Multimedia Traffic QoS Adaptation in UMTS Systems Through a Middleware Functionality

G. Araniti*, R. Agarwal**, A. Iera*, A. Molinaro***

* University “Mediterranea” of Reggio Calabria, Dept. D.I.M.E.T., Reggio Calabria, ITALY - {aranitig, iera}@ing.unirc.it
** The Indian Institute of Technology, Kanpur, INDIA - raviag@iitk.ac.in
*** University of Calabria, Dept. D.E.I.S., Cosenza, ITALY – molinaro@deis.unical.it

Abstract - In this paper the authors’ attention is on the specification of a Middleware functionality for dynamic QOS control in multimedia cellular environments, such as UMTS. The functionality implements a soft QOS paradigm, at the same time optimising the perceptual QOS provided to the end user according to her/his needs. It will be shown by exploiting suitable performance indexes and resource redistribution algorithms both an acceptable grade of service and a good degree of user satisfaction can be achieved, without an excessive increase in the network-signalling load.

I. INTRODUCTION

With the advent of multimedia applications into fixed and mobile systems the demands on transmission resources have greatly increased. As a result of the fast convergence of services such as audio, video, streaming data and web, the effective handling of resources has become an increasingly important criteria in the design of new and improved technologies; the focus being on increasing user satisfaction by providing him with an optimum Quality of Service (QOS), whilst adhering to the system constraints.

Nowadays, QOS is the keyword in the field of telecommunications. However, the meaning of QOS varies from network to network and person to person. What matters eventually is the quality of service perceived by the end user [1], and this is what is often referred to in literature as Perceptive QOS (P-QOS). With the advent of heterogeneous systems such as the 3rd generation UMTS and 4th generation all-IP mobile, effective QOS control has become a must. At the same time, this fast evolution towards the deployment of heterogeneous platforms consisting of terrestrial (both fixed and mobile) sections augmented by means of satellite links, has made the design and implementation of adaptive QOS management schemes increasingly complex.

Two possible approaches for the control of the QOS can be conceived in a heterogeneous scenario, like the one that will characterize the next generation of mobile systems. The first approach is based on the introduction of application-level control schemes to account for the QOS perceived by the user. This is a very attractive point of view, as the evaluation of the quality of service perceived by the user is usually not taken into account. Nevertheless, as it is often pointed out in literature [2], approaching QOS control at the application level consists in merely setting certain high level parameters whilst the control of the specified QOS is left to mechanisms implemented at the lowest layers. Thus, an alternative approach for QOS adaptation directly concentrations at the lowest layers of the system. This approach offers a kind of objective QOS control. It means that the QOS offered to the user is requested, measured, monitored and delivered in terms of objective parameters such as bit rate, CLR, mean delay or even BER, FER, SIR, etc. Surely, the latter approach offers a greater degree of dynamic control, which is easily quantifiable; however herein lies the risk for the developer in optimising certain low-level algorithms and protocols to guarantee a quality of service that does not take into consideration the quality as it is perceived by the user.

Thus a feasible solution recently emerging within the research community seems to be the deployment of a Middleware architecture between the network layer and the application layer. Such a middleware functionality within the Terminal equipment (TE) helps to maximize the End-to-End P-QOS by taking into account the priorities of the user whilst adhering to the system constraints, imposed by inherent bandwidth variations in a wireless link.

A second benefit is the provision of a soft QOS, which is made possible through the continuous renegotiation of the network resources made available to a consuming application. Moreover, the user would be able to transmit multimedia traffic without any service interruption, while moving from home to office and using different operators through different access systems (fixed or mobile). Such a concept is very close to the idea of Virtual Home Environment (VHE) [4], proposed within the 3rd and 4th generation mobile radio networks like UMTS [5].

