The Importance of Adaptive Applications in Mobile Wireless Networks

Michael Neophytou, Kyriakos Stavrou, Vasos Vassiliou and Andreas Pitsillides

Department of Computer Science, University of Cyprus

75 Kallipoleos Str, 1678 Nicosia, Cyprus

{csp4mn1, tsik, vasosv, cspitsil}@cs.ucy.ac.cy

Abstract. In this paper, we illustrate the importance of applications' adaptability and QoS awareness in mobile wireless networks. We utilize a "Mobile QoS Adaptation" strategy aiming to deal with the inconsistencies arising from mobility events in a wireless mobile environment. The performance of the specific adaptation strategy is evaluated through simulations under different traffic environments in a wireless network based on 3GPP's UMTS structure. As a result, we show that the performance of a network can be improved when adaptive applications are used by both mobile and fixed hosts.

1 Introduction

The task of a communication network is to provide traffic transportation services between end users by conveying data from one tier to another in a reliable and timely fashion. This task means that a network is a shared entity; its services should not be dedicated to only a pair of communicating entities but to as many entities possible during its operational lifetime. The requirement that a network should service as many users possible simultaneously highlights the importance of the "*user capacity*" concept. Network planners have always strived to design networks that achieve high user capacities but have always faced the same limitation: the finite amount of resources which a network has available for utilization in order to fulfill its task. Probably, the most valuable network resource in modern networks is link capacity and its value is even greatly appreciated in wireless networks. This is mainly due to the inherent difficulty to achieve efficient high speed modulation schemes for data transmission over the wireless medium. The limited frequency band size and the competition among various providers also contribute to the problem of radio resources scarcity.

Assuming abundance of all kinds of resources and under ideal circumstances, the solution to resource scarcity would be the infinite upgrade of all insufficiently dimensioned network components. In this way, no resource shortages would ever occur and consequently, infinite number of users could be successfully supported by the network. However, real-world limitations rule out such a solution as the concept of e.g. infinite capacity links is clearly theoretical and not realistic. A more realizable approach to this situation is network dimensioning at the maximum possible level provided modern expertise and available means e.g. use extremely high capacity links or high processing power servers, routers and switches. Although such a solution would be feasible, two factors appear as insurmountable obstacles to its implementation. The first factor relates to the fact that a large inter-network comprised of smaller scale sub-networks is as efficient, on an end-to-end point of view, as the most inefficient sub-network by which it is composed. Even if large parts of a large network are super dimensioned, the achieved performance would still be lower than expected if smaller parts of the network are still limited by low resource availability. Since upgrading all existing networks is impractical, this solution turns out infeasible. The second prohibitive factor is the high degree of resource waste that would occur in such a super-dimensioned network since for most of the time the resources available in the network would remain unused and high under-utilization levels would be experienced. Many studies in the past have shown that congestion conditions in a network are not a constant phenomenon but rather occur at discrete time points during its operational lifetime as a result of the unpredictable fluctuations of the amount of the offered load.

Bearing these in mind, the turn to intelligent resource management solutions seems inevitable. In this context, a proposal made in [1] is the implementation of a QoS adaptation scheme. The proposed adaptation scheme is based on the observation that several applications can operate with acceptable performance when the provided QoS fluctuates within a range of values. By taking advantage of this fact, a methodology is designed to adjust the QoS the network offers to an application in order to achieve better resource management and increase network capacity when possible, by allowing more concurrent user sessions. Apart from the above beneficial effect, the adaptation of delivered QoS (where this is possible) contributes to reducing the probability of forced call termination. It is well recognized that a trade-off exists between minimizing call blocking and forced termination probabilities. While a network operator strives for good performance in both cases, a choice must be inevitably made between them; it is widely acknowledged (especially from user perspective) that minimization of forced termination probability is far more desirable.

This paper is organized as follows. In section 2, a general description of the proposed QoS adaptation framework is given. Section 3 presents various configuration aspects used in the analysis setup. In section 4 the acquired performance evaluation results are presented and discussed while section 5 concludes this work.

2 QoS Adaptation Framework Description

The QoS adaptation framework used in this work was originally proposed in [1]. Similar work has also been proposed for WATM networks in [2]. The framework was extended, taking into consideration formal control theory, in [3]. The overall adaptation process engages both the undergoing application and the network. On one hand, the application must be involved as it has to adapt to changes in the available QoS while on the other hand, the network has to adapt to the alterations made to the QoS requirements of the supported applications. From the network point of view, the adaptation process includes procedures for QoS *upgrade, downgrade* and *satisfaction*. A visual representation of these actions is shown in Figure 1.

