QoS-aware Power Control and Handoff Prioritization in 3G WCDMA Networks

Tajje-eddine Rachidi, Amal Yasmine Elbatji, Mehdi Sebbane, and Hicham Bouzekri
Alakhawayn University in Ifrane,
PO.Box 1881, Ifrane 53000, Morocco
T.Rachidi@alakhawayn.ma

***Résumé*—L'intégration dans les systèmes de contrôle de puissance en boucle fermée dans les réseaux de 3ème génération à base du WCDMA (tel que le UMTS) des paramètres de Qualité de Service (QaPC) peux résulter en une meilleure distribution des ressources (notamment radio) au niveau des stations de base. De la même manière l'intégration des paramètres QoS dans la priorité des handoff (QaHO) peux aussi résulter en de meilleure performance comparé aux techniques dites « aveugles ». Cet article présente deux telles techniques (QaPC et QaHO) qui utilisent la classe de service, le débit ainsi que le descripteur de dégradation de service (SDD) [Lataoui 2000]. Les performances obtenues pour ces deux techniques combinées à l'aide du simulateur développé en [Elbatji 2003] sous diverses conditions de charges, de trafic et de stratégies d'admission sont présentées. Les résultats montrent que ces techniques combinées améliorent l'utilisation des ressources radio par 22% réduisent le blocage lors des handoffs par 12 %.**

***Abstract*--Quality of Service aware power control (QaPC) mechanisms that supersede to the traditional closed loop power control in Wideband Code Division Multiple Access (WCDMA) type of networks, such as the Universal Mobile Telecommunication System (UMTS), provide significant advantages over blind channel estimation mechanisms. These mechanisms integrate specific QoS requirements of users in power control decision yielding optimal use of resources available at the base. Similarly QoS aware prioritization of handoffs (QaHO) that leverage user QoS profile can also yield significant improvements over blind prioritization. This paper presents two such QaPC and QaHO mechanisms which are based on the class of service, the bitrate, and the Service Degradation Descriptor (SDD) [Lataoui 2000] as enabling QoS parameters. The performance of our combined QaPC and QaHO mechanisms obtained using the testbed described in [Elbatji 2003], under a variety of load and traffic scenarios, and admission strategies is also presented. The results show that, when measured against blind mechanisms, the combined QaPC and QaHO significantly improves contract upholding of premium service mobile users, as well as improve resource utilization by more than 22% while improving handoffs failures by 12 %.**

Keywords--QoS provisioning, Multimedia QoS support, Closed loop power control, Handoff prioritization

I. INTRODUCTION

The rational behind the Universal Mobile Telecommunication System (UMTS) evolution is the delivery of multimedia services characterized by stringent real time requirements, great sensitivity to delivery delay and packet loss, and the need for considerable *wireless* resources. UMTS, therefore, supports QoS provisioning through four

(4) basic classes of service [3GPP 2002a, 3GPP 2002b]:

Class1: Conversational (high sensitivity to delay and jitter);

Class2: Streaming (medium sensitivity to delay, and high sensitivity to jitter);

Class3: Interactive (low sensitivity to delay, high sensitivity to round trip delay time and Bit Error Rate (BER));

Class4: Background (no delay sensitivity, high sensitivity to BER).

Each of these classes imposes different QoS requirement on the UMTS network which must be maintained during the lifetime of the corresponding connections.

Provisioning QoS over WCDMA-based air interface cannot be fulfilled solely by proper Admission Control [Zheng 1997] and efficient Scheduling [Wong 2003]. This is due, on the one hand, to the inherent characteristics of the wireless link [Forman 1994, Satyanarayanan 1996], that is, user mobility and fading channel (time variations) [Rappaport 2001, Stuber 2001, Proakis 2000], high error rates, inherent interference limited characteristics of WCDMA [Ericsson 2001, Dahlman 1998], and low and varying bandwidth (2Mbps at most); and to the unexpected Soft Handoffs (SHOs) requests on the other hand. The former effects have been, until recently, catered for using closed loop power control mechanisms, that operate solely on the basis of channel gain, but that are not aware of QoS requirements of underlying connections. This *blind* mode of operation does not necessarily yield optimal power utilization, especially when other non premium connections in the system are willing to be *degraded*; that is, they are capable of adaptation, and willing to have their required bitrate/power reduced. While the later, that is unexpected SHOs, have been tackled using either reservation or prediction techniques [Soh 2003].

