Optimization of Application QoS Protocols for 3G/4G Mobile Networks

GORDANA GARDAŠEVIĆ
Department of Communications
Faculty of Electrical Engineering
Patre 5, Banja Luka
BOSNIA AND HERZEGOVINA
egardasg@etfbl.net

MILOJKO JEVTOVIĆ
Institute of Yugoslav
Academy of Engineering
Kneza Miloša 9/14, Belgrade
SERBIA
vladmijev@ptt.yu

PHILIP CONSTANTINOU
Mobile Radio Communications Lab
School of ECE, NTUA
9 Iroon Polytechniou, Athens
GREECE
fkonst@mobile.ntua.gr

Abstract: - The importance of Quality of Service (QoS) provisioning has become one of the central issues of 3G/4G mobile network design and analysis. Mobile multimedia applications have high demands in terms of available network resources, equipment design and QoS performances. The variety of new applications (TV program distribution, tele-engineering, high definition video telephony, medical applications, location-based services, etc.) deployed in modern mobile networks implies the necessity for precise and consistent analysis. Most of the research activities in this field have been oriented to the areas of system, network, transport and middleware support for QoS. On the other side, there is a lack of proposed architectures and protocols for implementing an overall adaptive application QoS support. The challenging task for research activities is the creation of an adaptive application QoS protocol suit in order to obtain optimal QoS performance. The selection of the appropriate metrics for performance evaluation represents an important step for establishing the application QoS profile. In this paper we identify key adaptation issues and propose the architectural framework for application level QoS adaptation.

Key-Words: - Mobile Networks, Quality of Service, Application, Protocols, Performance evaluation, Optimization

1 Introduction

The importance of Quality of Service (QoS) provisioning has become one of the central issues of 3G/4G mobile network design and analysis. The implementation of QoS in the actual network must be established on the "end-to-end" basis and must provide the service performance levels necessary for obtaining the required Quality of Experience (QoE) for the end-user. Mobile multimedia applications have high demands in terms of available network resources, equipment design and QoS performances. Actual research problems, as well as the development of new architectures and protocols have been referred as "systems beyond 3rd generation" or B3G. Future broadband mobile networks should be able to support simple and efficient access to diverse services regardless of access network or mobile terminal. The efficient employment of real-time applications in resource-constrained cellular environment remains an open field for research and analysis.

The concept of QoS optimization and adaptation should be primarily established on consistent QoS provisioning, for each particular service provided by network. Different application requirements represent important criteria for the optimal selection of QoS parameters. For example, real-time applications are very sensitive to delay, bandwidth and jitter. The proper evaluation of QoS parameters obtained by simulation and measurements plays also important role in creating optimal operating conditions in mobile networks. Thus, the creation of an adaptive protocol suite with the possibility to adapt to dynamic network changes and diverse user requirements represents the challenging task.

2 Problem Formulation

The variety of new applications (TV program distribution, tele-engineering, high definition video telephony, medical applications, location-based services, etc.) deployed in modern mobile networks implies the necessity for precise and consistent analysis. Each of these particular applications has its specific protocol stack with different parameter and attributes settings. Based on these settings, the QoS profile needs to be established and transmitted. Diverse applications share the same network resources and accordingly each service should be represented by properly chosen performance metrics. The selection of the appropriate metrics for performance evaluation represents an important step for establishing the application QoS profile. Additional problem represents the fact that applications within the same gruop of services have different traffic characteristics. The common set of QoS parameters involved in process of adaptation includes bandwidth, throughput, packet delay variation (jitter) and packet loss/error rate.

The necessity of creating the adaptive QoS application protocol is driven by the fact that today's mobile networks still employ the "best effort" QoS approach. This approach cannot satisfy the requirements of modern mobile networks.

Most of the research activities in this field have been oriented to the areas of system, network, transport and middleware support for QoS. The QoS mechanisms at the network and transport layer have been studied in depth, as well as resource reservation, admission control, routing, handoff procedures, service disciplines and traffic models.

Traditional transport protocols like TCP and UDP cannot support in entirety the QoS requirements of new mobile multimedia applications. Besides the effort to adapt and extent the usage of these protocols in cellular environment, the other protocols have been also considered such as DCCP (Datagram Congestion Control Protocol), SCTP (Stream Control Transport Protocol), etc.

The proposed architectures with adaptive mechanisms are provided mainly at the level of QoS signaling [1]. On the other side, there is the lack of proposed architectures and protocols for implementing an overall adaptive application QoS support.

There are several ongoing research activities, but a common standard has not been adopted yet.

3 Related Research Work

The research in this area is trying to address the key requirements needed for support of coordinated activities between application and critical system elements and resources.

The general framework for QoS adaptation can be divided into two broad groups: network level and application level QoS adaptation. In both cases, the adaptation process follows the principle of formal control theory [2]. This approach has been used as a framework for the concept of "network-aware" applications [3]. Receiver, sender or specific proxy module can initiate mechanisms for the network QoS adaptation. The application has to adapt to changes in the available QoS while the network has to adapt to the alternations made to the QoS requirements of the supported applications. Mobile multimedia applications belong to this broad group and require the QoS profile that enables predictable and acceptable performances. Actual research problems for this group of applications are related to optimal mapping of application level QoS parameters into network-level QoS parameters, high effective adaptation methods and low computational complexity.

