A Model for Threshold Comparison Call Admission Control in Third Generation Cellular Systems

Leonardo Badia, Michele Zorzi and Alessandro Gazzini
Dipartimento di Ingegneria, Università di Ferrara
via Saragat, 1 – Ferrara, Italy – {lbadia,mzorzi,agazzini}@ing.unife.it

Abstract—**In this paper, we present a study of Admission Control in 3G systems. In particular, the behavior of algorithms already presented in the literature is analyzed, with respect to their implementation in UMTS-like systems, and a model of trade-off between the QoS metrics, blocking and dropping probability, is presented. The obtained performance is discussed and analyzed under different points of view. Finally, possibilities to improve fairness and generality of these results open up when a more detalied model for mobility, data rate and discontinuous transmission (DTX) is considered.**

I. Introduction

Third Generation Mobile Communication Systems are based on Code Division Multiple Access (CDMA) multiplexing and access. CDMA systems allow an improvement of system capacity with respect to Frequency or Time Division Multiple Access systems, by avoiding the problems of channel allocation, being theoretically possible to manage all users with the same channel. This phenomenon is called *soft capacity* of a CDMA system, as opposed to the *hard capacity* of FDMA and TDMA systems, in which the maximum number of users is fixed by the amount of physical resources.

On the other hand, code-based multiplexing is limited by Quality of Service (QoS) requirements, essentially depending on the power levels between each user and the base station (BS) the mobile user is connected to. In order to have an acceptable power level at the receiver for each connection, a power control (PC) mechanism is implemented in CDMA systems, with the goal to appropriately tune the transmitted powers.

Even with PC, the system performance is still limited: the admission of new users to the system is possible, only at the price of a performance degradation for all active users. It is possible that a set of users with their QoS objectives does not admit a feasible solution, and if this is the case the PC algorithm diverges. Thus, in this case, a previously admitted user must be dropped in order to guarantee the requested service to other active users. Of course, from the users' point-of-view, cutting off an existing call is an event that should be avoided: typically the dropping probability is also used as a global QoS index.

Henceforth, the access of users to the system must be controlled, or congestion may arise and cause call dropping. That is, a check should be done to determine whether the system is near congestion, and if this is the case the new call requests should be refused. Of course this control must not be too conservative, because blocking a new call is still an undesirable event (though less annoying than dropping an existing call).

Call admission control (CAC) can be performed by following iterative real-time procedures [2] [5] [6], or by heuristic algorithms, that operate a threshold comparison [1] [3] [4] [7] [8] [9] [10].

Following [1] and [5], we can speak of *Feasibility-based CAC* (*F-CAC*) for the first class of algorithms, and divide the second class into *Number-* or *Interference-based CAC* (*N-CAC* and *I-CAC* respectively). This further division is based on the heuristic used (number of users vs. some measure of power).

The first approach leads to exact solutions, that however require long evaluation times, mainly because of the duration of the *testing phase* and other problems related to computational complexity. Moreover, algorithms of this kind present good accuracy, but also problems under the aspect of *fairness*, i.e., these algorithms act well if used to maximize the performance from the server's side, but their adaptation to requests of users' satisfaction is more difficult. For these reasons, no clear decision has been made so far as to how to select and implement these schemes in practical systems.

The second approach is more interesting from the point-of-view of a real implementation: in these "instantaneous" algorithms the evaluation time is much shorter, because the capacity of the system is approximated by using a heuristic based on measuring a quantity, that is simpler to obtain than the general description of the system, while of course being less accurate.

Lower computational complexity and better performance lead us to emphasize the role of CAC algorithms based on heuristics, especially for the case of power measurements (I-CAC). Our goal is to analyze the performance of this class of algorithms, in terms of blocking and dropping probability, and to discuss parameter optimization.

We present in this work a simple model that allows a description of admission control algorithms based on different metrics, and algorithms already presented in the literature are framed in this model. We will emphasize the uplink analysis, even though the results are still valid when heuristics for the downlink are considered. Moreover, we compare the performance of different algorithms by examining the trade-off between the two metrics and we adapt the heuristics to real cases of interest, in particular to third generation cellular networks. Finally, we propose an analysis of aspects that can significantly affect the performance, by showing that the global mean values are not always representative of the behavior of the network.