Our work aims at showing that user willingness to be degraded can be used to augment both traditional closed loop control mechanism for congestion (the effects inherent wireless link effects described above) handling, as well as, to improve handoff by reducing the rate of dropping of SHO requests.

A Lucent patented framework for modeling user willingness to be degraded as a new QoS parameter has been presented in [Lataoui 2000]. Therein, the Service Degradation Descriptor (SDD) is a number between 0 and 5; the larger the SDD is the more willing is the user/connection[1] to be degraded, and eventually dropped.

We adopt SDD in this work too, and use it together with the service class (1-4) and the bitrate as enabling QoS parameters for the mechanisms we seek to develop, that is a system that combines both QoS-aware closed loop power control techniques and QoS-aware prioritization of SHOs in WCDMA-based 3G networks. We also seek to quantify the benefits of the combination of our two QoS aware schemas from the perspectives of both the network provider (that is resource/bandwidth utilization) and the user (forced terminations, and rate of acceptance of SHOs).

Specifically, we aim at building mechanisms that **1.** cope with the inherent characteristics of the wireless link, and **2**. minimize the probability of dropping of Soft Hand offs (SHOs), while **3.** maintaining QoS requirements of active connections, and **4.** achieving high system utilization.

---

[1] The words user, subscriber and connection will be used interchangeably hereafter.

In [Bhatti 1998], a QoS information model for making adaptation decisions is described, and in [Choukair 2003] a run time adaptation of UMTS services to available resources is presented. Nonetheless, to our knowledge, there hasn't been much work specifically on QoS-aware closed loops and QoS-aware prioritization. We had, however, touched upon this issue in [Abid 2001]. Furthermore, the European Telecommunications Standard Institute (ETSI) specifications for WCDMA air interface suggest five actions to be taken respectively in the presence of congestion [Ericsson 2001, Dahlman 1998], that is five actions to cope with link degradation. These are:

Action 1: congestion control is activated which reduces the bit rate of non-real time applications to decrease the congestion level.

Action 2: if the 1st action is not sufficient, congestion control triggers the inter-frequency handover that moves some subscribers to less loaded frequencies.

Action 3: if the 2nd action fails, some subscribers can be handed over to a different operator.

Action 4: if the 3rd action approach fails, some subscribers will be handed over to a different system such as the Global System for Mobile Communication (GSM).

Action 5: consists of blocking subscribers of lower priority to protect the quality of the remaining ones.

Actions 1 and 5 aim, specifically, at rendering power control dependent on the QoS requirements. No specifics are given however. Furthermore, these actions can be considered as boundary conditions of our more general adaptation strategy that aims at redistributing system resources using extra information, i.e., user willingness to be degraded.

The remainder of the paper is structured as follows: We first describe call admission and handoff strategies we use, then we describe the combined SDD-based (QaPC and QaHO) mechanism. We then present the simulations carried out using the testbed of [Elbatji 2003] for performance evaluation. Results of the combined scheme are compared to a blind mechanism for congestion handling as specified in UMTS [Choukair 2003], which does not use other QoS attributes than the classification of the class into real time (RT) (i.e., classes 1 and 2) and non-real time (NRT) (i.e., classes 1 and 2) as in actions 1 and 5 described above. Finally, we present our conclusions and future works.

II. Admission Control and Handoff Strategies

User requests are processed on a FCFS basis. The decision of accepting or rejecting a request is based on the QoS profile attached to the request on the one hand and the maximum available power[2] in the system P_{max}. Various admission strategies are available:

A. *Strict admission strategy*

In this strategy, a connection *new* is accepted in the system at instant t only if $\sum P_i(t) + P_{new} \leq P_{max}$, where $P_i(t)$ is the power required by existing connection i, and P_{new} is the power required by connection *new*.