The document 3GPP TS 23.802 "*Architectural Enhancements for End-to-End QoS*" considers possible solutions to enhance the "end-to-end" QoS architecture and to enable improved "end-to-end" QoS in the case of interworking with IP network domains [4]. The general "end-to-end" reference model for QoS interworking is shown in Fig.1 [4].

This model gives a brief overview of some application-related tasks in a heterogeneous network infrastructure. Application nodes represent the interface between the domain specific nodes and backbone network. SIP (Session Initiation Protocol) has been chosen as application control protocol. As a direction for further research, some general recommendations related to QoS parameterization are emphasized.

One approach for application-level QoS provisioning is based on the combination of converters and description syntax that specifies QoS at the end-user, network and application level [5]. This paper proposed mechanisms that enable creation of application profile for efficient coordination between these three levels in QoS management.

An adaptive QoS framework for integrated cellular and WLAN networks was proposed by [6]. The model supports the delivery of adaptive real-time flows using the QoS reservation-based approach.

The EUQoS project has the objective to explore QoS technologies for the advanced QoS-aware applications (voice, video-conferencing, video-streaming, educational, tele-engineering and medical applications) over various networks infrastructures and domains [7]. The intention is to adapt QoS mechanisms to the need of forthcoming complex and heterogenous network environments.

In order to support QoS requirements for multimedia group communications in B3G networks, the *QoS Architecture for Mobile Multicast Multimedia Services* (Q_3M) has been developed [8]. This architecture is oriented towards the establishment of adaptive multicast in DiffServ (Differentiated Services) mobile environments, with the seamless mobility support for streaming multimedia applications.

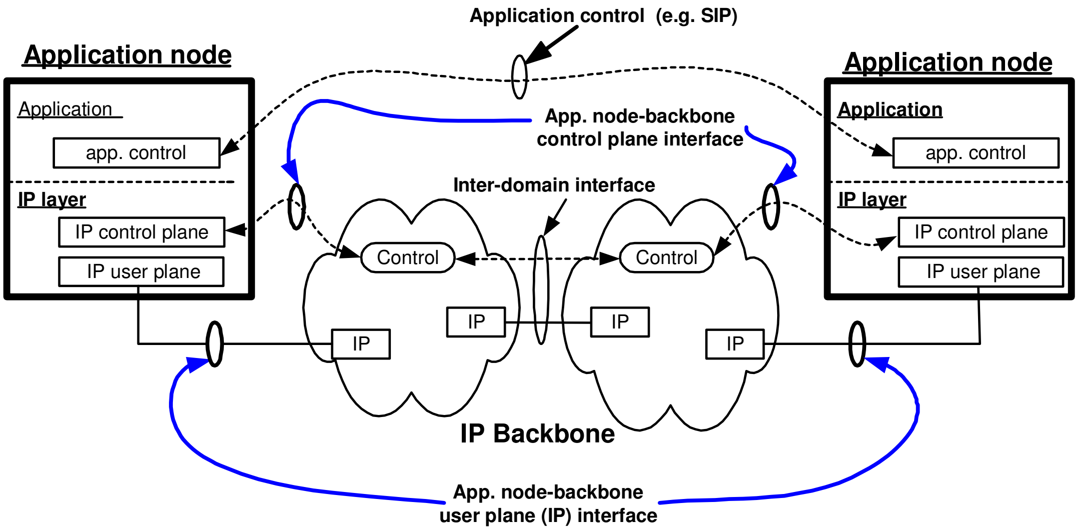

Fig.1. General end-to-end reference model for QoS interworking (TR 23.802) [4].

The process of adaptation at the application level incorporates several mechanisms: controlled quality degradation, adaptation of data format, the rate adjusting, implementation of seamless handoff procedures, as well as high QoS performances with low jitter, delay, low packet loss and guaranteed bandwidth. To achieve these requirements, different studies have been performed addressing the following methods: layered and multiple encoding, payload compression, correction, buffering, etc. [9-13].

4 Application level QoS

One of the key requirements for establishing the architecture of particular service is the creation of adequate framework for QoS performance analysis. The selection of apropriate application QoS protocol as well as mechanisms for its adaptation and optimization represents a broad field for research in today's mobile networks.

The optimization process at the application level should be applied on media components, because of aggregation of heterogenous traffic types.

It is very important to select the adequate performance metrics for particular application [14]. Applications can differ to a great extent when comparing the metrics for performance evaluation. For example, the performance metrics for real time video sharing application are the set-up delay for establishing connection and the delay experienced during the period of recording video on transmitting side until receive the content on the other end. For the similar application, video streaming, the low delay is not of vital importance, but the high bit rate available for transmission. The adaptation approach for streaming multimedia services is oriented towards using buffering and transcoding proxies.

The session establishment is based on specific application protocol and set of parameters exchanged between endpoints (network delay, bit-rate, PDP context, compression algorithms, buffering delay, etc.). Key Performance Indicators (KPI) enable the monitoring of system resources as well as the service flow (from the end-user viewpoint).