The remaining part of the paper is organized as follows. Section II describes the basic ideas of our QOS control architecture. In section III, we specify the conceived middleware by considering UMTS as the network base. In section IV, we present a brief description of the mapping between the middleware and the application layer to provide the subscriber with an optimum P-QOS whilst adhering to system constraints imposed by the lowest layers. Here we also give some simulation results.

II. THE CONCEPT OF MIDDLEWARE

A. The Middleware Functionality

Our generic model of middleware is best described in [3]. Here, we provide a brief overview of this general framework that will be specified for a UMTS System in Section III.

The middleware architecture consists of various modules which are implemented into different functional units. The modular structure of the middleware architecture is presented in Fig.1. The main functions performed by the middleware for adaptive QOS control are:

- monitoring network conditions,
- evaluation of user profiles and relevant mapping into requirements expressed in terms of network metrics,
- application resource control,
- information exchange among the composing blocks.
Due to the nature of the wireless link, bandwidth variations during a call are inherent. Such a variation in resources requires that the network dynamically provides information of its current status to the environment lying above. In our architecture the Uplink Control and the Downlink Control, which act as a network monitor, perform this information exchange. These monitoring units pass on the network signals to the Network Control Module. Thus the latter module is made aware of the system status and can carry out the resource negotiation/renegotiation by interacting with the network layer taking into account user preferences (derived from the User Profile Module). Clearly, the Network Control Module is the most crucial module of this architecture, as it must control the network resources, decide the rate adjustment, and transmit the control information to both the local and the remote Application Control Module. The Application Control Module implements a set of functions that directly interface the application. It performs actions aiming at modifying the application transmission parameters in accordance with user priorities (specified by the User Profile) and transparently to the underlying levels. The User Profile Module is dedicated to user profile data management. It furnishes the necessary data to the Network Control Module as well as the Application Control Module with the aim of optimising the QOS perception of the user.

The Downlink Detector Module has the task of monitoring the downlink channel. It must be able to realize a variation in the amount of resources available on a channel and forward the appropriate control information to the Network Control Module. Subsequently, the correct parameters for the allocation of the required channels are communicated back to the Downlink Detector Module, which interfaces with the underlying system primitives. The Uplink Detector Module monitors the uplink channel and behaves analogously to the Downlink Control Module. Consequently, for a correct functioning, suitable control channels between the end terminals need to be established.

B. Requirements

The middleware requires the availability of QOS control mechanisms from the underlying platform, which can be dynamically driven and controlled. This is not a problem, as both on the fixed and mobile network side several techniques for QOS differentiation have already been deployed. The only burden introduced by the middleware is the extra signalling load associated with the dynamic QOS adaptation carried out between end-users. However, this is a small price to pay if we consider the advantages associated with the idea of an end-to-end QOS maximising the user satisfaction, which the middleware accomplishes. As will be shown in the following sections, the signalling load can be properly minimised by means of suitable algorithms.

III. A MIDDLEWARE FOR UMTS

A. Feasibility

UMTS permits the user/application to negotiate the bearer characteristics that are the most appropriate for carrying the information [5][6]. Moreover, UMTS allows for the renegotiation of radio bearers during a transmission session; this renegotiation can be initiated either by the user/application or by the network. This is just the kind of provision the middleware requires to carry out a dynamic QOS adaptation. In a wireless multimedia scenario such as UMTS, where the customer access points can vary from session to session and even the network conditions could change during the same communication session, this possibility of bearer renegotiation alone provides more than sufficient motivation for the deployment of a middleware.

In UMTS, four QOS categories have been defined: Conversational, Streaming, Interactive and Background. A list of attributes such as Guaranteed Bit rate, Max SDU, Transfer Delay etc. describes the service provided by the UMTS network bearer to the user. Voice can be delivered over dedicated low delay data (LLD) bearers of the Conversational Class. The audio and video components of a multimedia call can be also delivered over low delay and low delay variation bearers of the Conversational Class. Applications such as multimedia streaming with stringent limitations on delay variations can exploit the Streaming Class whereas bursty traffic with non real time requirements, such as web browsing and background downloads can utilize the Background and Interactive Classes respectively.

