

An Integrated System for QoS Provisioning in 3G WCDMA Cellular Networks

T. Rachidi, M. Sebbane, A. Y. Elbatji, H. Bouzekri
Alakhawayn University in Ifrane,
Ifrane 53000,
Morocco
{T.Rachidi, M.Sebbane, A.Elbatji, H.Bouzekri}@aui.ma

Contact details of the first author:

T. Rachidi
Alakhawayn University in Ifrane,
PO BOX 1881,
Ifrane 53000, Morocco

Tel.:+212 55 86 2402
Fax.:+212 55 86 2424
Email:T.Rachidi@aui.ma

Abstract

This paper proposes an integrated Power-based system for QoS provisioning in 3G WCDMA cellular networks. The proposed system comprises a QoS-profile aware Call Admission Control (CAC), a QoS-profile driven congestion control, and a reservation-less real cell boundary aware handoff prioritization mechanism. Both Congestion control and handoff prioritization incorporate user service degradation descriptor (SDD) as a key enabling QoS parameters, in addition to the class of traffic and the required bandwidth. The former supercedes to the traditional closed loop power control while the latter implements a QoS-differentiated prioritization of Soft handoff without reservation.

A cluster of 7 cells underlying a road junction controlled by one Radio Network Controller (RNC) has been used to evaluate the proposed integrated system as a function of the offered load. Both cluster-wide forced terminations and network utilization have been measured, and the results are reported for different QoS Adaptation strategies. These measurements are contrasted to a non QoS-aware system, that is, a system where blind closed-loop power control and prioritization mechanisms are used. The results show that a fair dropping strategy (that is dropping solely based on cost -power/bit- of connection) to resolve congestion offers the best compromise between low dropping rate and maximum utilization, while prioritizing handoffs.

Keywords

QoS provisioning, 3G Wireless Networks, Handoff prioritization, QoS-profile driven Power Control.

An Integrated System for QoS Provisioning in 3G WCDMA Cellular Networks

T. Rachidi, M. Sebbane, A. Y. Elbatji, and H. Bouzekri
Alakhawayn University in Ifrane,
PO BOX 1881,
Ifrane 53000,
Morocco

Abstract—This paper proposes an integrated Power-based system for QoS provisioning in 3G WCDMA cellular networks. The proposed system comprises a QoS-profile aware Call Admission Control (CAC), a QoS-profile driven congestion control, and a reservation-less real cell boundary aware handoff prioritization mechanism. Both Congestion control and handoff prioritization incorporate user service degradation descriptor (SDD) as a key enabling QoS parameters, in addition to the class of traffic and the required bandwidth. The former supercedes to the traditional closed loop power control while the latter implements a QoS-differentiated prioritization of Soft handoff without reservation.

A cluster of 7 cells underlying a road junction controlled by one Radio Network Controller (RNC) has been used to evaluate the proposed integrated system as a function of the offered load. Both cluster-wide forced terminations and network utilization have been measured, and the results are reported for different QoS Adaptation strategies. These measurements are contrasted to a non QoS-aware system, that is, a system where blind closed-loop power control and prioritization mechanisms are used. The results show that a fair dropping strategy (that is dropping solely based on cost -power/bit- of connection) to resolve congestion offers the best compromise between low dropping rate and maximum utilization, while prioritizing handoffs.

Keywords—QoS provisioning in 3G Wireless Networks, Handoff prioritization, QoS-profile driven Power Control.

I. INTRODUCTION

The rationale behind the 3G Telecommunication System 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, for instance, supports QoS provisioning through four (4) basic classes of service (CL₁-CL₂) [1]. Each of these classes imposes different QoS requirement on the 3G network which must be maintained during the lifetime of the corresponding connections.

Provisioning QoS over WCDMA-based air interface cannot be fulfilled solely by efficient Scheduling [2]. This is due, on the one hand, to the inherent characteristics of the wireless link [3], that is, user mobility and fading channel (time variations) [4], high error rates, inherent interference limited characteristics of WCDMA [5], and low and varying bandwidth (2Mbps at most); and on the other hand due to the unexpected Soft Handoffs (SHOs) requests resulting from user mobility. Both phenomena, if not catered for, yield inevitably an over load on the wireless system causing increased forced terminations and/or QoS degradation.