B. *NRT Overload admission strategy*

In this strategy the base/system is allowed to accept connections even if

[2] Other system resources such as spreading sequences and buffers are assumed to exist in sufficient numbers/quantities.

the total power required by all connections exceeds the available power. In this case, NRT connections will have to be delayed by the scheduler. Specifically, a connection *new* is accepted in the system at instant t if and only if both conditions hold:

- $\sum P_{i/RT}(t) \leq P_{max}$ where $P_{i/RT}(t)$ is the power required by existing real time connection i, (that is class 1 and 2 connections in the system including eventually the new connection).

- $\sum P_i(t) \leq (1+\alpha) P_{max}$, where $0<\alpha<1$ indicates the maximum overload allowed for NRT connections, and i spans across all connections including the one under admission decision.

Fig. 1 shows the New Connection (NC) Admission handling process. A queued connection can be dropped from the queue if it reaches its timeout (set to 2 unit time by default). Fig 1 also shows that negotiation of QoS requirements (currently bandwidth only) takes place in the Admission Control entity. If negotiation fails and an NRT overload strategy is chosen, the NC can be accepted if it belongs to a NRT class, and reconsidered when power allows so.

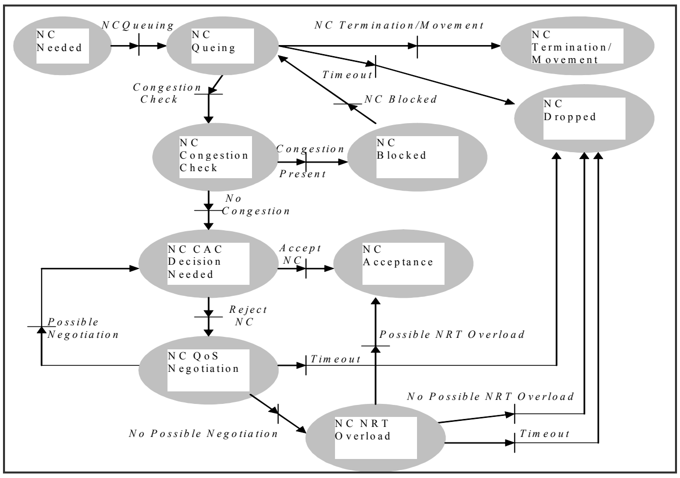

Figure 1. New connection (NC) admission handling

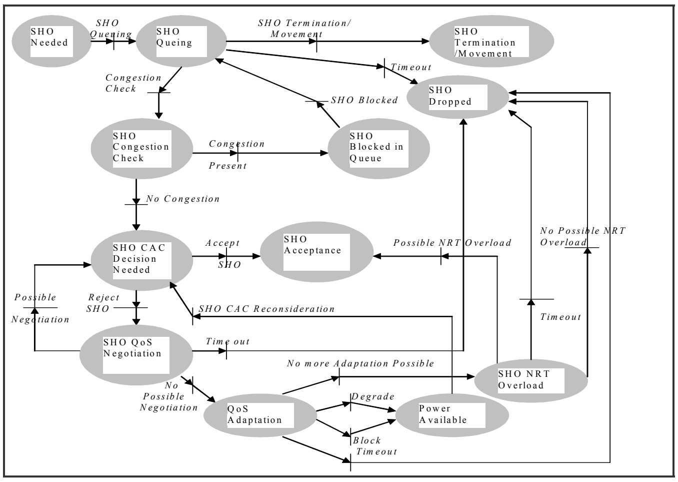

Figure 2. Handoff request handling

Fig. 2 shows the Soft Handoff (SHO) Admission handling process. A SHO undertake the same process as an NC. However, unlike an NC, there is no negotiation and QoS Adaptation (see next section) is triggered to provide the SHO with the necessary bandwidth at the expense of the existing connections.

III. COMBINED SDD-BASED QAPC QAHO PROVISIONING MECHANISM

In our approach, power is considered to be the only limiting resource. Other system resource such as spreading codes [Proakis 2000] and buffering capacity are considered to be available in sufficient quantities. The cost of a connection (i) at a given time is computed according to the following formula [Mueckenheim 2002]:

$$C_i(t) = Eb/No \ . \ 1/w \ . \ I_i(t)/H_i(t) \quad (1)$$

Energy to Noise ratio Eb/No is set by default to 18dB, but can be set to a different value to account for quality of User equipment. Intercell interference is not taken into consideration in the current version of the testbed. The Chip rate (W) is set to 3.84 Mchips.