Because the real time conversation is characterized by the fact that the end to end delay is low and that the limitation for delay variation are subject to human perception, this type of traffic is transferred over dedicated physical channels (DPCCH) established between end users. Streaming data is also an example of class requiring dedicated channels for data transfer. The DPCCH is further divided into Dedicated Physical Data Channel (DPDCH) for transfer of user data and Dedicated Physical Control Channel (DPCCH) for higher layer signalling information. The Interactive and Background Classes, which do not have real time delay restrictions utilize Physical Random Access Channel (PRACH) and Physical Downlink Shared Channel (PDSCH) for data transmission.

As stated previously, UMTS allows for the possibility of renegotiating the bearer characteristics. In the middleware architecture, the Network Control Module executes this renegotiation process. In an end-to-end QOS adaptation scenario, the bearer service negotiation must take place in every network entity constituting the link between the end users. The Admission Control Capability in the control plane of the UMTS system [6] maintains information of all
C. Application Level Mapping

During a call, whenever the characteristics of the underlying communication network change (this implying the reallocation of resources to the application) the role of the Application Control Module consists of dynamically adapting the application-level parameters in accordance with the user profile to the new network resources.

In detail, the Application Control Module includes a set of functions that map the application specific high-level parameters such as coding technique, resolution, bit rate, frame rate, colour depth etc. into parameters obtained from the Network Control Module, such as the available bandwidth. To maximise the P-QOS of the user, the mapping is carried out by referring to the User Profile. The User Profile Module is a database containing application specific information. The typical application specific user information is relevant to both the low-level priorities the user places on individual composing streams of the multimedia application and to priorities of application specific high-level parameters such as resolution, frame rate, colour depth etc. Low-level priorities are utilized by the Network Control Module during the bearer negotiation, while high-level priorities are exploited by the Application Control Module to decide how the bandwidth allocated to each bearer has to be subdivided among the high level parameters.

In the following, we outline a simple procedure utilized by the Application Control Module to compute the value, which is assigned to each high-level parameter and provide the end user with the best possible P-QOS. By using the application specific user profile, the Application Control Module defines the following functions:

\[ P\text{-QoS} = \frac{2}{N(N+1)} \sum_{i=1}^{N} \frac{(Value\_Assigned)_i}{(Max\_Value)_i} \]  

\[ Bandwidth = f(parameter_1, parameter_2, ..., parameter_N) \]  

\[ (Min\_Value)_i : (Value\_Assigned)_i : (Max\_Value)_i \]

where (Value Assigned)i is value assigned to the i-th high-level specific parameter, Min Value and Max Value are the minimum and maximum values permitted by the application for the i-th parameter, i = 1, ..., N, and N is the total number of high-level parameters. The Application Control Module invokes an inbuilt algorithm to maximise (1) under the constraints in (2) and (3). We are guaranteed a solution to the above problem since the function defined by (1) is maximised in an n-cubic volume (3), bounded by the expressions in (2). A feedback control is also provided between the user and the Application Control Module so that he/she may accept, renegotiate or reject the service.

In this paper, it is not our aim to optimise the mechanism used by the Application Control Module for the partition of the network resources among the high level specific parameters. Thus, in our simulations we adopt the sample mapping described above, although we are aware that further researches could bring to better alternative solutions.

IV. SIMULATION RESULTS

In this section we assess the performance of the middleware interfaced with a UMTS network simulator. The details of the network simulator are described in [3]. The QoS optimisation procedures illustrated in the previous sections are exploited. The evaluation of the user satisfaction is performed by referring to a videoconference application tool, which is available on the Web. This particular tool offers easy implementation of the Application Control Module of our
architecture. The application tool can be used to deliver voice, video and data traffic in an adaptive fashion. In addition, it allows the participants of a session to share a whiteboard. Each traffic flow generated by the application can be independently handled and delivered over different bearer services of the underlying UMTS network. The tool also allows us to choose the type of coding scheme for each component flow.