From the application point of view, the adaptation process involves aspects such as bandwidth, delay, delay variation and packet loss adjustments triggered by mobility events in the network. To this direction, the concept of "*bandwidth window*" is introduced. This approach enables application adaptability to variations in the available network bandwidth during mobile host's migration between wireless cells. Such an adaptation process is possible only when the application under question is able to operate with acceptable performance over a range of bandwidth without being forced to terminate.

It has been shown widely in the literature that several applications, such as video, can scale down their QoS requirements without being forced to terminate. The reduction could refer to a different frame rate, picture resolution size, color depth, etc. [4] [5]. It has also been shown that hierarchical encoding and variable compression techniques can also support bandwidth adaptation [6].

In addition to bandwidth adaptation, there are situations where delay adaptation is also possible even for delay-sensitive applications such as multimedia, provided that the delay values are within certain limits [6].

Variations of delay due to longer resultant paths after handoffs can be resolved, to some extent, by content buffering prior to usage by the application. A trade-off hence exists between greater roaming capabilities and lower experienced quality by the terminal user. Although quality degradation is clearly not desired, it should be considered as the only alternative to the even more undesirable event of forced call termination.

Network QoS adaptation involves the phases of *re-evaluation* and *update*. Re-evaluation occurs during host migration across wireless cells in addition to the handoff procedures to govern the decision of Mob-QoS upgrade, downgrade or satisfaction. The update process is then necessarily invoked to resolve any inconsistencies that might arise in the network after a handoff.

2.1 QoS Satisfaction

If during a new call request or a handoff process, the network is capable of supporting the QoS requirements both in the wireless and backbone segments, then the handoff process can be completed without any need for adaptation. This situation can be considered as the optimum scenario.

2.2 QoS Downgrade

If during a new call request or a handoff process, the network is not capable of supporting the QoS requirements of the application, QoS degradation must be performed to avoid call termination. Considering that the update process is performed during every handoff event, the resulting signaling overhead is likely to be substantial. Therefore, two different strategies are proposed. The "*Minor QoS degradation update strategy*" is a simplistic and fast approach which takes advantage of intra-cluster handoffs (where cluster refers to a group of wireless cells) by adjusting resource reservations only over the wireless links. However, the drawback of this approach is that reservations over the wired links remain unchanged consequently leading to an inefficient state of under-utilization.

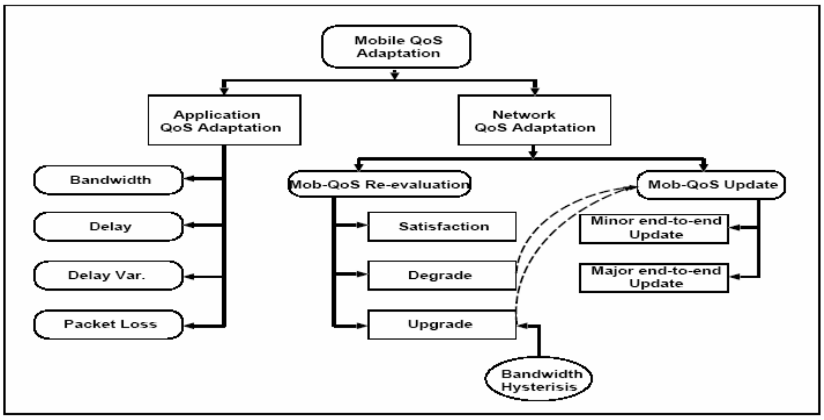


Figure 1: The Components of the Mobile QoS Adaptation Framework. [1]

On the other hand, the "*Major QoS degradation update strategy*" involves end-to-end resource allocation adjustment and is invoked post to inter-cluster handoff events. The facts that during inter-cluster handoffs a significant portion of the path is modified and this type of handoffs is less frequent constitute the execution of this update strategy necessary in spite of the signaling overhead it entails.

