State-Dependent Call Admission Control in Hierarchical Wireless Multiservice Networks

Shun-Ping Chung and Jin-Chang Lee

Abstract: **State-dependent call admission control (SDCAC) is proposed to make efficient use of scarce wireless resource in a hierarchical wireless network with heterogeneous traffic. With SDCAC, new calls are accepted according to an acceptance probability taking account of not only cell dwell time but also call holding time and system state (i.e., occupied bandwidth). An analytical method is developed to calculate performance measures of interest, e.g., new call blocking probability, forced termination probability, overall weighted blocking probability. Numerical results with not only stationary but nonstationary traffic loads are presented to show the robustness of SDCAC. It is shown that SDCAC performs much better than the other considered schemes under nonstationary traffic load.**

Index Terms: **Average call holding time, hierarchical network, overflowed traffic, overall weighted blocking probability.**

I. INTRODUCTION

The ever growing demand of the wireless/mobile communication has made the system capacity insufficiency worse. As a remedy, hierarchical wireless networks are proposed not only to provide higher capacity with small cells, but also to reduce handoff request load with large cells [1]–[3]. Also, UMTS/IMT2000 is expected to support four classes of services: Conversational, streaming, interactive, and background. Thus, it is needed to design a call admission control that maximizes the network utilization while supporting various services for multiservice wireless networks [4], [5].

Generally speaking, the forced termination of calls in progress is less desirable than the blocking of new calls. This fact has led to the development of various handoff prioritization schemes [6]–[8]. Basically, there are two schemes to prioritize handoff calls over new calls. The first scheme is to use handoff queue [7]. With handoff queue, upon arrival, if a handoff call can not be granted an idle channel by the serving base station, it will be put into a handoff queue to wait for the release of a channel. The second scheme is to use trunk reservation (TR) or guard channel [8]. With TR, some channels are reserved exclusively for handoff calls. It can reduce handoff blocking probability at the expense of new call blocking probability, and thus reduce the channel utilization. In this paper, we focus on the guard channel scheme. As is well known, a call with higher mobility and thus lower cell dwell time on average leads to more handoff requests during its lifetime [6], [7]. On the other hand, a call with longer call holding time on average also leads to more handoff requests during its lifetime. This may result in serious impact on the performance of wireless networks. As pointed out in [9], the average call holding time of voice calls is 3 min, while that of data calls, e.g., Internet browsing, is in the range of 20 min. Thus, it is worth taking the call holding time difference as well as cell dwell time into account for designing a call admission control (CAC) scheme.

A novel state-dependent call admission control (SDCAC) is proposed for the hierarchical cellular network with heterogeneous traffic taking account of call holding time differences among different services. Furthermore, SDCAC only allows the overflow of calls from microcells to macrocells. The goal is not only to maintain the forced termination probability at an acceptable level but also to reduce the new call blocking probability and thus improve the channel utilization. Importantly, the work here differs from [10] in several aspects. Firstly, the considered system is a two-layer hierarchical wireless network. Secondly, instead of the single-rate service, multi-rate conversational services and variable-bit-rate streaming services are taken into account. Thirdly, the handoff rate of each service is not assumed to be known, but obtained via an iterative algorithm. Furthermore, another three well-known CAC's are also studied for comparison with SDCAC: (1) Complete sharing (CS), (2) trunk reservation (TR), and (3) fixed acceptance probability (FAP). CS allows all classes of new and handoff calls equal access to bandwidth available at all times. With TR, some time slots are reserved for exclusive use of handoff calls. FAP is similar to SDCAC except that the acceptance probability is set to be a constant value, which is chosen to be 0.5 throughout the paper.

On the other hand, the new call arrival rate may change during the day, and the handoff arrival rates depend on both the number and movements of mobile terminals in adjacent cells. Thus, actual traffic and specifically the call arrival rate are seldom stationary, so that network performance analysis under nonstationary traffic load is also important in its own right [11], [12]. With those considered CAC schemes, the performance measures under nonstationary traffic are also conducted via computer simulation in the following. As shown in [4], by allowing takeback from macrocells to microcells of overflowed new and handoff calls to avoid overflowed traffic occupying excessive resources of the macrocells, the blocking probabilities in a hierarchical wireless network with single rate service can be reduced significantly. Therefore, we also consider SDCAC with takeback which incorporates the takeback mechanism into the proposed CAC.

Manuscript received August 13, 2004; approved for publication by Fotini-Niovi Pavlidou, Division II Editor, July 19, 2005.

S.-P. Chung is with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taiwan, R.O.C., email: chung@eniac2.ee.ntust.edu.tw.

J.-C. Lee is with the Chunghwa Telecommunication Laboratories, Taiwan, R.O.C., email: jimlee@cht.com.tw.

This work was supported in part by the National Science Council, Taiwan, R.O.C., under Grant NSC-92-2213-E-011-099.

1229-2370/06/$10.00 © 2006 KICS

Table 1. Some parameters for the focused system.

Notation	Definition
$C_1(C_2)$	Capacity (in slots) of the microcell (macrocell)
$G_1(G_2)$	Reserved time slots in the microcell (macrocell)
λ_{nk}	New call arrival rate of class-k calls
$\lambda_{hk}^{(1)}$	Handoff call arrival rate of class-k conversational calls in the microcell
$\lambda_{hk}^{(2)}$	Handoff call arrival rate of class-k calls in the macrocell
b_k	Time slot requirement of class-k conversational calls
$b_P(b_{K+1})$	Maximum (minimum) time slot requirement of streaming calls
b_A	Average number of assigned time slots of streaming calls
μ_k	Call service rate of class-k calls
$\eta_k^{(1)}$	Cell dwell rate of class-k conversational calls in the microcell
$\eta_k^{(2)}$	Cell dwell rate of class-k calls in the macrocell
n_k	The number of ongoing class-k calls in the focused cell
n	The total number of time slots used by conversational calls in the microcell
$n_{\min}$	The total number of minimum time slots used by overflowed conversational and streaming calls in the macrocell
$A_k^{(1)}(n)$	Acceptance probability of class-k new calls in the microcell
$A_k^{(2)}(n_{\min})$	Acceptance probability of class-k new calls in the macrocell

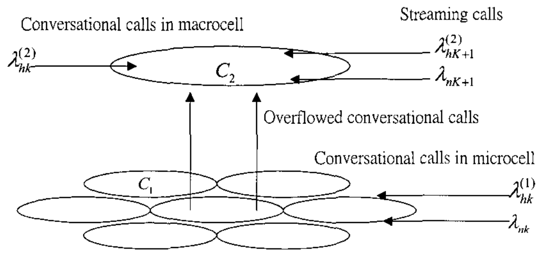


Fig. 1. Model for the focused system.

The remainder of the paper is organized as follows. In Section II, the system model for the considered two-layer network is described in detail. In Section III, the analytical method to compute the performance measures of interest is derived. Both analytical results and simulation results are shown to demonstrate the robustness of SDCAC compared to the other well-known schemes in Section IV. Section V concludes the paper.