While there are works addressing the inherent characteristics of the link [6,7], and except for our own initial works [8], none of them addressed degradation due to

inherent characteristics of the link in an integrated approach, where radio and interference constraints and QoS requirement are reconciled. That is, the inherent characteristics of the link have been addressed in a blind manner relying on the closed power loop mechanism ignoring QoS attributes and of underlying connections.

Unexpected SHOs, on the other hand, have been tackled using prioritization or reservation with/without prediction techniques. Mobility prediction techniques, in particular, use UE velocity estimation [9], to provide optimized resource reservation yielding efficient QoS provisioning with reduced forced termination. These techniques, however, either assume regular geometry of cells, or do not link the geometry to power (the fundamental resource in WCDMA) available to the cell/system. Of course cell boundary changes with the interference conditions as in Fig. 1, and must be accounted for in SHO in particular.

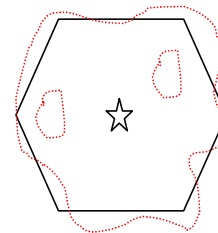


Fig. 1. Real handoff boundaries (shown in dotted lines) are determined by Signal to Interference and Noise Ratio (SINR), with possible handoffs occurring at the centre of the hypothetical cell.

Our research seeks to develop an integrated system where WCDMA wireless link characteristics, user mobility and user QoS requirements are reconciled in order to maximize wireless resource utilization, maintain QoS contracts and minimize forced terminations. In particular, our research aims at demonstrating that combining QoS profile driven power control, real boundary-aware SHO handling and QoS-aware Call Admission Control, in an integrated system across a cluster of cells can yield better performance, when user QoS profile is augmented with user willingness to have their QoS profile degraded. To this end, a patented Service Degradation Descriptor (SDD) representing user willingness to be degraded [10] is used as an enabling QoS attribute for driving the integrated system in general and the power control in particular.

SDD is a real number between 0 and 5. it describes how much the user is willing to get a degraded quality of service. The larger the SDD is, the more willing is the user to accept a degraded service and less the user is charged. A typical video telephony service can be degraded/adapted to current network conditions by using color/grey scale or by lowering the

quality of the coded speech. Connection request that do not set a value for SDD, will have this value automatically set to a default service value by the network provider.

In WCDMA based 3G networks, Base stations implement closed loop power control at the level of the Radio Resource Manager (RRM) on which we are trying to improve by taking into account QoS profile as an extra parameter for power control. In effect, by proposing our QoS profile driven power control (QPC), we are super-ceding to the basic closed loop control in RRM. Indeed, a traditional RRM only takes into account channel gain when making a decision to control base transmit power. In implementing power control, we, however, take into account not only channel gain, but also the negotiated¹ QoS requirements of existing users, namely the bit rate and the SDD representing willingness to be degraded.

When, due to the problems cited above, congestion occurs at a cell, QPC is triggered to address the overload in power requirements. It is this entity that copes with the link degradation in WCDMA based 3G wireless networks, by using QoS profiles of the active mobile users.

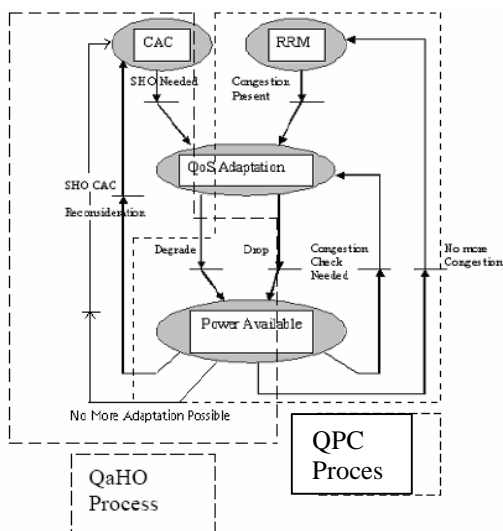


Fig. 2. Processes triggered to handle congestion and SHOs are based on a core QPC which SDD QoS descriptor, as well as class of service and bit rate.

