A Proactive Resource Management in 3G Networks

5 authors, including:

Abderrahmane Lakas
United Arab Emirates University
90 PUBLICATIONS **899** CITATIONS
SEE PROFILE

Farag Sallabi
United Arab Emirates University
45 PUBLICATIONS **384** CITATIONS
SEE PROFILE

Mohammed Boulmalf
Université Internationale de Rabat
91 PUBLICATIONS **693** CITATIONS
SEE PROFILE

Hesham Elsayed
Arab Academy for Science, Technology & Maritime Transport
5 PUBLICATIONS **6** CITATIONS
SEE PROFILE

A Proactive Ressource Management in 3G Networks

A. Lakas, F. Sallabi, M. Boulmalf, K. Shuaib, and H. Elsayed

College of Information Technology
UAE University
PO Box 17555, Al Ain, UAE
{alakas, f.sallabi, boulmalf, k.shuaib, helsayed}@uaeu.ac.ae

Contact details of the first author:

Abderrahmane Lakas
College of Information Technology
UAE University
PO Box 17555, Al Ain, UAE
Tel.: +971 50 4472688
Fax.: +971 03 7051681
Email: alakas@uaeu.ac.ae

Abstract

Providing quality of Service (QoS) over the 3G wireless networks such as the Universal Mobile Telecommunications System (UMTS) is one of the most active and challenging research area. 3G networks introduce IP services along with the traditional 2G phone system, and therefore represent an opportunity for real-time and multimedia applications such as the video conferencing. In this paper, we suggest a proactive QoS provisioning technique based on putting demands for network resources ahead of time as the mobile host moves from one cell to another. In this technique, the admission control algorithm prepares the mobile host's neighboring cells and therefore allowing real-time applications to move from one cell to another while maintaining and preserving its contracted QoS.

Keyword

Proactive provisioning; QoS; Admission Control; UMTS

A Proactive Resource Management in 3G Networks

A. Lakas, F. Sallabi, M. Boulmalf, K. Shuaib and H. Elsayed

College of Information Technology,
U.A.E. University, Al-Ain, P.O. Box: 17555, U.A.E
E-mail: {alakas,f.sallabi, boulmalf, k.shuaib, helsayed}@uaeu.ac.ae

***Abstract*—Providing quality of Service (QoS) over the 3G wireless networks such as the Universal Mobile Telecommunications System (UMTS) is one of the most active and challenging research area. 3G networks introduce IP services along with the traditional 2G phone system, and therefore represent an opportunity for real-time and multimedia applications such as the video conferencing. However, just like for wired IP, such applications require strict QoS requirements in order to be delivered reliably and timely. In addition to the traditional techniques used for providing QoS to multimedia applications, wireless networks requires new methods for maintaining QoS in a dynamic environment and mobile users. In this paper, we suggest a proactive QoS provisioning technique based on putting demands for network resources ahead of time as the mobile host moves from one cell to another. In this technique, the admission control algorithm prepares the mobile host's neighboring cells and therefore allowing real-time applications to move from one cell to another while maintaining and preserving its contracted QoS.**

Keywords: Proactive provisioning; QoS; Admission Control; UMTS

INTRODUCTION

Providing quality of Service (QoS) over the 3G wireless networks such as the Universal Mobile Telecommunications System (UMTS) is one of the most active and challenging research area. 3G networks introduce IP services along with the traditional 2G phone system, and therefore represent an opportunity for real-time and multimedia applications such as the video conferencing. However, just like for wired IP, such applications require strict QoS requirements in order to be delivered reliably and timely. In addition to the traditional techniques used for providing QoS to multimedia applications, wireless networks requires new methods for maintaining QoS in a dynamic environment and mobile users. In this paper, we suggest a proactive QoS provisioning technique based on putting demands for network resources ahead of time as the mobile host moves from one cell to another. In this technique, the admission control algorithm prepares the mobile host's neighboring cells and therefore allowing real-time applications to move from one cell to another while maintaining and preserving its contracted QoS.

