Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks

Yuguang Fang, *Senior Member, IEEE*, and Yi Zhang

***Abstract*—Call admission control (CAC) plays a significant role in providing the desired quality of service in wireless networks. Many CAC schemes have been proposed. Analytical results for some performance metrics such as call blocking probabilities are obtained under some specific assumptions. It is observed, however, that due to the mobility, some assumptions may not be valid, which is the case when the average values of channel holding times for new calls and handoff calls are not equal. In this paper, we reexamine some of the analytical results for call blocking probabilities for some call admission control schemes under more general assumptions and provide some easier-to-compute approximate formulas.**

***Index Terms*—Blocking probability, call admission control (CAC), mobile computing, resource allocation, wireless networks.**

NOMENCLATURE

C	Number of channels in a cell.
K	Threshold for new call bounding scheme.
m	Threshold for the cutoff priority scheme.
λ	Arrival rate for new calls.
λ_h	Arrival rate for handoff calls.
$1/\mu$	Average channel holding time for new calls.
$1/\mu_h$	Average channel holding time for handoff calls.
ρ	Traffic intensity for new calls (i.e., λ/μ).
ρ_h	Traffic intensity for handoff calls (i.e., λ_h/μ_h).
$p_{\rm nb}$	Blocking probability for new calls.
$p_{\rm hb}$	Blocking probability for handoff calls.
$p_{\rm nb}^a$	Blocking probability for new calls from the proposed approximation.
$p_{\rm hb}^a$	Blocking probability for handoff calls from the proposed approximation.
$p_{\rm nb}^t$	Blocking probability for new calls from the traditional approximation.
$p_{\rm hb}^t$	Blocking probability for handoff calls from the traditional approximation.
$u(x)$	Step function ($u(x) = 1$ for $x \geq 0$ and $u(x) = 0$ for $x < 0$).
α_i	Admission probability for new calls in Thinning Scheme II.
β_i	Admission probability for new calls in Thinning Scheme I.

Manuscript received August 7, 2000; revised August 14, 2001. This work was supported in part by the National Science Foundation under Faculty Early Career Development Award ANI0093241.

Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130 USA (e-mail: fang@ece.ufl.edu).

Y. Zhang is with the Lucent Technologies, North Andover, MA 01845 USA.

Publisher Item Identifier S 0018-9545(02)00428-0.

I. INTRODUCTION

THE future telecommunications networks (such as the third-generation wireless networks) aim to provide integrated services such as voice, data, and multimedia via inexpensive low-powered mobile computing devices over wireless infrastructures [21]. As the demand for multimedia services over the air has been steadily increasing over the last few years, wireless multimedia networks have been a very active research area. To support various integrated services with a certain quality of service (QoS) requirement in these wireless networks, resource provisioning is a major issue [8], [9]. Call admission control (CAC) is such a provisioning strategy to limit the number of call connections into the networks in order to reduce the network congestion and call dropping. In wireless networks, another dimension is added: call connection (or simply call) dropping is possible due to the users' mobility. A good CAC scheme has to balance the call blocking and call dropping in order to provide the desired QoS requirements [1], [4], [13], [14].

Call admission control for high-speed networks (such as asynchronous transfer mode networks) and wireless networks has been intensively studied in the last few years ([22]). Due to users' mobility, CAC becomes much more complicated in wireless networks. An accepted call that has not completed in the current cell may have to be handed off to another cell. During the process, the call may not be able to gain a channel in the new cell to continue its service due to the limited resource in wireless networks, which will lead to the call dropping. Thus, the new calls and handoff calls have to be treated differently in terms of resource allocation. Since users tend to be much more sensitive to call dropping than to call blocking, handoff calls are normally assigned higher priority over the new calls. Various handoff priority-based CAC schemes have been proposed [11], [23]; they can be classified into two broad categories.

1) *Guard Channel (GC) Schemes:* Some channels are reserved for handoff calls. There are four different schemes.
   a) The *cutoff priority scheme* is to reserve a portion of channel for handoff calls; whenever a channel is released, it is returned to the common poll of channels [9], [17].
   b) The *fractional guard channel schemes* (we call the new call thinning scheme I in this paper) is to admit a new call with certain probability (which depends on the number of busy channels). This scheme was first proposed by Ramjee *et al.* [19] and shown to be more general than the cutoff priority scheme.