Furthermore, when examining SHO requests QPC is also invoked to provide the necessary bandwidth for SHO requests that normally would not be accepted due to lack of resources/power. However, in doing this the class of service of the connection making the request is optionally used to provide a QoS differentiated prioritization: RT connections will always be accepted, if necessary at the expense of other connections, while NRT connections will be accepted only if enough capacity is available at the NodeB.

In effect, this is an unfair strategy which prioritizes RT handoffs by degrading the QoS of active users (to levels

agreed upon during admission negotiation) to provide the newly coming Handoff with the necessary resources/power.

On the other hand, new connection requests are queued at the Call Admission controller, where the decision takes place based on the available power at the base and the requested bandwidth.

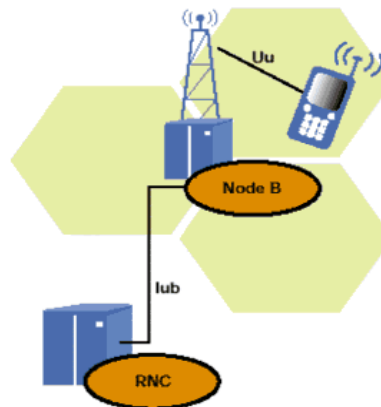


Fig. 3. RNC Architecture. One RNC controls many NodeBs.

The reminder of this paper is structured as follow: Section II describes the integrated system with its components, namely the QoS-profile driven power control, QoS- profile aware SHO prioritization and the power-based call admission controller. Section III presents the experimental setting as well as the simulations carried out for performance evaluation. Finally, Section IV concludes the paper and discusses directions toward future work.

II. THE INTEGRATED SYSTEM

A. System Generals

In our system, power is considered to be the only limiting resource. Other system resources such spreading codes [5] and buffering capacity are considered to be available in sufficient quantities.

The power required to provide the bit rate R_i for a connection (i) at a given time t is computed according to the following formula [11]:

$$P_i(t) = C_i(t) \cdot R_i(t) \quad (1)$$

Where $C_i(t)$ is the current cost of the connection, and is given by:

$$C_i(t) = \frac{Eb}{No} \cdot \frac{1}{W} \cdot \frac{I_i(t)}{H_i(t)} \quad (2)$$

Each connection request (i) 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.

¹ Our system optionally supports QoS profile negotiation at call setup, as well as non-real time (NRT) traffic overload.

Each active connection (i) experiences a continuous update of its cost $C_i(t)$ and $P_i(t)$ given that its channel gain $H_i(t)$, interference $I_i(t)$ and position $X_i(t)$ change over time.

The chip rate (W) is set to 3.84 Mchips, and Energy to Noise Ratio E_b/N_0 is set by default to 18 dB. It can also be set to a different value to account for quality of User Equipment. Currently, we do not take into account co-channel interference. The latter has a direct effect on Bit Error Rate (BER), which is not currently included in the QoS profile.

The interference at a given time $I_i(t)$ is the sum of interference exerted by all the existing users of the cluster on the target user. The central limit theorem is used to model $I_i(t)$ as a Gaussian process with zero mean and a given variance s^2 [5]. The s^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.

For simplicity reasons, each user is allowed to request one connection at a time. The scheduler (such as a Time-shift scheduler) is assumed to operate close to optimum in meeting all delay requirements.

The dwelling time of a connection is set according to the QoS profile of the connection, namely its bandwidth and class:

TABLE 1
DWELLING TIME SETTINGS

Bandwidth	Dwelling time
2Mbps	60 min ² (1h video on demand stream)
384Kbps	30 min (videophone call)
144Kbps	30 min
64, 16Kbps	4 min (typical voice call)

UE position $X_i(t)$ changes over time according to the speed assigned to the corresponding connection. The speed of each connection is set according to its bandwidth. Users are assumed to move with a constant velocity $V_i(t)$.