II. SYSTEM MODEL

We consider a two-layer cellular network with multiple classes of conversational and streaming services, where every N microcells are overlaid by a macrocell. The conversational class includes non-adaptive multimedia applications with strict bandwidth requirements or non-multimedia applications, where the bandwidth of an ongoing call must be kept constant during its lifetime [5]. The streaming class includes variable bit rate services, such as adaptive multimedia applications, where the bandwidth of an ongoing call could be time-varying during its lifetime [13], [14]. For simplicity, we consider a homogeneous cellular network so that we can focus on one particular cell in each layer [1]–[3]. We also assume that conversational calls are associated with shorter average call holding time, and streaming calls with longer average call holding time. For reducing handoff request load of streaming service with longer average call holding time, streaming calls are always assigned to the associated macrocell, while conversational calls are initially assigned to the associated microcell. Importantly, taking into account the characteristics of the streaming traffic, the bandwidth reallocation of existing streaming calls is allowed [13], [14]. A conversational call is assigned a fixed number of time slots periodically during its lifetime, and a streaming call is assigned time slots dynamically based on system state under the constraint that its allocated bandwidth is between its minimum and maximum time slot requirements during its lifetime. *Bandwidth reallocation* of existing streaming calls is allowed for accommodating more overflowed conversational calls or streaming calls by the macrocell under the constraint that no existing calls are blocked [13]. It should be noted that the bandwidth reallocation process only occurs in the macrocell and not in the microcell, since streaming traffic with longer average call holding time is supposed to be always handled by the macrocell. For fairness purposes, the remaining time slots of a macrocell are equally shared among the streaming calls. The number of time slots assigned to streaming calls in progress is assumed to change in a memoryless fashion, i.e., without considering how many time slots were allocated to them previously [14]. Naturally, the number of assigned time slots of any streaming call should range from b_{K+1} to b_p during its lifetime. In particular, the bandwidth reallocation of the existing streaming calls may occur at any new call arrival, call handoff, and call departure if necessary.

Assume K classes of conversational calls and one class of streaming calls, and streaming service is taken as the class-$(K+1)$ service. It is also assumed that all new call arrival and the handoff call arrival processes follow a Poisson distribution, and both the dwell time and the call holding time are exponentially distributed [1]–[3]. Some parameters used are defined in Table 1. The corresponding system model is shown in Fig. 1.

n_k is defined as the number of ongoing class-k calls in the focused cell. To show the capability of carrying more calls via bandwidth reallocation of existing streaming calls, the total number of minimum used time slots in the macrocell $n_{\min}$ is defined and computed with conversational calls being allocated their constant bandwidth requirement and streaming calls their minimum bandwidth requirement, that is, $n_{\min} = \sum_{k=1}^{K+1} b_k n_k$. Further, n is used to denote the number of time slots in use in the microcell, i.e., $n = \sum_{k=1}^{K} b_k n_k$.

The traffic handling policy of SDCAC is described as follows. First, focus on new calls. One of the key ideas of SDCAC is that, the guard channels, originally reserved for handoff calls, can be allocated to new calls based on some acceptance probabilities. Based on the associated acceptance probability, it is determined whether a new conversational (streaming) call will be served by the originating microcell (macrocell). If a new conversational call is rejected in the originating microcell, whether that call can be accepted by the overlaid macrocell is determined based on another acceptance probability. Of course, if an overflowed conversational call is accepted, the time slots assigned to existing streaming calls may be reallocated to accommodate the incoming conversational call if necessary.

After extensive numerical study and with reference to [10], taking account of call holding time, cell dwell time, and occupied bandwidth, the acceptance probability of class-k new calls is defined as $A_k^{(j)}(x) = \max\{0,\ f_k(x)\}$, $j = 1$ or 2, where

$$f_k(x)=\begin{cases} 1, & x \le (C_j - G_j) \\ \frac{\eta_k^{(j)}}{\mu_k}\left[\frac{(C_j-b_k+1)-x}{(G_j-b_k+1)}\right] & \\ +\left(1-\frac{\eta_k^{(j)}}{\mu_k}\right)\left[\cos\frac{2\pi(x+G_j-C_j)}{4(G_j-b_k+1)}\right]^{\frac{1}{2}}, & \text{otherwise} \end{cases} \quad (1)$$

where $x = n(n_{\min})$ if $j = 1(2)$. More specifically, if $\eta_k^{(j)} > \mu_k$, it implies that on average more than one handoff event would occur during one call holding time, $A_k^{(j)}(x)$ can be chosen to make more room for handoff calls, such that lower forced termination probability can be achieved. If $\eta_k^{(j)} = \mu_k$, $A_k^{(j)}(x)$ can be chosen to decrease linearly with the number of available channels. On the other hand, if $\eta_k^{(j)} < \mu_k$, it implies that on average less than one handoff event would occur during one call holding time, $A_k^{(j)}(x)$ can be chosen to allow a higher probability in accepting new calls for achieving lower new call blocking probability, while maintaining an acceptable forced termination probability for handoff calls. Importantly, it is worth mentioning that, as shown in [16], the handoff arrival rate can be approximated as $\lambda_{hk}^{(j)} \cong \frac{\eta_k^{(j)}}{\mu_k}\lambda_{nk}$ with small values of new call blocking and handoff blocking probabilities. Obviously, handoff arrival rate depends strongly on $\eta_k^{(j)}/\mu_k$. Because the handoff rate of each service is not assumed to be known, but obtained via an iterative algorithm. Thus, instead of defining the mobility parameter as the ratio of the handoff call arrival rate to the new call arrival rate $\lambda_{nk}/\lambda_{hk}^{(j)}$ as in [10], it is chosen as the ratio of the dwell time to the call holding time, i.e., $\eta_k^{(j)}/\mu_k$. Also, in terms of the boundary L and the area S of that considered cell, and mobile moving speed V, the mean of dwell rate can be described as $\eta_k^{(j)} = \frac{VL}{\pi S}$ [16]. Thus, the occurrence frequency of handoff events of each service can be characterized in SDCAC. On the other hand, it is noted that the associated parameters of $A_k^{(j)}(x)$ can be obtained empirically as in [15], where a variable threshold dynamic channel allocation applicable to wireless systems is presented. Specifically, the parameters in [15] are set based on traffic patterns of a particular cell under consideration, and a feedback loop is incorporated to adjust for unusual circumstances. To summarize, the key ideas of $A_k^{(j)}(x)$ in SDCAC is that, the guard channels, reserved preferentially for handoff calls, can be suitably allocated to new calls based on the system state and the characteristics of traffic in order to achieve better utilization of the wireless channels.

Handoff conversational calls in the microcell are accepted as long as there are enough idle time slots. Specifically, if the target microcell capacity C_1 minus n exceeds its constant time slot requirement, the handoff request of conversational call in microcell will be accepted. Moreover, an unsuccessful handoff conversational call in the microcell can overflow to the associated macrocell for a second chance to find required resource. If the handoff request can acquire enough bandwidth resource from that overlaid macrocell, it will be served to continue its communication; otherwise, it will be forced into termination. A handoff conversational (streaming) call will be accepted as long as the number of idle time slots is no less than its constant (minimum) time slot requirement. Specifically, if the target macrocell capacity C_2 minus $n_{\min}$ exceeds its constant (minimum) time slot requirement, the handoff request of conversational (streaming) call will be accepted. It is noted that the time slots assigned to existing streaming calls may be reallocated, if necessary, whenever a new arrival, handoff, or departure event occurs in the macrocell in question.

It is noted that the proposed policy can be flexibly extended by allowing the takeback process of conversational calls from macrocells to microcells with available resources as in [4] for obtaining better performance. In fact, it is taken into account in the nonstationary scenarios in this paper.

III. ANALYTICAL METHOD

In this section, with reference to [2], [5], [13], [14], and [17], an analytical method for computing the performance measures of interest in a two-layer network with conversational services, and streaming services are derived. The performance measures of interest are new call blocking probabilities P_{bk} and forced termination probability P_{fk} of conversational services, new call blocking probabilities P_{bK+1}, forced termination probability P_{fK+1} and average number of assigned time slots b_A of streaming services, weighted new call blocking probability P_{wb}, weighted forced termination probability P_{wf}, and overall weighted blocking probability P_B. The analyses of microcell layer, overflowed traffic, and macrocell layer are described in turn below.