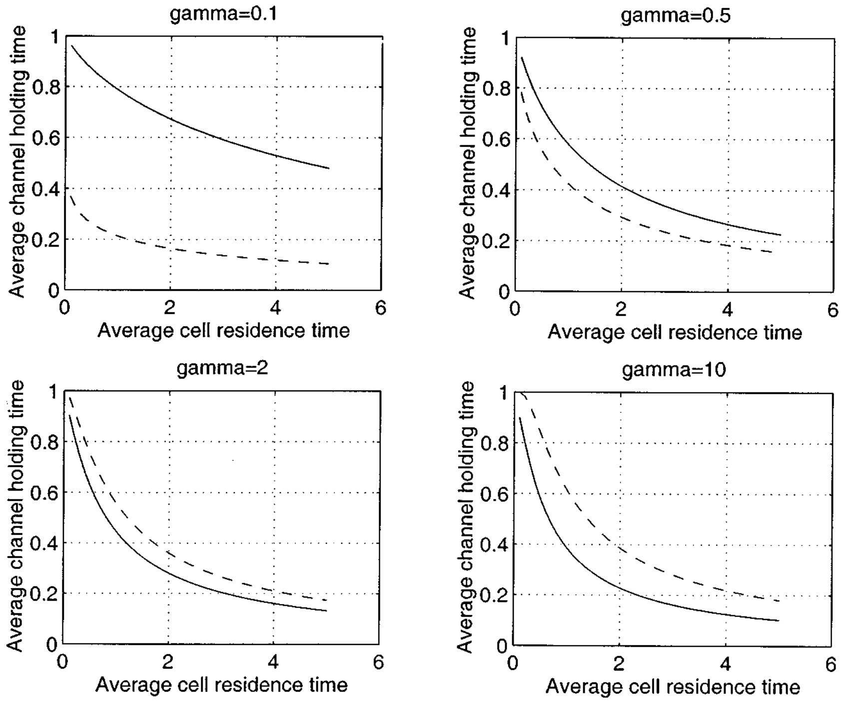


Fig. 1. Average channel holding times for new calls and handoff calls: solid line for new calls and dashed line for the handoff calls.

c) Divide all channels allocated to a cell into two groups: one for the common use for all calls and the other for handoff calls only (the *rigid division-based CAC scheme* [13]).

d) Limit the number of new calls admitted to the network (called the *new call bounding scheme* in this paper).

2) *Queueing Priority (QP) Schemes:* In this scheme, calls are accepted whenever there are free channels. When all channels are busy, new calls are queued while handoff calls are blocked [7], new calls are blocked while handoff calls are queued [3], [24], or all arriving calls are queued with certain rearrangements in the queue [1], [14].

Various combinations of the above schemes are possible depending on specific applications [1], [14]. In this paper, we concentrate on the guard channel schemes.

In the current literature, we observe that most performance analysis of CAC schemes was carried out under the assumption that the channel holding times for new calls and handoff calls are identically distributed (some with exponential distribution), i.e., all calls were assumed to be identically distributed with the same parameter. Thus, the one-dimensional Markov chain was used to obtain the blocking probabilities for new calls and handoff calls. However, recent studies ([5] and [6] and references therein) showed that the new call channel holding time and the handoff call channel holding time may have different distributions. Worse yet, they may have different average values. For example, Fig. 1 shows that the average channel holding times for new calls and handoff calls are different. In this figure, when the cell residence time is Gamma distributed with shape parameter varying at $\gamma = 0.1, \gamma = 0.5, \gamma = 2$, and $\gamma = 10$ (the latter two cases are in fact Erlang distributed), the average channel holding times for new calls and handoff calls computed using the formulas in [6] are significantly different for some cases. Thus, the one-dimensional Markov chain model for some guard channel CAC schemes assuming that the new calls and handoff calls are identically distributed may not be appropriate; the multidimensional Markov chain may be needed. Rappaport and his colleagues noticed such an observation and started a series of research works (e.g., [9], [20], and [18]). In [20], Rappaport used the generalized Erlang distribution to model some random variable (such as dwell time). In [18], Orlik and Rappaport proposed the sum of hyperexponential distribution to model the dwell time. The multidimensional Markovian chain theory has been extensively used in their research. This, of course, could solve the problem when the average channel holding times for new calls and handoff calls are different. However, another problem arises: the curse of dimensionality. As observed in [20], the dimension of states in the multidimensional Markov chain modeling increases very quickly. It will be desirable to study some approximate solutions to avoid solving a large set of flow equations.