The implementation of adaptation module at the application level has important advantage in comparison to implementation on lower layers. The lower levels are already involved in numerious system tasks and perform many functions. As a general rule, the adaptation procedures on lower layers are implemented in hardware elements.

The adaptation at the application layer can be implemented as a software module which makes it more easier for changes and restructuring. Moreower, the question is how many functions and processes can be performed at the application level, so the lower levels may perform other tasks.

The optimization procedures can be also applied. This is a motivation for the overall processing performances to be improved.

5 Proposed framework for QoS adaptation

The process of QoS adaptation is important for providing reliable and synchronized data transmission. Effects of adaptation are numerous: resource control is more efficient, the number of concurrent sessions is extended, and the overall performances are improved.

The idea behind the proposed adaptation framework is the creation of QoS profile transparent to lower layers, particularly the network layer.

This profile is established as a set of entities, where each of them performs specific tasks, as shown in Fig.2. These entities are dealing with traffic characterization, estimation, application policy and QoS syntax. The description of the proposed adaptation framework is given through the rest of paper.

5.1 Traffic characterization entity

The aggregation of different traffic types makes the problem of creating platform for QoS adaptation even more complex. Depending on application type, different traffic distributions can be exploited for the simulation purposes (normal distribution, lognormal, Pareto, Weibüll, etc). Some applications are very bursty and unpredictable in their nature, while others show no significant variations during the transmission. Traffic modeling has great impact on obtaining optimal QoS performances. These profile information will be mapped onto the set of traffic representatives.

The example of QoS test/reference profile is shown in Table 1. Applications with similar requirements can be aggregate to form one profile, where different traffic classes are based on the representative QoS attributes set.

Depending on chosen video format (frame rate, frame size, color depth, etc.), codec type (H.263, MPEG-4, H.264, etc.) and traffic distribution parameters, the QoS requirements may differ significantly.

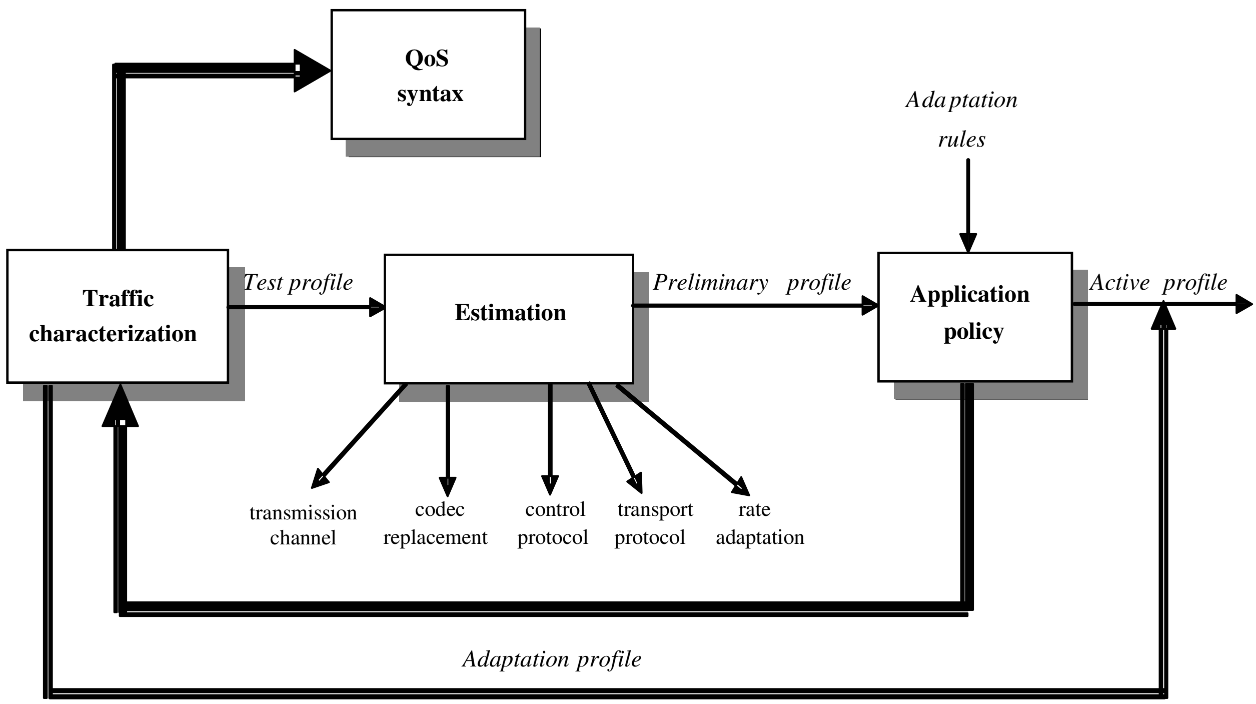

Fig.2. Application-layer QoS adaptation framework.

For example, videoconferencing is one of the most challenging applications to be supported since it requires high bandwidth, low latency and low jitter. The video traffic shows long-range dependency effects, and it is of significant importance to employ high quality traffic model, especially for simulation purposes.