Following a preliminary study phase, we select the NVDCT coding technique for the video component, due to its proven efficiency in utilizing the nominal bandwidth allocated to the component, at the same time providing acceptable levels of P-QOS to the user. In our study we exploited the evaluation method commonly used for voice based on Mean Opinion Score (MOS), here adapted to the perceived video quality evaluation. The output of this preliminary study is reported in Fig. 2, which shows the mapping between allocated bandwidth and resulting video MOS. It also shows the MOS corresponding to several speech coding rates taken from the 3GPP specifications [7].

Fig. 2. Mean Opinion Score (MOS) for video and speech

For a videoconference application the data service has the least priority; thus, its MOS levels are merely considered proportional to the bit rate. By using this preliminary study, we have derived a mapping between the user MOS and the corresponding network bandwidth utilized to achieve it. The mappings for speech, video and data bearers are shown in Table 1. The reader will notice that all bit rates comply with the rates specified for UMTS.

<table>
<thead>
<tr>
<th>Speech (Conversational class)</th>
<th>MOS Value</th>
<th>4.01</th>
<th>4.06</th>
<th>3.91</th>
<th>3.77</th>
<th>3.72</th>
<th>3.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit Rate[kbps]</td>
<td></td>
<td>12.2</td>
<td>10.2</td>
<td>7.95</td>
<td>6.7</td>
<td>5.9</td>
<td>4.75</td>
</tr>
<tr>
<td>Video (Conversational class)</td>
<td>MOS Value</td>
<td>4.42</td>
<td>4.12</td>
<td>3.76</td>
<td>3.42</td>
<td>3.24</td>
<td>3.20</td>
</tr>
<tr>
<td>Bit Rate[kbps]</td>
<td></td>
<td>384</td>
<td>128</td>
<td>64</td>
<td>57.6</td>
<td>32</td>
<td>28.8</td>
</tr>
<tr>
<td>Data (Background class)</td>
<td>MOS Value</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bit Rate[kbps]</td>
<td></td>
<td>384</td>
<td>256</td>
<td>128</td>
<td>64</td>
<td>32</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 1. MOS-Bandwidth mapping

The MOS corresponding to each bearer is indicated by its corresponding subscript; namely MOS_v for MOS Video, MOS_s for MOS Speech and MOS_d for MOS data. From this, we can evaluate an Overall MOS, which is an indicator of user satisfaction to the overall service. This parameter takes into account the component priorities set by the user and is defined as:

\[
MOS_{\text{Overall}} = \text{Priority}_v \cdot MOS_v + \text{Priority}_s \cdot MOS_s + \text{Priority}_d \cdot MOS_d
\]

The priorities are user priorities for individual information stream. Here, we also provide evaluation of Overall QOS, which is a measure of the efficiency of utilization of the network resources. The Overall QoS is defined as:

\[
QoS_{\text{Overall}} = \text{Priority}_v \cdot QoS_v + \text{Priority}_s \cdot QoS_s + \text{Priority}_d \cdot QoS_d
\]

and analogous expressions hold for speech and data components. The QoS index defined above for individual data streams is constrained within the interval [0,1]. Such a QoS index, defined in terms of a network level parameter, is useful to have a contemporary idea of a sort of “objective QOS” evaluated at the network layer corresponding to the “subjective QOS” at the highest levels (measured through the cited overall MOS). In our campaign we also evaluate the Grade of Service (GOS) in terms of call blocking and dropping probabilities as well as the signalling load introduced by the middleware bearer renegotiation requests.