The remainder of the paper is organized as follows. In section 2 we present an overview of the UMTS architecture. In section 3, we discuss the problem of handover and present the rational for using a proactive provisioning of QoS. Section 4 consists discusses the details of the proactive QoS provisioning algorithm. We present some performance results obtained from simulating our algorithm in section 5. Finally, we conclude this paper in the last section.

I. Overview of UMTS

The UMTS network architecture consists of three parts: the user equipment (UE), the UMTS terrestrial radio access network (UTRAN), and the core network (CN). Figure 1 depicts the packet switched (IP) service of the UMTS network architecture [7]. The UE consists of the mobile equipment, which is the radio terminal used for radio communication with UTRAN, and the user identity module (USIM). UTRAN consists of two components: the Node B and RNC (Radio Network Controller). Node B is functionally similar to the base station of the 2G systems. It converts the data flow between the UE-Node B interface and the Core Network-UTRAN interface. It also takes part in some of the radio resource management such as inner loop power control. The RNC is functionally similar to the BSC (Base Station Controller) in the 2G systems. It owns and controls the radio resources in its domain. RNC is responsible for the load and congestion control, and performs admission control and code allocation for new radio links to be established. It also makes the handover decision and outer loop power control. Each RNC controls one or more Node Bs. The core network consists of the routers and switches that connect the RNCs together, and connect the UMTS network to other networks like the Internet and the telephone network.

UMTS defines four QoS traffic classes: *conversational*, *streaming*, *interactive* and *background*. The conversational class provides strict delay guarantees, while background class offers no qualitative or quantitative guarantees. It can be at best referred to as a best effort class. When strict guarantees are necessary, for real-time voice and video applications, using conversational class seems to be a good choice. Slightly relaxed in terms of delays is the streaming class, to which all streaming applications can be mapped.

While interactive traffic follows a request-response pattern and can only justly provide qualitative guarantees, background class is similar to the best effort traffic consisting of bulk (e.g. ftp) and asynchronous traffic flows (e.g. e-mail) that will fall into this category. Table I below lists some of the attributes of the UMTS QoS classes.

With the emergence of new wireless technologies such as WLAN, and mobile networks that also carries data, new constraints appeared in the picture of QoS. Real-time applications are now subjected to new dynamics related to the over-the-air constraint represented by the variability of the offered bandwidth, the incurred delays and jitter. By moving from one cell to another one, the mobile experiences different QoS environments, because of the difference in the signal strength, the coverage, and the load of the visited cell. The handover process is limited to maintaining connectivity, and therefore, must be extended to also maintaining QoS that the mobile user contracted. The handover process, incurs extra delays due to the "*brake*" from the old point of attachment, and the "*make*" with the new point of attachment.

TABLE I. ATTRIBUTES RELATED TO UMTS CLASSES

UMTS QoS Class	Interactive	Conversational	Streaming	Background
Maximum Bit Rate	√	√	√	√
Maximum SDU Size	√	√	√	√
Delivery Order	√	√	√	√
SDU format information	√	√		
SDU Error Ratio	√	√	√	√
Residual Bit Error Ratio	√	√	√	√
Delivery of Erroneous SDUs	√	√	√	√
Transfer Delay	√	√		
Guaranteed bit rate	√	√		
Traffic priority			√	
Allocation priority	√	√	√	√

II. PROBLEM DESCRIPTION

The objective is to seek a solution that allows the preservation of the mobile's QoS during the handover operation. In order to do so, two ways can be considered: (1) either minimize the handover delays, or (2) do with these incurred extra delays by proactively transferring the QoS requirements to the neighboring cells ahead of time. This can be done by monitoring the mobile movement and request network resources from the neighboring cells when the handover operation is triggered. The following section details the operation of this proactive technique.

