3-2001

Admission Control Algorithms for Cellular Systems

Jon M. Peha
Carnegie Mellon University, peha@andrew.cmu.edu

Arak Sutivong
Carnegie Mellon University

Published In
Wireless Networks, 7, 117-125.

Accepted to appear in *ACM/Baltzer Wireless Networks*, 1999

Admission Control Algorithms for Cellular Systems

Jon M. Peha and Arak Sutivong
Carnegie Mellon University

Abstract

This paper evaluates call admission control algorithms for a cellular or microcellular system. Algorithms are evaluated based on two Quality of Service (QoS) metrics: the *new call blocking probability*, which is the probability that a new call is denied access to the system, and the *forced-termination probability*, which is the probability that a call that has been admitted will be terminated prior to the call's completion. Three novel algorithms are presented: the *Weighted Sum Scheme*, the *Probability Index Scheme*, and the *Hybrid Control Scheme*. The weighted sum scheme uses the weighted sum of the number of calls underway in various cells when making the admission decision. The probability index scheme computes a probability index, which reflects the forced-termination probability of a new call arrival, and admits those calls with low probability indexes. The hybrid control scheme combines these two approaches. These novel algorithms are compared with three known algorithms: the *Reservation Scheme* in which a specific number of channels are reserved in each cell for handoffs, the *Linear Weighting Scheme* in which the admission decision depends on the total number of calls underway in a group of cells, and the *Distributed Admission Control Scheme* in which the admission decision depends on the projected overload probabilities in the cell at which the new call arrives and adjacent cells. We show that the Hybrid Control Scheme yields the best performance, particularly during periods when load differs from the expected level. We also show that the simple Reservation Scheme performs remarkably well, often superior to more complex schemes that have been proposed.

1.0 Introduction

In a cellular or microcellular system (like Carnegie Mellon University's [1]), it is the network provider's responsibility to provide adequate Quality of Service (QoS) to all users. Two critical QoS metrics are *blocking probability* (P_b) and *forced-termination probability* (P_f), which are defined as follows. Since cell capacity is limited, some call attempts may be blocked. The probability that a new call is not admitted into the system is called the blocking probability. Even after a call is admitted, the network may terminate the call prematurely when a handoff is attempted into a cell that has no capacity available. The probability that a call that has been admitted will be terminated some time before call completion is called the forced-termination probability. Both of these QoS metrics are strongly influenced by the *call admission control* algorithm, which determines whether a new call should be admitted or blocked. Blocking more calls generally improves the forced-termination probability of those calls that are admitted, so there is a trade-off. But *which* calls should be blocked to minimize forced-termination probability? This paper presents novel heuristic call admission control algorithms, and it shows the extent to which both existing and novel algorithms can improve these QoS metrics.

The system we address handles a single class of devices, so all devices have similar data rates, quality of service objectives, and mobility rates. We also assume that blocked calls and handoffs are lost rather than queued, and fixed channel assignment (FCA) is used [2]. Furthermore, we assume that the network has no knowledge of the direction in which any specific device will travel, as in [3,4]. The network only knows aggregate mobility statistics on all devices, e.g. the expected fraction of devices in cell i that will travel in the given direction.

In a fixed channel assignment system, a new call is always blocked when there is no channel available in a cell at the time of arrival. However, it has been shown [3-10] that the network performance can be improved by imposing other admission criteria along with this inherent criterion. These criteria depend on the availability of capacity in one or more cells. Let C_i denote a cell i in the system and C_0 denote the *originating cell*, which is the cell at which the new call arrives. Let N be the maximum number of simultaneous calls a single cell can handle and N_i be the number of calls underway in C_i. Finally, let the *region of awareness* be the group of cells in which the originating cell knows the number of calls underway at the time of admission. Essentially, the size of the region of awareness represents the amount of information used in making the admission decision. An admission control algorithm decides whether to admit a new call based on the current value of N_i for each cell i in the region of awareness. For any algorithm, this function can be executed quickly with a look-up table using the N_i values as indices. This paper addresses how to determine the output values in that look-up table.

The paper is organized as follows. We describe the known admission control algorithms in Section 2. We then present the novel algorithms in Section 3. The performance comparisons are presented in Section 4. Finally, we conclude the paper in Section 5.

2.0 Known Algorithms

We briefly discuss the following three known admission control algorithms: the *Reservation Scheme*, the *Linear Weighting Scheme*, and the *Distributed Admission Control Scheme* in Sections 2.1, 2.2, and 2.3, respectively.

2.1 Reservation Scheme (RS)

The reservation scheme is the simplest scheme that we consider. The algorithm only looks at the originating cell in determining the admission, i.e., the region of awareness contains only the originating cell itself. Let N_h ($N_h \geq 0$) be the number of channels reserved specifically for call handoffs. New calls are admitted if $N_0 < N - N_h$ .Variations of the reservation scheme have been proposed in [5, 6].

2.2 Linear Weighting Scheme (LWS)

The linear weighting scheme uses the mean number of calls underway in all cells within a maximum number of hops D from the originating cell in determining the admission (i.e., the region of awareness contains all cells within D hops from C_0.) Let S be the set containing all cells in the region of awareness, and let N-N_h be the threshold. New calls are only admitted to the originating cell 0 if

$$\frac{1}{|S|}\sum_{i:i \in S} N_i < N - N_h$$

Note that the linear weighting scheme reduces to the reservation scheme when D = 0. The authors of [7, 8] utilize a similar idea to the linear weighting scheme as part of their admission control algorithm. The difference is that in [7, 8] each group of cells that is used to compute the average is predefined, and these groups are non-overlapping.

2.3 Distributed Admission Control Scheme (DACS)

The distributed admission control scheme [9] takes into consideration the number of ongoing calls in the originating cell and its adjacent cells in making the admission decision. The authors in [9] considered one-dimensional cellular arrays such as ones sometimes used in highways, in which each cell has at most two adjacent cells. However, the above approach can be extended to two-dimensional cellular systems [9].