TABLE 2
SPEED ASSIGNMENT

Bandwidth	Speed
2Mbps	0 Km/h
384Kbps	60 Km/h
144Kbps	100 Km/h
64Kbps	120 Km/h
16Kbps	160 Km/h

The Active Set $AS_i(t)$ of every existing connection also changes over time. This Set holds the cells of the cluster that can serve the user at any time. The Active Set of a user (i) changes according to the Signal to Interference and Noise Ratio (SINR_i). This Ratio (in dBm) is continuously monitored by each cell for all active connections of that cell:

$$SINR_i(t) = \text{Log } P_i(t) - 80 - S_j \text{Log } P_j(t) \quad (3)$$

where $P_i(t)$ is the required power (in mW) for the connection i and $P_j(t)$ is the the total power consumed by an active connection (j ?i) in the system/cluster.

In case the connection (i) has a low SINR (less than a threshold value), the NodeB will look in the Active Set of the connection $AS_i(t)$ to find the cell with the highest SINR for that connection and send the SHO request to that cell.

Congestion is triggered after a congestion indication persists for a predefined duration (set to two unit time by default). Congestion is indicated at instant t if $SP_i(t) > P_{max}$ where $P_i(t)$ is the power required by connection i, and P_{max} is the maximum power available at the base/system. When NRT overload is switched on, congestion indication is set when $S_{i \text{ is RT}} P_i(t) > P_{max}$, that is NRT traffic is not accounted for. It is worth mentioning that both mode are supported in the simulation model used for the 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.

B. QoS-profile driven power control

The rational behind our approach is to provide a basis for 1. Handling the channel degradation in the WCDMA radio access network by dynamically triggering a QoS Adaptation Algorithm, that supercedes to the power control loop and 2. Providing the incoming SHO requests which would be rejected by the Call Admission Controller (CAC) due to lack of resources, with the necessary resources by triggering the same QoS Power Control (QPC).

QPC is at the heart of our integrated system, and is triggered to make room for an incoming SHO, and in the presence of congestion. QPC resolves congestion in two phases. The two phases are applied differently in case of congestion and in case of SHO admission. In many ways,, it is an improvement of the algorithm presented in [8]. In accordance with the QoS framework defined in [10].

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 term of its bandwidth requirements as follow:

TABLE 3
DEGRADATION PHASE

Initial Bandwidth	After Degradation
384Kbps	144Kbps
144Kbps	64Kbps
64Kbps	16Kbps

2Mbps and 16Kbps connections are not degraded in this schema.

² All time units are converted to canonical unit time (ut) for simulation purposes.

The Dropping Phase: this phase is invoked only after all willing connections were degraded, but congestion still persists. In this phase dropping is based on:

$$F_i(t) = SDD_i \cdot C_i(t) \quad (4)$$

$F_i(t)$ is high for connections requiring much cost and at the same time more willing to be degraded. Connections with high $F_i(t)$ are dropped until congestion disappears.

C. Reservation-less SHO handling

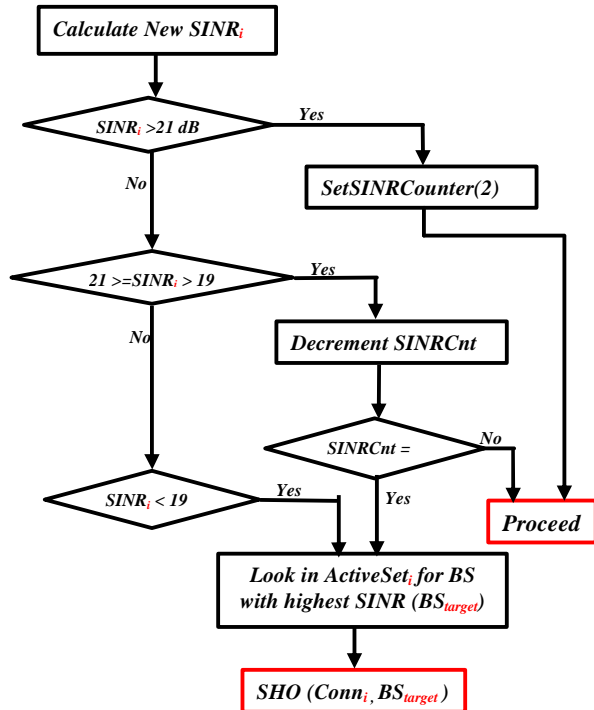


Fig. 4. SINR-driven SHO process.

