# Prioritized call admission control for web browsing services in 3G networks

Yusef R. Cáceres Zúñiga, Ernesto L. Andrade Neto,
Michel Daoud Yacoub

***Abstract*— This paper deals with a Call Admission Control (CAC) for web browsing services in an integrated DS-CDMA system. It considers the reverse link with heterogeneous traffic generated by voice and web browsing services. Two new CAC schemes that maintain the hand-off blocking probability below a specified threshold are presented. They are named Hard Partitioning and Soft Partitioning. As far as hand-off calls are concerned the latter always gives better results than the former. Both algorithms are suitable for cellular systems in which traffic load is variable in time.**

***Index Terms*— Quality of service, radio resource management, call admission control.**

## I. INTRODUCTION

THIRD generation (3G) wireless networks will provide enhanced high-speed data, multimedia and voice services in order to meet the users' need for integrated, flexible, cost-effective, and feature rich services. To enable these services it is essential for 3G wireless networks to support multiple classes of traffic with different characteristics and stringent quality of service (QoS) requirements such as delay, bit error rate, and throughput. However, the available air interface resources (i.e. bandwidth, power, codes) in 3G wireless networks are scarce. To accommodate these two contradictory demands (scarce resource and stringent QoS requirements), an efficient radio resource management (RRM) algorithm is needed. That algorithm should adapt to variable traffic load, user mobility, unreliable radio propagation channel conditions, and other changes in the system's conditions.

A key component of the RRM algorithms is the call admission control (CAC). The objective of CAC schemes is to regulate the operation of a network in such a way that ensures the QoS to the existing connections and at the same time accommodates in a optimum way the new connections request [1].

Several CAC algorithms have been proposed for 3G wireless networks (e.g. see [2], [3], [4].). Particulary, in [5], a two level access control for DS-CDMA systems carrying voice and web browsing services is developed. The first level of the access control is a call admission control scheme, which allows a user to access the system only if, after admission, the resulting number of users for each traffic class is feasible, i.e., the QoS specifications in terms of bit error rate (BER) and access delay for all users in the system can still be met. The second level of the access control is a flow control, which balances the system load such that real time services (voice) are given priority, while some bounded delay specification is met for delay tolerant applications (web browsing). This approach allows new and hand-off web connections to share all the residual capacity available after the voice contribution has been subtracted from the total available capacity. In such a case, there is no distinction between new calls and hand-off calls. However, it is usually desirable that hand-off calls are given priority over new calls, so as to reduce the forced termination probability of calls in progress that are entering a new cell.

In this work, the aim is to improve the solution proposed in [5], including two important performance measurements in connection-level quality of service (QoS): (i) new call blocking probability; and (ii) the hand-off blocking probability. Traditionally, there are two ways to prioritize hand-off calls: (i) reserving a certain capacity exclusively for hand-off purposes; and (ii) queuing hand-off requests according to various queuing disciplines. The former approach is used in this work to solve the call admission control problem with prioritized hand-off web connections.

The paper is organized as follows. In Section II, the system and traffic models is presented. In Section III and Section IV, two new forms of hand-off calls prioritization are introduced, namely and respectively, Hard Partitioning and Soft Partitioning. Finally, the conclusions are shown in Section V.

## II. SYSTEM MODEL

The system model considers a single cell DS-CDMA system that can support mobile terminals demanding voice and web browsing services. The voice source is modelled by a two-state ON/OFF Markov process. The transition rates from ON to OFF and from OFF to ON state are $\mu$ and $\lambda$, respectively.

Considering that there are $N_v$ voice users in the system the cumulative voice activity process is approximated by a discrete Markov chain. The transition probabilities for $n_v(n)$ active voice users in the $n$-th time slot are given by

$$P\{n_v(n+1)=k+1 \mid n_v(n)=k\} = \frac{\lambda_k}{\lambda_k+\mu_k}(1-\exp(-(\lambda_k+\mu_k)T_s)),$$
$$P\{n_v(n+1)=k-1 \mid n_v(n)=k\} = \frac{\mu_k}{\lambda_k+\mu_k}(1-\exp(-(\lambda_k+\mu_k)T_s)),$$
$$P\{n_v(n+1)=k \mid n_v(n)=k\} = \exp(-(\lambda_k+\mu_k)T_s),$$

where $\lambda_k=\lambda(N_v-k)$, $\mu_k=\mu k$ and $T_s$ is the time slot duration.