The interference at a given time $I_i(t)$ is the sum of interferences exerted by existing users on the target user at a given time within the same cell. The central limit theorem is used to model $I_i(t)$ as a Gaussian process with zero mean and a given variance σ^2 [Rappaport 2001]. The σ^2 was initially set to 0.5. However, it can be set to otherwise to account for multi-path. The channel gain at a given time $H_i(t)$ follows a Rayleigh distribution. This is modeled using a random process in the frequency domain [Rappaport 2001].

$C_i(t)$ is the cost (power per bit) for maintaining connection i in an interference limited environement; it embodies time variations of the channel [Elbatji 2003]. The total average power required by connection i operating at bitrate R_i is therefore:

$$P_i(t) = C_i(t) * R_i, \quad (2)$$

Our new combined approach complies with WCDMA [Ericsson

2001, Dahlman 1998] and 3GPP specifications [3GPP 2002a, 3GPP 2002b]. The rational behind it is to provide a basis for: **1.** Handling channel degradation in the WCDMA radio access network by dynamically triggering a **QoS Adaptation Algorithm** that supercedes to the closed loop power control, and **2.** Providing the incoming SHO requests, which would otherwise be rejected by the Call Admission Controller (CAC) due to lack of resources, with the necessary resources by triggering the same **QoS Adaptation Algorithm**. Fig. 3 describes our combined mechanism for handling congestion and SHOs.

The QoS adaptation algorithm is at the heart of the combined method, and is triggered to make room for an incoming SHO, and in the presence of congestion (that is the total power required by existing connections is less than the available power at the base: $\Sigma_i P_i < P_{max}$). Alternatively congestion is also defined as the lack of power for real time connections (RT) namely for class 1 and class 2 connections only. That is, $\Sigma_i P_{i \text{ in RT}} < P_{max}$.

It is worth mentioning that both modes are supported in the testbed used for evaluation, and that congestion is declared after 2 unit time (ut) persistence of congestion symptoms (lack of power). This confers to the congestion handling process stability with respect to temporary short fades.

A. QoS adaptation algorithm

The adaptation algorithm resolves congestion in two phases. The two phases are applied differently in case of congestion handling and in case of SHO admission. In many ways, it is an improvement to the algorithm suggested in [Abid 2001]. In accordance with the QoS framework defined in [Lataoui 2000], each connection request by the User Equipment (UE) includes a QoS profile. The profile comprises the required bit rate R_i, the traffic class CL_i, and the Service Degradation Descriptor SDD_i. The latter takes values between 0 and 5. The larger the SDD is, the more willing is a mobile user to get degraded/dropped.

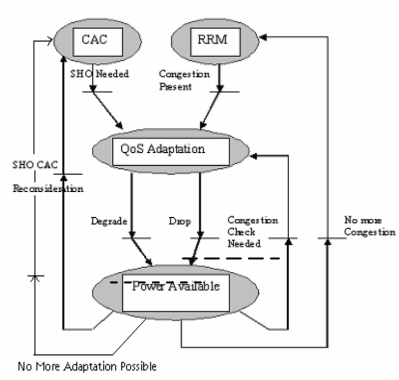

Figure 3. Processes triggered to handle congestion and SHOs are baesd on a core QoS adaptation algorithm which uses SDD QoS descriptor, as well as class of service and bitrate.

1) The Degradation Phase:

This phase is solely based on the SDD. Iteratively, the active connection that has the highest SDD is the connection that gets degraded in terms of its bandwidth requirements as follows: one such connection with 384Kbps bit rate requirement will be degraded to 144Kbps. Similarly 144Kbps is swapped for 64Kbps, and 64Kbps is swapped for 16Kbps. 2 Mbps and 16Kbps connections are not degraded in this schema.