This work is organized as follows: in Section II we present a simple model that allows a description of admission control algorithms based on different metrics, and in Section III the algorithms already presented in the literature are framed in the model. In Section IV we compare the performance of different algorithms and discuss parameter optimization and finally in Section V we introduce a trade-off model for the evaluation metrics that can be useful to meet the QoS requirements. Sec-

This work has been supported by Ericsson Research

tion VI concludes the paper.

II. Threshold Model for Heuristic Algorithms

In order to describe in an intuitive way the instantaneous call admission control algorithms, we present an abstract model, in which we assume the availability of a quantity I, that is a measure of the load of the system, as a basis for CAC heuristic algorithms. In the analyzed systems, a value I_{th} can be defined: this is a particular value of I that acts as a threshold between the normal behavior of the system and an overload condition. The aim of a call admission control algorithm is to keep I always below the threshold, and a simple way to operate this control is as follows. When a new user requests a call, the algorithm estimates the increase ΔI that the new user would cause to the current value of I. The admission control is operated by simply comparing $I + \Delta I$ with I_{th}. If

$$I + \Delta I < I_{th} \tag{1}$$

the new call is accepted, otherwise it is rejected. The index I can be alternatively replaced with a set of m parameters $\mathbf{i}$, in order to check different aspects of the system. This is equivalent to estimating the condition (1) in $\mathbb{R}^m$, where the sign "$<$" must be intended as a component-wise comparison, i.e., if $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ we have that $\mathbf{a} < \mathbf{b}$ if each entry of $\mathbf{a}$ is smaller than the corresponding entry in $\mathbf{b}$.

This description contains parameters that should be specified. To identify an algorithm, three choices must be made: the kind of I-parameters, the value of the threshold I_{th} and how to estimate the increase ΔI.

The identification of the index I is a basic selection, that as a matter of fact is almost equivalent to the choice of a specific class of algorithms. I must well represent the congestion level of the system, so a lower value of the index means a situation of easier access for the new call. For example, a trivial choice for the value I is the number of admitted users. In this case, the system is treated equivalently to a TDMA-FDMA system, in which the number of available channels for the calls is fixed, and a simple CAC rule can be implemented: if the system has at least one free channel (frequency or time slot) a new call can be admitted. This approach is known in the literature as *hard capacity* [1]: since the capacity in CDMA systems is soft (i.e., there is no obvious relationship between the number of users and the possibility of admitting one more), we can expect that the hard capacity approach is too conservative, since it must be based on worst-case considerations. In the literature, the role of the index I is often played by the power (e.g., received at the BS), or similar metrics of load.

The evaluation of the threshold is a related problem, that could heavily affect the performance of an algorithm if not properly done. The value for I_{th} can be chosen by an empirical approach, i.e., with test simulations. Moreover, external considerations related to the required QoS play an important role. For example a trade-off for the QoS is present in many cases: this is a general property, by which a more conservative choice of I_{th} is suitable in order to avoid undesired congestion in the system (that may lead to calls dropping), whereas a higher value of the threshold allows more users to be served at the same time. In the following part, we discuss the value of I_{th} by comparing different values of this parameter, only to show the sensitivity of the performance to the threshold.

The last choice that must be made in the algorithm specification, i.e., how to estimate ΔI, is the most important: in fact it represents a design question on which the performance highly depends, and it is at the same time a run-time decision whose actuation could be critical for the system. While the evaluation of the quantity I is a measurement, that can be almost always performed exactly without overloading the computational resource of the system, estimating ΔI implies a necessary approximation: this effect adds itself to the dispersion due to mobility, and may lead to errors in estimating the admission feasibility. Thus the admission control errors, i.e., admission of a call that can not be managed or rejection of a call that the system could have accepted, are mainly due to the imperfect estimation of the load increase of the system. On the other side, it must be observed that in the interactive algorithms it is possible to have a perfect evaluation of ΔI, by using a *negotiation phase* over a control channel: as mentioned previously, the basic problem of this approach is the time required for the evaluation. Moreover, it may happen that after the negotiation phase, the system dynamics have completely changed the state of the network and a re-evaluation is necessary.