It is also a common practice in the literature (see [3]) that the distinction between channel holding times for new calls and handoff calls is not made. We can find the average channel holding time for cell traffic (the merged traffic of new calls and handoff calls), use this parameter to form the exponential distribution to approximate the channel holding-time distribution, then apply the one-dimensional Markov chain model to find the call blocking probabilities. As we will show, this approximation may not be appropriate in some parameter range.

In this paper, we will examine a few CAC schemes under the assumption that the new calls and handoff calls may have different average channel holding times. We will present analytical results whenever possible and give some approximate results when computation is an issue. We will study the *new call bounding scheme*, in which a threshold is enforced on the number of new calls accepted into the cell, *the cutoff priority scheme*, in which a reserved number of channels are used for handoff calls, *the fractional guard channel scheme (the new call thinning scheme I)*, in which new calls will be selectively blocked when the cell traffic increases, and *the new call thinning scheme II*, in which the thinning of new calls is based on the number of new calls accepted into the cell. Simulations will be carried out to verify how approximations perform.

We point out that this paper represents only the first step toward general issues to be reexamined along this direction. As we notice, we have assumed that the channel holding times for new calls and handoff calls are independent and exponentially distributed but with different average values (the most important case). However, in reality, these assumptions may not be true. It is usually agreed that the new call and the handoff call have different channel holding-time distributions ([5] and [6] and references therein). Also, the handoff traffic may not be Poisson [5]. Performance analysis of CAC schemes under more realistic assumptions (using higher moments of cell traffic and channel holding times) has to be carefully carried out. We will present such a study in a subsequent paper.

Future generation wireless systems have shifted the focus on multimedia services and guaranteeing their QoS. Call connections may demand different amounts of network resource (channels). Thus, call admission control schemes can be designed to deal with multiclass services. The schemes (e.g., thinning schemes) can be generalized to handle such situations: permission probabilities can be chosen according to the resource utilization and amount of resource needed to support a call request. We can also use priority levels and multiple thresholds to handle different traffic classes. The details will be investigated in the future.

This paper is organized as follows. In the next section, we investigate some call admission control schemes and present some new analytical results. Simulation study will appear in the third section. We conclude this paper in Section IV.

II. Call Admission Control Schemes

In this section, we will study three call admission control schemes in wireless networks when the channel holding times for new calls and handoff calls are differentiated: the new call bounding priority, the cutoff priority scheme, and the new call thinning scheme. The analytical techniques and results can be easily extended to blocking performance for wireless multimedia networks with multiple prioritized traffic, in which corresponding call admission control schemes can be obtained. We can immediately observe that the analytical results are valid for wireless networks with two prioritized traffic.

Let $\lambda, \lambda_h, 1/\mu$, and $1/\mu_h$ denote the arrival rate for new calls, the arrival rate for handoff calls, the average channel holding time for new calls, and the average channel holding time for handoff calls, respectively. Let C denote the total number of channels in a cell. We assume that the arrival process for new calls and the arrival process for handoff calls are all Poisson, and the channel holding times for new calls and handoff calls are exponentially distributed, respectively. Although it has been observed [5], [6] that the handoff call arrival rate is closely related to the new call arrival rate, and that the channel holding times for new calls and handoff calls also depend on the cell residence time distribution, our study here is to show how call-blocking probabilities can be approximated when the channel holding times for new calls and handoff calls have different averages.

It has been observed that the channel holding times for new calls and handoff calls are distinct; even their average values are different. The current literature does not make such a distinction; the common assumption is that the channel holding time for the call arrivals (consisting of new calls and handoff calls) is exponentially distributed with parameters equal to the average channel holding time of new calls and handoff calls together, i.e., both new calls and handoff calls are distributed with the same distribution. We call this approximation the traditional approach. Due to such approximation, the one-dimensional Markov chain model can be used to derive analytical results for blocking performance. Of course, inaccuracy is expected due to such approximation. We will make such a distinction in this paper and derive some analytical formulas for blocking probabilities for both new calls and handoff calls.

A. New Call Bounding Scheme

In this scheme, we limit the admission of new calls into the wireless networks. The scheme works as follows: if the number of new calls in a cell exceeds a threshold when a new call arrives, the new call will be blocked; otherwise it will be admitted. The handoff call is rejected only when all channels in the cell are used up. The idea behind this scheme is that we would rather accept fewer customers than drop the ongoing calls in the future, because customers are more sensitive to call dropping than to call blocking. In this section, we give the analytical results for the new call blocking probability $p_{\rm nb}$ and the handoff call blocking probability $p_{\rm hb}$.