2) The Dropping Phase:

The dropping phase is invoked only when willing connections were degraded, but congestion persists. In this phase, dropping is based on:

$$F_i(t) = SDD_i * P_i(t), \qquad (3)$$

where P_i (t) is the power requited by connection i at time t. $F_i(t)$ is high for connections requiring much power and at the same time more willing to be degraded[3]. Iterating through class 4, 3, 2, then 1, connections with high $F_i(t)$ are dropped until congestion disappears.

To confer fairness to the dropping phase, connections with similar $F_i(t)$ are considered according to their cost C_i (t) first, then according to their arrivals time.

As mentioned earlier the QoS adaptation algorithm, just described, is invoked to provide the necessary bandwidth for SHO requests that normally would not be accepted due to lack of resources/power. Two cases are distinguished depending on the class of service of the SHO request:

- NRT: if the degradation of exiting connections is not enough to collect the necessary resources/power, the SHO request is rejected.

- RT: the necessary resources/power will always be provided by degradation first then dropping.

From an implementation point of view, the QoS adaptation algorithm doesn't degrade or drop a given user connection unless it is sure that the user undergoing SHO will be accepted. Actual degradation/dropping of connections takes place atomically.

B. Numerical Results and Discussion

The evaluation of the proposed combined approach is carried out using the WCDMA compatible testbed described in [Elbatji 2003]. This testbed allows for a variety of user/connection arrival patterns with UMTS compatible QoS profiles (classes, bit rates, speed, etc.) as well as SDD, to be injected. It also allows for a variety of admission and congestion signaling strategies to be setup. Using this testbed we benchmark our proposed combined approach, against a basic combined non QoS-aware congestion handling mechanism that conforms UMTS classical dropping [Choukair 2003], and a non QoS-aware SHO prioritization mechanism. Actions in this basic combined mechanism (BA) are triggered solely by power[4] availability regardless of QoS attributes of current connections (see Fig. 4).

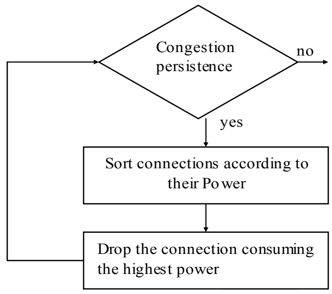

Figure 4. QoS provisioning in BA mechanism. SHO requests are treated like new connections.

Iteratively connections consuming largest amounts of power are dropped until the total power of remaining active connections becomes less than P_{max} (the maximum available transmit power to a UMTS Base). It is worth noticing that the cost (power/bit) is not used in BA either.

Specifically, we measure average dropping per class, SHO acceptance rate, and average bandwidth utilization for BA versus combined QaPC and QaHO under two load scenarios: a steady increase (A), and sudden increase (B) as in Fig. 4.

To this end, two series of experiments: series A and series B (5

[3] The connection that is much more willing to be degraded is the connection that should be first considered in the dropping phase.

[4] Other cell resources such as spreading sequences and buffers are assumed to be available in sufficient quantities.

runs exactly in each series) were carried out on a Pentium III (728 MHz) with 128MB of RAM running Windows XP using the testbed. These experiments consist in launching the testbed simulations for 300 unit time (ut), corresponding to 20 minutes real time, for each run. Then averaging all collected measurements for each series separately over the five runs.

TABLE I. MAIN EXPERIMENTAL SETTING PARAMETERS.

Physical Layer Parameters	E/N=18DB W=3.84Mcps P_{max} =35 W (for 100s of users) C_{max}= 2.5mW/bit
Traffic Parameters	CL=1, R=2Mbps, CD=60ut, V=0km/h CL=2, R=384Kbps, CD=30ut, V=60 km/h CL=2, R=144Kbps,CD=30ut,V=100 km/h CL=3, R=64Kbps, CD=4ut, V =120 km/h CL=1, R=16Kbps, CD=64ut, V=160 km/h Queuing timeout: 2 ut for all connections. SDD random
QoS adaptation triggering	2 ut congestion persistence
Degradation	384Kbps→144Kbps 144Kbps→64Kbps 64Kbps → 16Kbps. 2 Mbps and 16Kbps connections are not degraded
CAC strategy	NRT overload is 10% of the total available power
Admission negotiation	No