A. Analysis of Microcell Layer

Consider a microcell with K classes of conversational calls. New and handoff class-k conversational calls form Poisson processes with average arrival rate λ_{nk} and $\lambda_{hk}^{(1)}$, respectively. Therefore, the focused microcell can be modeled as a K-dimensional birth-death process with state vector $\vec{n} = (n_1, n_2, \cdots, n_K)$, where n_k is the number of ongoing class-k conversational calls in the microcell, $k = 1, 2, \cdots, K$. Let $\vec{n}_k^- = (n_1, \cdots, n_k - 1, \cdots, n_K)$ and $\vec{n}_k^+ = (n_1, \cdots, n_k + 1, \cdots, n_K)$. The feasible state space is $S_1 = \left\{\vec{n} : \sum_{l=1}^{K} b_l n_l \le C_1\right\}$. Thus, the arrival rate of class-k conversational calls in state $\vec{n}$ is $A_k^{(1)}(n)\,\lambda_{nk} + \lambda_{hk}^{(1)}$, $n = \sum_{k=1}^{K} b_k n_k$. The call holding time, i.e., the unencumbered call durations, is exponentially distributed with mean $1/\mu_k$. Specifically, it is the amount of time that a call would last if it could successfully complete without forced termination. The dwell time, i.e., the time spent by a mobile station in a cell, is also assumed to follow an exponential distribution with mean $1/\eta_k^{(1)}$. It is noted that the channel holding time, i.e., the time that the required channel is assigned to a call, is also exponentially distributed with mean $1/(\mu_k + \eta_k^{(1)})$, since the channel holding time is the minimum of the call holding time and the dwell time, both of which are exponentially distributed. It is easy to show that the handoff probability for class-k conversational calls in a microcell is $P_{hk}^{(1)} = \frac{\eta_k^{(1)}}{\mu_k + \eta_k^{(1)}}$. Denote the steady state probability distribution as $\pi(n)$. With the reasonable assumption that the incoming flow and outgoing flow of the focused microcell are balanced, the arrival rate of handoff class-k conversational call can be derived as

$$\lambda_{hk}^{(1)} = \left(1 - P_{hbk}^{(1)}\right) P_{hk}^{(1)} \lambda_{hk}^{(1)} + \left(\sum_{n=0}^{C_1} A_k^{(1)}(n)\,\pi(n)\right) P_{hk}^{(1)} \lambda_{nk} \tag{2}$$

where $P_{bk}^{(1)}$ and $P_{hbk}^{(1)}$ are the new and handoff blocking probabilities for class-k conversational calls in the microcell, respectively. After defining the needed parameters, the state transition diagram can be easily developed. The state transition diagram for an example cell with two classes of conversational calls is shown in Fig. 2.

According to the state transition diagram of our model, the associated equilibrium state equations can be easily formulated. By those equations, we can obtain the steady state probability distribution $\pi(n)$, where n is the number of used time slots in the focused microcell, and thus $P_{bk}^{(1)}$ and $P_{hbk}^{(1)}$, via an iterative algorithm as follows.

Iterative Algorithm:

1. Select a set of initial values for $\pi^{old}(\vec{n})$, and $\lambda_{hk}^{(1)}$, $k = 1, 2, \cdots, K$.
2. Compute $\pi^{new}(\vec{n})$ by solving the state equations of the focused system iteratively via Gauss-Seidel algorithm [20].
3. Compute $P_{bk}^{(1)}$ and $P_{hbk}^{(1)}$, using
   $P_{bk}^{(1)} = \sum_{n=C_1-G_1}^{C_1} \left[1 - A_k^{(1)}(n)\right] \pi(n)$ and
   $P_{hbk}^{(1)} = \sum_{n=C_1-b_k+1}^{C_1} \pi(n)$
   where $n = \sum_{k=1}^{K} b_k n_k$ for any $\vec{n} \in S_1$.
4. Compute $\lambda_{hk}^{(1)}$, $k = 1, 2, \cdots, K$, using
   $\lambda_{hk}^{(1)} = \left(1 - P_{hbk}^{(1)}\right) P_{hk}^{(1)} \lambda_{hk}^{(1)} + \left(\sum_{n=0}^{C_1} A_k^{(1)}(n)\,\pi^{new}(n)\right) P_{hk}^{(1)} \lambda_{nk}$
   where $P_{hk}^{(1)} = \frac{\eta_k^{(1)}}{\mu_k + \eta_k^{(1)}}$.
5. If $\left|\pi^{new}(\vec{n}) - \pi^{old}(\vec{n})\right| < \varepsilon$, for any $\vec{n} \in S_1$, terminate the algorithm, where ε is the stopping criterion. Otherwise, set $\pi^{old}(\vec{n}) = \pi^{new}(\vec{n})$, and go to step 2.

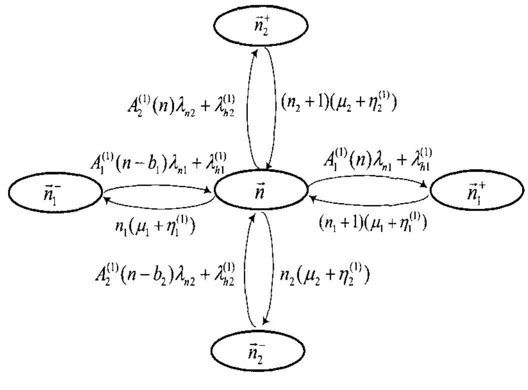


Fig. 2. State transition diagram for the considered cell.

Once $P_{hbk}^{(1)}$ is obtained, forced termination probability $P_{fk}^{(1)}$ of handoff class-k conversational calls in microcells can be computed as follows.

$$P_{fk}^{(1)} = \frac{P_{hk}^{(1)} P_{hbk}^{(1)}}{\left[1 - P_{hk}^{(1)}(1 - P_{hbk}^{(1)})\right]}. \tag{3}$$

B. Analysis of Overflowed Traffic

In this subsection, we take the mathematical modeling problem of overflowed traffic into account. The overflowed traffic consists of new and handoff conversational calls which can not be served by the corresponding microcell. The overflowed traffic model proposed in [2] is generalized in this subsection. An independent IPP process is used to model the overflowed traffic of class-k conversational calls to a macrocell, which is generated by the N identical microcells. Assume that both ON and OFF periods of the IPP are exponentially distributed. During ON periods, class-k conversational calls arrive according to a Poisson process, whereas during OFF periods, no class-k conversational calls arrive. The IPP process, which is used to describe the superposition of N identical overflowed traffic of class-k conversational calls from the underlying microcells, is characterized by the following infinitesimal generator $\vec{Q}_k$ and arrival rate matrix $\vec{\Lambda}_k$. $\vec{Q}_k$ and $\vec{\Lambda}_k$ are defined as follows.

$$\vec{Q}_k = \begin{bmatrix} -\omega_k & \omega_k \\ \gamma_k & -\gamma_k \end{bmatrix} \tag{4}$$

$$\vec{\Lambda}_k = \begin{bmatrix} 0 & 0 \\ 0 & (\lambda_{nk}^O + \lambda_{hk}^O) \end{bmatrix} \tag{5}$$