III. PROACTIVE PROVISIONING

The proactive provisioning algorithm consists of requesting network resources ahead of time from the neighboring cells in a provision of possible move of the mobile to any of the potential cells. By doing so, when the mobile is handed over to the control of the next cell, resources would have been already reserved. In addition, when the mobile moves to the next cell, the resources that are already reserved proactively in cells that are no longer neighboring cells are released.

This technique poses, however, few questions: first, the criteria in selecting the neighboring cells. Here, we have to define what the cells that we consider as neighbors are. Second, the notion of proactively reserving resources amounts to reserving resources in advance; resources that are not going to be used necessarily. This poses an issue of fairness when bandwidth is reserved and not used at the cost of other sessions that may not be admitted due to lack of resources. This can be solved by either minimizing the number of neighboring cells, or by making the reservation effective the latest possible. A depiction of proactive reservation is presented par the figure below.

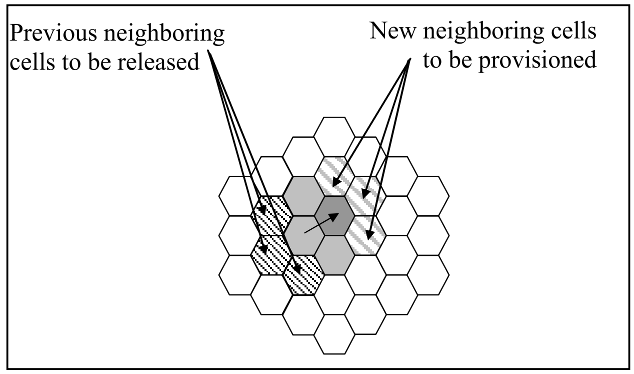

Fig 1. Proactive Provisioning

As for the notion of neighborhood, we have considered the following definitions:

1) *Geographical neighborhood*: in this case, neighboring cells are the ones that are sharing the same coverage border. In our environment, each cell will have six geographical neighboring cells.

2) *Potential neighborhood*: in this case, we define the neighboring cells as the ones that present a higher probability to be visited next by the mobile. Potential cells are a subset of geographical neighboring cells in a sense that the later have an equal probability to be visited next if we don't consider any other criteria. However, if we apply some predictive techniques we reduce the number of the neighboring cells by identifying the most probable cells among the neighboring cells to be visited next. Predictive techniques calculate this probability based, for instance, on the current and/or past path taken by the mobile while moving in the network. A variation of this technique is presented in[13].

3) *Signal strength-based neighborhood*: In this method the selection criteria consists of assigning a probability to be visited next, to the cells based on the signal strength and/or SNR. In this case, the signal strength/SNR scale is partitioned into ranges. Each range corresponds to a probability to be visited next (see Figure 3). Cells within the best signal strength/SNR will be selected as neighboring cells. In the figure below, the cells *Cell1*, *Cell3* and *Cell4* are considered the most probable neighboring cells to be visited next based on their current signal to noise ratio SNR.

Although it is very useful to assess these three methods for determining the neighborhood, it is not the purpose of this research to do so. The provisioning techniques that we adopted here is independent of the neighborhood notion. In actuality, our algorithm can use any of these techniques, which remains transparent to them. However, for the sake of proof of concept we have chosen to use the first approach – the geographical neighborhood – as our first step in this research.

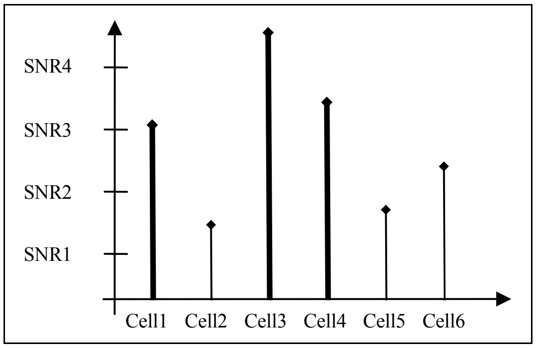

Fig 2. Signal strength-based neighborhood

A. Provisioning Algorithm

The proactive resource reservation is executed every time a mobile host carries out a hand-over from one cell to another one. The algorithm is executed in the following steps:

1) Determine the previous neighboring cells: in this step, we determine the previous neighboring cells: *PNC*. *PNC* is the set of neighboring cells of the cell that the mobile is leaving.