Fig. 2 indicates the transition diagram for the new call bounding scheme. Let K be the threshold for the new calls and $\lambda, \lambda_h, \mu, \mu_h$, and C as defined before. This diagram arises from the two-dimensional Markov chain with the state space

$$\mathcal{S} = \{(n_1, n_2) \,|\, 0 \le n_1 \le K, n_1 + n_2 \le C\}$$

where n_1 denotes the number of new calls initiated in the cell and n_2 is the number of handoff calls in the cell. Let $q(n_1, n_2; \bar{n}_1, \bar{n}_2)$ denote the probability transition rate from state (n_1, n_2) to state $(\bar{n}_1, \bar{n}_2)$. Then we have

$$\begin{aligned}
&q(n_1, n_2; n_1 - 1, n_2) = n_1\mu(0 < n_1 \le K, 0 \le n_2 \le C)\\
&q(n_1, n_2; n_1 + 1, n_2) = \lambda(0 \le n_1 < K, 0 \le n_2 \le C)\\
&q(n_1, n_2; n_1, n_2 - 1) = n_2\mu_h(0 \le n_1 \le K, 0 < n_2 \le C)\\
&q(n_1, n_2; n_1, n_2 + 1) = \lambda_h(0 \le n_1 \le K, 0 \le n_2 < C)
\end{aligned}$$

where (n_1, n_2) is a feasible state in $\mathcal{S}$. Let $p(n_1, n_2)$ denote the steady-state probability that there are n_1 new calls and n_2

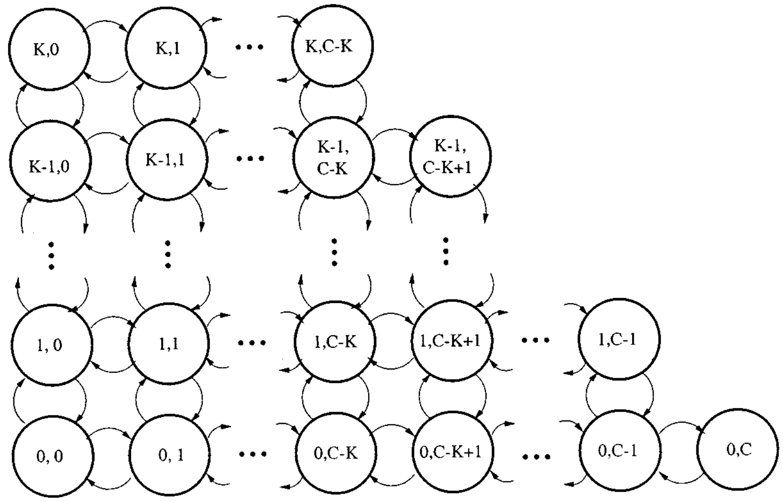


Fig. 2. Transition diagram for the new call bounding scheme.

handoff calls in the cell. Let $\rho = \lambda/\mu$ and $\rho_h = \lambda_h/\mu_h$. From the detailed balance equation, we obtain

$$p(n_1, n_2) = \frac{\rho^{n_1}}{n_1!} \cdot \frac{\rho_h^{n_2}}{n_2!} \cdot p(0,0), \quad 0 \le n_1 \le K, \quad n_1 + n_2 \le C, \quad n_2 \ge 0.$$

From the normalization equation, we obtain

$$p(0,0) = \left[\sum_{0 \le n_1 \le K, n_1+n_2 \le C} \frac{\rho^{n_1}}{n_1!} \cdot \frac{\rho_h^{n_2}}{n_2!} \right]^{-1} = \left[\sum_{n_1=0}^{K} \frac{\rho^{n_1}}{n_1!} \sum_{n_2=0}^{C-n_1} \frac{\rho_h^{n_2}}{n_2!} \right]^{-1}.$$

From this, we obtain the formulas for new call blocking probability and handoff call blocking probability as follows:

$$p_{\rm nb} = \frac{\sum_{n_2=0}^{C-K} \frac{\rho^K}{K!} \cdot \frac{\rho_h^{n_2}}{n_2!} + \sum_{n_1=0}^{K-1} \frac{\rho^{n_1}}{n_1!} \cdot \frac{\rho_h^{C-n_1}}{(C-n_1)!}}{\sum_{n_1=0}^{K} \frac{\rho^{n_1}}{n_1!} \sum_{n_2=0}^{C-n_1} \frac{\rho_h^{n_2}}{n_2!}} \quad (1)$$

$$p_{\rm hb} = \frac{\sum_{n_1=0}^{K} \frac{\rho^{n_1}}{n_1!} \cdot \frac{\rho_h^{C-n_1}}{(C-n_1)!}}{\sum_{n_1=0}^{K} \frac{\rho^{n_1}}{n_1!} \sum_{n_2=0}^{C-n_1} \frac{\rho_h^{n_2}}{n_2!}}. \quad (2)$$