According to the ITU-T Recommendation G.1010, limits are as following:

- packet delay: 150 – 400 [ms]
- jitter: < 1 [ms]
- packet loss: < 1%
- lip synchronization: < 80 [ms].

Test profile
traffic class codec type
max bit rate for uplink
max bit rate for downlink
max SDU size
delivery of erroneous SDUs
transmission delay
jitter
residual bit error ratio
frame error rate
traffic priority
traffic distribution parameters
operating environment (indoor, outdoor, vehicular, pedestrian)
aggregation impact factor

Table 1. An example of test profile.

5.1.1 Simulation results for videoconference application

For obtaining some QoS performances based on simulation of videoconferencing application, we used the industry-standard network simulation tool, OPNET Modeler [15].

OPNET Modeler has a well-defined user interface and rich set of modules for creating the required environment for specific simulation and providing the high level of accuracy.

OPNET offers several models for the application analysis:

- Application Demand;
- Standard Application;
- Custom Application;
- Application Characterization Environment;
- ACE Whiteboard.

Application Demand enables the simulation of "client-server" application traffic, with attributes allowing the manipulation of packets, bit-rate, type of service (ToS), etc.

Standard Application models represent the general network applications such as HTTP, Email, Voice, Videoconferencing, etc. This model is adequate for two-tier applications (peer-to-peer or client-server).

The Custom Application model uses application components (tasks and phases) to provide the user with the possibility to configure in detail the behavior of specific application (for example application-level signaling).

The ACE model enables the generation of an accurate traffic pattern of existing application. This model can be only used for those applications that are operational in real network conditions.

The ACE Whiteboard represents the environment for application design and is particularly useful for testing application before their implementing in actual network.

Our goal was to illustrate the complexity in modeling real-time application such as videoconferencing, and to show the necessity for applying the adaptation mechanisms. The videoconference, as one of the application offered for transmission in UMTS, requires significant network resources, as well as low delay, jitter and packet loss.

Two important metrics for performance characterization are available bandwidth and "end-to-end" delay. The procedure for modeling videoconferencing application represents the complex task, especially when simulating high time-varying parameters in mobile environments. On the other side, this application is also specific in terms of requirements for high bit-rate and synchronization during the transmission.

The simulation setup in OPNET is shown in Fig.3. The videoconference traffic is transmitted between two nodes inside the UMTS network. We measured QoS performances by varying specific system and application parameters.

The system parameters we included in our observation were background system utilization, throughput-based admission control. For application parameters, we measured "end-to-end" videoconference delay and video jitter as two critical QoS performances for this type of application, by changing the payload size, QoS class, distribution parameters, max available bit rate on uplink and downlink, etc.

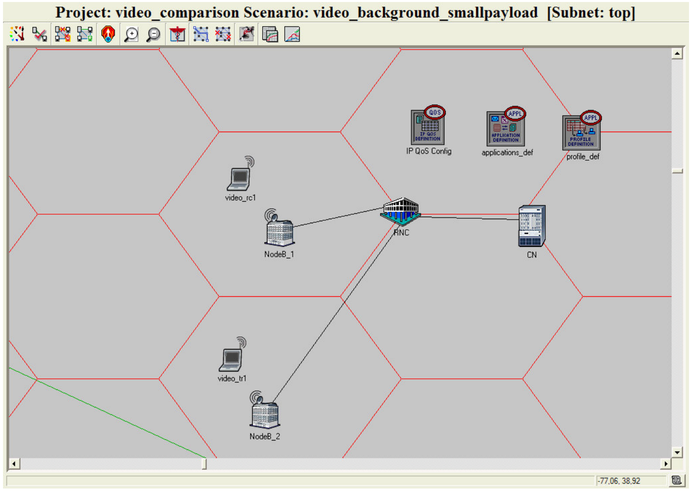

Fig.3. Simulation setup for videoconference performance testing.

We were testing the OPNET built-in (standard) videoconferencing model, Fig.4, but we also imported H.263 video trace-file from [16]. The videoconference profile enables the selection of attributes such as time distribution, number of sequence repetition, operational mode, etc. As a result, the videoconferencing traffic is generated.

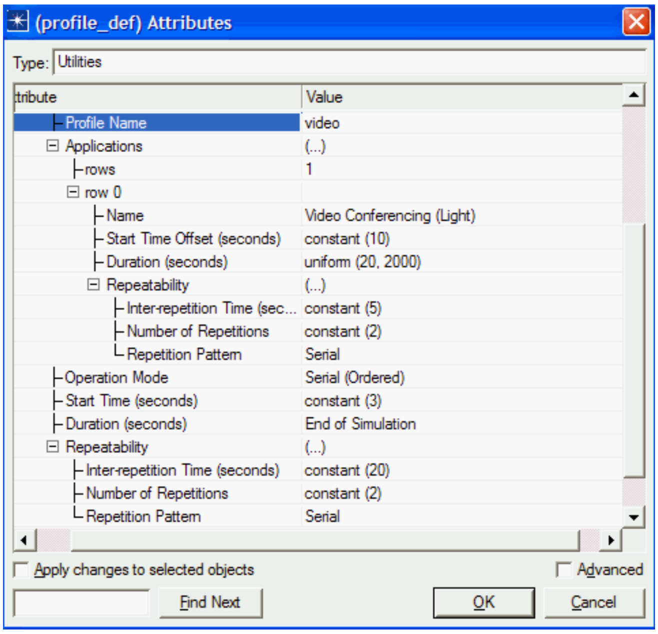

Fig.4. Parameters for videoconference profile.