2) Determine the new neighboring cells: in this step, we calculate the new neighboring cells: *NNC*. *NNC* is the set of neighboring cells of the cell that the mobile is entering.

3) Calculate the cells to provision: in this step we determine the new cells among the *NNC* to be provisioned proactively: *NNCp*. Note that not all new neighboring cells are to be provisioned since some of them are already provisioned part of the previous neighboring cells. The cells to be provisioned are defined as the following:

$$NNCp = NNC - \{NNC \cap PNC\}$$

4) Calculate the cells to release: *PNCr*. Not all previously provisioned cells are to be released; some of them are part of the new neighboring cells that needs to stay provisioned. The cells to be released are calculated according to the following:

$$PNCr = PNC - \{NNC \cap PNC\}$$

5) Provision the new cells. In this step, the cells in *NNCp* are granted the required resources after the admission control is applied. The process for acquiring resources in the neighboring cells proactively is described below according to specific policy.

6) Release the old cells. In this step, the cells that were proactively provisioned, and that are no longer part of the neighboring cells, are un-provisioned, and proactively acquired resources are relinquished.

B. Proactive Admission Control

The process of resource admission control in the context of proactive provisioning is applied differently. The applications that are classified as requiring guaranteed QoS are handled by the process admission control in such a way that since resources are acquired for them proactively, there is no need for them to pass the admission control test. Once a real-time session is admitted at the start of the session, resources are also reserved proactively for it ahead of time. Thus, during a hand-over from one cell to another, the session will use the pre-assigned resources. However, the allocation of resources for the neighboring cells has to be subjected to admission control.

C. Policy based Admission Control

The outcome of the admission control process being admission or rejection, mobile hosts subjected to this process may be admitted or rejected based on the load of the network. In the case where all the network resources are allocated the mobiles with guaranteed QoS will have priority over the best effort sessions. This priority grants a preferential treatment to guaranteed QoS to be admitted when admission control is applied. Furthermore, guaranteed sessions will have the right to preempt non-QoS sessions, whence resources are requested proactively in the neighboring cells. We use policies to flexibly define the behavior of mobile hosts roaming in the network. We adopt the policy framework of the IETF [16], which uses the concepts of a Policy Decision Point (PDP) and a Policy Enforcement Point (PEP). The components of the policy-based control that is used in our system are depicted in the figure below. It consists of a PDP, PEP, a policy repository, and a resource manager. The PEP is located on the base station, the PDP is located on a policy server that oversees and controls the policy enforcement of the QoS parameters, and the resource manager is typically located on the SGSN.

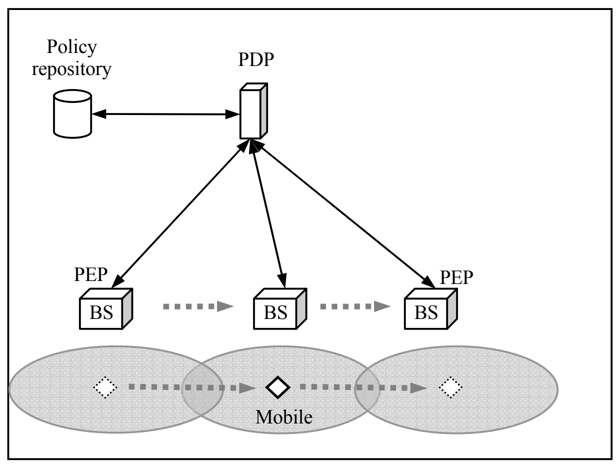

Fig 3. Policy-based system.

