Journal of Mobile Multimedia, Vol. 3, No. 2 (2007) 118–130

INTER-DOMAIN SLS NEGOTIATION FOR END-TO-END UMTS/IMS QoS

QUOC-THINH NGUYEN-VUONG, NAZIM AGOULMINE
Network and Multimedia Systems Research Group
IIE, University of Evry, Evry, France
nguyenv@iie.cnam.fr,nazim.agoulmine@iup.univ-evry.fr

Received March 23, 2006
Revised January 8, 2007

The Third Generation Partnership Project (3GPP) has defined an IP Multimedia Subsystem (IMS) for Third Generation (3G) networks. IMS offers services for any type of IP communication that can be provided by third parties independently of underlying access technology. The main challenge for IMS in the future is to support proper interworking with underlying resource management in order to provide the end-to-end QoS assurance. This problem has been addressed by the 3GPP as well as within several projects. However, the proposed solutions have concerned mainly the UMTS Radio Access Network (RAN) and core network domains while the resources at the border with the external IP domains are considered as static. In this paper, we introduce an enhancement of the 3GPP QoS architecture that allows a dynamic and automatic resource provisioning between the UMTS network and the external IP core network to respond to the variable demands of end users. This new mechanism provides the operators a more efficient resource utilization to support a larger number of IMS sessions in a cost-effective manner.

Keywords: IMS, PBM, E2E QoS, DiffServ, Negotiation agent, Bandwidth Broker

Communicated by: A. Hafid, P. Dini & K. Liu

1 Introduction

The 3G mobile networks, e.g. UMTS (Universal Mobile Telecommunications System) are designed in such a way to provide wide area coverage at high mobility and to offer high data rate services to mobile subscribers. The data capacity has been significantly improved with the arrival of HSDPA (High Speed Downlink Packet Access) and HSUPA (High Speed Uplink Packet Access), a smooth evolutionary path for UMTS networks. At the service layer, we can see the emergence of IMS [1, 2, 3, 4], a generic architecture to offer multimedia services in a flexible and cost-effective manner. This IMS architecture creates a new business model where third party service providers can provide new services to end-users independently of their access technology. IMS was first specified by 3GPP/3GPP2 and now is adopted by other standard bodies such as ETSI/TISPAN. IMS has been evolving so that multiple access technologies can be blended: GSM, WCDMA, CDMA2000, WLAN, WIMAX, DSL, broadband cable access, etc. In order to support services across any access technology, IMS uses SIP (Session Initiation Protocol) [5], standardized by the IETF, as the common underlying signaling protocol.

One of the main requirements for IMS system is to support the end-to-end QoS required by real-time services such as voice over IP, video streaming or any new multimedia service. To this end, the 3GPP has introduced a QoS control architecture instrumented by policies and aligned with the IETF Policy-based management (PBM) model. This architecture introduces a set of components in the UMTS domain to perform admission control and resource provisioning. Even though this approach introduces an efficient solution for the problem, several aspects have not been considered. One of them is the control of resource at the radio access, which was well studied in the IST Everest project [6]. The other is the interworking between the UMTS domain and the external IP network domain. The solution proposed by the 3GPP addresses only the matching between the requested resources in the UMTS domain and the available resources at the border with the external IP domain. However, it does not consider, at this stage, any inter-domain interactions which is the aim of this paper.

This paper is organized as follow: Section 2 is an overview of IMS architecture. Section 3 reviews the 3GPP end-to-end QoS framework to support IMS. Section 4 describes the motivation and the proposed architecture extension to support the end-to-end QoS. The performance evaluation and dynamic negotiation implementation are focused in Section 5, and a conclusion is given in the last section.

2 IMS Architecture for 3GPP networks

The integration of mobile communications and Internet technologies is one of the far-reaching consequences of IMS. In 3G mobile networks, IMS addresses the creation and the deployment of IP-based multimedia services such as voice over IP, video streaming, visioconference, instant and multimedia messaging, online gaming, etc. It specifies IP-based transport for real-time as well as non-real-time services and introduces a multimedia call model enabling communication sessions between multiple users and devices. It also provides the capacity for seamless integration of legacy services and is designed to interact with any heterogeneous technology. It supplies appropriate mechanisms to negotiate QoS and to collect the charging information. One key feature of IMS is that it can handle efficiently the resources to satisfy the QoS requirements of data flows from different media components within a service [7]. Therefore, IMS is promising for future service deployment.

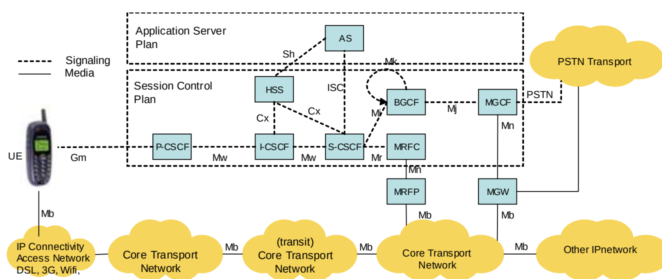

Fig. 1. 3GPP IMS Architecture.