Let P_{QoS} be the user-declared QoS, and the *overload probability* be the probability that a call is terminated during a handoff at any given time. An approximation of the overload probability is shown in [9]. In both one-dimensional and two-dimensional systems, a new call is admitted at time t_0 only when the following conditions are met:

Condition 1: At time $t_0 + T$, the overload probability of cell C_0 must be smaller than P_{QoS}.

Condition 2: At time $t_0 + T$, the overload probability of each cell adjacent to the originating cell must be smaller than P_{QoS}.

for some arbitrary value of T that must be determined experimentally. According to [9], the first condition is intended to maintain the desired P_{QoS} of calls that are underway in the system, and the second admission condition is intended to provide the desired P_{QoS} of the newly admitted call.

This approach requires approximating the overload probability at time t_0+T. This depends on the new call arrival rate into each cell α, and the expected number of calls in a cell $E[n]$. Let SUM denote the sum of the number of ongoing calls in the cells adjacent to the originating cell. For each cell C_i that is adjacent to C_0, let $E_i[n]$ be the total expected number of calls in the cells adjacent to C_i when the new call arrives, excluding those calls underway in C_0. For a marked mobile in the originating cell C_0, the authors derive a p_s that approximates the probability that this mobile remains in the same cell during a period of duration T, and a p_m that approximates the probability that mobile hands-off during the same period. Furthermore, it is assumed that the mobile is equally likely to move to either adjacent cell.

In [9], the overload probability is approximated as follows. For a given P_{QoS}, the value a is found such that $P_{QoS} = Q(a)$, where $Q(.)$ is the integral over the tail of a Gaussian distribution. In this work, we are only interested in the two-dimensional cellular systems.

To satisfy condition 1,

$$N_0 < \frac{1}{2p_s}\left[a^2(1-p_s) + 2N - \frac{p_m}{3}(SUM) - a\sqrt{a^2(1-p_s)^2 + 4N(1-p_s) - \frac{p_m^2}{9}(SUM) + 2\frac{p_m}{3}p_s(SUM)}\right]$$

To satisfy condition 2,

$$N_0 < \frac{1}{2\frac{p_m}{3}}\left[a^2\left(2-\frac{p_m}{3}\right) + 4N - 4\alpha T - 2E_i[n]\frac{p_m}{3} - 4N_ip_s - a\sqrt{a^2\left(2-\frac{p_m}{3}\right)^2 + 16N + 8\frac{p_m}{3}(\alpha T + N_ip_s - N) - 16N_ip_s^2}\right] \quad \text{for each } C_i \text{ adjacent to } C_0$$

3.0 Novel Algorithms

We propose three novel algorithms: the *Weighted Sum Scheme*, the *Probability Index Scheme*, and the *Hybrid Control Scheme* [10], which are described in Sections 3.1, 3.2, and 3.3, respectively.

3.1 Weighted Sum Scheme (WSS)

The weighted sum scheme uses the weighted sum of the number of ongoing calls in the originating cell and in other cells in determining the admission. Let n_i be the mean number of calls underway in cells that are distance i from the originating cell, and p_i be the weighting; $\sum_{i=0}^{\infty} p_i = 1, p_i \geq 0 \text{ for } i$. Let the admission threshold be N-N_h. New calls are admitted only when there is at least one channel available in a cell and

$$\sum_{i=0}^{\infty} p_i n_i < N - N_h$$

The optimal weights p_i can be determined experimentally. Note that the weighted sum scheme reduces to the linear weighting scheme when

$$p_i = \begin{cases} (1/|S|) & i \leq D \\ 0 & i > D \end{cases}$$

3.2 Probability Index Scheme (PIS)

To block calls that would experience high forced-termination probabilities, the probability index scheme returns an index (Pf_{id}) which roughly reflects the forced-termination probability of a new call. If the index does not exceed a set threshold ($Pf_{threshold}$) and there is a channel available in the originating cell, then the call is admitted; otherwise, it is blocked. Although there is no known method of calculating this forced-termination probability of an arriving call, we can make simplifying assumptions that make the problem tractable, thereby producing a reasonable index. Because of these simplifying assumptions, the index is not an exact calculation, but it does not need to be. An index is valuable if it has the property that a higher index implies a greater probability of forced-termination. We will show through simulation that this can be achieved.

We estimate the forced-termination probability of a call, which is the probability that if a call is admitted, the mobile will enter a cell with no available channels before the call completes. We first compute the index that estimates the forced-termination probability of the call at each hop m, $Pf_{id}(m)$, given that it was not forced-terminated in the previous m-1 hops. $Pf_{id} = \sum_m Pf_{id}(m)$. Computing $Pf_{id}(m)$ requires the following: the *handoff time distribution*, $f_{Tm}(t)$, which is the distribution of time when the m^{th} handoff takes place, the *location distribution*, $p_i(m)$, which is the probability that a marked mobile, beginning in cell C_0, is in cell C_i after m handoffs, and the *cell occupancy distribution*, $\pi_j(t)$, which is the probability that a given cell has j calls underway at time t. The simplifying assumptions follow. When calculating the probability that a call will be terminated in C_i, we assume that a call can never be terminated in cells other than cell C_i even if it enters a cell with no available channel, and that the probability that a mobile enters C_i in its m^{th} hop is independent of whether or not it ever entered a cell with no available channels before. In addition, we assume that the probability that a mobile will be handed off from C_i to C_j is independent of its previous locations. Furthermore, to reduce the amount of computation, we assume that the probability that a mobile finds no available channels in C_i during a handoff is independent of the actual time the mobile is handed off to C_i, but rather depends on the number of calls underway in C_i at the time the new call arrives and how far away C_i is from C_0. Finally, we assume that new calls arrive according to a Poisson process with rate α per cell. Call holding time is exponentially distributed with mean $1/\mu$. The time between handoffs is also exponentially distributed with mean $1/\gamma$. Consequently, the probability of forced-termination at the m^{th} hop can be calculated from $f_{Tm}(t)$, $p_i(m)$, and $\pi_j(t)$. We will find the approximations for $f_{Tm}(t)$, $p_i(m)$, and $\pi_j(t)$ in Section 3.2.1, and will then use them to compute Pf_{id} in Section 3.2.2. A comparison between Pf_{id} and the actual forced-termination probability is shown in Section 3.2.3