2.3 QoS Upgrade

Although the main goal targeted by the introduction of the QoS adaptation framework is to minimize the probability of forced call termination and/or blocking due to insufficient resources, the event of QoS upgrade is also introduced in order to better exploit resource abundance when this is available. However, a limiting factor called "*bandwidth upgrade hysteresis*" is considered even if the possibility of upgrade exists. This factor is considered in order to prevent unnecessary invocations of the upgrade process when no significant improvement will be made to justify the overhead caused by the upgrade process. In contrast to the QoS downgrade case, an end-to-end update process is required in the event of QoS upgrade for both types of handoffs since extra resource allocation should be performed to support the increased service quality received by the mobile host.

3 The Impact of QoS Adaptation on Network Performance

3.1 Network Topology

The network topology used in this work is based on the generic structure of a 3G UMTS network in order to examine the behavior of the QoS adaptation strategy under investigation, in the context of a modern wireless network.

A UMTS network is roughly decomposed into the Core network and UMTS Terrestrial Radio Access Network (UTRAN). The former includes multiple SGSN and GGSN nodes while the latter includes multiple RNC nodes each controlling several Node-Bs (i.e. base stations). A more detailed description of UMTS structure and node functionality can be found in [7]. The topology layout used is shown in Figure 2.

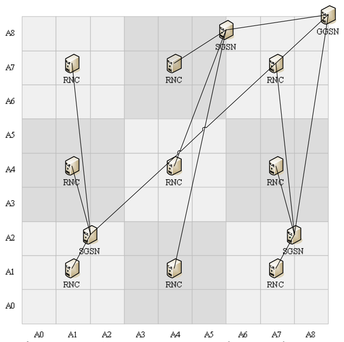


Figure 2: Custom network topology.

The number of mobile and fixed hosts introduced in the network was set to 27 and 9 respectively. The 27 mobile hosts were positioned in the 81-cell topology while the 9 fixed hosts were attached to the 9 terminal switches evenly distributed over the entire simulation area. The mobility of each mobile host is controlled independently by choosing uniformly between neighboring base stations as the target BS for the handoff.

3.2 Network Traffic

Call Type	BW_{min} (kbps)	BW_{max} (kbps)	BWΔ (kbps)	AIT (min)	AHT (min)
Mobile hosts					
VOI	4	12	2	2	3
VTE	32	384	88	2.22	3
MWB	384	2000	404	6.25	3
IMM	1000	4000	750	6.67	3
Fixed hosts					
TVD	4780	12780	2000	33.34	60
HDV	2048	4096	512	33.34	30

Table 1: Call types associated with mobile and fixed hosts.

A total of six call types were defined for use during the simulations. The selection of these types was based on [8] where a detailed and in-depth analysis of application categorization and characterization is performed in the context of Enhanced UMTS networks. These applications are plain Voice calls (VOI), Video Telephony (VTE), Multimedia Web Browsing (MWB), Instant Messaging for Multimedia (IMM), TV Programme Distribution (TVD) and High Definition Video Telephony (HDV). The attributes of each application/call type are also summarized in Table 1. These include the minimum and maximum BW requirements, the BW adaptation window (BWΔ), the average interarrival time (AIT) and the average holding time (AHT). The bandwidth adaptation window size used during both the upgrade and downgrade processes is set to the same value for all call types and is equal to $(BW_{max} – BW_{min}) / 4$.

Calls of the first four types (VOI, VTE, MWB and IMM) are initiated by mobile hosts and can terminate on either mobile or fixed hosts (MH-to-MH and MH-to-FH). Calls of the last two types (TVD and HDV) are initiated and terminated only on fixed hosts (FH-to-FH).

3.3 Evaluation Objectives

The objective of the QoS adaptation strategy evaluation is to explore the performance of the strategy under different wireless environments defined in [8]. The three environments simulated were an Office environment (OFF), Business City Center (BCC) and Vehicular (VEH). The OFF environment is chosen as a typical representative of indoor environments which are characterized by slow user speeds and relatively well defined mobility paths. The BCC environment represents an environment with base stations placed outdoors also covering internal building areas. The topological layout of these environments is usually modeled using the Manhattan model. Finally, the VEH environment applies to scenarios in urban and suburban areas with coverage cells having larger area and higher transmission power levels. Although the discrimination and characterization of alternative environments is performed based on a wide range of criteria (e.g. user mobility patterns, building density and location) the primary aspect of interest during our study is the variable user density in each environment and the distribution percentage of each call type in a given user population. The actual user populations values used during the simulation experiments are summarized in Table 2.

Table 2: User population and call type distribution per environment.