This observation can be related even to the algorithms that use heuristics to decide whether or not the new call is to be admitted. In fact the use of a single parameter I (or a limited set of parameters) significantly reduces the time for the evaluation of the load increase. On the other hand, the more complicated the heuristic, the longer the computational time. This is heavier if the value of I is related to quantities of the network that require a long time to be correctly estimated. Then, in choosing a specific I, one must make sure that the estimation of ΔI requires a sufficiently short time to be performed, depending on the network requirement. Therefore it might happen that an algorithm with a conceptually good heuristic to describe the traffic load exhibits low performance due to the unavoidable approximations in the estimation of the load increase.

III. Algorithms' Identification with the Model

This Section is devoted to revisiting already existing algorithms within the proposed model, and briefly discusses how the above mentioned problems are managed. As shown in Section II, a very intuitive way to translate the relation (1) in a realistic admission control environment is to identify I with the number of users, so that ΔI is simply equal to 1. This algorithm, called in the following *Hard Capacity Call Admission Control* (HC-CAC) algorithm [1], avoids the problems related to the long time needed to estimate the load increase. Another pleasing aspect of this approach is that the additional BS control software is very easy to implement, since almost no overhead for hardware and software is introduced. The weak point of this algorithm is essentially that it does not take into account that different users produce different load increases. Moreover, the choice of the threshold I_{th} remains critical: it is dangerous to select a high threshold for the system, because the real weight is not equal for all users. Thus, the selection of I_{th} should be done in a conservative way, thereby leading to a waste of capacity.

To find an alternative to HCCAC, we can refer to a scheme presented in [1] and [5]: in these works, different threshold comparison algorithms are grouped in two basic types, N-CAC and I-CAC, as defined in Section I. The first type is equal to the already mentioned HCCAC, whereas the second type uses a more convincing specification of the I parameter, i.e., it refers to an interference heuristic. A metric that represents the interference present in the system could be the total power measurement at the base station, that is a well known measure of the global load of the system [4] [9].

In [3] two algorithms are described, in which the decision is made by comparing the total power received (or transmitted) by the BS in which the new user should be admitted with an appropriately chosen threshold. The difference between them is related to the different measurements used (in the first case the BS measures the power that it receives from all mobiles, in the second case it measures the power that it transmits to the mobiles), and implies that the *Received Power Call Admission Control* (RPCAC) algorithm is more suitable for our goals. In fact, because of the definition, RPCAC is a really instantaneous algorithm, whereas in the *Transmitted Power Call Admission Control* (TPCAC) algorithm an additional setup time is required, because the transmitted power levels have to be set in an assigned number of iterations, so it provides decisions only after a testing phase. Note that RPCAC is based on the uplink, whereas TPCAC studies the downlink: however, the heuristic nature of the algorithms allows us to limit the analysis to only one direction. For this and for the better conceptual simplicity and performance, we will analyze in the following the RPCAC algorithm as representative of power-based CAC. Another algorithm is proposed in [4], where the adopted metric is γ^{-1}, where γ is the SIR, intended as signal-to-interference ratio measured on the reverse link by the BS to whom the mobile requests connection, so that the threshold model for the algorithm is equivalent to the instantaneous feasibility of the target SIR.

Note that algorithms of this kind avoid in a smart way the problem of the computation of the increase ΔI, that is equal to 0. This fact, that may appear counterintuitive, means that the admission control is done only by evaluating the load of the system before the new mobile is added. This allows to operate CAC correctly, by simply choosing a threshold I'_{th} that does not represent the edge of the overload situation, but it includes a margin, e.g., if I_{th} is the real threshold that represent the congestion, we can choose $I'_{th} = I_{th} - I_m = 0.9 I_{th}$. Approximately then, if the received power is below the threshold I'_{th}, the system can provide service to at least another user, without computing the actual additional load given by the new call.

This algorithm better exploits the heuristic model, because the numerical evaluation of the I-parameters is easier. However, there is still a trade-off in the choice of I_m: if no margin is adopted, it is possible to switch from being below threshold to overload without protection. If the margin is too large, capacity may be wasted, or, on the other side, the system can remain too conservative if the normal operating point is between $I_{th} - I_m$ and I_{th}: in fact, in this range no admission is performed, even if the new user requirement is low. As a final comment, note that the use of power metrics corresponds to a small additional overhead in the system, since in common situations the power level measurements are already part of the system procedures.