Obviously, when $K = C$, the new call bounding scheme becomes the nonprioritized scheme. As we expect, we obtain

$$p_{\rm nb} = p_{\rm hb} = \frac{\frac{(\rho+\rho_h)^C}{C!}}{\sum_{n=0}^{C} \frac{(\rho+\rho_h)^n}{n!}}.$$

As we mentioned earlier, in most literature the channel holding times for both new calls and handoff calls are identically distributed with the same parameter. In this case, the average channel holding time is given by

$$\frac{1}{\mu_{\rm av}} = \frac{\lambda}{\lambda+\lambda_h} \cdot \frac{1}{\mu} + \frac{\lambda_h}{\lambda+\lambda_h} \cdot \frac{1}{\mu_h} = \frac{\rho+\rho_h}{\lambda+\lambda_h}. \quad (3)$$

From this, the traffic intensities for new calls and handoff calls using the above common average channel holding time $1/\mu_{\rm av}$ are given by

$$\hat{\rho} = \frac{\lambda}{\mu_{\rm av}} = \frac{\lambda}{\lambda+\lambda_h}(\rho+\rho_h)$$
$$\hat{\rho}_h = \frac{\lambda_h}{\mu_{\rm av}} = \frac{\lambda_h}{\lambda+\lambda_h}(\rho+\rho_h).$$

Applying these formulas in (1) and (2), we obtain similar results for new call blocking probability and handoff call blocking probability following the traditional approach (one-dimensional Markov chain theory), which obviously provides only an approximation. We will show later that significantly inaccurate results are obtained using this approach, which implies that we cannot use the traditional approach if the channel holding times for new calls and handoff calls are distinct with different average values. We observe that there is one case where these two approaches give the same results, i.e., when the nonprioritized scheme is used: $K = C$. This is because we have the following identity: $\hat{\rho} + \hat{\rho}_h = \rho + \rho_h$.

As a final remark, this scheme may work best when the call arrivals are bursty. When a big burst of calls arrives in a cell (for example, before or after a football game), if too many new calls accepted, the network may not be able to handle the resulting handoff traffic, which will lead to severe call dropping. The new call bounding scheme, however, could handle the problem well by spreading the potential bursty calls (users will try again when the first few tries fail). On another note, as we observe in wired networks, network traffic tends to be self-similar ([15]). Wireless network traffic will behave the same considering more data services will be supported in the wireless networks. This scheme will be useful in the future wireless multimedia networks.

B. Cutoff Priority Scheme

Instead of putting limitation on the number of new calls, we base on the number of total on-going calls in the cell to make a decision whether a new arriving call is accepted or not. The scheme works as follows.

Let m denote the threshold, upon a new call arrival. If the total number of busy channels is less than m, the new call is accepted;

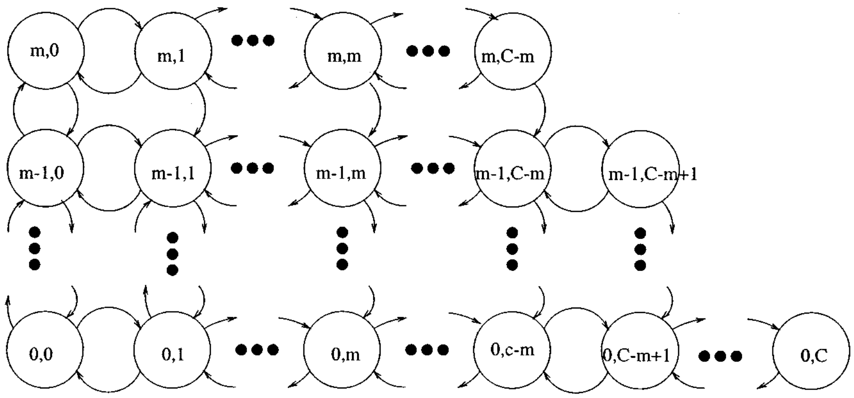


Fig. 3. Transition diagram for the cutoff priority scheme.

otherwise, the new call is blocked. The handoff calls are always accepted unless no channel is available upon their arrivals. This scheme has been studied in many papers [9], [24], [16], and analytical results for call blocking probabilities are obtained under the assumption that the average new call channel holding time and average handoff call channel holding time are equal so that one-dimensional Markov chain theory can be used. When the average channel holding times for new calls and handoff calls are different, the approach will not work.