A *PDP* represents a controlling entity that applies policies to the PEP. The PDP persistently monitors the state of the available resources and of the mobile's preferences and uses this information to evaluate the policies' conditions. If the circumstances are such that the "if" condition of a policy becomes true, then the PDP decides to enforce the actions defined in the "then" part of the policy. Our PDP is also responsible for selecting the best configuration according to the user preferences and the current available resources. A *PEP* represents the controlled entity upon which policy decisions are being enforced (by the PDP), yielding a constrained behavior of the PEP. A *Policy Repository* contains (inactive) policies written in a policy specification language. The repository allows policies to be flexibly downloaded into a PDP, possibly at run-time.

Its main advantage is that policies are platform independent. In our system, the repository for instance contains defines the conditions which invoke the proactive handoff policies. It determines how to execute a handoff. To retrieve the appropriate policies, the PDP matches the application preferences from the mobile and bind them to the actions to be taken should policy applies.

We expect that the policy repository will typically reside in the fixed Internet, thus enabling a user to consistently apply the same policies to all of his devices. A PDP can generally use external information sources to come to its decisions. The resource information that determines whether a policy should be applied or not, is dictated by the *Resource Manager*. It is responsible for monitoring available resources, such as availability of networks, available bandwidth, signal strengths, signal to nose ratio, packet loss of a channel …etc.

IV. Performance Assessment

This section contains a description of the simulation environment, and some of the results using the UMTS simulator developed at the IT College, UAE University. We have extended the simulator to include a module for proactive provisioning.

A. *Simulation description*

The UMTS reference scenario adopted during our simulation campaign consists of 37 hexagonal cells where we assume that each cell has six neighboring cells. Mobile users move from one cell to another one following one of the six directions and roam freely across the whole network. We have considered both cases of pedestrian and vehicular users. Pedestrian users move at speeds generated randomly with a mean speed of 3 km/h for pedestrian users, and with mean speed of 60 km/h for the vehicular ones. The users enter the system so as to maintain a given load of the system – maximum load factor-, and are assigned a residence time in the system that is exponentially distributed. The residence time of a user in a cell is exponentially distributed, its mean value given by the ratio between the radio coverage diameter of a cell and the user average velocity. Handover events are distributed according to a decreasing exponential function with mean equal to the sojourn time. For the sake of simplicity we have chosen two classes of services: Best effort users (BE), and Real-Time users (RT). The RT class of service is defined by a guaranteed service and has priority to access the network resources over BE class of users.

TABLE II. QoS Classes

Class of Users	Description	QoS
BE	Best-Effort	No guarantee on delay Possible jitter, Packet loss
RT	Real-Time	Strict guarantee on delay Minimal to no delay jitter Minimal to no packet loss

The traffic generated by each user in the system can be of the following types: voice, videoconference, real-time (RT) streaming, best-effort (BE) streaming. The real-time service provides users guaranteed services, while best-effort service does not provide any guarantees over the quality and timeliness of the connection. In the presented simulation campaigns we considered two scenarios:

Scenario1: In this scenario, we varied the maximum load factor and measured the impact of the network load on the drop probability of both classes of services, RT and BE. We collected the performance results in term of session's drops incurred by the admission control, the handover and the network load control.

Scenario2: in this scenario, we varied the ratio of RT sessions over the BE sessions, and measured the impact of coexisting of these two classes on their respective drop probability. We started loading the system with 100% BE traffic, and increasingly loading the system by RT traffic. For this scenario, we chose to maintain the system at a maximum load factor of 0.9.