The IMS architecture is depicted in Figure 1. The components of the IMS architecture can be grouped according to their role: Call handling components, Media processing components and Gateway components [1]. The call handling components are composed of a set of Call Session Control Function nodes (CSCFs), Application Server (AS) and Policy Decision Function (PDF). The functions of CSCFs include those of SIP proxies whose tasks are to establish, to modify and to release media sessions with guaranteed QoS and AAA charging support. The PDF performs QoS authorization, session binding, gating control and charging correlation for IP based real time services. However, it should be noted that the PDF in the IMS architecture does not evaluate the policy rules. In addition, the AS can be connected to the IMS to provide advanced services.

The Media Resource Function (MRF) can be split up into the Media Resource Function Controller (MRFC) and the Media Resource Function Processor (MRFP). It provides media stream processing resources for media mixing, media announcement, media analysis and media transcoding. For interoperability with the PSTN (Public Switched Telephone Network), IMS defines an architectural element called Media Gateway Control Function (MGCF). This latter performs the translation between the SIP messages and ISUP (Integrated Services Digital Network User Part) messages. It also controls the Media Gateway that performs the full duplex translation between the Internet voice flow and the legacy telephony voice flow.

3 3GPP End-to-End QoS framework

The term "Quality of Service" sums up all quality features of a communication as perceived by a user for a specific service. In order to achieve the end-to-end QoS, it is necessary to maintain a level of QoS all along the path from the source TE (Terminal Equipment) to the destination TE crossing various administrative domains. In the context of UMTS/IMS services, the involved domains will be UMTS Bearer Service domain, external IP domain, IMS domain and/or other UMTS Bearer Service domains. The 3GPP proposes the use of DiffServ [8, 9, 10] to support QoS in the underlying IP networks. Furthermore, the provisioning of QoS is performed by the PBM framework standardized by the IETF [11, 12, 13].

DiffServ provides a scalable aggregate approach to categorize traffic into different classes that are subjected to a specific treatment, known as PHB (Per Hop Behavior). IETF defines three main groups of classes: EF (Expedited Forwarding), AF (Assured Forwarding) and BE (Best Effort). The EF class aims to provide low loss, low delay and low jitter guaranteed services. The AF class gives different forwarding assurances in terms of loss, delay and jitter. Finally, the BE class corresponds to the traditional best-effort services.

The policy architecture introduced by the IETF and adopted by the 3GPP defines several components to enforce high level policies in the network. It is composed of a set of Policy Enforcement Point (PEP), a Policy Decision Point (PDP) and a Policy Repository component. The PEP component is a policy decision enforcer located in the network and system equipments. The PDP is a decision-making component that governs the logic of the overall management system based on the high level directives of the administrator/operator based on the agreed SLA (Service Level Agreement) with his customers.

3.1 Policy based E2E 3GPP QoS framework to support IMS

The QoS framework in UMTS R5/R6 [14] introduces new components which have been aligned with the IETF PBM architecture to control the QoS for multimedia services supported within IMS domain. The E2E QoS architecture framework is illustrated in Figure 2. The framework extends the UMTS Bearer Service (BS) with QoS control capabilities over three different segments: RAN, Iu bearer service and core network (CN). In the UMTS BS, the mobile terminal can interact directly with the GGSN using a session level protocol referred as Packet Data Protocol (PDP), which allows the UE to set up a tunnel to the GGSN with a QoS assurance.

In comparison with the IETF PBM, the Policy Decision Function (PDF) in 3GPP policy framework is equivalent to a PDP and the GGSN includes the functionalities of a PEP. The other new logical component in 3GPP policy framework is the Application Function (AF)

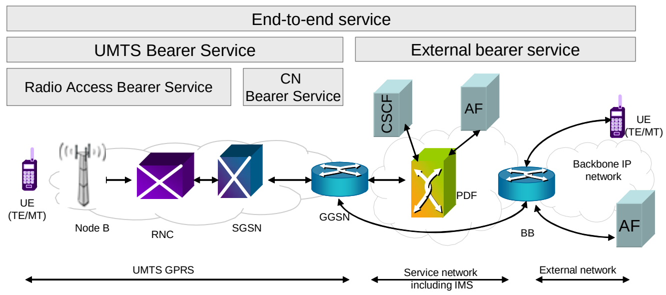

Fig. 2. 3GPP End-to-End QoS Architecture.

introduced in R6, which is in charge of offering services that require the control of underlying IP resources. In R5 this element is defined as a proxy SIP (P-CSCF), but in order to achieve the independency from the IMS architecture, the P-CSCF is of AF type.

Application Function Role: Typically, when the AF receives a session signaling message initiating a new AF session, it maps QoS-related application level parameters (e.g. obtained from the Service Description Protocol) into the policy set-up information, and sends this information to the PDF via the Gq interface. The AF requests one authorization token from the PDF for this AF session. Once it receives the token, it will initiate the IMS session.