In our integrated system every NodeB monitors the SINR of all its active connections continuously. This value should be kept above a threshold value ($SINR_{Target}$) to guarantee a good quality (the $SINR_{Target}$ is set to the default value 21dB). If the $SINR_i$ of the connection i is higher than 21, the $UnderSINRcounter_i$ is set to 2.

If $SINR_i$ is less than $SINR_{Target}$ but still higher than the $SINR_{Threshold}$ (the $SINR_{Threshold}$ is set to 19 dB), the $UnderSINRcounter$ for that connection is decremented (it is by default initialized to 2).

If $SINR_i$ is less than the $SINR_{Threshold}$, or the $UnderSINRcounter$ equals zero, the NodeB looks in the *Active Set* of that connection for the $NodeB_{Target}$ with the highest SINR value for that connection. Once $NodeB_{Target}$ is found, a SHO request is sent to the concerned NodeB. Fig. 4 illustrates the process.

As mentioned earlier, QPC is invoked to provide the necessary bandwidth for SHO requests that normally would not be accepted due to lack of resource/power. Two strategies are distinguished in case degradation with QPC does not yield the necessary resources:

In **Unfair_DQAA** strategy [8], in case of lacking resources after degradation, the SHO request is processed depending on its class of service: **1. NRT**: The SHO request is rejected. **2. RT**: the necessary resources/power will always be available by degrading first then dropping. dropping is based on power ($F_i(t) = SDD \cdot P_i(t)$). Starting with Class 4, 3, 2 then 1, connections with high $F_i(t)$ are dropped until congestion disappears.

In **Fair_DQAA** strategy, however, dropping is based on cost ($F_i(t) = SDD \cdot C_i(t)$). Regardless of the classes, connections with high $F_i(t)$ are dropped until congestion disappears.

D. Call Admission Control

The role of admission control is to decide if a new connection can be accepted: if yes, then the problem becomes one of using which channel and what power level. To achieve higher spectrum utilization, the power control algorithm used (QPC) should converge to the optimal usage of resources, while the QoS-aware prioritization of handoffs (QaHO) should leverage users QoS profiles. Our proposed CAC makes its decision to whether to accept or reject the new user's request based solely on its QoS profile and the current available power in the target cell. Fig. 5 describes the new connection request process.

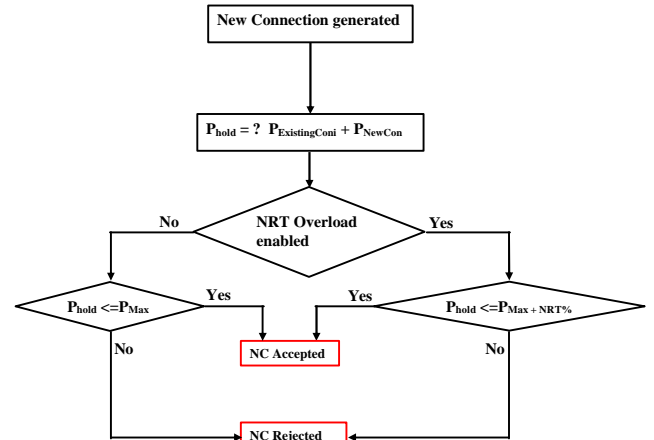


Fig. 5. Connection admission process, with support for NRT traffic overload.