Let $\lambda, \lambda_h, \mu, \mu_h,$ and C be defined as before; and let m denote the cutoff threshold. As in the previous section, we can use the two-dimensional Markov chain to model the system. Let (n_1, n_2) denote the state, where n_1 and n_2 denote the numbers of new calls and handoff calls in the cell, respectively. The state diagram is shown in Fig. 3 with the following transition rates:

$$
\begin{aligned}
q(n_1,n_2;n_1-1,n_2) &= n_1\mu(0<n_1\le m, 0\le n_1+n_2\le C)\\
q(n_1,n_2;n_1+1,n_2) &= \lambda(0\le n_1<m, 0\le n_1+n_2<m)\\
q(n_1,n_2;n_1,n_2-1) &= n_2\mu_h(0\le n_1\le m, 0<n_1+n_2\le C)\\
q(n_1,n_2;n_1,n_2+1) &= \lambda_h(0\le n_1\le m, 0\le n_1+n_2<C).
\end{aligned}
$$

We observe that in some states, such as those when $n_1+n_2>m$, the flows no longer have the symmetric nature. It is doubtful whether the detailed balance equations are valid. Indeed, we do not have the product form for this scheme when $\mu \ne \mu_h$. Let $u(x)$ denote the step function, which is defined as follows: $u(x)=1$ when $x\ge 0$ and $u(x)=0$ when $x<0$. Let $\bar{u}(x) = 1-u(x)$. Then, from Fig. 3, we obtain the following global balance equations:

$$
\begin{aligned}
&[\bar{u}(n_1+n_2-m)\lambda + \bar{u}(n_1+n_2-C)\lambda_h + n_1\mu\\
&\quad + n_2\mu_h]p(n_1,n_2)\\
&\quad = u(n_2-1)\lambda_h p(n_1,n_2-1)\\
&\qquad + \bar{u}(n_1-m)\mu p(n_1+1,n_2)\\
&\qquad + \bar{u}(n_1+n_2-C)(n_2+1)\mu_h p(n_1,n_2+1)\\
&\qquad + \bar{u}(n_1+n_2-1-m)\lambda p(n_1-1,n_2),\\
&\qquad\qquad 0\le n_1\le m, n_1+n_2\le C. \quad (4)
\end{aligned}
$$

Thus, we have to solve these global balance equations to find the steady-state probability distribution $p(n_1,n_2)$, from which blocking probabilities can be obtained, as done when multidimensional Markov chain theory is used.

However, as we mentioned before, solving the global balance equations may be computationally intensive when the state dimension is large. It will be useful to find some approximation for the call blocking probabilities. We now present an approximation based on the following idea: we attempt to reduce the two-dimensional Markov chain model to a one-dimensional Markov chain model by normalizing the average service time for each stream so that the average service time becomes identical for both streams. In this way, we can use the one-dimensional Markov chain theory to find the call blocking probabilities. This idea is based on the following observation: the blocking probability for each stream depends on the traffic intensity. The higher the traffic intensity, the higher the blocking probability. By normalizing the average service time (say, making the average service time to the unity), the arriving traffic for that stream will be scaled appropriately. This normalization process does not change the traffic intensity. Hopefully, the resulting blocking probability can be approximated. We have not been able to analytically show how good this approximation is. We will, however, show that this approximation provides much better performance than the traditional approach.