Call type	OFF		BCC		VEH	
	Num	%	Num	%	Num	%
VOI	36	24.0	8	25.8	5	42.0
VTE	24	16.0	5	16.1	2	16.6
MWB	30	20.0	8	25.8	2	16.6
IMM	36	24.0	9	29.1	2	16.6
TVD	16	10.6	1	3.2	1	8.2
HDV	8	5.4	0	0.0	0	0.0
Total	**150**	**100.0**	**31**	**100.0**	**12**	**100.0**

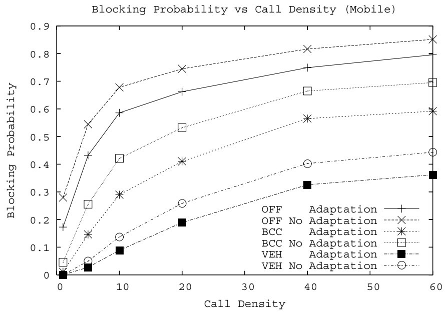
4 Performance Results

The beneficial effect of having applications participating in the QoS adaptation process was verified by the simulations. This section presents all simulation results and compares the three environments defined in the objectives. Interested readers can find results for a random topology and a different traffic mix in [1].

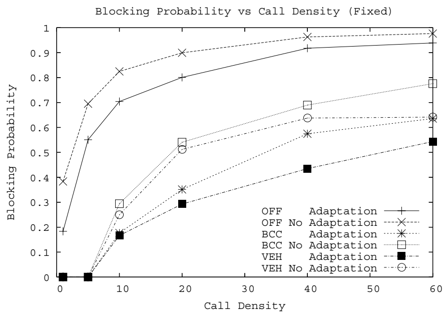
4.1 New Call Blocking Probability

Figure 3a depicts the blocking probability of mobile hosts' calls for the three different environments simulated and for a range of call densities. It is clear that the blocking probability of new calls originating from mobile hosts increases proportionally to population density of the environment. For the office environment (OFF), this probability reaches 85% (about 3 times the respective probability of the least dense vehicular environment). Blocking probability is directly correlated to call density and this is intuitive since increased call density means that an increased number of call admission requests is made to the network. Network resources are heavily used and hence, the probability of successful admittance of the requests is decreased.

The same trend appears in Figure 3b which shows the respective new call blocking probability for fixed host (FH-to-FH) calls. The densest environment (OFF) exhibits the highest blocking percentage because network resources are allocated by many low demanding mobile users and therefore high demanding fixed users experience high blocking probability.



(a)



(b)

Figure 3: New Call Blocking Probability. (a) Mobile calls, (b) Fixed calls.

This probability is reduced in less dense environments (BCC and VEH) because of (i) the lesser total number of users in the network (and consequently the more available resources) (ii) the lesser number of fixed users requesting admittance. In particular, in the BCC and VEH environments there exists only one of the possible two types of fixed calls and in spite of their high bandwidth requirements the fact remains that their percentage of the total population is lower compared to OFF.

Figures 3a and 3b provide one additional important piece of information. They illustrate the difference in the blocking probability for all environments if no adaptation was used. Adaptation causes a reduction in new call blocking in the order of 5% for the OFF and 10% for the BCC and VEH environments for both types of users.

4.2 Call Dropping Probability

Figure 4 presents the call dropping probability for mobile calls. The trends depicted are anticipated since call dropping probability appears increased in the case of the OFF environment while the VEH environment exhibits the best performance. User density plays an important role (as was the case with the blocking probability) given that the more users are active in the network the less available resources exist in order to fulfill QoS requirements of mobile hosts during handoffs. The QoS adaptation strategy comes into play when QoS satisfaction is impossible during user migration between cells but its application can not guarantee zero dropping rates; however, applying QoS adaptation clearly lowers dropping rates experienced in no adaptation scenarios. The reduction is 10% in all operating environments.

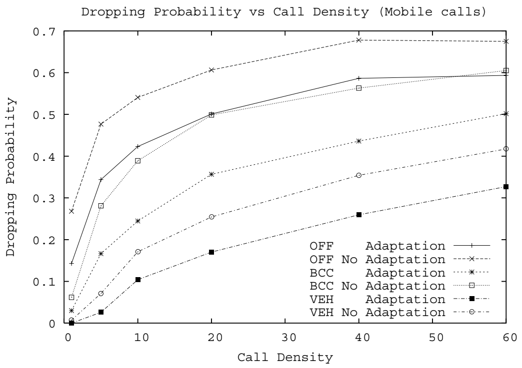


Figure 4: Call Dropping Probability.