B. *Simulation results*

The results of the simulations carried out based on the scenarios described above are consistent with the expectation of the proactive provisioning technique. First, they show that as we increase the number of real-time call injected in the network, the number of the best-effort calls that remains active in the system decreases proportionally so as to keep the load factor constant. Figure 4 shows that as we vary the ratio of the real-time calls over the best-effort calls, the best-effort calls progressively exists the system to leave in order to accommodate the newly activated real-time calls with the network resources. In this case, we have considered a ratio of 1 as being 100% best-effort calls in the system and 0% real-time calls. A ratio of 0.9 corresponds to 90% of best-effort calls and 10% of real-time calls. Similarly, a ratio of 0.2 corresponds of 20% of best-effort calls and 80% of real-time calls. We also notice that while the number of best-effort calls that remain active in the system decreases, the number of active real-time calls remains constant. This is explained by the increasing pressure put by the latter on the best-effort calls in acquiring more and more of the network resources that they request at the cost of best-effort calls.

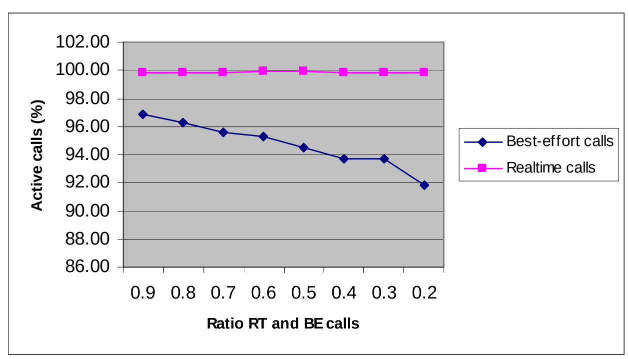

Fig 4. Effect of the ratio RT-BE calls on the calls active in the system

Second, the results show that if we consider the variation of the ratio of best-effort/real-time calls on the drop probability, we found that as more and more real-time calls are arrive into the system, the drop probability of the best-effort calls increases dramatically. This result complements the previous test, by clarifying why the best-effort calls leave the system as the real-time calls enter it. This is explained, by the constraints posed by the real-time calls on the best-effort calls by acquiring the network resources forcing the best-effort calls to terminate. Note that, there are three contexts where a call can terminate: First during the hand-over. When a user moves from one cell to another one, a call may not find enough resources to continue the session. Real-time sessions will be given higher priority over best effort when it comes to terminate a call because of lack of resources.

In addition, a call may be preempted when another call with higher priority is handed over to a new cell and where resources are not available. And, third: during admission control when a call is newly activated in a cell that is congested with real-time calls. In this case also, preferential treatment is granted to real-time sessions over best-effort sessions.

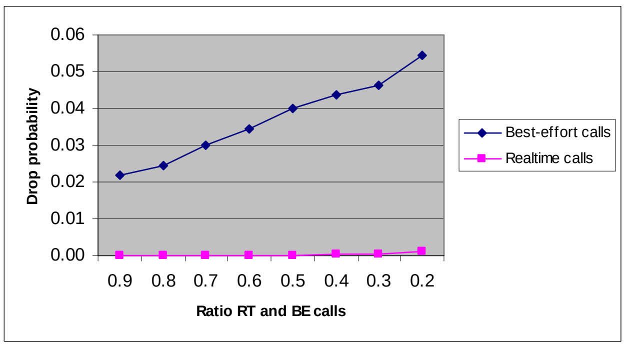

Fig 5. Effect of the ratio RT-BE calls on the drop probability

Figure 5 depicts the increase of drop probability when we increase the ratio of best-effort calls over real-time calls. Here, we also notice the drop probability of the real-time calls which is constant and almost nil. This can be easily explained by the nature of the proactive provisioning technique. Real-time calls are not only granted higher priority over best-effort calls on acquiring and keeping the network resources, but also, by doing so proactively. When, a real-time call moves to a cell and acquires resources, it also requests resources in the neighboring cells. This aggressive way of seizing the network resources forces the call drops of best-effort users to high numbers.