Policy Decision Function Role: The PDF role is to coordinate the events and to react according to the operator policies and AF decisions. If the policy setup information received from the AF is consistent with the operator policy rules, the PDF will use them to calculate the proper QoS authorization (e.g. bandwidth) and store the Service Based Local Policies for this AF session. The PDF will generate an authorization token for each AF session on request from AF. The authorization includes limits on QoS for the set of IP flows and restrictions on individual IP flows. These decisions are then communicated to the GGSN through policy objects via Go interface by using the COPS-PR protocol (Common Open Policy Service with a PRovisioning profile)[15]. The token is forwarded to the UE through a SIP signaling message and then reused during PDP context activation.

Policy Enforcement Point Role: The PEP is defined in terms of "gate" which is kind of filter that is implemented in the GGSN. A "gate" operates an unidirectional flow of packets (upstream or downstream direction) on which it enforces the decisions of the PDF in terms of QoS resource authorization.

3.2 QoS mapping between UMTS and external IP domains

The reservation of resources for the end-to-end QoS requires a Translation/Mapping function [16] among different domains: IMS, UMTS and external IP domains. The GGSN has to perform proper mapping between the IP QoS information (requested by the IMS session) and the UMTS QoS information (provided by the UMTS CN). Recall that the UMTS provides a connected-oriented service with 4 different QoS classes: conventional, streaming, interactive and background [17]. The conversational class requires strict delay while the background class provides no quality or quantity guarantee. The streaming class is slightly relaxed in terms of delay while the interactive traffic follows the request-response pattern. Usually, a number of DiffServ QoS classes can be mapped to one UMTS QoS class, which is called many-to-one mapping [18]. An example of mapping UMTS and DiffServ QoS classes is shown in Table 1.

Table 1. UMTS-DiffServ QoS class mapping

UMTS QoS class	DiffServ QoS class	Mapping criteria
Conversational	EF	QoS guarantee: low delay and jitter
Streaming	AF1	low delay
Interactive	AF2	relatively low delay and loss probability
Background	BE	best effort

If the requested QoS exceeds the authorized QoS, the UMTS BS Manager will downgrade the requested UMTS QoS information to the authorized QoS information. An UMTS call will be accepted if the amount of unused bandwidth of its mapped class of service in the external IP domain is greater than that required by the call. Typically, the UMTS/IMS provider can allocate a total bandwidth C_i in the external IP domain to ensure the services of class i. If all the bandwidth C_i is used, all the incoming calls in class i will be refused even if the bandwidth of lower priority classes may be still available. If the admission control is simply based on the maximum data rate of each requested service c_i^{max}, the $(n+1)^{th}$ service of class i will be accepted if $C_i - C_{used} > \alpha c_{(n+1)}^{max}$ where $C_{used} = \sum_{j=1}^{n} c_j^{max}$ is the estimated used bandwidth, $\alpha \leq 1$ is a multiplexing coefficient and n represents the number of current connections. When $\alpha = 1$, QoS will be perfectly supported since the reserved bandwidth for each connection is equal to its peak data rate. When $\alpha < 1$, more connections will be accepted however they may suffer from the quality degradation. Otherwise, the used bandwidth of multiplexed sources can be estimated by an effective bandwidth [19] instead of the sum of maximum data rate. One solution to avoid the connection failure is to overdimension the external IP domain resources. However, this approach is not cost-effective and cannot fully protect against the congestion. Therefore, a dynamic bandwidth allocation management in the external IP domain is the motivation for our work.

4 Proposed Extension to support End-to-End QoS

4.1 Motivation

As discussed in the previous sections, the resource provisioning in the existing QoS framework addresses only the resources in the UMTS bearer service domain, considering that resources in external domain are fixed and based on the contractual agreement. In this work, we will address the QoS provisioning and all its perspectives. The policy based management approach permits to hide the complexity of provisioning resources thus to accelerate its realization. With this crucial advantage, it is possible to replace the pre-established SLA whose parameters are static by new dynamic SLA whose parameters can be negotiated or adapted according to the customer requirements (in this case the UMTS providers). The proposed extension aims to provide a flexible and scalable framework where the bandwidth can be provided on-demand based on the activated services at the boundary between the UMTS bearer network and the external IP network. We assume that the external IP network implements DiffServ and that the network control is achieved by using a policy based management solution called Bandwidth Broker (BB)[6]. The BB concept is a centralized solution to the QoS management problem of DiffServ domain. The BB controls the network resources in its domain by giving different priorities to aggregated flows following DiffServ principles and also interacts with QoS management in neighbor domains, e.g. UMTS domain in this work, for QoS negotiation and reservation. Our proposal is to extend the 3GPP QoS framework, in particular, the PDF component with a new interface allowing it to request bandwidth on-demand to the BB in the external IP network through a dynamic SLA.