IV. Simulations and Performance Evaluation

Simulations have been performed with a simulator of the UMTS system, in which some user dynamics have been implemented. The simulation environment presents a users' deployment based on a structure of 3×3 hexagonal cells wrapped onto itself so as to have no "border effect" [1]. In radio channel propagation, in addition to path loss, both fast fading and shadowing have been taken into account: fast fading with the well known multi-oscillator Jakes' simulator [11] and shadowing with Gudmunson's correlation model [12]. Doppler frequency for the Jakes' simulator has been set equal to $f_c + (v/\lambda)$ hertz, where v is mobile speed, $\lambda = f/c$ the wavelength, equal to 0.16 in the simulator, and f_c a constant term equal to 2 Hz [2]. The parameter of the log-normal distribution of the shadowing is $\sigma = 4dB$.

We consider users with speed that is Gaussian distributed and independently re-determined every 0.1 s, with assigned mean and variance: the direction of movement is also changed by choosing a rotation angle uniformly between $-\pi/4$ and $+\pi/4$. Users are generated, or already connected users terminate their call, following a birth–death Poisson process. When a new user arrival time is calculated, its mobility parameters are also randomly determined, by selecting one of four mobility classes with equal probability. In practice, we have indeed four Poisson processes, which correspond to four kinds of user: stationary (mean speed $v_m = 0.0$m/s, with standard deviation $v_s = 0.0$ m/s), slow ($v_m = 4.0$m/s, $v_s = 0.5$ m/s), medium ($v_m = 8.0$m/s, $v_s = 1.0$ m/s), fast ($v_m = 12.0$m/s, $v_s = 1.5$ m/s). When users are admitted Power Control tries to guarantee a minimum SIR $\gamma^{tar} = 4.5$dB.

The examined algorithms for Admission Control are RPCAC and HCCAC, each with different threshold values. As seen before, the heuristics are the received power for RPCAC and the number of users for HCCAC. Thus, different values of the threshold P_{th} in the RPCAC algorithm are different received power levels, normalized to the average power contribution that a MS, $0.5d$ away from the BS (d is cell radius), gives to P_{tot} when it transmits at maximum power. In HCCAC results instead, the threshold N_{th} corresponds to the number of users per cell. A curve, introduced only for the sake of comparison, is called *Admit all* and corresponds to giving access to every user that requests it (i.e., $I_{th} = \infty$). So, no calls are blocked and only the probability of dropping can be calculated.

The studied metrics are the probability of blocking a generic user that requests to be admitted (P_b), the probability of being dropped for a user already in the system, due to overload of the network (P_d) [3], and a weighted combination of these two metrics, i.e., $P_b + 10P_d$, being call dropping generally considered more annoying than call blocking during the admission phase. These metrics are evaluated as a function of the mean load of each cell in the network, expressed in erlang/cell.

Let us consider Figures 1–6: they represent P_b, P_d and the linear combination $P_b + 10P_d$ of both RPCAC and HCCAC algorithms, for different threshold values. As it can be observed, when P_{th} for RPCAC or N_{th} for HCCAC varies, blocking and dropping probability present an interesting trade-off, because if the threshold is decreased, i.e., the system becomes more conservative, P_b increases whereas P_d decreases. However, Figures 3 and 6 show that the linear combination $P_b + 10P_d$ for the two algorithms only weakly depends on the chosen threshold.

We can notice that a power-based decision has better performance than the hard capacity approach: in fact, with the same load, RPCAC obtains lower values of the $P_b + 10P_d$ metric: e.g., if the load is 3 erlang/cell, the two algorithms with the best threshold have a metric equal to 0.02 and 0.03, respectively.

[1] The consistence of the results for the cases of 4×4 and 5×5 cells has also been verified.

[2] This additional constant term allows to consider the environment mobility, i.e., to assign a non-zero Doppler frequency even to stationary users with $v = 0$.

[3] In our simulations, trace is kept of every user's SIR, and if the SIR of a user remains below a threshold $\gamma^{th} = \gamma^{tar} - 0.5$dB (i.e., $\gamma_i < \gamma^{th}$) for a specified amount of time, congestion is detected and that user is dropped.