In each run, the testebed is loaded with 100 initial connections to bring it to an initial close to congestion state. P_{max} (see Table 1) as well as other physical layer parameters have been carefully chosen to yield congestion around 100 connections. Subsequent connections are thrown in according the following traffic models: call requests are generated for series A according to Poisson distribution with a rate of 2 connections/ut during the 300 ut. As for series B, a burst of 5connections/ ut is generated between ut 50 and 100 (see Fig. 4). The initial position in the cell of a new call, as well as its class of service CL, and its SDD are generated randomly. For each call, the bitrate R, the speed V, and the call duration (CD) are assigned according to the class. For the purpose of all simulations a 10% overload corresponding to NRT traffic is used. A summary of the traffic model as well, physical layer main parameters, and testbed admission strategy is given in Table 1.

Although Fig. 5 shows the arrival patterns for both series A and B, due to space shortage we will show only the graphs corresponding to series A, that is simulation under a steady arrival pattern. The results obtained for series B will nonetheless be given.

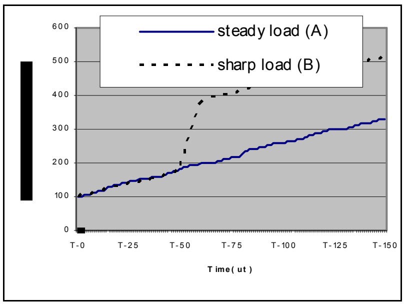

Figure 5. Connection arrival scenarios A: steady and B: unexpected sharp load at 50 ut.

Fig. 6 summarizes the average connection dropping per class for the combined QaPC and QaHO versus BA, for series A. In BA, premium traffic of class 2 is heavily penalized when congestion occurs, while with the combined QaPC and QaHO, premium traffic that is classes 1 and 2 experience less dropping, thus maintaining QoS requirements for critical traffic. Similar results are obtained under series B.

Moreover, as shown in Fig. 7, an improvement of 12% is obtained for the combined QaPC and QaHO for SHO requests. This improvement reaches 19% for series B.

As illustrated in Fig. 8, combined QaPC and QaHO gives a total 22% more resource/power utilization than BA with series A, and a staggering 25% more utilization is obtained for series B.

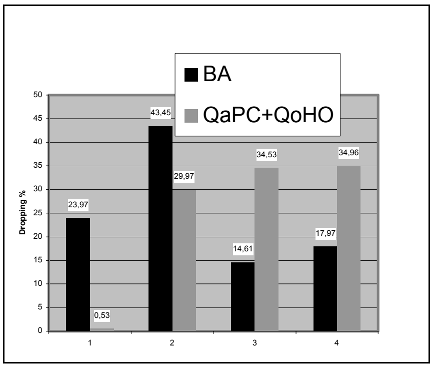

Figure 6. Percentage average dropping rate per class obtained for series A, for combined QaPC and QaHO versus BA.

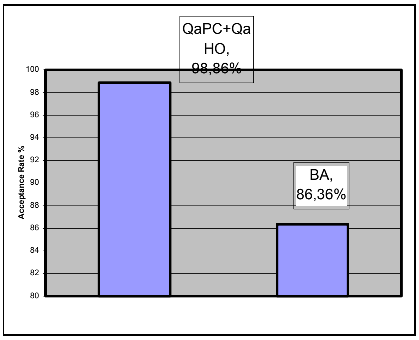

Figure 7. SHO acceptance rate for combined QaPC and QaHO vs. BA obtained with series A.

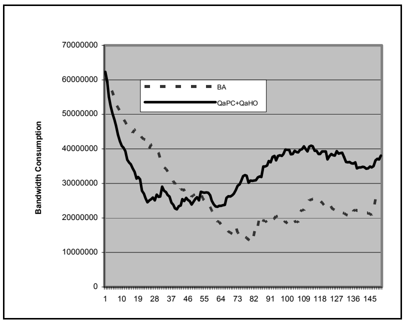

Figure 8. Resource utilization combined QaPC and QaHO versus BA across time.

The results clearly show that our combined QoS aware congestion and SHO handling is superior to the mechanism suggested in [Choukair 2003], in all respects.