The motivations behind this proposal are:

- The support of a new horizontal QoS negotiation between bearer service operators beside the already existing vertical QoS enforcement.
- The enhancement of the session admission probability at the PDF when performing admission control mechanism, therefore the connection setup failure is reduced.
- A better control of the resource utilization against the cost for the UMTS/IMS service providers.

4.2 Proposed architecture

The proposed end-to-end inter-domain SLS (Service Level Specification) negotiation architecture is depicted in Figure 3. The proposed architecture introduces a new component called IP Core Network Bandwidth Broker (IP CN BB) and a new interface between the PDF and this component. The IP CN BB is responsible for managing the resources in the external IP core network. The new interface between the PDF and the IP CN BB allows the former to control the available resources in the IP core network operator.

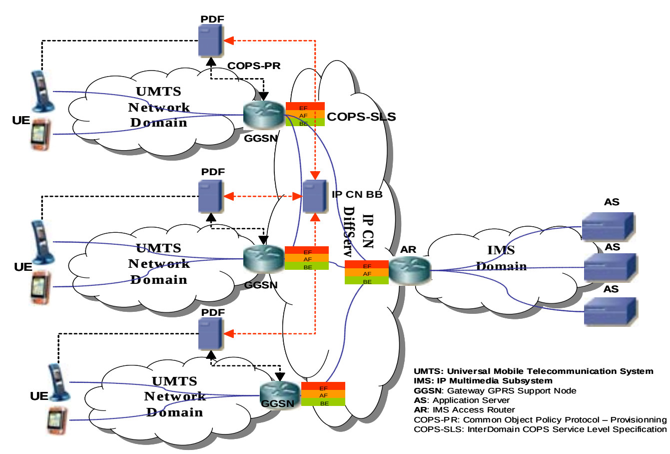

Fig. 3. Proposed Inter-domain QoS Architecture

If the UMTS provider can dynamically negotiate the resources such as the bandwidth allocation for each QoS class, the global bandwidth, etc., with the external IP CN, it can increase its resource utilization to best serve its end users. The proposed negotiation interface between PDF and IP BB CN called Qx is based on web services. The reason behind this choice is to provide an open interface using web-based protocols which are widely used in different contexts to avoid the interoperability problems.

4.3 Dynamic service level agreement

The SLA between the UMTS/IMS providers and the external IP CN providers aims to fulfil the business objective of both parties. From the UMTS BS provider's point of view, the business objectives are to satisfy the customers' SLA while optimizing its resource utilization. From the point of view of the UMTS/IMS service providers, it is important to be able to adjust the resources available for them in the external IP network based on their on-going business activities and thus to align the reserved resources with the demand. If the demand is increased above the reserved resources, the interface Qx will allow the UMTS bearer manager to increase its reserved resources accordingly. However, if the activity is decreased under the reserved resources, then it will be necessary to reduce the reserved resources to be cost-effective in a competitive market. The UMTS bearer/IMS provider will be therefore happy thanks to the concept of dynamic SLA as he can adjust the resources in the external IP CN corresponding to its business activities. However, not all parts of this SLA are dynamic, it still has static part. The static part describes the administrative involvement as well as the basic service characteristics (e.g., minimum bandwidth allocation, number of DiffServ classes, QoS of each class, etc.). The information such as peer identities, security, and billing is also included in the static part. This phase can be achieved using paper or electronically. The second phase is to specify the list of parameters that could be dynamically negotiated. Different strategies can be deployed by the external IP CN provider to reach its business objectives.

Being aware of this, the UMTS BS providers can request online the changes of the dynamic part of their contract without having to renegotiate the entire SLA. The advantage of this approach is that the IP CN provider, in case of acceptance, can deliver in a very short time the new required service through his PBM system. Hence, this renegotiation process is performed in a completely automatic manner without any human intervention, reducing the risk of misconfiguration and the time to service.

There are two aspects which should be considered when deploying this solution in a realistic context: (1) An accurate anticipation of resource demand evolution is required to perform proactive resource provisioning instead of real-time provisioning in an inter-domain context which is difficult to achieve. (2) The resource allocation adjustment can lead to an unstable situation if the PDF increases or decreases the resources so often, which is additionally constrained by the time to service limitation of the IP CN BB.

There are different manners to attain these two objectives. We have developed an intelligent approach where all processes related to the negotiation and renegotiation of SLS are performed by a negotiation agent as part of the PDF. This negotiation agent aims to help the PDF to anticipate the demands for IMS services from the P-SCSF and to initiate the negotiation with the external IP CN BB to adjust the available resources accordingly.