where ω_k and γ_k are the state transition rates from OFF to ON and ON to OFF, respectively. $\lambda_{nk}^O + \lambda_{hk}^O$ is the average arrival rate of overflowed traffic in state ON. The associated parameters ω_k, γ_k, λ_{nk}^O, and λ_{hk}^O can be calculated via a moment matching method. At first, for each class of conversational calls, the associated mean and variance of the overflowed traffic from one microcell is computed with a fictitious overflowed group. The fictitious overflowed group is supposed to accommodate any unacceptable call that overflows from the focused microcell. Thus, the distribution of the busy time slots in the fictitious overflowed group is the distribution of calls overflowing from the focused microcell. The number of time slots of the overflowed group is chosen to be sufficiently large compared with the load offered to one microcell. In particular, assume that when the number of used time slots in a microcell is greater than $(C_1 - G_1)$ but not greater than $(C_1 - b_k)$, class-k conversational calls request to overflow to the overflowed group according to a Poisson process with average rate $[1 - A_k^{(1)}(n)]\lambda_{nk}$. Also, when there are not enough time slots to accommodate an arriving class-k conversational call, the average arrival rate to the overflowed group is $\lambda_{nk} + \lambda_{hk}^{(1)}$. In contrast, there are no arrivals to the overflowed group in all the other states when arriving calls can be served by the microcell. By solving the associated state equations, we can derive the mean and the variance of the overflowed traffic for each class of conversational calls resulting from one underlying microcell. Let α_k and ν_k represent the mean and variance of the number of class-k conversational calls in the overflowed group. It is noted that the two parameters of class-k conversational calls in the N microcells are the same, because the N microcells overlaid by one focused macrocell are assumed to be statistically identical and independent of one another. Let the mean and variance of the composite overflowed traffic seen by one overlaid macrocell be denoted as α_{Tk} and ν_{Tk}, respectively. Obviously, since the overflowed stream seen by one overlaid macrocell is the aggregation of N overflowed stream each of which results from a underlying microcell, α_{Tk} (ν_{Tk}) can be obtained by summing up the mean (variance) of the underlying N microcells. Since N microcells are statistically identical to each other, the value of α_{Tk} (ν_{Tk}) is just N times that of α_k (ν_k). Specifically, they can be computed by

$$\alpha_{Tk} = N\alpha_k \tag{6}$$

$$\nu_{Tk} = N\nu_k. \tag{7}$$

After obtaining α_{Tk} and ν_{Tk}, a moment matching method is adopted to evaluate the IPP parameters $(\lambda_k, \gamma_k, \omega_k)$ that characterize overflowed class-k conversational calls, where λ_k represents the traffic intensity in ON state, $1/\gamma_k$ the average ON time, and $1/\omega_k$ the average OFF time. Specifically, the three IPP parameters $(\lambda_k, \gamma_k, \omega_k)$ of overflowed class-k conversational calls can be computed based on the following equations.

$$\lambda_k = \mu_k \frac{(2\delta_2 - \delta_1)(\delta_1 - \delta_0) - \delta_1(\delta_2 - \delta_1)}{(\delta_1 - \delta_0) - (\delta_2 - \delta_1)} \tag{8}$$

$$\omega_k = \mu_k \frac{\delta_0(\frac{\lambda_k}{\mu_k} - \delta_1)}{\frac{\lambda_k}{\mu_k}(\delta_1 - \delta_0)} \tag{9}$$

$$\gamma_k = \omega_k \frac{(\frac{\lambda_k}{\mu_k} - \delta_0)}{\delta_0} \tag{10}$$

where $\delta_n = \frac{M_{(n+1)}}{M_{(n)}}$, $M_{(n)}$ is the n-th factorial moments of overflowed traffic in the overflowed group. Once λ_k, γ_k, and ω_k are determined, the average overflowed arrival rates of new and handoff class-k conversational calls can be calculated as follows.

$$\lambda_{nk}^{O} = \lambda_k \frac{1}{\alpha_{Tk}} \left(\frac{\lambda_{nk}}{\mu_k} P_{bk}^{(1)} N \right) \tag{11}$$

$$\lambda_{hk}^{O} = \lambda_k \frac{1}{\alpha_{Tk}} \left(\frac{\lambda_{hk}^{(1)}}{\mu_k} P_{hbk}^{(1)} N \right). \tag{12}$$

C. Analysis of Macrocell Traffic

Next, the analytical method for the macrocell incorporating the corresponding overflowed traffic model is presented. Since conversational calls are allowed to overflow to the macrocell layer, there are K classes of overflowed conversational calls as well as one class of streaming service involved in the analysis of macrocell. The overflowed traffic of each class of conversational calls is approximated as an independent IPP process. Let s_k denote the state of IPP traffic resulting from the overflowed class-k conversational calls. Specifically, the IPP is in ON (OFF) state, if $s_k = 1\,(0)$. Similarly, the handoff arrival rate $\lambda_{hk}^{(2)}$, $k = 1, 2, \cdots, K+1$, are also computed *iteratively*. Let n_k represent the number of ongoing class-k calls, $k = 1, 2, \cdots, K+1$. The focused macrocell can be modeled as a $(2K+1)$-dimensional Markov chain. The state vector of the focused macrocell is defined as $\vec{m} = (n_1, n_2, \cdots, n_{K+1}, s_1, s_2, \cdots, s_K)$. The feasible state space is defined as S_2, i.e., $S_2 = \{\vec{m} : \sum_{i=1}^{K+1} b_i n_i \le C_2\}$. Let $q(\vec{m})$ be the steady state probability in state $\vec{m}$. Let $S_2(n_{\min}) = \{\vec{m} : \sum_{i=1}^{K+1} b_i n_i = n_{\min}\}$ and $q(n_{\min})$ be the steady state probability that the total number of minimum used time slots by conversational and streaming calls is $n_{\min}$. $q(n_{\min})$ can be computed as follows.

$$q(n_{\min}) = \sum_{\vec{m} \in S_2(n_{\min})} q(\vec{m}). \tag{13}$$

Let $q_{ki}(n_{\min})$ denote the probability that the total number of minimum used time slots of the concerned macrocell is $n_{\min}$ and $s_k = i$, and further $\vec{q}_k(n_{\min}) = [q_{k0}(n_{\min}), q_{k1}(n_{\min})]$. In terms of $q(\vec{m})$, they can be expressed as follows.

$$q_{k0}(n_{\min}) = \sum_{s_k=0, \vec{m} \in S_2(n_{\min})} q(\vec{m}) \tag{14}$$

$$q_{k1}(n_{\min}) = \sum_{s_k=1, \vec{m} \in S_2(n_{\min})} q(\vec{m}). \tag{15}$$

Moreover, the state equations of the associated $(2K+1)$-dimensional Markov chain can be developed, and can be solved for equilibrium state probabilities by a Gauss-Seidel iterative approach. As is well known, the state probability distribution seen by an arriving call is identical to that by an outside observer only if the arrival process is Poisson. It is noted that the overflowed traffic is IPP. Therefore, it is necessary to derive both the outside observer's distribution $q(n_{\min})$ of the number of minimum used time slots, and the arriving call's distribution $\Delta(n_{\min})$ of the number of minimum used time slots. The arriving call's distribution $\Delta_k(n_{\min})$ can be computed as follows.

$$\Delta_k(n_{\min}) = \frac{\vec{q}_k(n_{\min})\,\vec{\Lambda}_k\,\vec{e}}{\sum_{n=0}^{C_2} \vec{q}_k(n_{\min})\,\vec{\Lambda}_k\,\vec{e}}. \quad (16)$$

In the above equation, $\vec{e}$ is 2×1 all-ones vector. It is worth mentioning that the blocking probabilities of handoff class-k conversational traffic among the macrocells, and new and handoff streaming traffic are computed in terms of the outside observer's distribution $q(n_{\min})$ since those arrival processes are assumed to be Poisson. In contrast, the blocking probabilities of conversational overflowed traffic are computed in terms of the arriving call's distribution $\Delta_k(n_{\min})$ since those arrival processes are assumed to be IPP. Next, the probability that a new class-k conversational call desires to overflow but is rejected by the macrocell layer is given by

$$P_{obk}^{(2)} = \sum_{n_{\min}=C_2-G_2}^{C_2} \left[1 - A_k^{(2)}(n_{\min})\right] \Delta_k(n_{\min}). \quad (17)$$

The probability that a handoff class-k conversational call desires to overflow but is rejected by the macrocell layer is given by

$$P_{ohbk}^{(2)} = \sum_{n_{\min}=C_2-b_k+1}^{C_2} \Delta_k(n_{\min}). \quad (18)$$

The probability $P_{hbk}^{(2)}$ that a class-k conversational handoff call is rejected in a macrocell and hence forced into termination can be computed as follows.