IV. CONCLUSION AND FUTURE WORKS

We have presented a QoS aware mechanism for power control and Handoff in 3G WCDMA networks. We have used bitrate, service class and Service Degradation Descriptor as enabling QoS parameters.

Numerical results obtained using a WCDMA-and UMTS compatible testbed, show that our proposed QoS aware mechanism significantly improves QoS contract upholding for premium mobile users, as well as increase resource utilization, while improving SHO acceptance. Current investigations are focusing on integrating BER and queue length as extra enabling QoS parameter for our approach; as well as, evaluating this mechanism in presence of distributed admission strategies.

REFERENCES

[Abid 2001] M.R.Abid, T.Rachidi, A.Bensaid, S.Gruhl, and M.Soellner “Adaptive fuzzy call admmision controller for UMTS”, the 5th world conference on Sysmetics, Cybernetics and Informatics SCI, Vol. XVI, 2001, pp. 93-99.

[Bhatti 1998] S.Bhatti and G.Knight "Notes on a QoS information model for making adaptation decisions" 1998.

[Choukair 2003] Z. Choukair, and S. Sfar, "Run time adaptation of UMTS services to available resources", Proc. of the 17th Int. Conf. on Advanced Information Networking and Applications, IEEE 2003.

[Dahlman 1998] E.Dahlman and P.Bening "WCDMA –the radio Interface for future mobile multimedia communications", IEEE Transactions on Vehicular Technology, Vol 47, No 4, November 1998.

[Elbatji 2003] A. Y. Elbatji, T. Rachidi, and H.Bouzekri "A testbed for the evaluation of QoS provisioning in WCDMA based 3G wireless Networks", in Proceedings of the IASTED International Conference on Communication Systems and Networks, , CSN, September 8-10, 2003, Benelmadena, pp.31-36

[Ericsson 2001] Ericsson Radio System, "Basic concepts of WCDMA radio access network", 2001, ericsson.com.

[Forman 1994] G. H. Forman, and J. Zahorjan, "The challenges of mobile computing," IEEE Computer, vol. 27, no. 4, pp. 38-47, April 1994.

[Lataoui 2000] O.Lataoui, T.Rachidi, and L.G Samuel "QoS management architecture for packet switched 3rd generation mobile systems" , in proceedings of INTEROP, Mai 2000, p.365.

[Mueckenheim 2002] J. Mueckenheim, and S. Gruhl, Quality of Service Scheduling Method for UMTS Downlink, Lucent Technologies. Personal communication, 2002.

[Proakis 2000] J. G. Proakis, "Digital communications", McGraw-Hill; 4th edition, 2000.

[Rappaport 2001] T. S. Rappaport, "Wireless communications: principles and practice" Prentice Hall PTR, 2001.

[Satyanarayanan 1996] M. Satyanarayanan, "Fundamental challenges in mobile computing", in Proceedings of the Fifteenth Annual ACM Symposium on Principles of Distributed Computing, ACM, 1996.

[Soh 2003] Wee-Seng Soh and Hyong S. Kim, "QoS provisioning in cellular networks based on mobility prediction techniques", IEEE communications Magazine,Vol 41, No. 1., January 2003, pp. 86-92.

[Stuber 2001] G. L. Stuber, "Principles of mobile communications", Kluwer Academic Publishers; 2nd edition, 2001.

[Wong 2003] William K. Wong, Haiying Zhu, and Victor C. M. Leung "Soft QoS provisioning using the token bank fair queuing scheduling algorithm", IEEE Wireless Communications, Vol. 10, No.3, June 2003, pp.8-16.

[Zheng 1997] Q. Zheng, T. Yokatani, and T. Ichihashi "Connection admission control for hard real-time communication in ATM networks", Proceedings of the 17th International Conference on Distributed Computing System, IEEE 1997.

[3GPP 2002a] 3GPP "QoS concept and architecture", http://www.3gpp.org, ETSI 23.107 v5.5.0 (2002-2006).

[3GPP 2002b] 3GPP "UTRAN overall description", http://www.3gpp.org, ETSI 25.401 v5.4.0 (2002-2009).