3.2.1 Approximations of $f_{Tm}(t)$, $p_i(m)$, and $\pi_j(t)$

Because time between handoffs is exponentially distributed, the distribution of the time at which the m^{th} handoff takes place follows an m-Erlang distribution, or $f_{T_m}(t) = \frac{\gamma e^{-\gamma t}(\gamma t)^{m-1}}{(m-1)!}$. The next step is to determine the location distribution, $p_i(m)$. While the call is underway, let the *state* of the system be the location of this marked mobile as long as it remains within the region of awareness. To reduce state space, once a mobile leaves the region of awareness, it permanently enters the *out-of-region state* (denoted by *O*). Any further hand-offs, if successful, will go from state O to state O. When the call finally completes, the mobile enters the *completion state* (denoted by *C*). Let *P* be the state transition matrix, d_{ij} be the probability that a mobile enters C_j when it leaves C_i. Finally, $\frac{\gamma}{\mu+\gamma}$ and $\frac{\mu}{\mu+\gamma}$ are the probabilities that the mobile moves and that a call is successfully completed, respectively, given that there is a change of state.

$$P_{ij} = \begin{cases} \frac{\gamma}{\mu+\gamma} d_{ij} & j \notin \{O, C\} \text{ and } i \notin \{O, C\} \\ \frac{\mu}{\mu+\gamma} & j = C \text{ and } i \notin \{O, C\} \end{cases} \qquad P_{Oj} = \begin{cases} \frac{\gamma}{\mu+\gamma} & j = O \\ \frac{\mu}{\mu+\gamma} & j = C \\ 0 & otherwise \end{cases}$$

$$P_{Cj} = \begin{cases} 1 & j = C \\ 0 & otherwise \end{cases} \qquad p_i(m) = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \end{bmatrix} P^m \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow \text{i}^{\text{th}} \text{ location}$$

Finally, we compute the cell occupancy distribution, $\pi_j(t)$. In the absence of blocking and forced-termination, the cellular network would be a Jackson network with a queue for each cell. The cell occupancy distribution for each cell could then be determined independently assuming Poisson arrivals. In the actual system, arrivals are not exactly Poisson, and the arrival rate is lower because the admission control algorithm would block new call arrivals when the risk of forced-termination is great, leaving only handoffs. We approximate this system as follows. When the number of free channels exceeds threshold *A* known as the *adjusting factor*, then calls arrive according to a Poisson process whose rate is the sum of the rate from new arrivals and the rate from handoffs. When the number of free channels is less than *A*, then arrivals are Poisson with rate that equals the arrival rate from handoffs only. Note that this is similar to reserving a certain number of channels in the reservation scheme. The adjusting factor which can be selected experimentally should be set such that the probability index closely represents the actual forced-termination probability.

Both the steady-state and transient analyses are essential in computing the probability index. The transient analysis is used to compute the occupancy distributions $\Pi_i(t)$ of cells *i* that are within the region of awareness. The occupancy distributions of these cells depend on the number of calls underway in their cells at the time the new call arrives. However, since we do not know the number of calls underway in cells that lie outside of the region of awareness, we cannot use the transient analysis to compute the occupancy distribution. Instead, we use the steady-state distribution. In essence, cells that are outside the region of awareness are assumed to have reached the steady state by the time the marked mobile gets there.

At the steady state, the net arrival rate into a cell is given as follows:

$$\lambda_j = \begin{cases} \left(\frac{\gamma+\mu}{\mu}\right)\alpha & 0 \le j < N - A \\ \left(\frac{\gamma}{\mu}\right)\alpha & N - A \le j \le N \end{cases}$$

The net departure rate per call is $U = \gamma+\mu$ For a given cell, let π_j denote the probability of a cell carrying *j* calls at steady state, which can be computed as follows:

$$\pi_0 = \left\{ 1 + \sum_{j=1}^{N-A} \frac{\lambda_a^j}{j!U^j} + \sum_{j=N-A+1}^{N} \frac{\lambda_a^{(N-A)}\lambda_b^{(j-(N-A))}}{j!U^j} \right\}^{-1}$$

$$\pi_j = \begin{cases} \dfrac{\lambda_a^j}{j!U^j}\pi_0 & 0 \le j \le N-A \\ \dfrac{\lambda_a^{(N-A)}\lambda_b^{(j-(N-A))}}{j!U^j}\pi_0 & N-A < j \le N \end{cases} \qquad \text{where} \qquad \lambda_a = \left(\frac{\gamma+\mu}{\mu}\right)\alpha, \quad \lambda_b = \left(\frac{\gamma}{\mu}\right)\alpha$$

In computing the transient distribution, the net arrival rate into cell C_i (currently carrying j calls) is given as follows:

$$\lambda_j = \begin{cases} \alpha + \gamma \sum_{k:k \in S} d_{ki}N_k & 0 \le j < N-A \\ \gamma \sum_{k:k \in S} d_{ki}N_k & N-A \le j \le N \end{cases}$$

where S is the set containing cells that are adjacent to C_i. The net departure rate per call is $U = \gamma+\mu$. For a given cell C_i, let $\pi_j(t)$ denote the probability of a cell carrying j calls at time t and let $\Pi_i(t) = [\pi_0(t)\ \pi_1(t) \ldots\ \pi_N(t)]$. Furthermore, let $Q(t)$ denote the transition rate matrix. The transient analysis is achieved by solving