To demonstrate the resulting performances for the videoconferencing, we selected two different QoS parameter profiles. The ToS (Type of Service) chosen for the simulation purpose was interactive multimedia and background.

The setup parameters for interactive multimedia profile are shown in Fig.5.

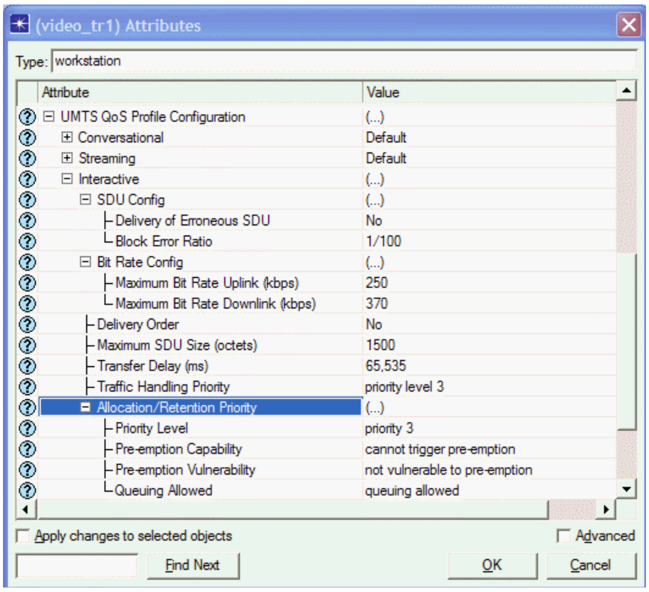

Fig.5. Setup parameters for 'interactive multimedia' profile.

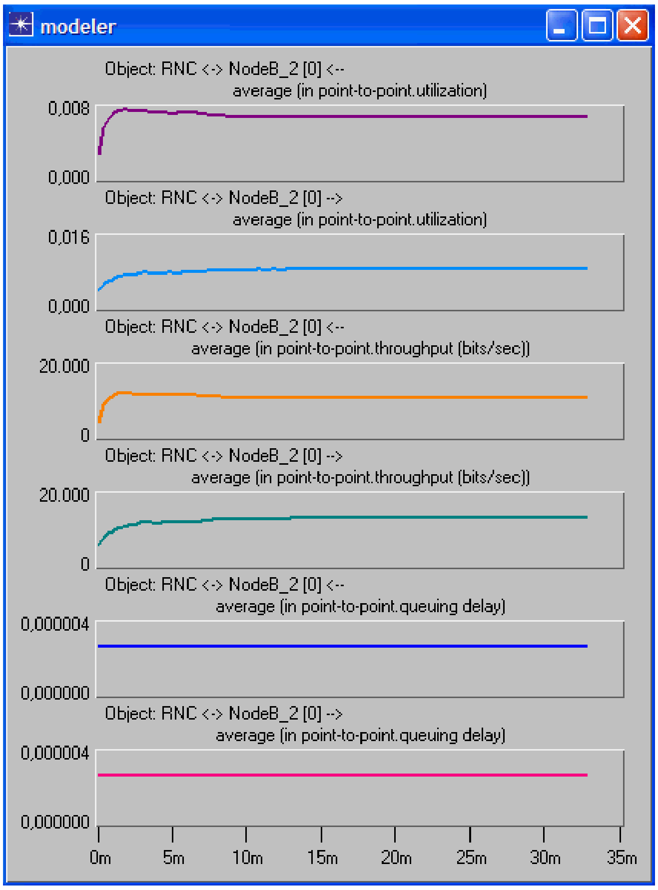

Fig.6. Simulation results for RNC (Radio Network Controller): utilization, throughput, and queuing delay.

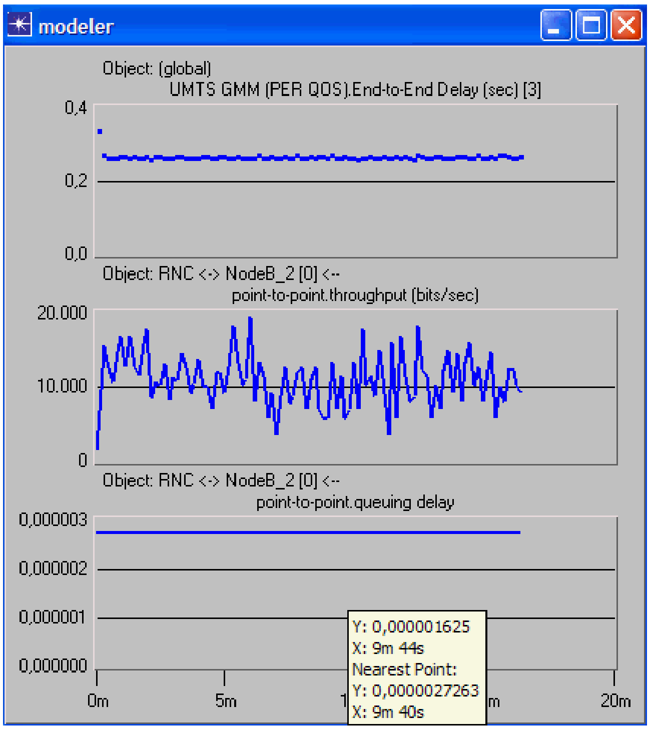

Fig.7. Simulation results for RNC: UMTS GMM "end-to-end" delay.