4.4 SLS negotiation agent

The purpose of the SLS negotiation agent is to negotiate, on behalf of the PDF, the values of the dynamic SLS parameters with the external IP network provider. To this end, the PDF maintains a management database containing information about the on-going IMS sessions and the corresponding allocated resources. The negotiation agent monitors this database and tries to detect any important change in the customer behavior that will affect the performance of the system in terms of resources allocation. The negotiation agent uses the information in the management database to anticipate the required QoS and bandwidth at the gateway between the GGSN and the external IP network. Each time the negotiation agent realizes a new IMS session and its requested resources, the negotiation agent predicts again the traffic evolution in this class so as to decide whether it is necessary to renegotiate the SLA or not. If the anticipation shows a risk of resources shortage, the negotiation agent will initiate a negotiation with the external IP CN BB to request more resources for the under provisioned class of service. In contrast, if it detects a steady over provisioning in one or several classes of service, it will initiate a negotiation to reduce the reserved resources. The negotiation decision is triggered if the forecasted required QoS satisfies the predefined policy rules. Once the decision is made, the negotiation agent will be responsible for specifying the new SLS parameters to negotiate.

Table 2. Negotiation messages between the PDF and the external IP CN BB

Message Type	Parameters
SLSRequest	action="ChangeBWRequest"
	Userid="UMTSBearer1"
	SessionSecurityKey="1223837474423410"
	UnitType="'Mb""
	type="EF" value="200"
	type="AF" value="0"
	type="BE" value="2"
SLSResponse	action="ChangeBWResponse"
	Userid="UMTSBearer1"
	SessionSecurityKey="1223837474423410"
	UnitType="'Mb""
	type="EF" Proposedvalue="200"
	type="AF" Proposedvalue="0"
	type="BE" Proposedvalue="2"
SLSConfirmation	action="ChangeBWConfirmation"
	Userid="UMTSBearer1"
	SessionSecurityKey="1223837474423410"
	UnitType="'Mb""
	type="EF" Confirmedvalue="200"
	type="AF" Confirmedvalue="0"
	type="BE" Confirmedvalue="2"

The interactions between the PDF and the external IP CN BB are achieved through the exchange of SOAP/HTTP messages [20]. The messages exchanged between these components are coded in XML [21] and are performed in three phases: request, response and confirmation as specified in Table 2. The exchanged messages contain the list of parameters to negotiate as well as their values. In this case, the SLS parameters are the class of service and the corresponding value of the requested bandwidth. These requests can be either an increase or decrease in bandwidth of one or several QoS classes according to the prediction. Upon receiving the negotiation request message from the PDF, the IP CN BB responds with the possible adjustment values which can be different from those expected by the PDF. Finally, the PDF confirms the accepted values. Note that the negotiation messages may also contain a session security key that has been obtained during the opening of a secured session.

5 Performance Evaluation of the Proposal

5.1 Model description

In this section we present an example of dynamic negotiation between the UMTS/IMS provider and the external IP provider to enhance the performance of end-to-end QoS. Consider that there are three different QoS classes (EF, AF and BE) in the DiffServ external IP network. It means that we map streaming and interactive classes in UMTS to AF class in DiffServ (see Table 1). Usually the UMTS/IMS provider allocates a fixed amount of bandwidth for each class of service from the external IP network. With the introduction of new interface between the PDF and external IP BB, the PDF can dynamically adjust the reserved

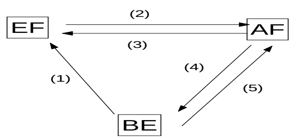

Fig. 4. Bandwidth adjustment flows

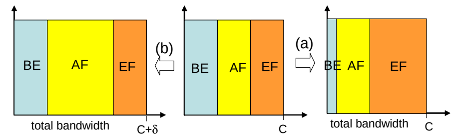

Fig. 5. Example of bandwidth adjustment

bandwidth among classes of service. The bandwidth adjustment flows in different EF, AF and BE classes are illustrated in Figure 4. For example, if PDF predicts an increase of EF traffic, it will request the external IP BB to increase the EF bandwidth by taking the bandwidth from the BE class (arrow 1) and the AF class (arrow 3) when the bandwidth of BE class equals to zero. The bandwidth adjustment between different prioritized classes of service is illustrated by arrow (a) in Figure 5. In contrast, if the PDF detects a reduced bandwidth utilization ratio in EF class, it will demand the external IP BB to increase the AF bandwidth by taking the over-provisioning part of the EF bandwidth (arrow 2).

Furthermore, the PDF can also request the external IP BB to increase or decrease the total allocated bandwidth according to the traffic pattern prediction. The arrow (b) in Figure 5 illustrates an increase of total allocated bandwidth in the external IP domain to suit the increasing demand of services in AF class. In our simulation, we consider this total bandwidth unchanged and the dynamic negotiation is to adjust the reserved bandwidth among the classes to improve the bandwidth utilization.