Here is how our approximation works. Let $\rho = \lambda/\mu$ and $\rho_h = \lambda_h/\mu_h$. We use the following approximate model: the new call arrival stream is Poisson with arrival rate ρ and with service rate (corresponding channel holding time for new calls) 1 (the unity). The handoff call arrival stream is also Poisson with arrival rate ρ_h and service rate 1. Let p_j^a denote the probability that there are j busy channels in steady state $(j=0,1,\ldots,C)$ for the approximate model. Then, we can obtain the following stationary distribution for the approximate model:

$$
p_j^a = \begin{cases} \frac{(\rho+\rho_h)^j}{j!}p_0, & j\le m\\ \frac{(\rho+\rho_h)^m\rho_h^{j-m}}{j!}p_0, & m+1\le j\le C. \end{cases}
$$

where

$$
p_0^a = \left[\sum_{j=0}^{m}\frac{(\rho+\rho_h)^j}{j!} + \sum_{j=m+1}^{C}\frac{(\rho+\rho_h)^m\rho_h^{j-m}}{j!}\right]^{-1}.
$$

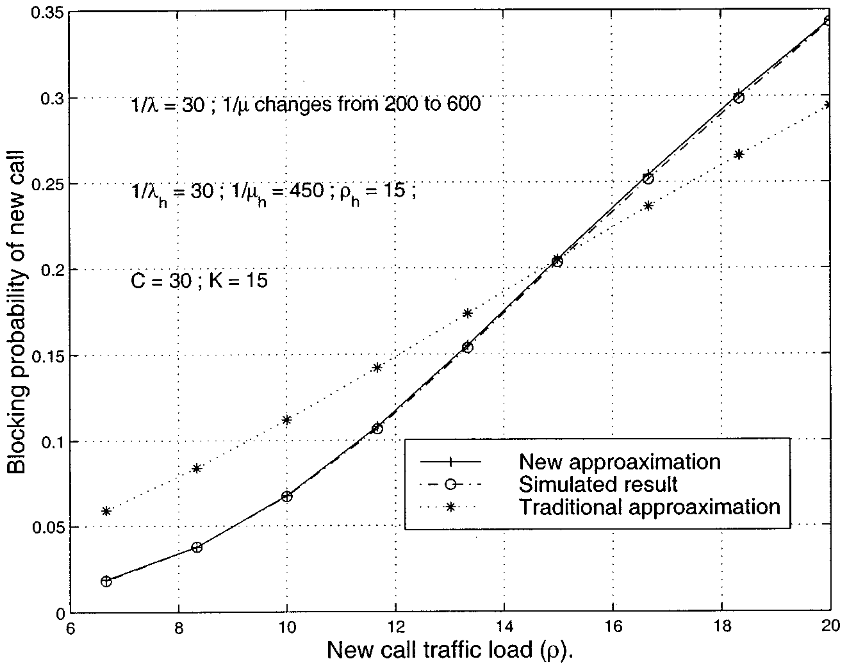


Fig. 4. Call blocking probability in the new call bounding scheme.

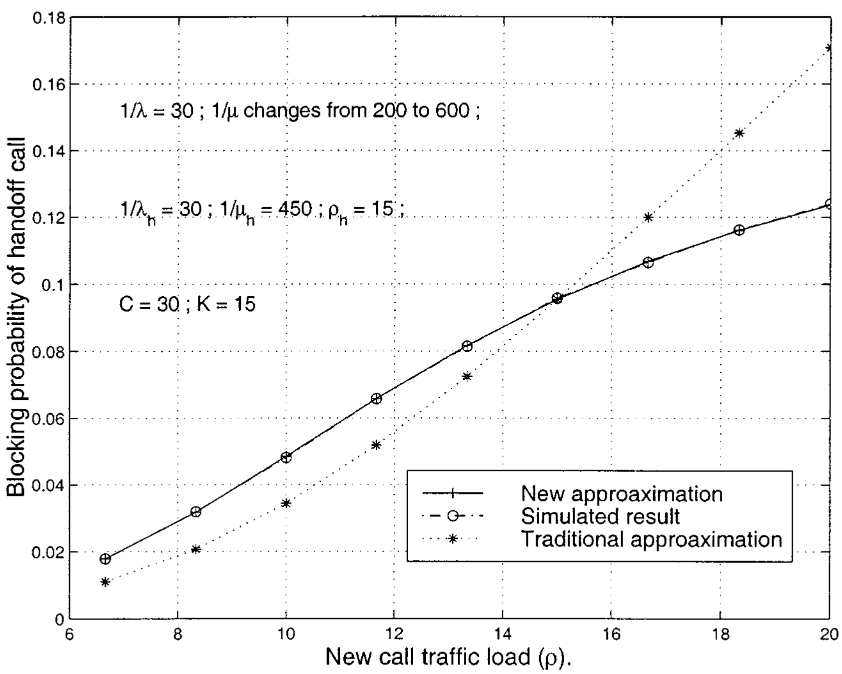


Fig. 5. Handoff call blocking probability in the new call bounding scheme.