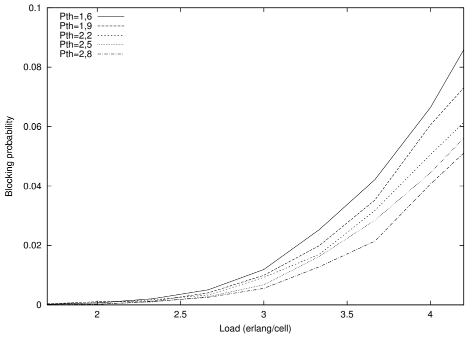

Fig. 1. Block probability for RPCAC

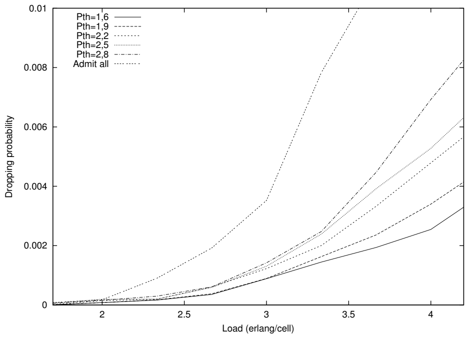

Fig. 2. Drop probability for RPCAC

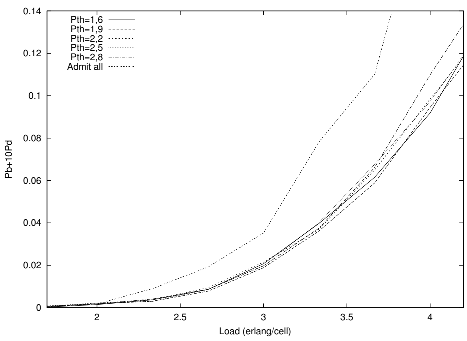

Fig. 3. $P_b + 10P_d$ for RPCAC

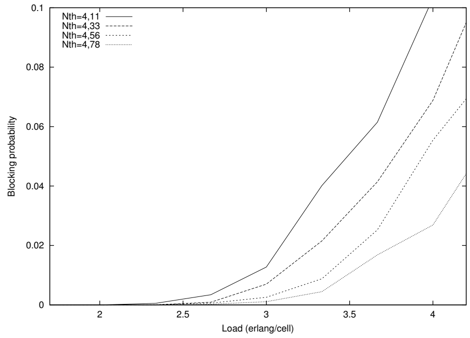

Fig. 4. Block probability for HCCAC

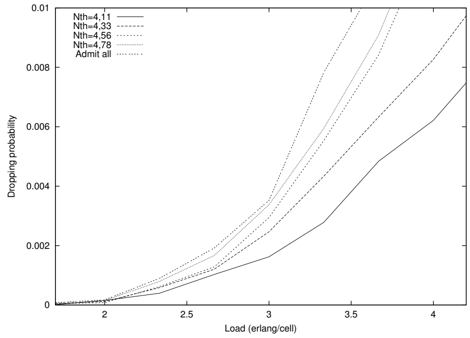

Fig. 5. Drop probability for HCCAC

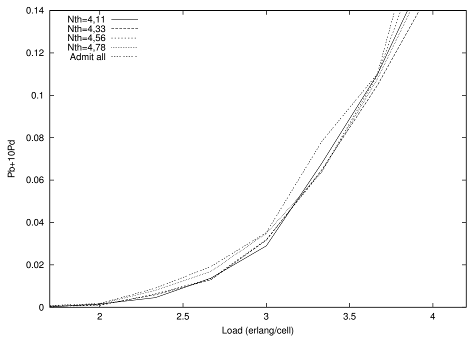

Fig. 6. $P_b + 10P_d$ for HCCAC

If the load is higher, RPCAC outperforms HCCAC even more: if the load is 4 erlang/cell, RPCAC presents, as its best performance, a metric equal to 0.09, while HCCAC has almost 0.17. We can conclude that under this aspect RPCAC algorithm has a consistent gain with respect to HCCAC. It can also be seen that the choice of the threshold does not affect the global performance (the linear combination has almost the same values in both RPCAC and HCCAC cases), but implies great variability of the point in which the trade-off between blocking and dropping is cut. Moreover, Figures 1–3 and 4–6 show that a power-based Admission Control allows a significant improvement with respect to the "Admit All" situation (i.e., call blocking is performed well, so that a great number of dropping events is avoided), whereas a hard capacity approach manages the already mentioned trade-off between blocking and dropping, but decreases only marginally their global effect: this means that the blocking is often done in an erroneous way and to decrease the call dropping a large number of calls must be blocked.