$$P_{hbk}^{(2)} = \sum_{n_{\min}=C_2-b_k+1}^{C_2} q(n_{\min}). \quad (19)$$

Furthermore, the forced termination probability $P_{fk}^{(2)}$ of handoff class-k conversational calls in a macrocell can be computed by

$$P_{fk}^{(2)} = \frac{P_{hk}^{(2)} P_{hbk}^{(2)}}{\left[1 - P_{hk}^{(2)}(1 - P_{hbk}^{(2)})\right]}. \quad (20)$$

Next, the analysis of streaming service is conducted. As to the number of time slots assigned to one streaming call, there are two cases. One case is that b_P time slots are assigned to each streaming call. The other case is that the number of time slots assigned to each streaming call may not be equal to b_P and may differ at most by one. Specifically, the average number of assigned time slots of streaming calls can be computed as follows.

$$b_A = \sum_{\vec{m}\in S_2, n_{K+1}=0}^{\beta} b_p q(\vec{m}) + \sum_{\vec{m}\in S_2, n_{K+1}=\beta+1}^{\lfloor C_2/b_{K+1}\rfloor} \phi q(\vec{m}) \quad (21)$$

where $\beta = \left\lfloor \left(C_2 - \sum_{l=1}^{K} b_l n_l\right) \Big/ b_P \right\rfloor$, and $\phi = \left(C_2 - \sum_{l=1}^{K} b_l n_l\right) \Big/ n_{K+1}$.

The blocking probability of new streaming calls is given by

$$P_{bK+1} = \sum_{n_{\min}=C_2-G_2}^{C_2} \left[1 - A_{K+1}^{(2)}(n_{\min})\right] q(n_{\min}). \quad (22)$$

The handoff blocking probability of handoff streaming calls is given by

$$P_{hbK+1} = \sum_{n_{\min}=C_2-b_{K+1}+1}^{C_2} q(n_{\min}). \quad (23)$$

Furthermore, the forced termination probability of handoff streaming calls is given by

$$P_{fK+1} = \frac{P_{hK+1} P_{hbK+1}}{\left[1 - P_{hK+1}(1 - P_{hbK+1})\right]}. \quad (24)$$

D. Analysis of the Overall Performance Measures

The overall performance measures are now derived. Since the streaming calls can be served only by the macrocell, their overall performance measures can be obtained from the previous subsection. The new call blocking probability P_{bk} of new class-k conversational calls, which is the probability that a new class-k conversational call is rejected by both microcell and macrocell layers, can be computed as follows.

$$P_{bk} = P_{bk}^{(2)} P_{obk}^{(2)}. \quad (25)$$

Furthermore, the forced termination probability P_{fk} of class-k conversational calls can be computed as follows.

$$P_{fk} = \frac{\left(1 - P_{bk}^{(1)}\right) P_{fk}^{(1)}}{(1 - P_{bk})} \left[P_{ohbk}^{(2)} + \left(1 - P_{ohbk}^{(2)}\right) P_{fk}^{(2)}\right] + \frac{P_{bk}^{(1)}\left(1 - P_{obk}^{(2)}\right)}{(1 - P_{bk})} P_{fk}^{(2)}. \quad (26)$$

Obviously, the forced termination probability of class-k conversational calls is the probability that a originally accepted call is interrupted because of a handoff failure in microcell layer without successful overflowing to a macrocell or a handoff failure in macrocell layer. Let ρ_k denote as the offered load of class-k calls and further be defined as $\rho_k = \frac{b_k \lambda_k}{\mu_k}$. It is noted that b_k is replaced by b_A as the offered load of class-$K+1$ streaming calls ρ_{K+1} is computed. The weighted new call blocking probability P_{wb} and the weighted forced termination probability P_{wf} can be computed as follows.

$$P_{wb} = \sum_{k=1}^{K+1} \rho_k P_{bk} \Big/ \sum_{k=1}^{K+1} \rho_k \quad (27)$$

$$P_{wf} = \sum_{k=1}^{K+1} \rho_k P_{fk} \Big/ \sum_{k=1}^{K+1} \rho_k. \quad (28)$$

To evaluate the combined effect of P_{wb} and P_{wf}, the overall weighted blocking probability is defined as follows.

$$P_B = \alpha P_{wb} + (1 - \alpha) P_{wf} \quad (29)$$

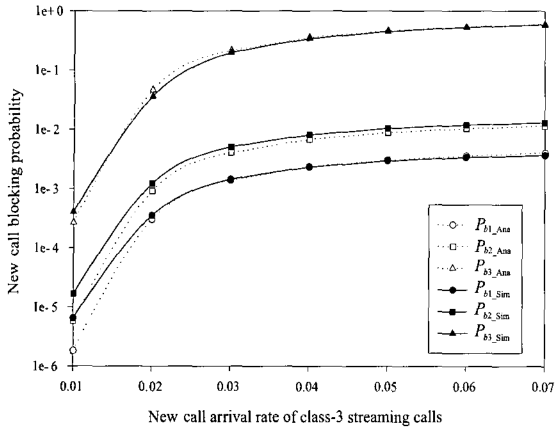


Fig. 3. New call blocking probability.

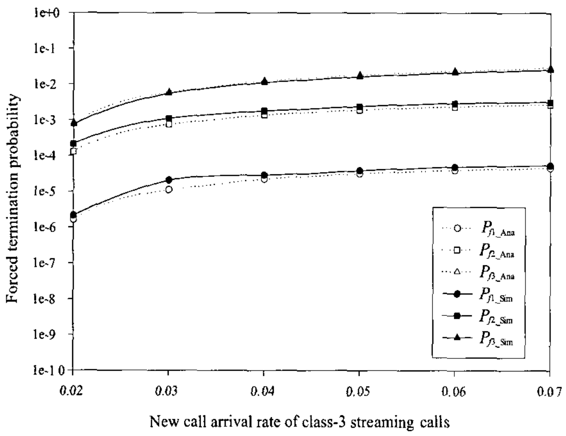


Fig. 4. Forced termination probability.

where α is a weighted factor, and $0 \leq \alpha \leq 1$. The value of α depends on the relative importance of new call blocking probability with respect to forced termination probability, which can be adjusted by the network operator. Throughout the paper it is assumed to be 0.3.

IV. NUMERICAL RESULTS

The numerical results for a two-layer system supporting two classes of conversational calls and one class of streaming calls are presented. Besides the analytical results, we also develop a discrete-event simulation program to collect the corresponding simulation results as in [18] and [19]. The simulation results are used to verify the accuracy of the analytical results. The simulation program is written in C language. At least 10^8 new call arrivals are generated in each simulation run. In the simulation model, a wrapped mesh of 37 macrocells and 259 microcells is considered to eliminate the boundary effect occurring in an unwrapped topology. It is assumed that new calls are generated in each cell according to a Poisson process and each new call is assigned a call holding time with an exponential distribution. The dwell time for which a user resides in the cell in question is also exponentially distributed. It is assumed that each call moves from a cell to any of the six neighboring cells with the same probability $1/6$. Also, to emphasize on the

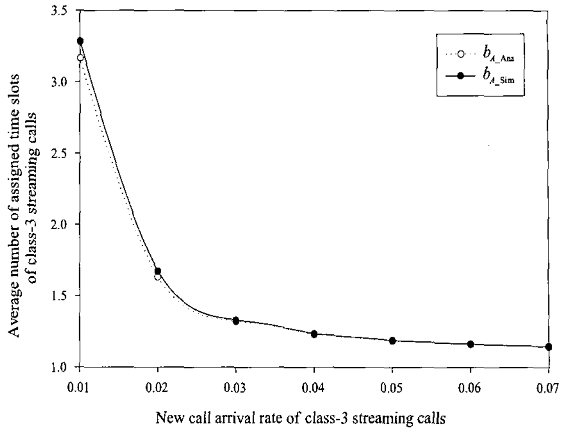


Fig. 5. Average number of assigned time slots of class-3 *streaming* calls.