III. SIMULATION PARAMETERS AND NUMERICAL RESULTS

To evaluate our integrated System, we augmented the testbed of [13] with distributed capabilities and used it to analyze the different QoS adaptation strategies, namely the Unfair_DQAA, the Fair_DQAA, and compared them to a Basic Algorithm (BA), that is a blind non-QoS aware strategy

with no SHO prioritization. To resolve congestion, BA uses a one-phase process where connections with highest power are dropped first, while both Unfair_DQAA and Fair_DQAA undergo a two-phase process with a degradation phase and dropping phase. They both undertake the same degradation phase, but a different dropping phase as explained earlier.

During the simulation, connections are generated according to a user arrival pattern (UAP) with random QoS profiles. Each profile comprises the required bit rate (2Mbps, 384Kbps, 144Kbps, 64Kbps, and 16Kbps), the traffic class (1, 2, 3, or 4) and the Service Degradation Descriptor (SDD) (0-5). Traffic parameters such as speed and dwelling time are set according to the class and bandwidth of the connection (see table 4).

In our simulation, we assume a cluster of 7 NodeBs underlying a road junction (see Fig. 6). Every NodeB is initiated with an N^3 number of connections to bring the system to a close to congestion state ($N = \{10, 20, 30, 40, 50, 60, 70, 80 \text{ or } 90\}$). Then connections are thrown in according to the UAP. The available power at each NodeB (P_{max}) is set to a value suitable to the load (see Fig. 7). The physical layer parameters of each NodeB are given in table 4.

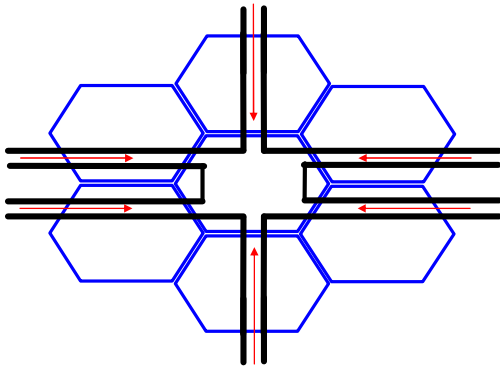


Fig. 6. A road junction covered by a cluster of 7 NodeBs which are controlled by one RNC is used for simulation.

TABLE 4
KEY PARAMETERS SETTING

Physical Layer Parameters	E/N=18dB W=3.84Mcps SINR _{threshold} =19dB SINR _{target} =21dB P_{max} as in Fig 7. CT = IUL dependant
Traffic Parameters	CL=1, R 2Mbps, CD=60ut, V= 0Km/h CL=2, R 384Kbps, CD=30ut, V= 60Km/h CL=2, R 144Kbps, CD=30ut, V= 100Km/h CL=3, R 64Kbps, CD=4ut, V= 120Km/h CL=4, R 16Kbps, CD=64ut, V= 160Km/h Queuing timeout 2ut for all connections SDD random
QoS Adaptation strategy	<ul style="list-style-type: none"> • Fair_QDAA • Unfair_DQAA • Basic Algorithm (BA) • 2ut congestion persistence
Degradation	384Kbps → 144Kbps 144Kbps → 64Kbps 64Kbps → 16Kbps

³ Initial Users' Load (IUL)

	2Mbps and 16Kbps connections are not degraded
CAC strategy	<ul style="list-style-type: none"> • NRT overload (optional) • Bandwidth Negotiation (optional)

The performance measures of interest for us are the Normalized average Dropped Connections and power utilization for every NodeB, as well as for the whole cluster, under varying loads and QoS Adaptations algorithms.

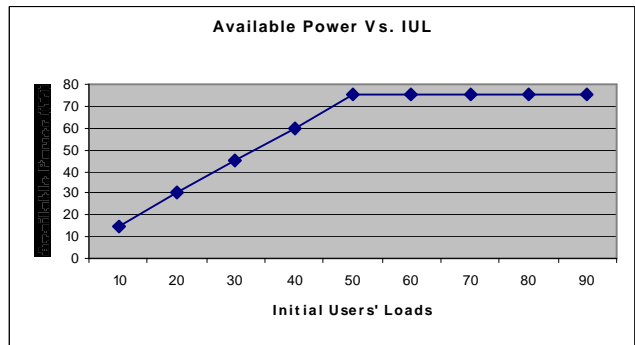


Fig. 7. Initial cell power is adapted to the load to guarantee occurrence of congestion.

