks using other access technologies in the vicinity of the user end-point. By correlating this information with the information about the momentary load of the networks in the QIF and with the user

The 2nd International Conference on Wireless
Broadband and Ultra Wideband Communications (AusWireless 2007)
0-7695-2842-2/07 $25.00 © 2007 IEEE

specific policy and charging policies from the PCRF, the user end-point is announced to which networks could connect in order to enhance the quality of the service.

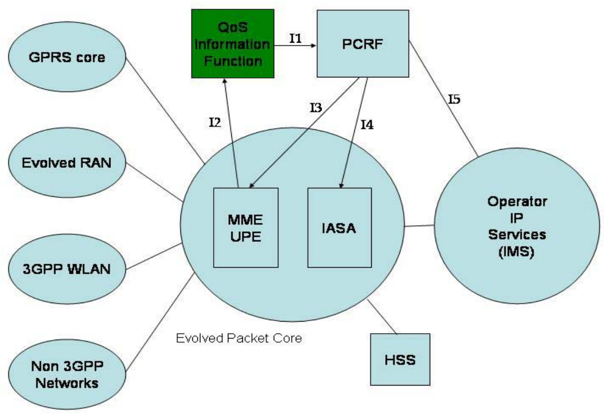


Figure 3. Enhanced System Architecture Evolution

By using the mechanisms of IMS, the user is geographically localized in a network secure manner, also offering the possibility to enhance the resources used accordingly to the specific service requests.

Two typical scenarios illustrating the usage of the architecture are further analyzed: the registration scenario and the QoS service provisioning scenario.

### 3.1. Registration Scenario

This scenario, as depicted in Figure 4, describes how a default signaling channel is allocated, when a new terminal registers with the network, in a technology transparent manner.

When a new user end-point registers to a technology specific access network, information is sent to the MME/UPE (1), which can determine the identity of the user for specific network types e.g. UMTS or cannot for other networks e.g. WLAN. In both situations the information has to be passed to the PCRF (2), in order to reserve a user specific or a general minimal default resource as to make the signaling possible. First the resources have to be reserved on the anchor point of the access networks (3, 4), then confirmed as to be reserved on the MME/UPE and enforced through the network to the user (5, 6).

The user specific policy might ensure more resources for the default resources than the anonymous default reservation. Therefore after the user registers with the IMS infrastructure (7), a reallocation of the default resources is considered, if the momentary network capacities permit it. The PCRF receives the user information from the IMS structure (8) and taking into account the information from the QIF it enforces it on the anchor point and on the access network to the user (9, 10).

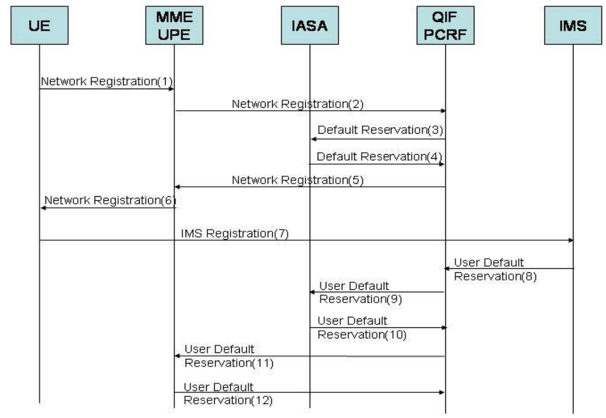


Figure 4. Registration Scenario with the Enhanced SAE

By separating the anonymous registration from the user registration a better allocation of the resources is obtained. Also due to the fact that all the traffic passes through the IASA a filter could be added. This can help in restricting the access of anonymous users to other domains than the one controlled by the operator. Therefore the service provider could secure the network from being used by unregistered parties.

If in the future the default resource allocation is decided to be used as a bearer of data for third party services, the allocation of the resources could be done dynamically, using a profile inserted in the registration messages and evaluated by the PCRF. For example if the operator decides that for a specific connected user, a specific bandwidth should be available for non-signaled services, after the user registers with the IMS infrastructure, this service could be offered, by using this re-allocation mechanism. Also this resource could be dynamically adjusted by subsequent registration requests.