4.3 Call Upgrade Probability

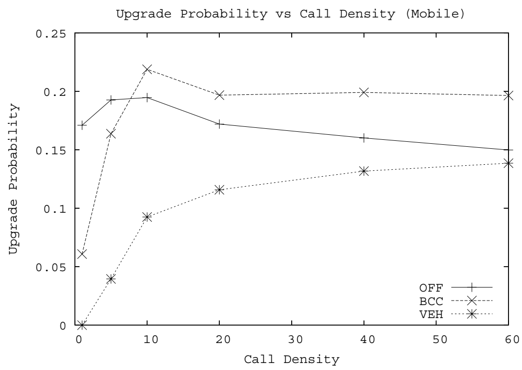


Figure 5: QoS Upgrade Probability (Mobile calls).

Figure 5 depicts the upgrade probability resulting from the simulations. By observing the plot, one notices that QoS upgrade is more probable in the BCC compared to the OFF environment. This was expected since availability of more resources (due to the lower number of users) enables QoS upgrade of mobile hosts whenever available resources exist. A discrepancy concerning the VEH environment is also obvious. Based on the previous reasoning, one would expect higher upgrade probability to be experienced in this environment. However, QoS upgrade appears to be least probable in the sparse VEH environment with the very low value of about 10%. Nevertheless, this abnormal behavior can be justified considering that low user population in this environment compared to the other two entails that the network is able to satisfy the initial QoS requirements of each call upon its establishment and therefore no need for future upgrade remains.

As far as fixed calls were concerned, no calls were upgraded in all simulated environments. This can be attributed to the fact that due to their high demands once fixed calls are accepted in the network, they will either settle with the initially granted QoS or will suffer degradation.

4.4 Call Downgrade Probability

The downgrade probabilities for mobile and fixed calls are shown in Figures 6a and 6b respectively. For mobile users, the results are rather intuitive. Specifically, Figure 6a shows that the OFF environment has the highest downgrade probability for the largest range of call densities. This can be attributed to the large user population of the specific environment and consequently to system's effort to minimize forced call terminations by degrading their QoS instead.

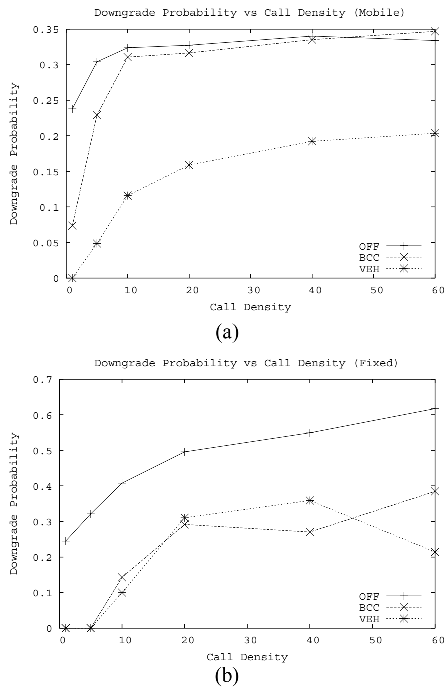


Figure 6: QoS Downgrade Probability. (a) Mobile calls, (b) Fixed calls.

Although in the BCC environment the situation does not differ much, the dropping probability is lower compared to the OFF environment. This can be justified by observing the system's upgrade probability. Specifically, BCC has the highest upgrade probability leading the system to often downgrades in order to minimize call dropping rate. For VEH, low dropping probability together with the low probability for upgrade, minimizes the necessity of downgrades as the network is already capable of effectively serving the offered load.

For the fixed users, the only environment that has the same behavior as for mobile users is the OFF one. This is due to the fact that the network can't effectively serve its workload and hence tries to save resources whenever it has the ability to. The other two environments have significantly lower downgrade probability which is mainly due to their small number of fixed users.

For both mobile and fixed users, the downgrade probability increases with the call density. The rationale behind this is that increased call density leads to more call requests and increases need for resource allocation. While trying to allocate these requested resources the network downgrades the QoS offered in order to decrease the dropping probability.