For every Adaptation algorithm, and for every load, 5 simulations have been carried out and the results averaged for the whole cluster. Fig. 8 and Fig. 9 illustrate RT and NRT normalized average dropping rate obtained for every algorithm with different normalized user load.

As expected, unfair DQAA proves its affinity towards RT connections, while the fair DQAA performs better than BA for RT traffic (Fig. 8), but surprisingly BA outperforms Fair DQAA strategy for NRT traffic (Fig. 9).

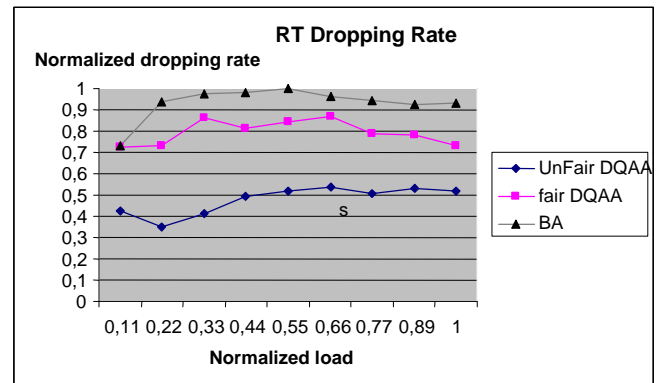


Fig. 8. Normalized RT connections dropping rates as a function of normalized load.

In Fig. 10, however, we can see that the Fair_DQAA achieves, as predicted, the lowest dropping rate (20% less than Unfair DQAA) when considering all classes of services.

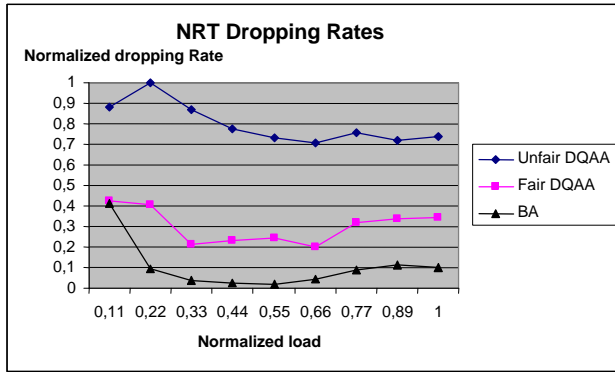


Fig. 9. Normalized NRT connections dropping rates as a function of normalized load.

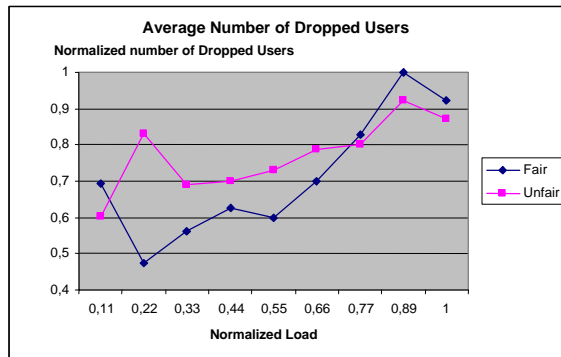


Fig. 10. Normalized connection dropping rates (all classes alike) as a function of normalized load.

Fig. 11 presents the graph of the normalized bandwidth utilization comparing the Fair and Unfair DQAA. As expected, the Unfair DQAA accomplishes (7%) better utilization than Fair DQAA, mainly due to dropping with class 4 and 3 connections.