effect of the call holding time difference, all classes of users are assumed to have the same moving velocity. Class-1 conversational service is assumed to be narrowband with $b_1 = 1$, $\lambda_1 = 0.05\ \text{s}^{-1}$, $\mu_1 = 1/180\ \text{s}^{-1}$, $\eta_1^{(1)} = 0.00245\ \text{s}^{-1}$, and $\eta_1^{(2)} = 0.000926\ \text{s}^{-1}$, and class-2 conversational service is wideband with $b_2 = 4$, $\lambda_2 = 0.01\ \text{s}^{-1}$, $\mu_2 = 1/300\ \text{s}^{-1}$, $\eta_2^{(1)} = 0.00245\ \text{s}^{-1}$, and $\eta_2^{(2)} = 0.000926\ \text{s}^{-1}$. The parameters of the class-3 streaming service are assumed to be $b_P = 5$, and $b_3 = 1$, $\mu_3 = 1/1800\ \text{s}^{-1}$, and $\eta_3^{(2)} = 0.000926\ \text{s}^{-1}$. It is noted that we assume that the average call holding time of class-3 streaming cells is ten times that of class-1 conversational calls. We also assume $C_1 = 40$, $C_2 = 60$, $G_1 = 8$, $G_2 = 12$, and $N = 7$.

First of all, simulation results are shown to verify the accuracy of analytical results. The results are shown in Figs. 3 to 5, where λ_3 is varied from 0.01 to 0.07 in steps of size 0.01. The analytical and simulation results for new call blocking and forced termination probabilities are shown in Figs. 3 and 4, respectively. Also, the results for the average number of assigned time slots of class-3 streaming calls are shown in Fig. 5. As you can see, the analytical results are close to the simulation results.

Second, the performance measures of interest with CS, FAP, TR, and SDCAC for the studied cases are compared in Figs. 6 to 9, where λ_3 is varied from 0.01 to 0.07 in steps of size 0.01. The results for weighted new call blocking probability are shown in Fig. 6. It is observed that TR results in higher weighted new call blocking probability and CS in lower one. Additionally, the results for weighted forced termination probability are shown in Fig. 7. It is observed that TR results in lower weighted forced termination probability and CS in higher one. Obviously, this is due to the fact that CS allows all classes equal access to bandwidth available at all times. No priority is given to handoff calls. It results in identical new call blocking probability and forced termination probability. As the traffic load is light, both of those blocking probabilities with CS are smaller than those with other schemes. As the traffic load increases, the forced termination probability increases rapidly with CS. In contrast, with TR, where guard channels are reserved for exclusive use of handoff calls, new calls are more likely to be rejected. It results in higher new call blocking probabilities but lower forced

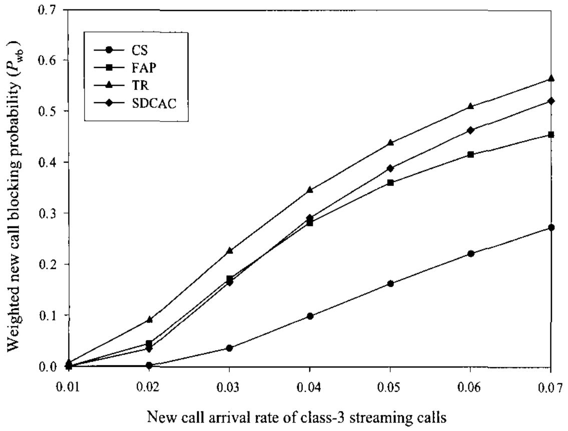


Fig. 6. Weighted new call blocking probability of the considered schemes.

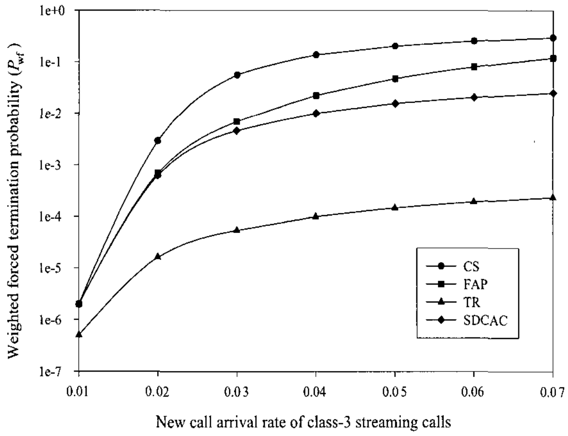


Fig. 7. Weighted forced termination probability of the considered schemes.

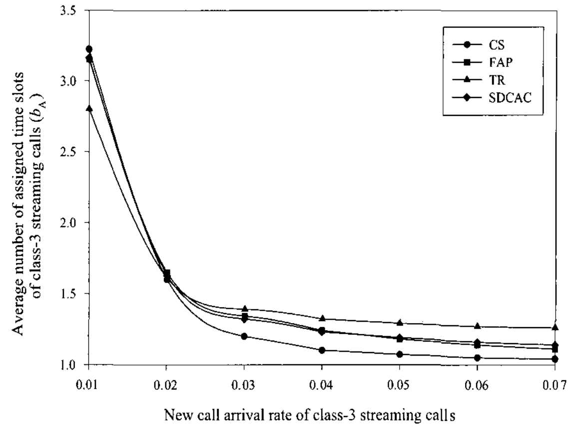


Fig. 8. Average number of assigned time slots of class-3 streaming calls of the considered schemes.

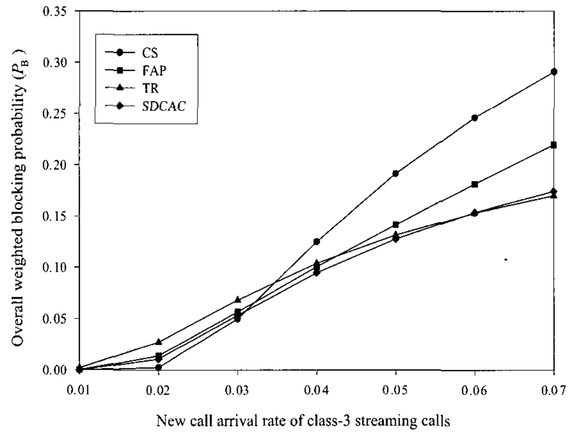


Fig. 9. Overall weighted blocking probability of the considered schemes.

termination probabilities. Higher new call blocking probability implies lower system utilization. Obviously, new calls being more likely rejected in light traffic load is an obvious disadvantage using TR. From Fig. 7, it is found that the weighted forced termination probabilities with CS and FAP are greater than that with TR or SDCAC in the studied case with $\lambda_{n3} = 0.07\ \mathrm{s}^{-1}$. This outcome with CS and FAP is undesirable from the viewpoint of users, since the weighted forced termination probability may be too high to be acceptable for users. It is the common drawback with CS and FAP. With SDCAC, if the offered load is too high, lower forced termination probabilities can be obtained at the expense of higher new call blocking probabilities. On the other hand, if the offered load is light, both weighted new call blocking probabilities and weighted forced termination probabilities can be kept low by using SDCAC. Specifically, with SDCAC, it is found that $P_{wb} = 0.0002$ and $P_{wf} = 0.000002$ with $\lambda_{n3} = 0.01\ \mathrm{s}^{-1}$, as well as $P_{wb} = 0.52$ and $P_{wf} = 0.025$ with $\lambda_{n3} = 0.07\ \mathrm{s}^{-1}$.