With OPNET Modeler, different statistics can be collected, Fig.6, 7 and 8. This is important for proper network dimensioning, in order to achieve better resource management and overall performances.

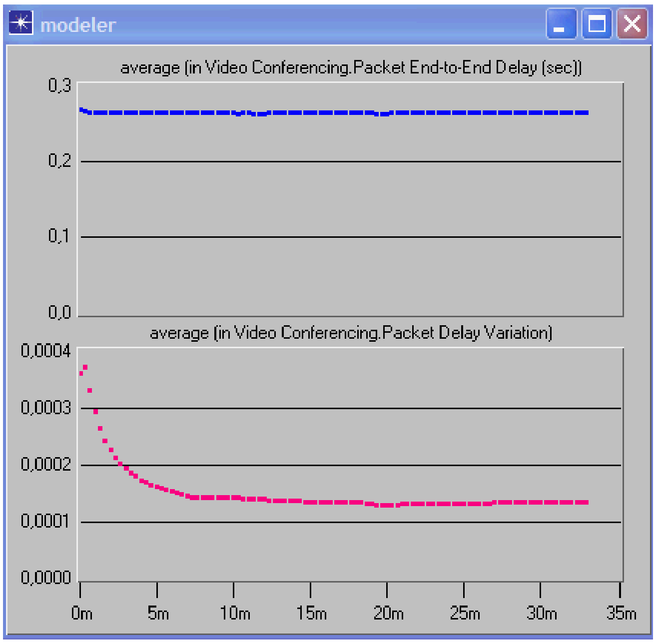

Fig.8. "End-to-end" delay and packet delay variation for videoconference.

For the simulation results, different statistic parameters can be selected. It is possible to select global and local data.

Global statistics provides information about overall network or system, Fig.9. On the other side, local statistics is attached to particular objects, links or modules.

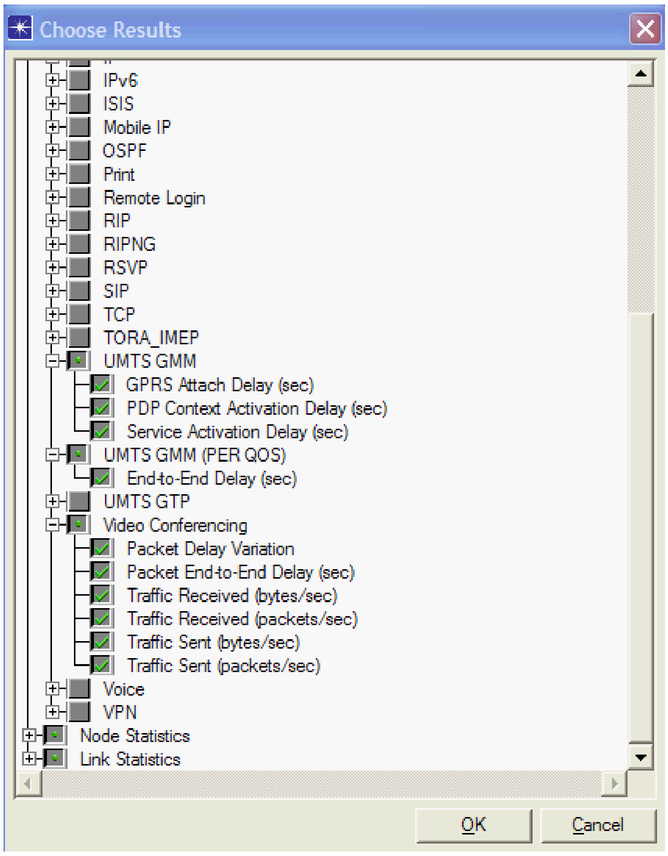

Fig.9. Global statistics for simulation results.

The profiles are established in the following way:

- First profile (*video_background_smallpayload*) has following attributes: incoming/outgoing stream frame size = 1000 bytes; ToS = background (1).
- Second profile (*video_interactive_bigpayload*) has following attributes: incoming/outgoing stream frame size = 9000 bytes; ToS = interactive multimedia (5).
- Third profile (*video_ interactive_tracefile*): ToS = interactive (5).

The results obtained from Fig.10 and 11 show clearly that the possibility for adaptation of traffic characterization profile, as well as the correct system behavior prediction is of vital importance in order to obtain optimal QoS performances. Different traffic profiles have different QoS performance results. The results of simulation imply that even the QoS profiles from the same application class may have significantly different characteristics.