UNICAMP- Univ. Est. de Campinas, Fac. de Eng. El´etrica, Depto. de Comunicações, C.P. 6101, 13081-970 Campinas, SP, Brazil. E-mail: yusef@decom.fee.unicamp.br, michel@decom.fee.unicamp.br

The stationary probability of state where $k$ voice users are active is

$$P\{n_v = k\} = \frac{\binom{N_v}{k}\lambda^k\mu^{-k}}{\sum_{j=0}^{N_v}\binom{N_v}{j}\lambda^j\mu^{-j}} \tag{1}$$

From (1), a mean value of active voice users $n_v$ is obtained in an exact manner as

$$E[n_v] = \lambda N_v/(\lambda+\mu).$$

The web browsing sessions arrive in random times $T_1, T_2, \ldots$ according to a Poisson process with parameter $\lambda_s$. Each web browsing session consists of a random number of packet calls ($N_{pc}$), which is geometrically distributed with mean $\mu_{N_{pc}}$. The reading time between two consecutive packet calls is a geometrical distributed random variable with mean $\mu_{T_{pc}}$. The packet call is comprised of a random number of IP packets $N_p$, which is Pareto distributed. The inter arrival time between packets is a geometrical distributed random variable with mean $T_i$. Finally, the web packet length is fixed to $m$ voice packets. This model, proposed in [6], can be used to represent either the reverse traffic for web downloads or large file uploads.

To represent the mobility, a simple model is used. The dwell time (the amount of time during which a mobile terminal stay in a cell during a single visit) is assumed to follow an exponential distribution with mean $\mu_d$.

The general aspects of the two level access control algorithm, as presented in [5] are depicted in Figure 1. The first level is the call admission controller, which makes decisions based on: (i) the aggregate average web load measurement ($\rho^m(n)$ ), using a sliding observation window, and (ii) the admission threshold ($T_\rho$). This kind of load measurement is used to overcome the errors introduced by the high variability of the web traffic heavy tailed distribution. The load outage condition is defined as the probability that, after admission, the web load in the next slot $\rho(n+1)$ is higher than a predefined target load $\rho_t$, which is the mapping of a delay constraint. Then, the admission threshold is the minimum value of $\Delta\rho = \rho_t - \rho^m$ such that the load outage probability is lower that a fixed value.

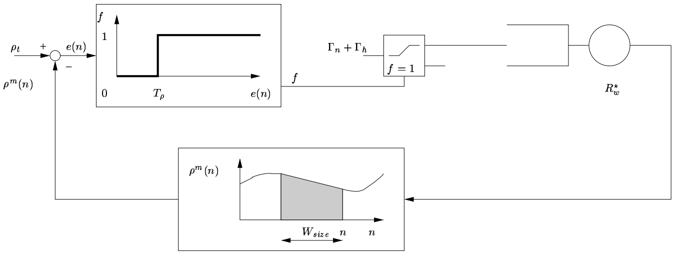

Fig. 1. Access control as developed in [5].

The admission control algorithm admits all new and hand-off web connection requests in the next slot if $\rho_t - \rho^m(n) \geq T_\rho$ holds. Otherwise, it rejects all new and hand-off web connections for the next time slot. This approach allows new calls and hand-off web connections to share all the residual capacity available after the voice contribution has been subtracted from the total available capacity. However, when a hand-off call occurs, the CAC has to ensure that the new cell has sufficient resources to support the new requested services at the desired QoS level. The issue of hand-off prioritization, not investigated in [5], will be explored in detail in the following sections.

The second level is the flow control algorithm, which predicts the residual capacity available for web sessions in the next slot $\hat{R}^\star_w(n+1)$, expressed by

$$\hat{R}^\star_w(n+1) = \frac{W(1-\eta) - SIR_v\hat{n}_v(n+1)R_v}{SIR_w}$$

based on the current measurements on the number of active voice user $n_v(n)$, i.e. $\hat{n}_v(n+1) = n_v(n)$. The predicted capacity is distributed in a round-robin fashion to allocate resources for the active connections in the next slot.