### 3.2. QoS Provisioning Scenario

A typical resource reservation scenario using the same network oriented QoS provisioning in the access network is depicted in Figure 5. When a resource allocation request arrives from one of the user end-points (1), the IMS signaling infrastructure makes a request to the PCRF (2). The PCRF, after combining the user information, with the set of policies and with the momentary load of the network received from the QIF, decides for a specific resource class and enforces it into the IASA (3). Also this enforcement request is sent to the MME/UPE (4), signaling that the resources


0-7695-2842-2/07 $25.00 © 2007 IEEE

that are reserved are less than the user required if necessary.

The MME/UPE allocates the resources to the terminal on the access network; also if the user has the possibility to use other interfaces and they are in an inactive state, based on the Tracking Area of the end-point it creates a list of the interfaces which can be found by the terminal (5). The usable networks are than passed through a filter of the QIF (6), in order to keep only the ones that are highly probable of sustaining the resources required and send to the IMS infrastructure (7). The confirmation for the low level QoS that was reserved and the information about other networks that could enhance the service quality are sent back to the user end-point (8). At this moment the service could be started with a low resources allocation.

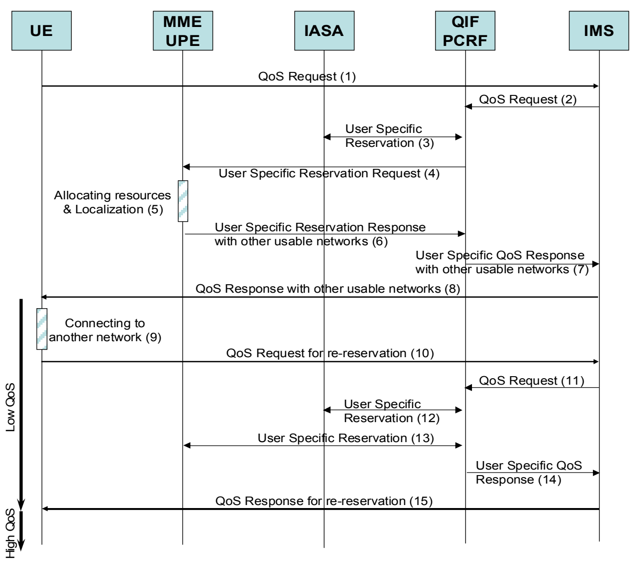


Figure 5. QoS Provisioning Scenario

In the meantime, using the information received in the QoS request, about possible free networks in the vicinity, the end-point connects to one network, authenticates and receives a default reservation as it was previously described (9). This user end-point decision should not consider only the information received, but also the signal strength of the network it wants to connect to and the mobility of the user end-point. When the connection to the secondary network is completed the UE sends a new QoS Request for the same service session (10). The request is then processed by the PCRF and QIF and enforced on the complete data path from the user end-point to the SAE.

### 3.3. Policy Provisioning

A policy description, as received from the signaling core, can be a set of codecs and their usage as in the case of present scenarios. From this set of codecs the bandwidth and the service class are deduced and sent from the IMS Core to the Evolved Packet Core.

Information containing possible networks to which the user could connect is received from the MME/UPE. This enables the PCRF to select a set of policies that could be enforced if the resources permit it. The PCRF receives the degree of occupation of the networks from the QIF. By combining this information with the policies from the PCRF a reply is sent to the user. It contains a set of networks and which resources could be used if the user connects to that network.

This enables the user end-point to receive information about the neighbor networks and their possibility to sustain its services. Also information about networks of other operators can be sent by using this mechanism if roaming service is required and desired.

The information of network locations can be introduced statically in the MME/UPE. It does not have to match exactly the networks that are seen by the user at one specific moment, but it has to cover that set in order to give a larger set of possibilities to the user end-point. Thus, the networks that are considered by the PCRF should cover a larger area than the exact location of the user. This enables the user to select the network to connect based on its mobility and on other local criteria.

This information is particularly useful for users that are already connected to a network and the network cannot sustain anymore the service requirements e.g. the signal of the network is rapidly decreasing due to an increase of the distance between the access point and the user end-point. This way the handover decision does not contain only the signal strength and the mobility parameters of the user, but also the degree of QoS availability in the network to which the user end-point is going to connect.