Finally, the chosen threshold range and steps allow us to conclude that setting the threshold in HCCAC is more problematic than in RPCAC. In fact, even with a finer granularity of the steps (each curve in Figures 4–6 differs from the previous by almost 5%, whereas the curves in Figures 1–3 sweep the global range with steps greater than 10%), blocking and dropping probabilities are more variable in the HCCAC case. Thus, we can conclude that the correct choice of the threshold is more difficult in HCCAC than in a power-based approach.

V. Analysis of Parameter Setting

We now propose a slightly different way to model and discuss the algorithms' performance. Because of the threshold nature of the algorithms, if the load varies, different classes of behavior can be identified. In particular, when the load is relatively low with respect to the threshold, the performance metrics are close to 0, regardless of the chosen threshold. If we suppose to study the algorithms by increasing the load from 0 to higher values, we can think of the initial phase, in which the load is low, as a *startup phase*.

By looking for example at Figure 1 (or Figure 4), we can notice that, by increasing the load, a *threshold effect* can be observed, i.e., the curves are almost equivalent for low values, and rise separately when the load increases, such that lower thresholds present the rise point earlier. This can be called, because of its approximatively linear behavior, *quasi-linear phase* of the algorithm. In this phase, different thresholds lead to significantly different values of the performance metrics, and so this part needs to be discussed and analyzed in more detail.

We can expect that after the quasi-linear phase, if the load is further increased, the algorithm encounters a *saturation phase*, in which P_d remains constant, whereas P_b tends to 1. The saturation corresponds to a situation in which most new users are blocked (and the greater the offered load, the higher the blocking probability), so the network load due to admitted users is roughly constant and only depends on the threshold value. In the performed simulations, the saturation is only partially visible, because if the threshold is correctly chosen, it occurs relatively late, when P_d is already too high. Based on these considerations, the threshold management has to be referred only to the quasi-linear phase, that is the region of interest.

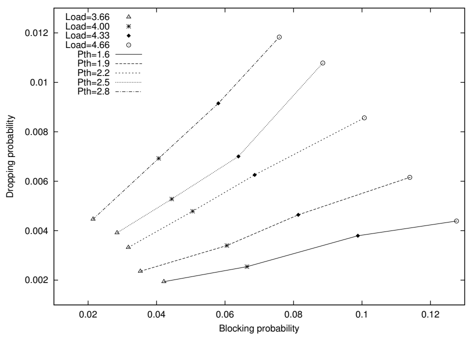

Fig. 7. P_b vs. P_d in RPCAC for various threshold values

P_{th}	1.6	1.9	2.2	2.5	2.8
m	0.029	0.049	0.076	0.092	0.136
m^{-1}	34.5	20.4	13.2	10.8	7.4

TABLE I

SLOPE OF P_b VS. P_d DEPENDING ON P_{th}

Figures 1 and 2 show only the variability of these parameters with respect to the network load. More interesting observations can be done by examining Figure 7 in which RPCAC is considered and P_d versus P_b is represented, for various loads and thresholds: it can be observed that the curves $P_d = P_d(P_b)$, for fixed values of the threshold, are almost straight lines. This corresponds to a *linear region* of behavior: although these metrics are not linear as a function of the offered load, if they are plotted one versus the other it can be seen that, as the network load increases, the relative increases of the probabilities of blocking and dropping are almost linearly related.