5.1.1 User traffic model

We assume that the arrival of calls in each class of service follows the Poisson process with parameter λ_i (i is the index name of the considered class) and the duration of each call follows the exponential distribution of mean μ_i. For the sake of simplicity, we assume that the required bandwidth for each call in the same class, c_i, is identical and constant and the total bandwidth in each class is C_i. If the reserved bandwidth in each class is static, the dynamic of bandwidth usage can be modelled by a M/M/K Markov chain where $K = C_i/c_i$. The connection failure probability is then formulated as probability where K servers in the Markov chain are all occupied [22]: $P_{failure} = P(K) = \frac{\rho^K}{K! \sum_{i=0}^{K} \frac{\rho^i}{i!}}$ where $\rho = \lambda_i \mu_i$

5.1.2 Traffic pattern and simulation parameters

In order to evaluate the dynamic negotiation, we adopt a traffic pattern which fluctuates according to three main periods: office hours (9h00-18h00), evening hours (18h00-23h00) and night hours (23h00-9h00). In our simulation, the arrival rate of calls per minute for three classes of service, λ_i, is assumed to follow the values presented in Figure 6.

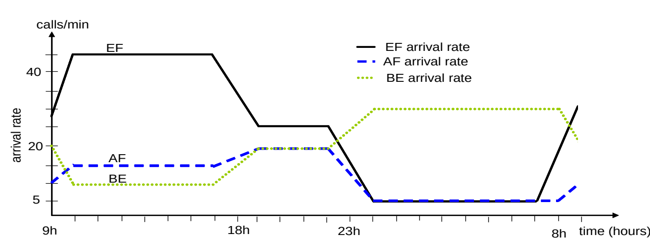

Fig. 6. Daily traffic arrival rate

The idea of this traffic pattern choice is based on real network activities. In general, the number of conversational calls in EF class is large during office hours and then decreases during evening and night. The number of streaming and interactive calls such as video on-demand, video-conference in AF class is average in office hours, large in evening and small at night. On the contrary, the number of P2P/FTP/messaging connection in BE class is large during night hours. The other simulation parameters are presented in Table 3. We assume that in the competitive market context, if the operator cannot accept a call due to lack of available resources in external IP domain, the operator must compensate the clients for each refused call.

Table 3. Simulation parameters

IMS-UMTS Services	Conversational	Streaming-Interaction	Background
Mapped DiffServ Classes	EF	AF	BE
Reserved bandwidth C_i(Mb)	$C_{EF} = 30$	$C_{AF} = 100$	$C_{BE} = 70$
c_i(Mb)	$c_{EF} = 0.1$	$c_{AF} = 0.3$	$c_{BE} = 0.2$
μ (mins)	$\mu_{EF} = 5$	$\mu_{AF} = 20$	$\mu_{BE} = 20$
Cost (euros/min)	0.5	0.3	0.1
Refused call compensation (euros)	1	0.5	0

5.1.3 Dynamic negotiation process

The dynamic bandwidth negotiation between the PDF and the external IP BB consists of two phases: traffic pattern prediction and determination of increasing/decreasing amount of bandwidth. The traffic prediction can be simply based on the average used bandwidth during a previous period and/or its standard deviation. Otherwise, the traffic prediction can be achieved by using the adaptive linear prediction techniques [23]. For the sake of simplicity in our simulation, we anticipate the traffic evolution based on the average used bandwidth $\overline{B}$ over the last 20 minutes. The choice of too short duration for averaging may impact the instability due to the high fluctuation of average bandwidth quantity. If the averaging duration is too long, the prediction may not reflect fast enough the changes in the traffic pattern.

Since the dynamic negotiation may cause the instability and inscalability problem, the negotiation should be initiated on-event. In our simulation, the negotiation is triggered when $C_i \leq th_1\overline{B}$ or $C_i \geq th_2\overline{B}$ where th_1 and th_2 are two chosen thresholds greater than 1. These thresholds may be dynamic based on the standard deviation of the traffic over the last 20 minutes. The choice of these thresholds may be different for different classes and an optimal choice is out of the scope of this paper. At each minutes, we recalculate the average used bandwidth in EF and AF classes. The proposed dynamic algorithm is as follow:

Algorithm:

- Calculate $\overline{B}_{EF}$ and $\overline{B}_{AF}$ over the last 20 minutes

- If $C_{EF} \leq th_{EF1}\overline{B}_{EF}$ then EF takes $(th_{EF1}\overline{B}_{EF} - C_{EF})$ bits from C_{BE}. If the BE does not have enough, EF takes all the remaining C_{BE} bandwidth and then takes $(th_{EF1}\overline{B}_{EF} - C_{EF} - C_{BE})$ bits from the AF class

- If $C_{EF} > th_{EF2}\overline{B}_{EF}$ then EF gives $(C_{EF} - max\{th_{EF2}\overline{B}_{EF}, EF\_MIN\})$ bits to the AF class where EF_MIN is the minimum reserved bandwidth of EF class

- If $C_{AF} \leq th_{AF1}\overline{B}_{AF}$ then AF takes $(th_{AF1}\overline{B}_{AF} - C_{AF})$ bits from C_{BE}. If the BE does not have enough, it gives all it has

- If $C_{AF} > th_{AF2}\overline{B}_{AF}$ then AF gives $(C_{AF} - th_{AF2}\overline{B}_{AF})$ bits to the BE class

In our simulation, we have chosen $th_{EF1} = 1.3, th_{EF2} = 1.5, th_{AF1} = 1.2$ and $th_{AF2} = 1.3$ and $EF\_MIN = 6Mb$.