$\frac{d}{dt}\Pi_i(t) = \Pi_i(t)Q(t)$ [10] where

$$Q(t) = \begin{bmatrix} q_{00} & q_{01} & q_{02} & \cdots & q_{0k} & \cdots & q_{0N} \\ q_{10} & q_{11} & q_{12} & \cdots & q_{1k} & \cdots & q_{1N} \\ \vdots & & & & & & \vdots \\ q_{k0} & q_{k1} & q_{k2} & \cdots & q_{kk} & \cdots & q_{kN} \\ \vdots & & & & & & \vdots \\ q_{N0} & q_{N1} & q_{N2} & \cdots & q_{Nk} & \cdots & q_{NN} \end{bmatrix} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots & 0 & 0 \\ U & -\lambda_1 - U & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & 2U & -\lambda_2 - 2U & \lambda_2 & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & kU & -\lambda_k - kU & \lambda_k & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & (N-1)U & -\lambda_{N-1} - (N-1)U & \lambda_{N-1} \\ 0 & 0 & 0 & 0 & 0 & NU & -NU \end{bmatrix}$$

and q_{mn} is the transition rate from state m to state n (i.e., from having m calls to n calls underway in the cell). By solving the differential equation, we find that

$$\Pi_i(t) = \left[\pi_0(t)\ \pi_1(t) \ \cdots \ \pi_i(t) \ \cdots \ \pi_N(t)\right] = e^{Q(t)t}\Pi_i(0)$$

where $\Pi_i(0)$ is the initial distribution of calls in cell C_i and $e^{Q(t)t} = \sum_{k=0}^{\infty} \frac{1}{k!}t^k Q^k(t)$, which can be solved relatively easily by using a similarity transformation [11].

3.2.2 Calculation of Pf_{id}

We can now use $f_{Tm}(t)$, $p_i(m)$, and $\pi_i(t)$ to calculate the index. As defined earlier, Pf_{id} is the probability index, and $Pf_{id}(m)$ is the index reflecting the chance of a call being forced-terminated in its m^{th} hop, assuming that it was not terminated in previous hops. In addition, let H be the maximum number of hops taken into consideration, and *MAXSTATE* be the total number of cells in the region of awareness + an out-of-region state. Finally, let $P_{full}(C_i,N_i)$ be the probability that cell C_i is full when the mobile enters it, given that there are N_i calls underway in C_i at the time of new call arrival. When doing the transient analysis, we reduce the amount of computation by approximating $P_{full}(C_i,N_i)$ to be the probability $\int_0^{\infty} f_{Tm_i}(t)\pi_N(t)dt$ that cell C_i is full when the mobile makes its m_i^{th} handoff, where m_i is the minimum number of handoffs required for a marked mobile to get from C_0 to C_i. (The minimum number of handoffs required to get back to C_0 is 2.) Pf_{id} is computed as follows.

$$Pf_{id}(m) = \begin{cases} \displaystyle\sum_{i=0}^{MAXSTATE-1} P_{full}(C_i, N_i)p_i(m) & \text{when m = 1} \\ \displaystyle\left(\sum_{i=0}^{MAXSTATE-1} P_{full}(C_i, N_i)p_i(m)\right)\prod_{k=1}^{m-1}(1-Pf_{id}(k)) & \text{when m > 1} \end{cases}$$

$$Pf_{id} = \sum_{m=1}^{H} Pf_{id}(m)$$

A call should only be admitted when $Pf_{id} < Pf_{threshold}$.

3.2.3 Comparison between Pf_{id} and actual P_f

We now show the relationship between the probability index (Pf_{id}) and the actual forced-termination probability (P_f), as obtained by simulation, in which a call is blocked if and only if it arrives in a cell with no available channels, or its index > $Pf_{threshold}$ = 0.012. In this scenario, new calls arrive according to a Poisson process with rate α = 3.2 calls/min per cell. Both the call holding time and the mean time between handoffs are exponentially distributed with mean 5 minutes. Each cell can carry up to 20 simultaneous calls. Figure 1 shows the average forced-termination probability of calls that have a Pf_{id} in the range between 0.007 + (i-1) 0.0005 and 0.007 + i 0.0005, for every integer i from 1 to 10. These simulation results are accurate within ±2% with 95% confidence. Despite the simplifying assumptions used in the calculations of the index, the figure shows that the probability index is an excellent approximation of the actual forced-termination probability of a call.

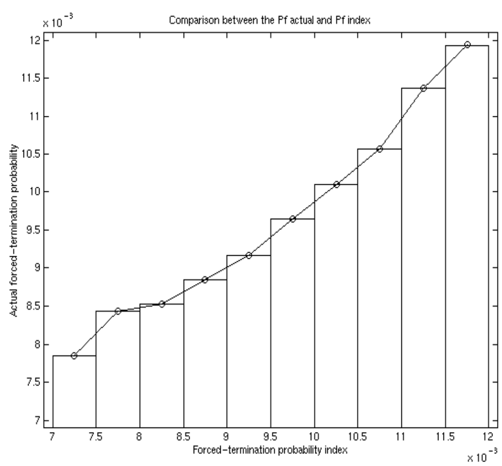

Figure 1. Comparison between P_f actual and P_f index

3.3 Hybrid Control Scheme (HCS)

The probability index scheme blocks calls that would face a high probability of forced-termination but does not consider how admitting a call might affect the forced-termination probability of those calls that are already underway. For example, the probability index scheme might allow the occupancy level in the originating cell to grow dangerously high, as long as calls beginning in that cell will not be terminated as they move to surrounding cells. Consequently, this approach can cause other calls to be forced-terminated when they move. The weighted sum scheme is better able to prevent such problems, but it does not explicitly consider the forced-termination probability of the new call. Consequently, there are reasons to consider a hybrid of the two schemes, which only admits calls when both criteria are met:

1) $N_{weighted} < N - N_h$

2) $Pf_{id} < Pf_{threshold}$

4.0 Performance Comparisons

All six algorithms presented in Sections 2 and 3 were compared via simulation. Two additional algorithms, which are derivatives of the distributed admission control system (DACS), were also included. Recall that DACS has two conditions that must be met for a call to be admitted. We have therefore included DACS1, which only enforces the condition designed to protect calls that are already underway, and DACS2, which enforces only the condition designed to protect the new call arrival. The simulation employs a 9x9 grid of hexagonal cells which are folded over onto the surface of a torus to avoid edge effects, allowing meaningful data to be collected over all cells, as described in [2]. Each cell can carry up to *N* simultaneous calls. New calls arrive according to a Poisson process with rate α per cell, call holding time is exponentially distributed with mean $1/\mu$, and the time between handoffs is also exponentially distributed with mean $1/\gamma$. Furthermore, we assume that the mobile device is equally likely to be handed off to each of the six neighboring cells during a handoff. Finally, we assume that the region of awareness contains at most seven cells (i.e., the originating cell and its six neighboring cells). This is to limit the extent that information has to be passed around between cells (and thus limit the usage of the scarce bandwidth) during the admission. In comparing the performance of the algorithms, we first compare these algorithms' ability to meet QoS requirements at a known load, and then explore the performance they can achieve when load differs from its expected level, as will inevitably be the case.