User end-point requests and network resource availability is sent using the service signaling. This enables the user end-point to send QoS requirements, which, by this mechanism, can be resolved in a network secure manner. It also allows information to be sent from the network to the user end-point containing the reserved resources and the possible other resources available. The network is completely controlled by the operator, thus reducing the risk of possible attacks and the error recovery.

By using this complex mechanism, services have at the beginning at least a minimal resource reservation.

Afterwards, if the network momentary capacities permit it, supplementary resources could be allocated according to the request of the parties involved. This leads to a provisional client satisfaction and to a multiple class behavior of the QoS provisioning.

Multiple-card end-points can effectively use their network capacities in order to get access to more resources in a centralized coordinated manner controlled by the network operator. Thus the operator can control tighter the data traffic through the core. This increases the degree of usability of the network and decreases the delay due to exception mechanisms in the QoS provisioning. Also the multi-card terminals do not have to bind the second interface to some random network, expecting some user-space service to use the provided resources. By using the double reservation protocol here presented, the second card can be connected only after some resources are required, thus reducing the power consumption of the user end-point.

## 4. Conclusions

In this paper we presented an enhancement to the 3GPP SAE able, to optimize the QoS provisioning in a single operator scenario. It is improving the service user perception, especially for multi-card end-points and for tiered services. Transparent inter-operator scenarios have to be studied, in order to ensure roaming possibilities.

A simple model was proposed for describing the user policies and the mechanism of processing these policies both on the IMS and on the Evolved Packet Core. A future study should present how the policies are provided by the user end-point to the Evolved Packet Core, the entities in the IMS Core that should modify and transmit the messages and how the selection is done by the PCRF, QIF and MME/UPE inside the core.

We have then proposed some additions to the 3GPP specification, in order to optimize the service provisioning. From both the operator and the user end-point, the services are better provided using a network QoS provisioning, which ensures a better QoS management from the operator and a better mechanism for exception handling as in the former 3GPP proposal.

A QoS Information Function (QIF) in the Evolved Packet Core able to register the load of the access networks was considered. This function is able to receive information about the various networks of the operator and to provide filters for selecting the ones that have enough resources for the user required services. The Mobility Management Entity (MME) has knowledge of the user location by using a Tracking Area. An extension to this Tracking Area mechanism was introduced. It contains not only the networks of the tracked type e.g. UMTS, but also the networks that can be reachable by the user, by means of other technologies e.g. WLAN.

A correlation between the QIF information and the information on the available networks provided by the extended Tracking Area provides a better network selection for multi-card user end-points. In order to obtain this correlation and the enforcement of the resources, two interfaces were considered. The first one, between the PCRF and the MME/UPE, is able to enforce policies on the mobility access and to inform about these enforcements. The second one, between the PCRF and the IASA, is able to enforce the policies in a transparent manner as seen from other parties involved in the service.

The concepts here presented apply to the access networks of an operator. It remains also as an open issue the optimizations that have to be considered for the user end-point in order to benefit from these core enhancements.

## 5. References


[1] 3GPP TR 23.228, “3GPP System Architecture Evolution – Report on Technical Options and Conclusions”, November 2006.

[2] 3GPP TS 23.203, “Policy and Charging Control Architecture”, December 2006.

[3] J Rosenberg, et al., “SIP: Session Initiated Protocol”, RFC 3261, June 2002.

[4] 3GPP TS 23.107, “Quality of Service (QoS) Concept and Architecture”, June 2005.

[5] 3GPP TR 23.836, “Quality of Service (QoS) and Policy Aspects of 3GPP – Wireless Local Area Network (WLAN) Interworking”, December 2005.

[6] 3GPP TS 23.228 “IP Multimedia Subsystem (IMS)”, December 2006.

[7] D. Vingarzan, P. Weik, T. Magedanz, “Introducing the FOKUS Open Source IMS Core System for Multi Access Multimedia Service Environments”, IIR’s Telecoms Signalling World Forum 2006, London, October 2006.

[8] K. Knuettel, D. Witaszek, T. Magedanz, “The IMS playground @ FOKUS – an open testbed for generation network multimedia services”, Tridentcom 2005, February 2005.

[9] 3GPP TS 23.402 “3GPP Architecture Enhancements for non-3GPP accesses”, May 2007.



The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless 2007)
0-7695-2842-2/07 $25.00 © 2007 IEEE



IEEE COMPUTER SOCIETY