The problem becomes more evident when transmitting heterogeneous traffic inside the mobile network. In this case becomes necessary to apply some of adaptation mechanism (the rate adjusting, controlled quality degradation, adaptation of data format, implementation of seamless handoff procedures, etc.).

The proper selection of application parameters is in close relation to efficient use of network and system parameters.

The following aspects need also to be taken into account:

- The impact of user-perceived QoS.
- Comparison of parameter performances for different applications/services.
- The traffic characteristics of diverse applications.

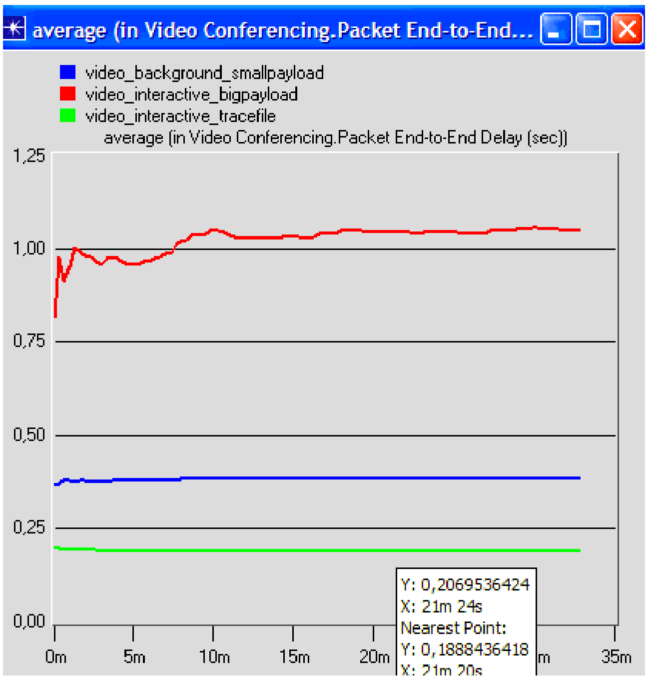

Fig.10. "End-to-end" delay for three different scenarios.

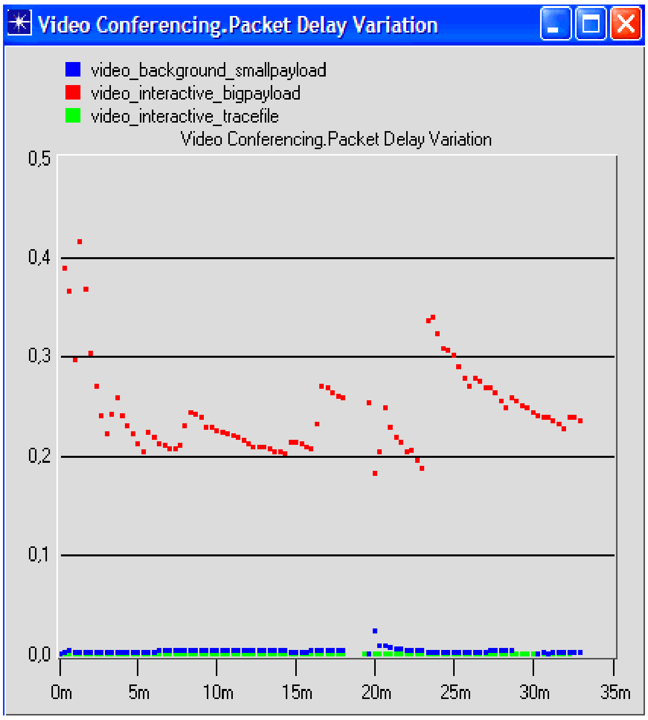

Fig.11. Packet delay variation for three different scenarios.

5.2 Estimation entity

The generated test profile represents the input for the estimation entity. The main task of this entity is to estimate attributes critical for particular bearer service. The general bearer performance attributes are "end-to-end" packet transfer delay, "end-to-end" delay variation, throughput, and packet loss/error rate.

Based on test profile generated in previous entity, estimation entity provides minimum QoS requirements extracted from raw data describing transmission environment.

These requirements represent some form of preliminary profile recommendation. For example, the recommendation may suggest the use of DCH (Dedicate Channel) for real-time video transmission channel, as well as may suggest the codec replacement, header compression or the introduction of new QoS parameters.

Additionally, the recommendation may include the suggestion for suitable application-level control protocol and transport protocol. The exit from estimation entity is the preliminary profile that will be used for the adoption of appropriate application policy.

5.3 Application policy entity
This entity is in charge of establishing adaptation rules that will be used for correction of initial test profile attributes.

Mechanisms involved in this entity provide policies to dynamically regulate the behavior of system components involved in obtaining adaptation profile. Based on this, adaptation rules are selected.

The policy entity should establish appropriate performance metrics for the application QoS profile evaluation. The complexity factor can be added to preliminary profile to indicate the "QoS-aware" status of particular profile or specific QoS mode that can be attached to particular profile. The profile with this status will have higher priority in further processing.

5.4 QoS syntax entity
QoS syntax entity implements the profile description in specification language that enables dynamic and transparent correspondence with user and network requirements.