We are also interested in comparing the performance of the algorithms under different user *mobilities*. The mobility is the mean number of handoffs before a call completes (γ/μ) in the absence of forced-termination. We will examine two levels of mobilities--low mobility and high mobility. The low mobility case can be used to describe the mobility of users in typical macro-cellular systems. In macro-cellular systems, even though users can travel at a high speed (e.g. highway speed), there will be fewer handoffs because cell size is typically larger. For example, if the call holding time is 5 minutes and the user travels at 60 mph in a cell whose diameter is 5 miles, there will be, on average, 1 handoff during the connection life-time. The high mobility case, on the other hand, can be used to describe the mobility of users in typical micro-cellular systems like [1]. For example, if the call holding time is 5 minutes and the user travels at 3 mph in a cell whose diameter is 0.05 miles, there will be, on average, 5 handoffs during the connection life-time.

4.1 Scenario 1 (Known Load)

This scenario reflects the case where load is predictable: $N = 20$ channels, system load is 70% (14 Erlangs) when mobility = 5, and load is 80% (16 Erlangs) when mobility = 1. Note that when user mobility is low, the average forced-termination probability will be small, so comparable performance is achieved at the higher load [12]. Algorithms can be evaluated by comparing their *feasible regions* [13]. Each axis in a feasible region represents one dimension of QoS metrics, which in this case, are blocking probability (P_b) and forced-termination probability (P_f). A point (P_f, P_b) is within the feasible region if that QoS or better can be achieved. We make comparisons by showing the lower bound of the feasible region for each algorithm, so a lower curve is better.

Before making these comparisons, we must determine appropriate parameters for the various algorithms. We first determine the most effective weights in the weighted sum scheme. Since the region of awareness consists only of the originating cell and adjacent cells, $p_i = 0$ for i > 1. Without loss of generality, let $p_0 = p$ and $p_1 = 1-p$. Figures 2 and 3 show the feasible regions of the weighted sum scheme under different values of weight *p* for the high mobility and low mobility cases, respectively. Figures 2 and 3 demonstrate that no single value of *p* is optimal under all circumstances, e.g., no value of *p* produces a feasible region that subsumes all other feasible regions. However, over a wide range of *p* values near the optimum (0.4-1.0 for high mobility and 0.6-1.0 for low mobility), the weighted sum scheme achieves roughly the same results, so it is not necessary to find the exact optimal value. Based on Figure 2 (high mobility case), since the lower bound of the corresponding feasible region of $p = 0.5$ is below other curves most of the time, we chose $p = 0.5$ to be the optimal weight for the weighted sum scheme. Similarly, based on Figure 3 (low mobility case), $p = 0.7$ is chosen to be the optimal weight.

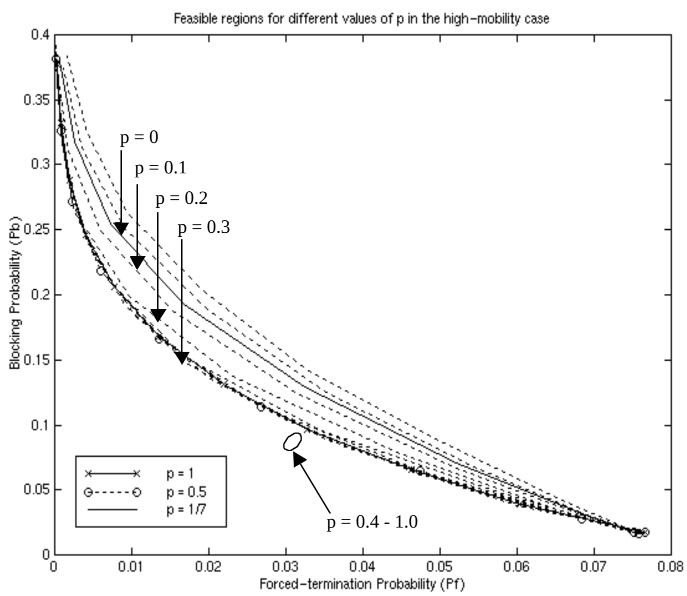

Figure 2. Comparison of feasible regions for different values of *p* in the high-mobility case

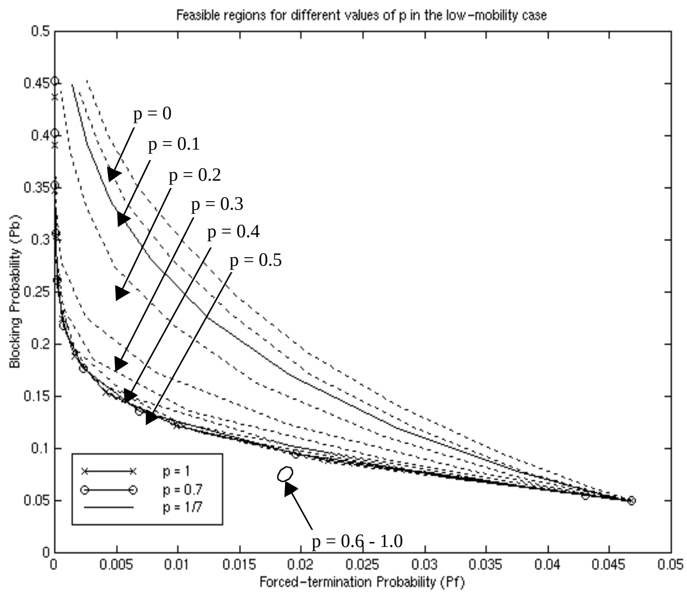

Figure 3. Comparison of feasible regions for different values of *p* in the low-mobility case

Like the weighted sum scheme, the distributed admission control scheme also requires the parameter (*T*) to be set experimentally. Figure 4 shows the feasible regions for various values of *T*. Although no value of *T* is optimal in all respects, in this scenario, T=1/10γ seems to be an effective choice.