4.5 QoS Satisfaction Probability

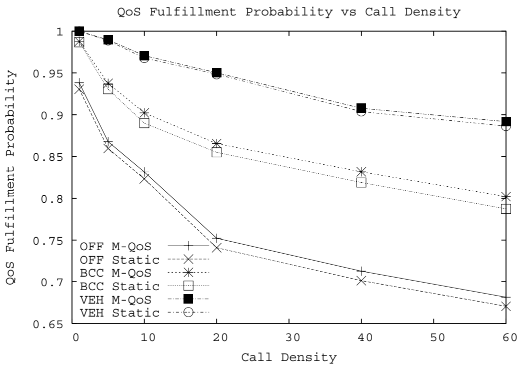


Figure 7: QoS Satisfaction Probability.

The QoS satisfaction probability presents no surprising results given the dropping, upgrading and downgrading probabilities of the environments under study. Specifically, in the OFF environment calls are often dropped in spite of downgrading a non negligible subset of them. This explains the fact that the network is rarely able to fulfill the requested QoS. In the BCC environment, the dropping probability is lower than the OFF environment and higher compared to the VEH one. This directly explains QoS satisfaction behavior experienced in this environment. Finally, the VEH environment manages to effectively fulfill the requested QoS as it does not drop calls often, as can be seen from Figure 7 while rarely downgrading on-going calls.

In all cases, satisfying the QoS values supported in the previous cell (Mob-QoS) has greater probability (2-3%) than satisfying the original QoS defined during call setup (static).

5 Conclusions

An overview of the results obtained in this work indicates that in a number of different operating environments (different traffic mix) and for different applications, adaptability of the latter and adaptation techniques utilization in the network improves performance of all metrics studied. However, for all metrics the recorded values worsen when the population of the environment increases. This is an anticipated behavior since the potential for effective adaptation is closely correlated to the availability of resources (when upgrade adaptation is performed) and applications' tolerance to QoS degradation (when downgrade adaptation is performed). When the network operates under heavy load, it is quite natural that upgrade adaptations will be very rare (if possible at all) since resources are being heavily used to support the active users. At the same time, downgrades become more and more probable as the network strives to regain and reallocate its resources. This downgrade process cannot continue indefinitely and at some point the adaptation strategy proves inadequate to prevent forced call termination.

Having the opportunity to observe the performance achieved by both mobile and fixed users, we conclude that the latter experience better QoS service compared to the former at least when referring to the least desirable event of forced termination. This can be attributed to the fact that fixed users have greater QoS requirements and although these requirements reflect the high demanding nature of the applications they execute, there are at the same time wider margins for adaptation should the original QoS requirements be unfulfilled.

The general conclusion drawn for the previous discussion is that the proposed QoS adaptation strategy surely performs well and improves network performance, but when the network operates at extreme conditions in terms of offered load, the strategy itself is not sufficient for large scale, efficient resource management.

Acknowledgements

This work has been partly supported by the Research Promotion Foundation of Cyprus under the project BINTEO and by the University of Cyprus under the project ADAVIDEO.

References

[1] V. Vassiliou, "Performance Evaluation of a QoS Adaptation Framework for Mobile Networks," *In Proc. 11th European Wireless Conference (EW 2005)*, Nicosia, Cyprus, April 2005.

[2] G Lin. and C-K Toh, "Performance Evaluation of a Mobile QoS Adaptation Strategy for Wireless ATM Networks," *In Proc. IEEE International Conference on Communications ICC 1999,* Vancouver, Canada, June 1999.

[3] V. Vassiliou and A. Pitsillides, "QoS Adaptation Control in Virtual Circuit Switched Mobile Networks," *In the Proc. of the International Conference on Intelligent Systems And Computing: Theory And Applications (ISYC 2006)*, Ayia Napa, Cyprus, July 2006.

[4] J.Liu, J.Lin, W. Shih, A. Yu, J. Chung and Z. Wei. "Algorithms for Scheduling Imprecise Computations", *IEEE Computer*, 24(5):58–68, May 1991.

[5] M. Schwartz, "Network Management and Control Issues in Multi-Media Wireless Networks", *IEEE Personal Communications Magazine,* 2(3):8–16, June 1995.

[6] C. Diot. "Adaptive Applications & QoS Guaranties", *In Proc. of International Conference on Multi-Media Networking*, pages 99–106, September 1995.

[7] 3G TS 23.101 3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects, General UMTS Architecture (version 3.0.1)".

[8] J. Ferreira and F. Velez, "Enhanced UMTS Services and application characterization", Telektronik, Vol. 101, No. 1, 2005.