Furthermore, the average number of assigned time slots of class-3 streaming calls increases as the arrival rate of class-3 streaming calls decreases. This is because capacity is shared among more class-3 streaming calls with an increase of arrival rate of class-3 streaming calls and thus less time slots can be assigned to each class-3 streaming call. From Fig. 8, it can be seen that the results for the average number of assigned time slots are similar with any of the four CAC schemes. Specifically, it is found that with CS $b_A = 1.04$, with FAP $b_A = 1.11$, with TR $b_A = 1.26$, and with SDCAC $b_A = 1.14$ in the studied case with $\lambda_{n3} = 0.07\ \mathrm{s}^{-1}$.

Now, the results of overall weighted blocking probability for various CAC schemes are shown in Fig. 9. It is observed that CS leads to the smallest overall weighted blocking probability in the light traffic load (i.e., $0.01 \leq \lambda_{n3} \leq 0.03$), SDCAC in the medium traffic load (i.e., $0.03 < \lambda_{n3} \leq 0.06$), and TR in the heavy traffic load (i.e., $\lambda_{n3} > 0.06$). Specifically, given $\lambda_{n3} = 0.02\ \mathrm{s}^{-1}$, P_B is 0.003, 0.014, 0.027, and 0.011 with CS, FAP, TR, and SDCAC, respectively. Additionally, given $\lambda_{n3} = 0.07\ \mathrm{s}^{-1}$, P_B is 0.291, 0.220, 0.170, and 0.174 with CS, FAP, TR, and SDCAC, respectively. It is found that P_B with CS and FAP increase 67% and 26%, respectively, compared to the counterpart for SDCAC. Moreover, a huge increase in P_B as the load increases is also a drawback of CS and FAP. As λ_{n3} increases from 0.02 to 0.07, the increase in P_B with CS and FAP is 81% and 30%, respectively, compared to that with SDCAC. In contrast, because the time slots reserved for handoff calls lead to smaller weighted forced termination probabilities, TR results in undesirably lower system utilization in light traffic regions. Specifically, given $\lambda_{n3} = 0.02\ \mathrm{s}^{-1}$, compared to P_B

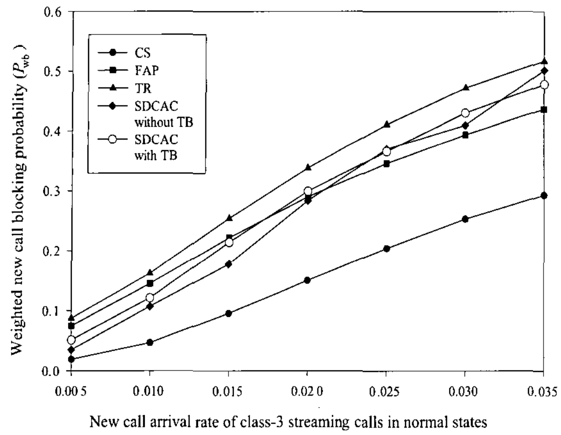


Fig. 10. Weighted new call blocking probability of the considered schemes.

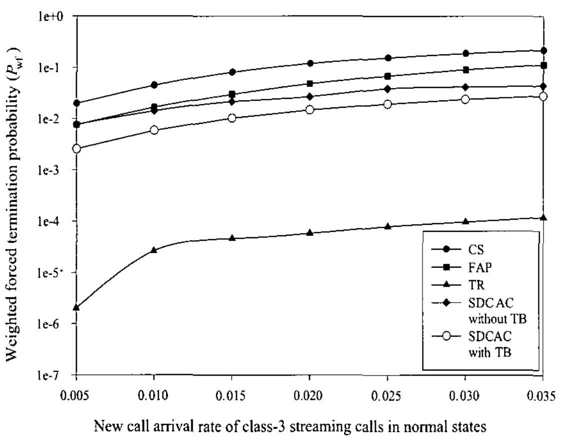


Fig. 11. Weighted forced termination probability of the considered schemes.

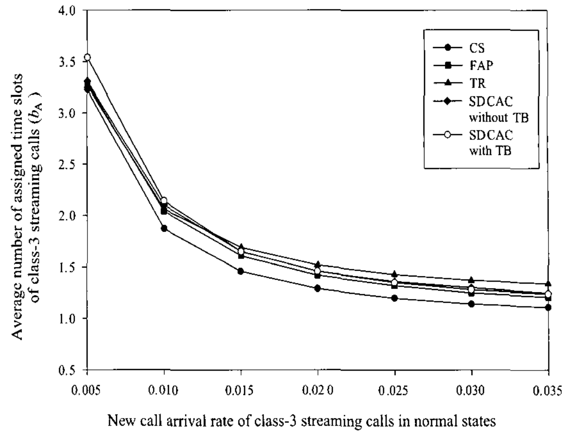


Fig. 12. Average number of assigned time slots of class-3 streaming calls of the considered schemes.

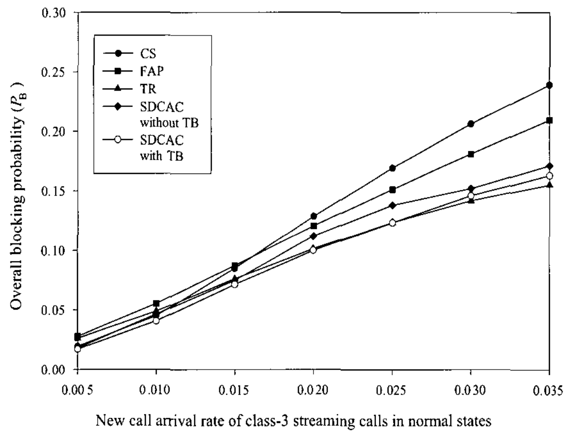


Fig. 13. Overall weighted blocking probability of the considered schemes.

with SDCAC, there is a 145% increase in that with TR. Generally speaking, in terms of overall weighted blocking probability, SDCAC performs the second best in both light and heavy traffic regions. SDCAC appears to act as CS in the light traffic load and as TR in the heavy traffic load.

Last but not least, the nonstationary scenarios are studied *via computer simulation*. Under nonstationary scenarios, each cell alternates in two states: Busy and normal, independent of one another. A cell in the busy state is assumed to have higher new call arrival rate than that in the normal state. The duration of each state is exponentially distributed. The transition rate from normal state to busy state is 1/3600 sec and that from busy state to normal state is 1/600 sec. In the normal state, the new call arrival rates are $\lambda_1 = 0.03$ s^{-1}, $\lambda_2 = 0.007$ s^{-1}, and λ_3 is varied from 0.005 to 0.035 in steps of size 0.005. The new call arrival rates of conversational and streaming calls in the busy state are 5 times those in normal state. Other parameters are the same as those mentioned before. To avoid overflowed conversational calls occupying excessive resources of the macrocells, we also consider SDCAC with takeback which allows conversational calls in the macrocells to be taken back to microcells if possible. As in [4], for reducing undesirable handoff executions, the takeback process is assumed to take place at the time of the cell boundary crossing. In other words, when resources become available in the microcell, the process of moving the conversational calls from time slots in the macrocell to those in the microcell is delayed until it crosses the microcell boundary. The results are shown in Figs. 10 to 13. Similar conclusions can be found in those figures. Furthermore, in terms of P_B, it is found that the overall weighted blocking probability of SDCAC with takeback is the lowest as λ_3 is not greater than 0.025 and that of TR is the lowest as $\lambda_3 > 0.025$. It is noted that the overall weighted blocking probability of SDCAC with takeback is close to that of TR as $\lambda_3 > 0.025$. Specifically, given $\lambda_{n3} = 0.005$ s^{-1}, P_B is 0.020, 0.028, 0.027, 0.018, and 0.017 with CS, FAP, TR, SDCAC without takeback and SDCAC with takeback, respectively. Additionally, given $\lambda_{n3} = 0.035$ s^{-1}, P_B is 0.24, 0.21, 0.16, 0.17, and 0.16 with CS, FAP, TR, SDCAC without takeback and SDCAC with takeback, respectively. With $\lambda_{n3} = 0.005$ s^{-1}, it is found that P_B with CS, FAP, and TR increase 18%, 65%, and 59%, respectively, compared to the counterpart for SDCAC with takeback. With $\lambda_{n3} = 0.035$ s^{-1}, it is found that P_B with CS, FAP increase 50% and 59%, respectively, compared to the counterpart for SDCAC with takeback. Moreover, a huge increase in P_B with the increased load can also be found for CS and FAP in the studied cases. As λ_{n3} increases from 0.02 to 0.07, the increase in P_B with CS and FAP is 51% and 25%, respectively, compared to that for SD-