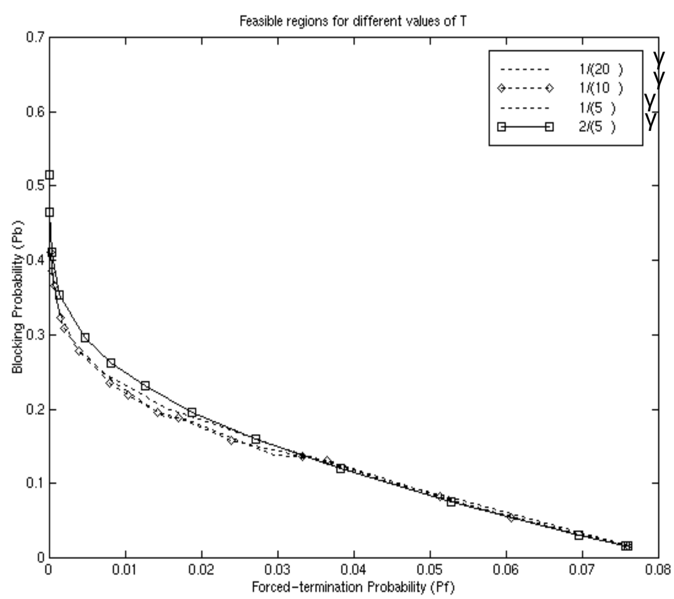

Figure 4. Comparison of feasible regions for different values of *T*

Figures 5 and 6 show the feasible regions of all the algorithms in the high mobility case and low mobility case, respectively. The performance of the reservation scheme, the weighted sum scheme, the hybrid control scheme, and DACS1 are comparable and all perform well compared to the rest. DACS1 seems to perform slightly better than the other three. However, given the simplicity and effectiveness of the reservation scheme, when the load is known a priori with confidence, it may be desirable to use the reservation scheme. As shown earlier, the performance of the weighted sum scheme suffers when *p* is poorly chosen, as is the case with the linear weighting scheme.

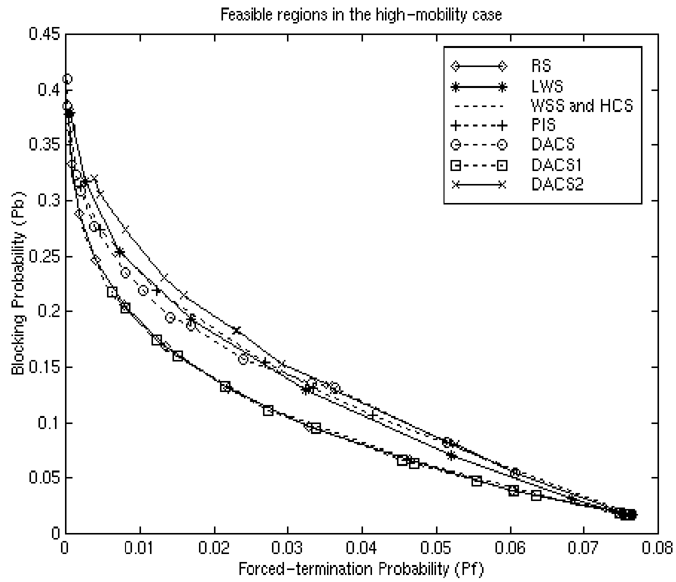

Figure 5. Comparison of feasible regions in the high-mobility case

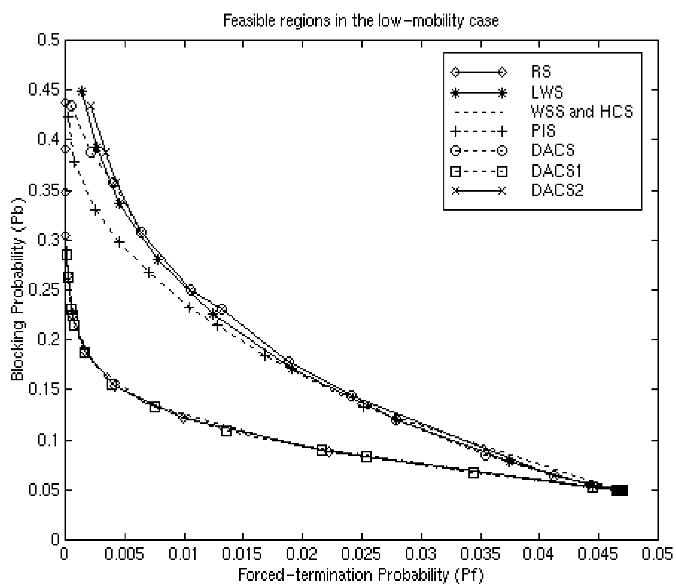

Figure 6. Comparison of feasible regions in the low-mobility case

Note that the hybrid control scheme results in the same feasible region as the weighted sum scheme. This is because when the load is known, the optimal admission decision under a hybrid control scheme is to set the threshold such that calls with any probability index values are admitted. This indicates that using the probability index scheme is not helpful when the load is known in advance with high confidence.

4.2 Scenario 2 (Varying Load)

This scenario reflects the situation in which the load is not known in advance. The network is designed to operate well at a given *nominal load*, but load can vary from this expected level. As shown earlier, the relative performance of the algorithms in scenario 1 does not change significantly under different mobilities. Consequently, we will focus on the high-mobility case, which is applicable to the emerging micro-cellular networks. We will look at performance as a function of load. Since there are two QoS metrics, in order to unambiguously compare the performance of the algorithms, we need to find a single metric with which all the algorithms can be compared. We define a *benefit per cell* (*BPC*) to be load*{(the probability that a call will complete successfully, i.e., without being blocked or terminated prematurely) - (the probability that a call is admitted but is terminated prematurely)*K} = $(\alpha/\mu)*\{(1-P_b)(1-P_f) - (1-P_b)(P_f)*K\}$, where K: $K > 0$ indicates how strongly users will object to experiencing unwanted termination of their calls.