CONCLUSION

In this report, we have presented a new technique for proactive QoS provisioning in wireless network. This technique allows real-times sessions to have priority in using and maintaining the network resources over non real-time calls. The proactiveness consists of provisioning the neighboring cells ahead of time so as to allow real-time sessions to moves around without interruption of disruption in the quality contracted originally. We have presented the algorithms implemented along with measurements carried out to test the performance of this technique. The results showed the benefits of this technique; however, it also brought forward more questions as to the fairness (or unfairness) with regards to the non real-time calls in accessing the network resources. Therefore, we intend to pursue this research to optimize this technique by introducing parameters that allow some degree of fairness among real-time and non real-time calls. Initial analysis, suggests the use of more classes, the introduction of re-negotiable QoS, and a fairness factor which will determine a threshold beyond which real-time calls will have lesser and lesser priority over the non real-time calls. The fairness factor, a function of the drops incurred by the increase in the real-time activity, will help in regulating the ratio of non real-time traffic over the real-time traffic.

REFERENCES

[1] Ye; Papavassiliou, S.; Anastasi, G.; Puliafito, A., "Strategies for dynamic management of the QoS of mobile users in wireless networks through software agents", Computers and Communication IEEEs, 2002. Proceedings. ISCC 2002. Seventh International Symposium on , 2002 Page(s): 369 -374

[2] Wu, Si; Wong, K.Y.M.; Li, Bo, "A dynamic call admission policy with precision QoS guarantee using stochastic control for mobile wireless networks", Networking, IEEE/ACM Transactions on , Volume: 10 Issue: 2 , April 2002 Page(s): 257 –271.

[3] Das, S.K.; Chatterjee, M.; Kakani, N.K. , "QoS provisioning in wireless multimedia networks" Wireless Communications and Networking Conference, 1999. WCNC. 1999 IEEE, 1999 Page(s): 1493 -1497 vol.3.

[4] Chan, Jonathon; Seneviratne, Aruna; Landfeldt, Bjorn; "The Challenges of Provisioning Real-Time Services in Wireless Internet", Telecommunications Journal of Australia, vol 50 no. 3 pages 37 – 48.

[5] 3GPP, Quality of Service (QoS) Concept and Architecture, TS 23.107, V5.8.0, 2003-03.

[6] 3GPP, Network Architecture, TS 23.002, V6.0.1, 2003-03

[7] 3GPP, UTRAN Overall Description, TS 25.401, V6.0.0, 2003-03.

[8] F. Sallabi, A. Karmouch, "Resource Reservation Admission Control Algorithm with User Interactions " Proceedings of the Globecom'99, December 5-9, 1999, Rio de Janeiro, Brazil.

[9] H. Holma and A. Toskala, "WCDMA for UMTS", John Wiley & Sons, Ltd., 2002.

[10] A. Solana, A. Bardaji and F. Palacio, "Capacity Analysis and Call Admission Techniques for CDMA Packet Transmission Systems, 4th IEEE International Workshop on Mobile and Wireless Communications Network (WMWC2002); Stockholm- Sweden, September 9-11, 2002.

[11] 3GPP, Radio Frequency system scenarios, TR 25.942 V6.0.0, 2002-12.

[12] Yu. Cheng, W. Zhang, DiffServ resource allocation for fast handoff in wireless mobile internet, IEEE Communications Magazine 40 (5) (2002) 130–136.

[13] Wee-Seng Soh and Hyong S. Kim, "QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques," (one of the five papers selected from WTC'02) *IEEE Commun. Mag.*, pp. 86-92, Jan. 2003.

[14] S. Sroka and H. Karl, "Performance Evaluation of a QoS-Aware Handover Mechanism", In Proc. 8th IEEE Symp. on Computers and Communications, pp. 117-124, Kemer, Turkey, July 2003.

[15] A. Lakas and M. Jaseemuddin, Managing and Maintaining QoS in Mobile and Wireless Networks, Nortel Networks Design Forum, December 2000.

[16] Durham, D., Boyle, J., Cohen, R., Herzog, S., Rajan, R, Sastry, A, "The COPS (Common Open Policy Service) Protocol, RFC: 2748, January 2000, ftp://ftp.funet.fi/rfc/rfc2748.txt.