CAC with takeback. On the other hand, it is noted that the performance of P_B for SDCAC with takeback is better than that of SDCAC without takeback. That is, it is found that the overall weighted blocking probability of SDCAC with takeback is lower than that of SDCAC without takeback in those studied cases. It results from the fact that lower weighted forced termination probabilities, which are shown in Fig. 11, can be obtained by using SDCAC with takeback. Obviously, the improvement for SDCAC with takeback is achieved by adding the takeback capability for conversational calls. Importantly, since actual traffic and specifically the call arrival rate are seldom stationary, an appropriate CAC should lower the impact of nonstationary traffic on the overall performance. Among those considered schemes, SDCAC with takeback performs better than the others due to its adaptability to the time variation of traffic load.

Overall speaking, SDCAC outperforms the other schemes in terms of overall weighted blocking probabilities and the robustness (thus improve the network utilization). It is clear that improvement can be achieved via proper assignment of reserved channels to new calls. SDCAC performs better because it can adapt to the occupied bandwidth and traffic characteristics. In fact, SDCAC appears to act as CS in the light traffic region and as TR in the heavy traffic region. However, the shortcoming of SDCAC is that, compared to the other schemes, SDCAC does need to obtain more system related information, e.g., call holding time, cell dwell rate, etc., which can be collected empirically as in [15].

V. CONCLUSION

SDCAC is proposed to make efficient use of scarce wireless resource while supporting different services in hierarchical networks with heterogeneous traffic. With SDCAC, new calls are accepted according to an acceptance probability taking account of call holding time, cell dwell time and occupied bandwidth. An iterative algorithm is developed to calculate performance measures of interest. Numerical results are presented to show the robustness of SDCAC compared to the other schemes. Overall speaking, SDCAC not only performs well in the medium traffic region but also appears to act as CS in the light traffic region and as TR in the heavy traffic region. SDCAC performs better than the other considered schemes due to its adaptability to the time variation of traffic load.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] B. Jabbari and W. F. Fuhrmann, "Teletraffic modeling and analysis of flexible hierarchical cellular networks with speed-sensitive handoff strategy," *IEEE J. Select. Areas Commun.*, vol. 15, no. 8, pp. 1539–1548, Oct. 1997.

[2] L.-R. Hu and S. S. Rappaport, "Personal communication systems using multiple hierarchical cellular overlays," *IEEE J. Select. Areas Commun.*, vol. 13, no. 2, pp. 406–415, Feb. 1995.

[3] S. H. Wie, J. S. Jang, B. C. Shin, and D. H. Cho, "Handoff analysis of the hierarchical cellular system," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 2027–2036, Sept. 2000.

[4] M. Umehira, M. Nakura, H. Sato, and A. Hashimoto, "ATM wireless access for mobile multimedia: Concept and architecture," *IEEE Pers. Commun.*, vol. 3, no. 5, pp. 39–48, Oct. 1996.

[5] D. S. Eom, M. Sugano, M. Murata, and H. Miyahara, "Call admission control for QoS provisioning in multimedia wireless ATM networks," *IEICE Trans. Commun.*, vol. E82-B, no. 1, pp. 14–23, Jan. 1999.

[6] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *IEEE Pers. Commun.*, vol. 3, no. 3, pp. 10–31, July 1996.

[7] X. Wu, D. Ghosal, and B. Mukherjee, "Benefits of queued handoff in a multi-tier architecture," in *Proc. IEEE GLOBECOM 2000*, San Francisco, USA, Nov. 2000, pp. 1396–1401.

[8] D. Hong and S. S. Rappoport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Technol.*, vol. 35, no. 3, pp. 77–92, Aug. 1986.

[9] M. Cheng and L. F. Chang, "Wireless dynamic channel assignment performance under packet data traffic," *IEEE J. Select. Areas Commun.*, vol. 17, no. 7, pp. 1257–1269, July 1999.

[10] Y. C. Kim, D. E. Lee, B. J. Lee, Y. S. Kim, and B. Mukherjee, "Dynamic channel reservation based on mobility in wireless ATM," *IEEE Commun. Mag.*, vol. 37, no. 11, pp. 47–51, Nov. 1999.

[11] O. T. W. Yu and V. C. M. Leung, "Adaptive resource allocation for prioritized call admission over ATM-based wireless PCN," *IEEE J. Select. Areas Commun.*, vol. 15, no. 7, pp. 1208–1225, Sept. 1997.

[12] L. Santos, L. S. Ferreira, V. Garcia, and L. M. Correia, "Cellular planning optimisation for non-uniform traffic distributions," in *Proc. IEEE PIMRC'98*, Boston, USA, Sept. 1998, pp. 1198–1202.

[13] S.-P. Chung and J.-C. Lee, "Random early blocking with engineered threshold in wireless multimedia Networks," in *IEEE VTC 2002*, Vancouver, Canada, Sept. 2002, pp. 1432–1436.

[14] A. Lombardo, S. Palazzo, and G. Schembra, "Management of wireless ATM networks loaded by mobile ABR source traffic," in *Proc. IEEE GLOBECOM'99*, Rio de Janeiro, Brazil, Dec. 1999, pp. 2758–2762.

[15] A. B. Mathew, A. D. Cheria, and S. Govindaraj, "A variable threshold method for dynamic channel allocation in wireless systems," in *Proc. MoMuc 2000*, Tokoy, Japan, Oct. 2000, pp. P-2-1–P-2-5.

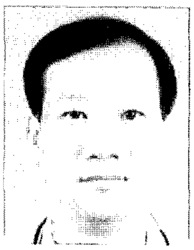
[16] B. Jabbari, "Teletraffic aspects of evolving and next-generation wireless communication networks," *IEEE Pers. Commun.*, vol. 3, no. 6, pp. 4–9, Dec. 1996.

[17] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*, Johns Hopkins University Press, 1981.

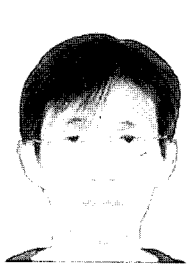
[18] Y.-B. Lin and V. K. Mak, "Eliminating the boundary effect of a large scale personal communication service network simulation," *ACM Trans. Modeling Computer and Simulation*, vol. 4, no. 2, pp. 165–190, Apr. 1994.

[19] J.-Y. Jeng and Y.-B. Lin, "Equal resource sharing scheduling for PCS data services," *Wireless Networks*, vol. 5, no. 1, pp. 41–55, Jan. 1999.

[20] D. Gross and C. M. Harris, *Fundamentals of Queueing*, Wiley, 1985.



Shun-Ping Chung received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1984, and the M.S. degree and the Ph.D. degree from University of Pennsylvania, Philadelphia, PA, USA, in 1988 and 1990, respectively, all in electrical engineering. After graduating in 1990, he joined the department of Electrical Engineering at National Taiwan University of Science and Technology where he is now an associate professor. His current research interests are in the areas of broadband wireline/wireless networks.



Jin-Chang Lee received the M.S. degree and the Ph.D. degree in electrical engineering from National Taiwan University of Science and Technology, Taipei, Taiwan, in 1997 and 2002. After graduating, he joined the Chunghwa Telecom Laboratories as an associate researcher. Presently, his research interests focus on cellular communication networks with particular emphasis on call admission control, traffic modeling, and performance analysis.