We first find the admission parameter(s) for each algorithm to maximize *BPC* at the nominal load, with the exception of the hybrid control scheme [10]. For the hybrid control scheme, we use the same optimal weight p and N_h used in the weighted sum scheme, and we select a $Pf_{threshold}$ such that the forced-termination probability is limited when the load increases beyond expected, but the performance at low load is close to that of the weighted sum scheme. Table 1 shows the parameter(s) for each algorithm.

Table 1: Optimal admission threshold(s) required by each algorithm when Load = 70% and K = 5

Admission Control Scheme	Admission Threshold(s)
RS	$N_h = 5$
LWS	$N_h = 7$
DACS	$a = 1.9$ and $T = 1/(10\gamma)$
DACS1	$a = 4.0$ and $T = 1/(10\gamma)$
DACS2	$a = 2.0$ and $T = 1/(10\gamma)$
PIS	$Pf_{threshold} = 0.0255$ and $A = 15$
WSS	$p = 0.5$ and $N_h = 6$
HCS	$p = 0.5$, $N_h = 6$, $Pf_{threshold} = 0.035$, and $A = 14$

Figure 7 shows performance as a function of load with these parameters held constant. Compared to the reservation scheme, all the algorithms with the exception of the DACS2 have lower blocking probabilities when load is significantly lower than expected (less than 11 Erlangs or 55% utilization), and have lower forced-termination probabilities when load is significantly higher than expected (greater than 16 Erlangs or 80% utilization). This is desirable from both the service provider’s and the users’ perspective [7]. From the service provider’s perspective, when the load is low, more calls can be admitted while maintaining a reasonable forced-termination probability, which in turn results in a higher total revenue. From the users’ perspective, when the load is high, users want to ensure that their calls are not forced-terminated. This can be achieved by blocking new call arrivals, which in turn results in a higher blocking probability.
More interestingly, the forced-termination probability curves of the probability index scheme and the hybrid control scheme are relatively flat, especially at high loads. Clearly, this is desirable since this implies that the forced-termination probability is relatively insensitive to load. This helps the service provider determine and guarantee the maximum forced-termination probability of the system, independent of actual load. In addition, the hybrid control scheme has comparable blocking and forced-termination probabilities to those of the weighted sum scheme when load is light. However, when load increases beyond the nominal load, the hybrid control scheme has a much smaller forced-termination probability, but only a slightly higher blocking probability. This is a result of running the probability index scheme on top of the weighted sum scheme. The probability index scheme does not significantly affect the weighted sum scheme until the load is higher than expected, but tends to flatten out the forced-termination probability for the hybrid control scheme at higher loads, which is a significant advantage.

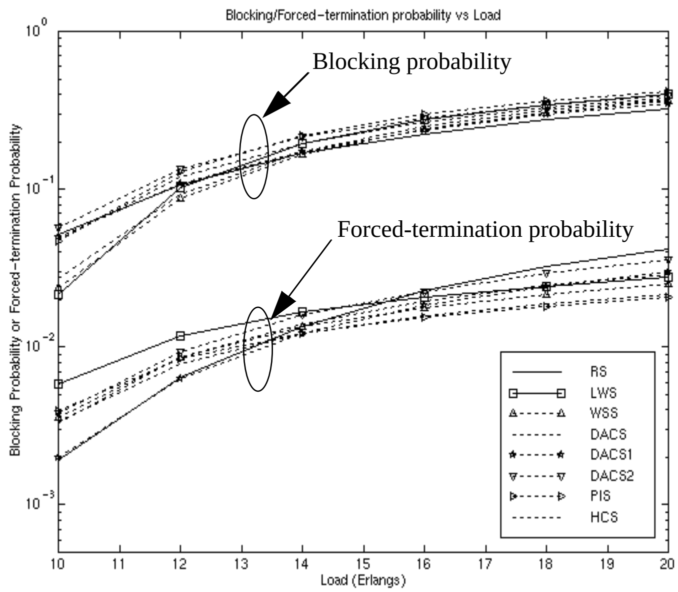

Figure 7. Comparison of block probability and forced-termination probability as load varies.

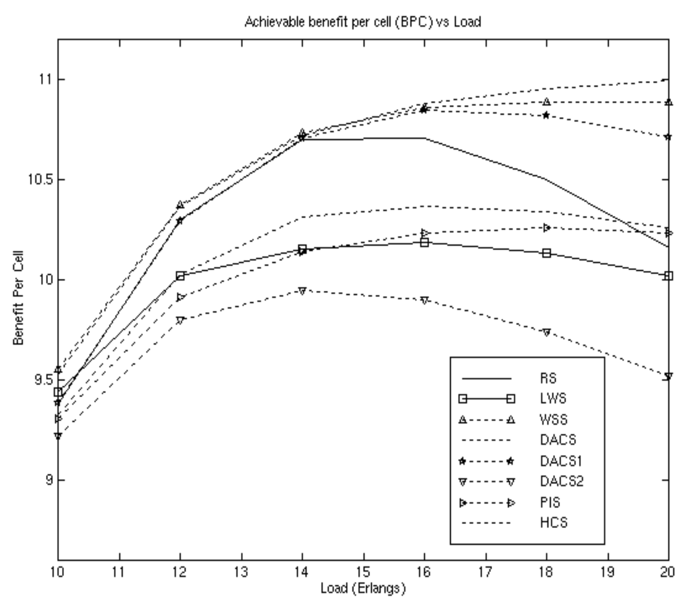

Figure 8. Comparison of the achievable BPC as load varies

5.0 Conclusion

We presented three new heuristic admission control algorithms for a cellular system carrying a single class of traffic: the *Weighted Sum Scheme*, the *Probability Index Scheme*, and the *Hybrid Control Scheme*. In the process of developing the probability index scheme, we have developed a way to calculate an index that reasonably reflects forced-termination probability. This may be especially useful when there is heterogeneous traffic, such that one class can tolerate a greater forced-termination probability than another, and can therefore have a different index threshold, $Pf_{threshold}$. Recent allocations of unlicensed spectrum [14-16] make it increasingly likely that microcellular systems such as [1] will emerge that support diverse applications, ranging from wearable computers to videoconferencing.