## III. HARD PARTITIONING CAC

The Hard Partitioning CAC algorithm, as proposed here, divides the available capacity into two separate portions to maintain the hand-off blocking probability below a pre-defined threshold, while keeping the QoS requirements of all web browsing services. In this scheme, each type of call can only use its assigned capacity. Figure 2 depicts the block diagram for the Hard Partitioning implementation.

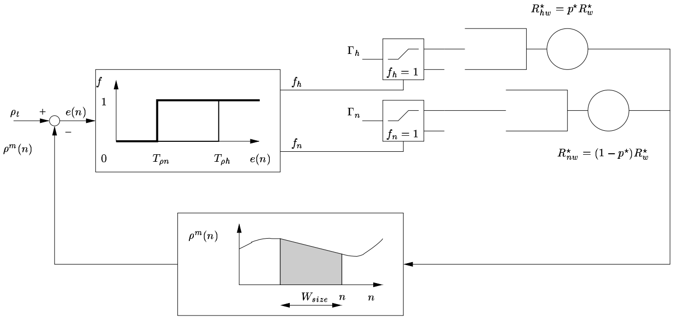

Fig. 2. Hard partitioning access control.

In [5], the aggregate average rate for the web users is given by

$$R^\star_w = E\Big[\sum_{i=1}^{n_w} R^i_w\Big] \leq \frac{W(1-\eta) - SIR_v E[n_v]R_v}{SIR_w} \tag{2}$$

where $n_v$ ($n_w$) represents the number of active voice (web) calls, $R_v$ ($R_w$) denotes the transmission rate for voice (web) calls, $SIR_v$ ($SIR_w$) is the target signal to interference ratio for voice (web) calls, $W$ represents the system bandwidth and $\eta$ is the ratio between the background noise level and the total received interference.

We reserve part of the capacity in (2) exclusively for hand-off web calls ($R^\star_{hw}$) by means of a partitioning factor ($p^\star$). The analysis begins by defining $p$ as

$$p = \frac{R_{hw}}{R^\star_w}, \qquad R_{hw} \in [0, R^\star_w].$$

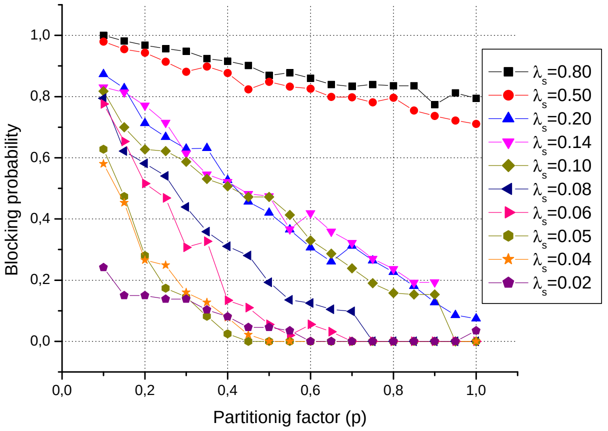

Fig. 3. Blocking probability as a function of the partitioning factor for hard partitioning strategy.

Denoting by $T_n$ the instant in which the $n$-th web session arrives, the hand-off blocking probability ($P_{hb}$) is given by

$$P_{hb}(R_{hw} = p \cdot R_w^\star) = P\{\rho(T_n) > \rho_t - T_\rho\}$$

The above definition enables one to determine the partitioning factor $p^\star$ as the minimum value of $p$ such that the hand-off blocking probability ($P_{hb}$) is below a certain threshold ($\phi$), i.e.

$$p^\star = \min\{p : P_{hb}(R_{hw} = p \cdot R_w^\star) \leq \phi\} \tag{3}$$

Thus, in order to reduce hand-off failures, the capacity reserved only for hand-off procedures can be determined by $R_{hw}^\star = p^\star R_w^\star$ and, as consequence, the capacity reserved exclusively for new web sessions is given by $R_{nw}^\star = (1 - p^\star)R_w^\star$.

Since analytical solutions for the blocking probability, and consequently the optimum value for partitioning factor, are hard to obtain simulations are used to determine these parameters.

Figure 3 shows the blocking probability of hand-off web sessions where the aggregate average rate is given by $pR_w^\star$. The aim is to vary the partitioning factor $p$ so that the required hand-off blocking probability in (3) is achieved. Table I summarizes the parameters of the wireless model used in the simulation.