5.1.4 Evaluation criteria

Three main criteria for performance evaluation are considered in this work. These are the connection set-up failure probability, the bandwidth utilization ratio and the UMTS/IMS provider benefit. The connection failure probability $P_{failure}$ is the probability where the PDF refuses a connection request due to lack of resources in a specific class at the border between the UMTS network and the external IP CN to support IMS services. The bandwidth utilization is the ratio between the average used bandwidth and the total bandwidth reserved by the UMTS/IMS Bearer in the external IP CN as part of the SLA. The provider benefit is considered as the provider's average daily income after subtracting the compensation.

5.2 Numerical results

Figure 7 represents the traffic prediction based bandwidth adjustment between EF, AF and BE classes during 24 hours started at 9h00 in the morning. Firstly, we observe the increase of reserved bandwidth of EF and AF classes due to the increase of incoming calls in these two classes. Subsequently, the bandwidth of BE is reduced. After that in the evening hours the EF bandwidth is reduced since its traffic decreases whereas the AF bandwidth is significantly increased to support the intensive arriving traffic. During the night time, most of bandwidth of EF and AF yield to BE class so that the BE can support its services. In the early morning, we see again the proactive provisioning of bandwidth in EF and AF classes to support the imminent traffic arrival.

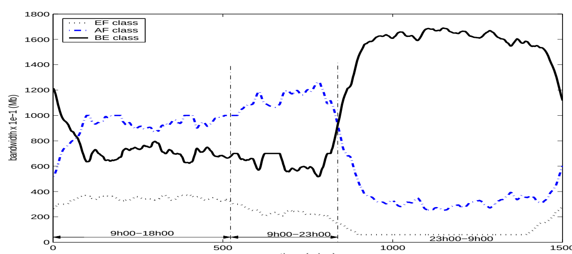

Fig. 7. Negotiation based bandwidth adjustment

Table 4. Connection set-up failure probability $P_{failure}$ (%)

	Class	9h00-18h00	18h00-23h00	23h00-9h00
Static case	EF	0.17	0	0
	AF	0.62	16.75	0
	BE	0.57	18.52	42.92
Optimal case	EF	0	0	0
	AF	0	0	0
	BE	0	31.99	0
Proposed approach	EF	0	0	0
	AF	0	0	0
	BE	3.99	68.18	6.91

Table 4 represents the connection failure probability for each class of service and Table 5 shows the total bandwidth utilization ratio according to three considered periods.

The static case is the case where there is no dynamic negotiation between the PDF and the external IP CN BB, i.e. the reserved bandwidth for each class remains unchanged. In this case, we notice a high connection failure probability for the AF traffic during evening hours and for BE traffic during night hours. Meanwhile, the total bandwidth utilization during evening and night hours is not fully used. Particularly, during the night period much

Table 5. Total bandwidth utilization ratio(%)

	9h00-18h00	18h00-23h00	23h00-9h00
Static case	80.6	90.8	53.2
Optimal case	84.7	99.2	87.3
Proposal case	82.5	86.7	91.6

bandwidth is reserved for EF and AF classes while no significant traffic presents in these two classes. A dynamic adjustment is therefore required to solve this inefficient allocation of resources.

The optimal case is the case where the bandwidth negotiation among the classes of service is achieved in *real-time*. It means that if the EF/AF calls arrive and there is no available resources for these calls, the PDF will ask immediately the external IP BB to reallocate the available bandwidth in the BE class to support them. However, this solution is not realistic in terms of implementation because the reconfiguration may take a certain time and the solution can not scale regarding the large number of reconfiguration requests. We observe that even in this optimal case, the connection failure is experienced during the evening hours due to the lack of total available bandwidth, i.e. the bandwidth utilization is about 100%.

With the proposed dynamic negotiation algorithm, the performance is significantly improved compared to the static case. Firstly, during the office hours, the proposed solution allows to avoid the connection failure for the EF and AF services. Secondly, when the AF traffic increases considerably due to, for example, a live soccer match visioning during the evening hours, the provisioned AF bandwidth is not sufficient. In the static case, the failure probability of AF class during this period is up to 16.75%. Instead, the proposed solution permits to avoid this failure. Lastly, during the night period, the negotiation mechanism allows to reduce significantly the connection failure in BE class compared to the static case.

From the UMTS/IMS providers' point of view, the proposed approach allows them to best serve their clients and more importantly to increase their benefit. Our simulation shows that in the static case the provider's average daily income is 232300 euros after paying the compensation for refused calls. By adopting the proposed dynamic negotiation approach, the provider benefit will be up to 245800 euros since the connection failure is perfectly eliminated (see Table 4) and the resources are efficiently utilized. It means that this UMTS/IMS provider can earn 13500 euros (i.e. 6% of total benefit) more every day.