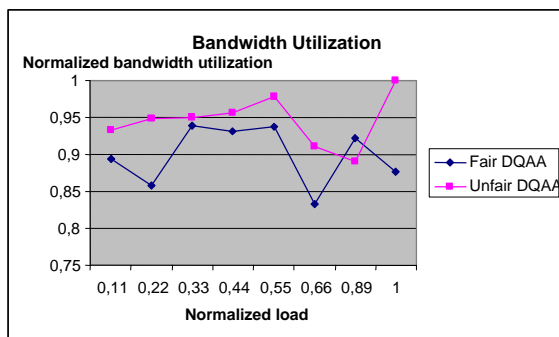


Fig. 11. Normalized power utilization as a function of normalized load.

Considering all experimental results and keeping in mind all the underlying assumptions, the Fair_DQAA proves its strength for minimizing the dropping probability of WCDMA networks while yielding a good utilization compared to BA. It is true that the Unfair_DQAA achieves better power utilization, but this is mainly because of its unfairness toward low bandwidth connections when dealing with congestion.

IV. CONCLUSIONS AND FUTURE WORK

We presented in this paper an integrated Power-based model that uses real cell boundaries for QoS provisioning in 3G WCDMA cellular networks. We have specifically emphasized on the use of the QoS-aware components to accomplish effective decisions, and efficient congestion and SHO handling, while achieving good power/resource utilization. The experiments demonstrate that a fair congestion control mechanism that is based solely on the cost (power/bit) of maintaining a particular connection offers the best tradeoff between utilization and QoS provisioning (dropping rate).

Our current efforts are directed towards the use of road information and mobility prediction techniques to investigate whether potential performance improvements. We are also incorporating the Bit Error Rate (BER) parameter into the system as another key element in the QoS profile for optimal resource utilization.

REFERENCES

- [1] 3GPP "QoS Concept and Architecture", <http://www.3gpp.org> ETSI 23.107 v5.5.0 (2002-2006).
- [2] Ericsson Radio System, "Basic concepts of WCDMA radio access network", 2001. <http://www.ericsson.com>
- [3] E. Dahlman and P. Bening "WCDMA –the radio interface for future mobile multimedia communications", IEEE Trans. on Vehicular Technology, Vol 47, No 4, Nov 1998.
- [4] G. H. Forman, and J. Zahorjan, "The challenges of mobile computing," IEEE Computer, vol. 27, no. 4, pp. 38-47, April 1994.
- [5] G. L. Stuber, "Principles of mobile communications", Kluwer Academic Publishers; 2nd edition, 2001.
- [6] M.Xiao, and N.Shroff "Distributed admission control for power-controlled cellular wireless systems" IEEE/ACM Transactions on Networking, Vol. 9, NO.6, December 2001.
- [7] Z. Choukri and S.Sfar, "Run Time Adaptation of UMTS Services to Available Resources", Proc. 17th IEEE Int. Conf. on Advanced Information Net. and Applications, 2003.
- [8] T. Rachidi, A. Y. El Batji, M. Sebbane, and H. Bouzekri, "QoS-aware Power Control and Handoff prioritization in 3G WCDMA Networks", in Proc. of IEEE WCNC04, Atlanta March 21-25. 2004.
- [9] Wee-Seng and Hyong S. Kim, "QoS Provisioning in Cellular Networks Based on Mobility Prediction Techniques", IEEE Comm. Magazine, Jan, Vol. 41, N^o1, pp.86-92. 2003
- [10] O. Lataoui, T. Rachidi, L. G. Samuel, S. Gruhl, and Ran Hong Yan, "A QoS management architecture for packet switched 3rd generation mobile systems", in proceedings of INTEROP00, May 17, Las Vegas, p. 365. 2000
- [11] J. Mueckenheim, and S. Gruhl "Quality of service scheduling method for UMTS downlink", Lucent Technologies. Personal communication, 2000.
- [12] "Radio Resource Control," UMTS.v100.50.120.601. Ed. Urs Bernhard. 2000.
- [13] A. Y. ElBatji, T. Rachidi, and H. Bouzekri, "A Testbed for the Evaluation of QoS Provisioning in WCDMA based 3G Wireless Networks", in Proc. of the Int. Conf. on Communication Systems and Networks, IASTED, CSN, Sept 8-10, Benelmadena, pp.31-36, 2003.