It is important to notice that, the trade-off between new calls and hand-off blocking probability is established as a function of the partitioning factor $p$. To exemplify a simple case, assume that the hand-off web session arrival rate is given by $\lambda_{hs} = 0.04$ and the value of $\phi = 0.1$, then from Figure 3 the optimum value for the partitioning factor $p^\star$ is 0.4. Consequently, the partitioning factor for new calls is $1 - p^\star = 0.6$. Again, for Figure 3 the blocking probability of new web sessions is 0.27 when their arrival rate $\lambda_{ns}$ is 0.10. Also notice that, an excessive new call blocking probability is a result of giving too much priority to hand-off calls.

Based on the arguments developed in [5], the following CAC algorithm for the hard partitioning approach is defined:

Step 1:Measure the new total offered load averaged over the last slots:

$$\rho_h^m(n) = \frac{\frac{1}{\min\{n, W_{size}\}} \sum_{j=\max\{0, n-W_{size}\}}^{n} (N_h^{arr})_j}{p^\star R_w^\star},$$

$$\rho_n^m(n) = \frac{\frac{1}{\min\{n, W_{size}\}} \sum_{j=\max\{0, n-W_{size}\}}^{n} (N_n^{arr})_j}{(1 - p^\star) R_w^\star},$$

where the number of hand-off (new) packets arriving in slot $j$ is denoted by $(N_h^{arr})_j$ $((N_n^{arr})_j)$ and the aggregate average rate reservation for web session users is given by (2).

Step 2:Admit all hand-off connection requests in the next slot, if

$$\Delta\rho_h(n) = \rho_t - \rho_h^m(n) \geq T_{\rho h}$$

holds. Otherwise, reject all hand-off connections for the next time slot.

Admit all new connection requests in the next slot, if

$$\Delta\rho_n(n) = \rho_t - \rho_n^m(n) \geq T_{\rho n}.$$

Otherwise, reject all web connections for the next time slot.

As a final observation, the above scheme is able to adapt to dynamic load conditions due to fluctuations on the hand-off web sessions arrival rate ($\lambda_s$) and maintains the hand-off blocking probability below a fixed target value. In order to track variations in the hand-off web session arrival rate, the capacity reserved only for hand-off web sessions can be adjusted dynamically by selecting the appropriate partitioning factor $p$ using Figure 3.

## IV. SOFT PARTITIONING CAC

The Soft Partitioning CAC algorithm, also divides the available capacity into two portions. But, if a hand-off connection attempt fails to enter into the system using the capacity reserved for hand-off it attempts connection again, now trying to use the system capacity reserved for new calls. Therefore, it is expected that borrowing capacity reserved for new calls reduces the hand-off blocking probability, while satisfying the QoS (BER and delay) of all ongoing calls. The Figure 4 depicts the block diagram for the Soft Partitioning implementation.

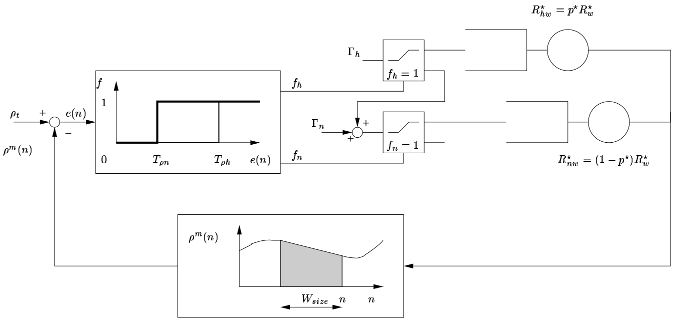

Fig. 4. Soft partitioning access control.