The policy represents the specification of method or action that meets service requirements. This entity should enable the decomposition of protocol functionality into components in a way that they can be re-assembled dynamically.

6 Conclusion
With the introduction of new packet-optimized radio technologies (High Speed Downlink/Uplink Packet Access - HSDPA, HSUPA) as well as new radio access network architecture (3GPP Long Term Evolution) in mobile networks, the variety of multimedia applications become available. The "end-to-end" QoS support has been recognized as one of the key requirements for successful employment of modern mobile networks. The necessity of creating the adaptive QoS application protocol is driven by the fact that today's mobile networks still employ the "best effort" QoS approach. This approach cannot satisfy the requirements of modern mobile networks.

Based on the above-mentioned facts it can be concluded that there is a need for an in-depth study of application level QoS protocols and mechanisms.

The process of QoS adaptation is important to provide reliable and synchronized data transmission. The effects of adaptation are numerous: resource control is more efficient, the number of concurrent sessions is extended, and the overall performances are improved. The implementation of adaptation module at the application level has important advantage in comparison to implementation on lower layers. The lower levels are already involved in numerious system tasks and perform many functions. As a general rule, the adaptation procedures on lower layers are implemented in hardware elements.

The challenging task for research activities is the creation of adaptive application QoS protocol suit for obtaining optimal QoS performance. Particularly, the main issues are:
- how to create an adaptive application-level QoS set of parameters and attributes;
- how to achieve the optimal mapping between application QoS parameters and system components.

In this paper we are identifying the key adaptation issues and propose the architectural framework for application level QoS adaptation. This framework is established as a set of entities, where each of them performs specific tasks. These entities are dealing with traffic characterization, estimation, application policy and QoS sintax. The next step in our work will be the implementation and simulation verification of the proposed framework.

7 Acknowledgments
This work has been supported by the Laboratory of Mobile Radio Communications, School of Electrical and Computer Engineering, National Technical University of Athens, Greece.

References:

[1] T. Guenkova-Luy, A. Kassler, and D. Mandato, "End-to-End Quality of Service Coordination for Mobile Multimedia Applications", *IEEE Journal of Selected Areas in Communications (JSAC)*, Vol. 22, No. 5, June 2004.

[2] J. Bolliger, and T. Gross, "A Framework-Based Approach to the Development of Network-Aware Applications", *IEEE Transactions on Software Engineering*, Vol.24, 1998.

[3] J. Cao, K.M. McNeill, D. Zhang, and J.F. Nunamaker, Jr., "An overview of network-aware applications for mobile multimedia delivery", *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, Track 9 , Jan. 2004.

[4] 3rd Generation Partnership Project: Technical Report 23.802, V1.2.0, "Architectural Enhancements for End-to-End QoS", Sept. 2005, http://www.3gpp.org/ftp/Specs/html-info/23802.htm

[5] A. Thomas, "Supplying legacy applications with QoS: a description syntax at application, end-user and network level", *Proceedings of the Software Engineering and Applications,* Track 374-059*,* 2002.

[6] X. G. Wang, G. Min, J. E. Mellor, K. Al-Begain, and L. Gua, "An adaptive QoS framework for integrated cellular and WLAN networks", *Computer Networks*, Vol.47, No.2, Feb. 2005, pp. 167-183.

[7] http://www.euqos.eu/

[8] Project: "Q3M - QoS Architecture for Mobile Multicast Multimedia Services", http://www.workingonweb.com/q3m/

[9] F. Houéto, and S. Pierre, "Quality of service and performance issues in multiservice networks subject to voice and video traffics", *Journal of Computer Communications*, Vol.28, No.4, March 2005, pp. 393–404.

[10] Y. Fakhri, J. Vidal, D. Aboutajdine, B. Nsiri, "Downlink throughput Maximization in Multicarrier Wireless Communications Systems", *WSEAS Transactions on Communications*, Iss.8, Vol.5, August 2006, ISSN 1109-2742.

[11] R. Ko, C.He, "Using Composite MultiMedia Document Structure and Request Merging with On-Demand Broadcasting to Deliver QoS within Wireless Networks", *WSEAS Transactions on Communications*, Iss.8, Vol.5, ISSN 1109-2742, August 2006.

[12] Z. Bojkovic, D. Milovanovic, "Cross-layer Quality of Service for video wireless multimedia delivery: some challenges and principles", *Proceedings of the 5th WSEAS Int. Conf. on Data Networks, Communications & Computers*, Bucharest, Romania, Oct. 2006.

[13] C. Goudemand, F.Coudoux, M. G. Gazalet, P.Corlay, "QoS Optimization for Video on Demand Applications over Asymmetric Digital Subscriber Line", *WSEAS Transactions on Communications*, Iss.12, Vol. 4, December 2005, ISSN 1109-2742.

[14] D. Soldani, and M. Li, R. Cuny, QoS and QoE Management in UMTS Cellular Systems, John Wiley & Sons, Ltd, 2006.

[15] http://www.opnet.com, OPNET Technologies homepage.

[16] www.tkn.tu-berlin.de/research/trace/ltvt.html