We then compared the performance of the novel algorithms with the three known algorithms: the *Reservation Scheme*, the *Linear Weighting Scheme*, and the *Distributed Admission Control Scheme*. An ideal algorithm would be easy to implement, it would be effective at meeting QoS requirements at a given load (e.g., feasible regions), and it would react well when load varies from the expected value.

When the load is known in advance with confidence, four algorithms achieve comparable performance: the reservation scheme, the weighted sum scheme, the hybrid control scheme, and DACS1. Although DACS1 can slightly outperform the reservation scheme, none of the algorithms can *significantly* outperform the reservation scheme, which is the simplest since it does not require knowledge about the states of neighboring cells. Similar results were found when applying a genetic admission control algorithm in non-Markovian systems, where algorithms with small regions of awareness performed almost as well as systems with larger regions of awareness [17]. This implies that the simplest algorithm is surprisingly useful, allowing networks to decrease the usage of bandwidth to exchange information from cell to cell.

In real systems, load will vary from the expected level. When this occurs, the hybrid control scheme yields the best performance. This makes the hybrid algorithm the best choice for improving performance. However, this algorithm is more complicated than the reservation scheme, so there is a design trade-off here. The weighted sum scheme may be a useful compromise, since it is simpler than the hybrid, and is almost as effective.

We also showed that incorrect use of information on the number of calls underway in the neighboring cells can result in a poor overall performance. For example, with the linear weighting scheme, we found that considering the current utilization of cells other than the originating cell actually degrades performance. This also implies that there may be motivation to reexamine the choice of admission control scheme used in [7, 8]. There may also be motivation to reexamine condition 2 of the distributed admission control scheme (DACS) [9]. Based on our results, condition 2 degrades the overall performance of DACS. Perhaps condition 2 should be altered or dropped, or perhaps this condition simply does not extend well to a two-dimensional cellular system.

References

[1] A. Hills and D. B. Johnson, “Seamless Access To Multiple Wireless Data Networks. A Wireless Data Network Infrastructure At Carnegie Mellon University,” *IEEE Personal Communications*, vol. 3, no. 1, February 1996, pp. 56-63.
[2] H. Salgado, M. Sirbu, and J. M. Peha, “A Narrow Band Approach to Efficient PCS Spectrum Sharing Through Decentralized DCA Access Policies,” *IEEE Personal Communications,* vol. 4, no. 1, February 1997, pp. 24-34.
[3] A. Aljadhai and T. F. Znati, “A Predictive Bandwidth Allocation Scheme for Multimedia Wireless Networks,” *Proceedings of IEEE/ACM/SCS Conference on Communication Networks and Distributed Systems Modeling and Simulation*, Phoenix, January 1997, pp. 95-100.
[4] D. A. Levine, I. F. Akyildiz, and M. Naghshineh, “A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept,” *IEEE/ACM Transactions on Networking*, vol. 5, no.1, February 1997, pp. 1-12.
[5] E. C. Posner and R. Guerin, “Traffic Policies in Cellular Radio that Minimize Blocking of Handoff Calls,” *Proceedings of ITC-11*, Kyoto, Japan, September 1985, pp. 153-163.
[6] D. Hong and S. Rappaport, “Traffic Modelling and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoffs Procedures,” *IEEE Transactions on Vehicular Technology*, vol. 35, no. 3, August 1986, pp. 77-92.
[7] A. S. Acampora and M. Naghshineh, “An Architecture and Methodology for Mobile-Executed Handoff in Cellular ATM networks,” *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 8, October 1994, pp. 1365-1374.
[8] A. S. Acampora and M. Naghshineh, “Control and Quality-of-Service Provisioning in High-Speed Microcellular Networks,” *IEEE Personal Communications*, Second Quarter 1994, pp. 36-43.
[9] M. Naghshineh and M. Schwartz, “Distributed Call Admission Control in Mobile/Wireless Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 4, May 1994, pp. 711-717.
[10] A. Sutivong, “Call Admission Control Algorithms for a Cellular System: Proposal and Comparison,” Masters Thesis, Carnegie Mellon University, Pennsylvania, 1996.
[11] C. Chen, *Linear System Theory and Design*, Harcourt Brace College Publishers, Orlando, 1984.
[12] S. Rappaport, “The Multiple-call Hand-off Problem in High Capacity Cellular Communications Systems,” *IEEE Transactions on Vehicular Technology*, vol. 40, no. 3, August 1991, pp. 546-557.
[13] J. M. Peha and F. A. Tobagi, “Cost-Based Scheduling and Dropping Algorithms to Support Integrated Services,” *IEEE Transactions on Communications,* vol. 44, no. 2, February 1996, pp. 192-202.
[14] J. M. Peha, “Spectrum Management Policy Options,” *IEEE Communications Surveys*, vol. 1, no. 1, Fourth Quarter 1998.
[15] D. P. Satapathy and J. M. Peha, “Spectrum Sharing Without Licensing: Opportunities and Dangers,” in *Interconnection and the Internet: Selected Papers From the 1996 Telecommunications Policy Research Conference*, G. L. Rosston and D. Waterman editors, Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 1997, pp. 49-75.
[16] D. P. Satapathy and J. M. Peha, “Etiquette Modifications For Unlicensed Spectrum: Approach and Impact,” *Proc. IEEE Vehicular Technology Conference*, May 1998, pp. 272-6.
[17] A. Yener and C. Rose, “Genetic Algorithms Applied to Cellular Call Admission: Local Policies,” *IEEE Transactions on Vehicular Technology*, vol. 49, no. 1, February 1997, pp. 72-79.