TABLE I

NETWORK MODEL PARAMETERS

| Parameter | Notation | Value |
|---|---|---|
| System bandwidth | $W$ | 3.75 MHz |
| Noise rise coefficient | $\eta$ | 0.1 |
| Target SIR voice | $SIR_v$ | 7 dB |
| Target SIR www | $SIR_w$ | 10 dB |
| Number of voice users in the system | $N_v$ | 20 |
| Mean packet inter arrival time | $t$ | 0.01 s |
| Discrete unit of time for packet inter arrival | $t_i$ | 0.001 s |
| Probability of a new packet arrival | $p_i$ | 0.2 |
| Probability of a new packet call initiation in the next time slot | $p_c$ | 0.0001667 |
| Mean number of packet calls per session | $\mu_{N_{pc}}$ | 5 |
| Mean inter arrival time between packet call initiations | $\mu_{T_{pc}}$ | 5 |
| Pareto distribution parameter | $\alpha$ | 1.1 |
| Pareto distribution parameter | $b$ | 2.27 |
| Length of a packet measured in number of basic packets | $m$ | 20 |
| Time slot = frame duration | $T_s$ | 0.02 s |
| Load outage probability constraint | $\psi$ | 0.025 |
| Target web load | $\rho_t$ | 0.1 |
| $i$-th voice call rate | $R_v^i$ | 9.6 Kb/s |
| Voice activity factor | $a$ | 0.4 |
| Mean dwell time | $\mu_d$ | 10 min |
| Sliding window size | $W_{size}$ | 30000 slots |

Because hand-off web sessions blocked in its assigned capacity are steered towards new web sessions capacity, the threshold for new web sessions ($T_{\rho n}$) have to estimated considering such a phenomenon.

In [5], the new web load for the next slot $\rho_{new}(n+1)$ is given by

$$\rho_{new}(n+1) = N_{new}(n+1)\frac{R_{pc}}{R^{\star}_{wn}}$$

where the number of new packets calls offered in slot $n+1$ is denoted by $N_{new}(n+1)$, the average rate offered per packet call is represented by $R_{pc}$ and the aggregate reserved rate for new web session is denoted by $R^{\star}_{wn}$.

For the Soft Partitioning CAC algorithm, the number of new packet calls offered in slot $n+1$ ($N_{new}(n+1)$) comprises:

- new sessions arrivals ($N_{ns}(n+1)$) with Poisson distribution of rate $\lambda_{ns}T_s$,
- new packet call initiations from previously admitted sessions ($N_{bs}(n+1)$) currently inactive. This random variable follows a binomial distribution with parameters $N_i$ and $p_c$,
- hand-off session arrivals ($N_{hs}(n+1)$) blocked in its assigned capacity with Poisson distribution of rate $\lambda_{hs}P_{hb}T_s$, where $P_{hb}$ is given by (3).

That is,

$$N_{new}(n+1) = N_{ns}(n+1) + N_{bs}(n+1) + N_{hs}(n+1).$$

From this result and following the same arguments given in [5], the load outage probability is determinate by (4).

Thus, the threshold for new web session $T_{\rho n}$ is calculated by

$$T_{\rho n} = \min\{\Delta\rho : P_{out-load} \leq \psi\}.$$

The soft partitioning admission control algorithm is defined as follows:

Step 1:Measure the new total offered load averaged over the last slots:

$$\rho_h^m(n) = \frac{\frac{1}{\min\{n, W_{size}\}}\sum_{j=\max\{0, n-W_{size}\}}^{n}(N_h^{arr})_j}{p^{\star}R_w^{\star}},$$

$$\rho_n^m(n) = \frac{\frac{1}{\min\{n, W_{size}\}}\sum_{j=\max\{0, n-W_{size}\}}^{n}(N_n^{arr})_j}{(1-p^{\star})R_w^{\star}},$$

where the number of hand-off (new) packets arriving in slot $j$ is denoted by $(N_h^{arr})_j$ ($(N_n^{arr})_j$) and the aggregate average rate reservation for web session users is given by (2).

Step 2:Admit all hand-off connection requests in their reserved capacity for the next slot if

$$\Delta\rho_h(n) = \rho_t - \rho_h^m(n) \geq T_{\rho h}$$

holds. Otherwise, admit those requests in the assigned new calls capacity for the next slot if the following inequality

$$\Delta\rho_n(n) = \rho_t - \rho_n^m(n) \geq T_{\rho n}$$

is feasible. Otherwise, reject all hand-off connections for the next time slot.

Admit all new connection request in the next slot, only if

$$\Delta\rho_n(n) = \rho_t - \rho_n^m(n) \geq T_{\rho n}$$

is satisfied.