A. Performance Analysis

A comprehensive simulation campaign has been conducted to assess the effectiveness of our QOS control approach based on the MOS evaluation. Through a software tool we simulate the presence of an urban area served by two overlapping UMTS coverage levels: microcells and macrocells, with a coverage radius of 300 and 900 meters respectively. Pedestrian (slow) and vehicular (fast) users freely roam and establish multimedia calls in these cells. In our simulation, the mobility of a user within a cell is not modelled, as perfect power control is assumed; the user mobility is modelled as a succession of transitions from microcell to micromcell. The dwell time within each cell is exponentially distributed with mean equal to the cell radius over the intentional velocity ratio. In the curves shown the intentional velocities are assumed to be 3Km/h and 60Km/h for slow and fast users respectively. The multimedia call duration is modelled exponentially distributed with an average duration of 15 minutes. Due to length constraints, only some of the most interesting results are presented here. The curves shown are plotted for 50 percent fast user scenario with videoconference occupying 20 percent of all calls initiated. The remaining part of the background traffic is equally partitioned into speech, video, and data monomedia services. Similar interesting results are confirmed for overloaded conditions, i.e. when the percentage of videoconference calls is 80 % of the total calls and the vehicular user percentage varies from 20 to 80 %. Table 2 specifies the values assumed by further parameters.

<table>
<thead>
<tr>
<th>Access technique</th>
<th>W-CDMA/FDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel bandwidth</td>
<td>5 MHz</td>
</tr>
<tr>
<td>Chip rate</td>
<td>3.84Mcps</td>
</tr>
<tr>
<td>Frame length</td>
<td>10 ms</td>
</tr>
<tr>
<td>Power control</td>
<td>Ideal</td>
</tr>
<tr>
<td>Number of users per shared channel</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 2. System parameters.
blocking and dropping probability for videoconference users. The traditional hard QoS approach provides an unacceptable grade of service, especially under high traffic load.

![Fig. 3. Blocking and Dropping Probability vs. offered load](image1)

![Fig. 4. Overall QoS vs. offered load](image2)

![Fig. 5. Overall MOS vs. offered load](image3)

As observed in figures 4 and 5, the price to pay for an increased grade of service is a value of overall QOS lower than 1. Nevertheless, by adopting our soft QoS approach the offered overall QOS does not fall to extremely low values even for high traffic loads. Also the user satisfaction, measured through the overall MOS remains within highly acceptable limits. The level of MOS offered by Alg 2 is lower than that of Alg 1. This is an expected result since Alg 1 implements the soft QoS paradigm by always trying to provide the end user with the maximum achievable MOS. However, in the choice of the most effective renegotiation strategy we have to consider also the results summarised in figure 6. It shows the average number of end-to-end signalling exchanges per handover request. This number is always 1 for the hard QoS approach (only 1 signalling exchange is required, i.e. the one to trigger the handover), while it increases when our middleware implement the soft QoS paradigm (one more exchange for each renegotiation).

![Fig. 6. Avg. no. of signalling exchanges per handoff request](image4)

The results obtained verify our intuitive idea that by means of Alg 2 (which tries to maintain the QOS always close to a let’s say “operational level”) the signalling is minimised without a significant reduction in the perceived MOS when compared to Alg 1. This suggest that coupling our middleware with a renegotiation policy like Alg 2 allows to effectively control both the quality of service perceived by a multimedia user and the grade of service the network offers her/him, with a reduced amount of signalling overload.

V. CONCLUSIONS

A research addressing the implementation of a soft-QOS paradigm in an evolutionary heterogeneous system, such as UMTS, has been conducted. The most promising aspect investigated seems to be a dynamic QOS control mechanism, implemented by a middleware functionality, based both on the resource availability in the network and on the QOS perceived by the user. It has been shown that through suitable algorithms for bearer service renegotiation, the amount of signalling load in the network can be suitably minimised.

ACKNOWLEDGMENT

Special thanks go to Dr. Carmelo Mallamace for his valuable contribution to the MOS evaluation activities.

REFERENCES