In short, the proposed solution performs better than the static case and the obtained performances are not too far from the optimal ones. Though the evaluation is based on the simplified assumptions, it confirms strongly the interest of having the dynamic SLA between the PDF and the external IP domains to manage efficiently the resources to reduce the connection failure, to increase the resource utilization and then to increase the cost-benefit.

6 Conclusions

This paper has presented the general architecture of IMS as well as the interactions of its components with the UMTS architecture. We have emphasized the importance of considering QoS from end-to-end and introduced the 3GPP approach. This approach aims to support IMS services through the introduction of policy components in the UMTS domain that interacts with the IMS components. We have highlighted that the work addressed so far in this architecture only handles QoS issues in the UMTS domain and does not consider the interaction between the UMTS domain and the IP core network domain. We have proposed an extension of this architecture with a new interaction interface and component that allow the dynamic and automatic resources negotiation between the UMTS/IMS provider and the external IP CN provider to proactively support the end to end required QoS. The proposed solution adapts the available resources in the external IP CN according to the demands of IMS services and therefore reduces the probability of resource shortage as well as the service degradation. The solution also well aligns the resources with the demand and then avoids over-provisioning of resources which causes high cost for the IMS/UMTS operator. The performance tests have shown that this approach provides for IMS and UMTS providers a real advantage in the control of their resources to fulfil the customers demands. The IP network providers can really take advantage of this approach as they can provide to their business customers (mobile network operators, IMS providers, etc.) facilities that can help them all to improve their quality of service while controlling their cost in a highly competitive market.

References

1. 3GPP (2005), *IP multimedia subsystem (IMS)-stage 2,* TS23.228 v7.2.0, Tech. Rep., Dec 2005.
2. D.Attal (2005), *IMS: Internet age telephony,* ALCATEL, Tech. Rep., 2005.
3. T. Magedanz, D. Witaszek and K. Knuettel (2005), *The IMS playground @ fokus an open testbed for next generation network multimedia services,* in Proc. of 1st TridentCom, 2005.
4. K. Knightson, N. Morita and T. Towle (2005), *NGN architecture: Generic principles, functional architecture, and implementation,* in IEEE communication, October 2005.
5. IETF (2002), *SIP: Session initiation protocol,* Tech. Rep., June 2002.
6. E. IST-2002-001858 (2004), *End-to-end QoS over B3G systems: QoS management in the core networks and QoS mapping,* Deliverable D08, Tech. Rep., July 2004.
7. UMTS Forum (2005), *3G/UMTS towards mobile broadband and personal internet,* White Paper, Tech. Rep., October 2005.
8. S. Blake et al. (1998), *An architecture for differentiated services,* Tech. Rep., December 1998.
9. K. Nichols, S. Blake and F. Baker and D. Black (1999), *Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers,* Tech. Rep., June 1999.
10. V. Jacobson, K. Nichols and K. Poduri (1999), *An expedited forwarding PHB,* RFC2598, 1999.
11. R. Yavatkar, D. Pendarakis, R. Guerin (2000),*A Framework for Policy-based Admission Control,* IETF RFC 2753, Jan. 2000.
12. A. Westerinen et al. (2001),*Terminology for Policy Based Management,* RFC 3198, Nov. 2001.
13. D.C. Verma (2000), *Policy-based networking-architecture and algorithms,* New Riders Publishing, Tech. Rep., November 2000.
14. 3GPP (2005), *End-to-end QoS concept and architecture (R6),* TS23.207 v6.6.0, Sept. 2005.
15. J. Boyle et al. (2000), *The COPS (common open policy service) protocol,* RFC2748, Jan. 2000.
16. 3GPP (2005), *Policy control over Go interface,* TS29.207 v6.5.0, Tech. Rep. Sept 2005.
17. 3GPP (2005), *Quality of service (QoS) concept and architecture (R6),* TS23.107 v6.3.0, June 2005.
18. R. Chakravorty, M. Drienzo, I. Pratt and J. Crowcroft (2003), *A framework for dynamic SLA-based QoS control for UMTS,* in Wireless Communications IEEE, 2003.
19. R. Guerin, H. Ahmadi, M. Naghshineh, (1991) *Equivalent capacity and its application to bandwidth allocation inhigh-speed networks,* IEEE JSAC, Sept. 1991, Vol. 9, No. 7, pp. 968 - 981.
20. W3C (2003), *Soap version 1.2 part 1: Messaging framework,* Tech. Rep., June 2003.
21. W3C (2004), *Extensible markup language (XML) 1.1,* Tech. Rep., February 2004.
22. G. Bolch et al. (1998), *Queuing networks and Markov chains: Modelling and performance evaluation with computer science applications,* Wiley- InterScience, Tech. Rep., August 1998.
23. M. Lopez-Guerrero et al (2003),*On the dynamic allocation of resources using linear prediction of aggregate network traffic,* Computer Communications , Vol. 26, No. 12, pp. 1341-1352, 2003