The system performance using the Soft Partitioning CAC algorithm is obtained with simulation. Table II shows that the use of the Soft Partitioning approach improves the performance of the system compared to the Hard Partitioning scheme. These better results are obtained with an increase on the blocking probability of new calls. On the other hand, it is observed that the improvement in performance is highly dependent on the traffic load generated by the new calls.

$$P_{out-load|(N_i,T_i,N_{nb},N_{ns},N_{hs})} = \begin{cases} 1, & \frac{m(N_{bs}+N_{ns}+N_{hs})}{(\Delta\rho)R^{\star}_{nw}T_i} \geq 1, \\ 0, & (1-\frac{m(N_{bs}+N_{ns}+N_{hs})}{(\Delta\rho)R^{\star}_{nw}T_i}) \geq \frac{1}{b}, \\ 1-b^{\alpha}(1-\frac{m(N_{bs}+N_{ns}+N_{hs})}{(\Delta\rho)R^{\star}_{nw}T_i})^{\alpha}, & \text{otherwise.} \end{cases} \tag{4}$$

TABLE II

COMPARISON OF THE HAND-OFF BLOCKING PROBABILITY AS A FUNCTION OF THE PARTITIONING FACTOR $p$ FOR THE HARD AND SOFT PARTITIONING SCHEMES

| Web session arrival rate | $p^{\star}$ | Hand-off Blocking Probability | |
|---|---|---|---|
| | | Hard Partitioning | Soft Partitioning |
| $\lambda_{hs} = 0.06$ and $\lambda_{ns} = 0.14$ | 0.40 | 0.1338 | 0.0704 |
| | 0.45 | 0.1102 | 0.0751 |
| $\lambda_{hs} = 0.05$ and $\lambda_{ns} = 0.14$ | 0.30 | 0.1449 | 0.0579 |
| | 0.35 | 0.0821 | 0.0048 |
| $\lambda_{hs} = 0.04$ and $\lambda_{ns} = 0.14$ | 0.30 | 0.1602 | 0.0722 |
| | 0.35 | 0.1271 | 0.0718 |
| | 0.40 | 0.0773 | 0.0662 |
| $\lambda_{hs} = 0.02$ and $\lambda_{ns} = 0.14$ | 0.30 | 0.1379 | 0.0459 |
| | 0.35 | 0.1034 | 0.0115 |
| | 0.40 | 0.0805 | 0.0110 |

The observation carried out in Section III as far as dynamic adjustment is concerned is also applicable here.

## V. CONCLUSION

The admission control problem with prioritization of hand-off web sessions for 3G systems is analyzed. Two new algorithms are developed, named Hard Partitioning and Soft Partitioning, which keep the hand-off blocking probability below a pre defined threshold. According to the simulations, the Soft Partitioning algorithm always yields better results than the Hard Partitioning scheme. These better results are obtained with an increase of the blocking probability of new calls over the hand-off calls. It has been observed that the achieved improvement is highly dependent on the traffic load. On the other hand, the proposed algorithms are suitable for dynamic traffic load variations because their parameters can be dynamically adapted according to such variations.

## REFERENCES

[1] N. Dimitriou, G. Sfikas, and R. Tafazolli, "Call admission polices for UMTS," in *IEEE Vehicular Technology Conference*, 2000, pp. 1420–1424.

[2] C.-J. Ho, J. A. Copeland, C.-T. Lea, and G. L. Stüber, "On call admission control in DS/CDMA cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 6, pp. 1328–1343, 2001.

[3] W. S. Jean and D. G. Jeong, "Call admission control for mobile multimedia communications with traffic asymmetry between uplink and downlink," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 1, pp. 59–66, 2001.

[4] S. Kim, T. Kwon, and Y. Choi, "Call admission control for prioritized adaptive multimedia services in wireless/mobile networks," in *IEEE Vehicular Technology Conference*, 2000, pp. 1536–1540.

[5] C. Comaniciu, N. B. Mandayam, D. Famolari, and P. Agrawal, "Wireless access to the World Wide Web in an integrated CDMA system," in *IEEE Transactions on Wireless Communications* , vol. 2, no. 3, pp. 472–483, 2003.

[6] ETSI, "Selection procedures for the choise of radio transmission technologies of the UMTS," 1998.