On the other hand, as was to be expected, varying the threshold for fixed load leads to a trade-off between P_d and P_b. If we consider a variation of the load, it can be important to know in this case how the trade-off changes in terms of relative variations of P_b versus P_d. To evaluate them, let us consider the slope of the lines of Figure 7: their approximate values are tabulated in Table I. A slope equal to m means that a variation $+\delta$ of P_b roughly corresponds to a $+m\delta$ increase for P_d: in other words the call drop "annoyance" increases m^{-1} times proportionally to the one of the block in admission. If we set of a "blind ratio" k between the weight of dropping and blocking, i.e., a drop event is evaluated as annoying as k blocks in admission, a way to evaluate the correct threshold could be to choose a threshold that satisfies the equation: $k = m^{-1}$. Even though these results and considerations have been derived for the RPCAC algorithm, we expect them to apply to other heuristic CAC policies as well. Further research [13] can extend this work, and the proposed framework appears to be also in good agreement with the preliminary results for other algorithm, like the *Looking Around* [8].

VI. CONCLUSIONS AND FUTURE WORK

In this paper, Call Admission Control instantaneous threshold algorithms have been modeled and discussed, and their application to Third Generation systems has been emphasized. Moreover, a threshold model has been developed, with a framework useful both to describe the behavior of the network and to set the threshold parameters of instantaneous algorithms directly, if the QoS utility function of the server is known.

Further ways to optimize the performance can be found if a better model of the system is available, i.e., different QoS requirements for the users, or more classes of service, by means of priority, data rate, mobility are considered. In particular, the framework could be identified when not only the global blocking and dropping probability, but other QoS parameters such as fairness or generality of the results are considered.

REFERENCES

[1] Y. Ishikawa, N. Umeda, "Capacity design and performance of call admission control in cellular CDMA systems", *IEEE Journal on Selected Areas in Communications (JSAC)*, Volume 15 Issue 8, pp 1627–1635, October 1997.

[2] M. Andersin, Z. Rosberg, J. Zander, "Soft and safe admission control in cellular networks", *IEEE/ACM Transactions on Networking*, Volume 5, Issue 2, pp 255–265, April 1997.

[3] C.-Y. Huang, R.D. Yates, "Call admission in power controlled CDMA systems", *Proceedings of VTC Fall 1996*, IEEE VTS Vehicular Technology Conference, Volume 3, pp 1665–1669, 1996.

[4] Z. Liu, M. El Zarki, "SIR-based call admission control for DS-CDMA cellular systems", *IEEE JSAC*, Volume 12, Issue 4 , pp 638–644, May 1994.

[5] L. Nuaymi, P. Godlewski, C. Mihailescu, "Call admission control algorithm for cellular CDMA systems based on best achievable performance", *Proceedings of VTC Spring 2000*, IEEE VTS Vehicular Technology Conference, Volume 1, pp 375–379, 2000.

[6] D. Kim, "Efficient interactive call admission control in power-controlled mobile systems", *IEEE Transactions on Vehicular Technology*, Volume 49, Issue 3, pp 1017–1028, May 2000.

[7] J. Kuri, P. Mermelstein, "Call admission on the uplink of a CDMA system based on total received power", *Proceedings of ICC'99*, Volume 3, pp 1431–1436, June 1999.

[8] Y. Ma, J.J. Han, K.S. Trivedi, "Call admission control for reducing dropped calls in code division multiple access (CDMA) cellular systems", *Proceedings of INFOCOM 2000*, IEEE Computer and Communications Societies Conference, Volume 3, pp 1481–1490, 2000.

[9] H. Holma, J. Laakso, "Uplink Admission Control and Soft Capacity with MUD in CDMA", *Proceedings of VTC Fall 1999*, IEEE VTS Vehicular Technology Conference, Volume 1, pp 431–435, 1999.

[10] J. Outes, L. Nielsen, K. Pedersen, P. Mogensen, "Multi-cell admission control for UMTS", *Proceedings of VTC Spring 2001*, IEEE VTS Vehicular Technology Conference, Volume 2, pp 987–991, May 2001.

[11] W. C. Jakes, *Microwave Mobile Communications*, New Jersey, IEEE Press, 1993.

[12] G. Stüber, *Principles of mobile communication*, Boston, Kluwer, 1996.

[13] L. Badia, M. Zorzi, "A Framework for Call Admission Control Threshold Setup and Evaluation of the Performance in WCDMA Systems", accepted for publication in the *Proceedings of VTC Spring 2003*, April 2003.