From this stationary distribution, we obtain the blocking probabilities for new calls and handoff calls as follows:

$$p_{\rm nb}^{a} = \frac{\sum_{j=m}^{C} \frac{(\rho+\rho_h)^m \rho_h^{j-m}}{j!}}{\sum_{j=0}^{m} \frac{(\rho+\rho_h)^j}{j!} + \sum_{j=m+1}^{C} \frac{(\rho+\rho_h)^m \rho_h^{j-m}}{j!}} \tag{5}$$

$$p_{\rm hb}^{a} = \frac{\frac{(\rho+\rho_h)^m \rho_h^{C-m}}{C!}}{\sum_{j=0}^{m} \frac{(\rho+\rho_h)^j}{j!} + \sum_{j=m+1}^{C} \frac{(\rho+\rho_h)^m \rho_h^{j-m}}{j!}} \tag{6}$$

We will use these to approximate the call blocking probabilities for the cutoff priority scheme. We observe the following: when $m = C$, the result becomes exact for a nonprioritized scheme. When the average channel holding times for new calls and handoff calls are equal, the approximation also leads to the exact result.

If we use the traditional approach, we do not distinguish the new call channel holding time and the handoff call channel holding time. In this case, the average channel holding time is given by (3). The corresponding result is given by the equation at the bottom of the next page.

C. New Call Thinning Schemes

The new call thinning schemes are schemes in which a new call is admitted with certain probability. The idea behind these schemes is to smoothly throttle the new call stream as the network traffic is building up. Thus, when the network is approaching the congestion, the admitted new call stream becomes thinner. Due to the flexible choice of new call admission probabilities, these schemes can be made very general. In this section, we study two thinning schemes. The first one uses the information about the total number of busy channels, which leads to the *fractional guard channel scheme*. The second scheme utilizes the number of channels occupied by the new calls.

We start with the study of the first scheme (Thinning Scheme I). This brilliant idea behind this scheme was first proposed by Ramjee *et al.* [19]. Let $\beta_i (i = 0, 1, \ldots, C-1)$ denote the nonnegative numbers less than or equal to unity. The new call thinning scheme works as follows: when the number of busy channels is i, an arriving new call will be admitted with probability β_i. An arriving handoff call will always be admitted unless there are no channels available, in which case all calls will be blocked. Obviously, when $\beta_0 = \cdots = \beta_{m-1} = 1$ and $\beta_m = \cdots = \beta_C = 0$, this scheme becomes the cutoff priority scheme. We also observe that when $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_C$, the new call stream becomes thinner and thinner when the number of busy channels is increasing.

The exact analysis can be carried out as in the last section, using the two-dimensional Markov chain theory. Here, we only present the approximate results for call blocking probabilities for these schemes. Once again, let $\lambda, \lambda_h, \mu, \mu_h$ be defined as before, and let $\rho = \lambda/\mu$ and $\rho_h = \lambda_h/\mu_h$. Let p_j^a denote the probability that there are j busy channels in steady state ($j = 0, 1, \ldots, C$). We can obtain the following stationary distribution for the approximate model:

$$p_j^a = \frac{\prod_{i=0}^{j-1}(\beta_i \rho + \rho_h)}{j!} p_0, \quad 1 \leq j \leq C$$

where

$$p_0^a = \left[\sum_{j=0}^{C} \frac{\prod_{i=0}^{j-1}(\beta_i \rho + \rho_h)}{j!} \right]^{-1}.$$

From this stationary distribution, we obtain the blocking probabilities for new calls and handoff calls as follows:

$$p_{\mathrm{nb}}^a = \sum_{j=0}^{C} (1-\beta_j) p_j, \beta_C = 0$$
$$p_{\mathrm{hb}}^a = p_C.$$

Obviously, when the new call channel holding time and the handoff call channel holding time have the same average, i.e., $\mu = \mu_h$, the result becomes the exact one obtained in [19].

A variation of Scheme I is to admit the new calls based on the number of new calls currently in service; we call it *Thinning Scheme II*. Let $\alpha_i (i = 0, 1, \ldots, C)(\alpha_C = 0)$ be nonnegative numbers not exceeding unity. A new call is admitted with probability α_i if there are i new calls currently in service, and all calls will be blocked if all channels are busy. Obviously, if $\alpha_0 = \alpha_1 = \cdots = \alpha_{K-1} = 1$ and $\alpha_K = \cdots = \alpha_C = 0$, then this scheme becomes the new call bounding scheme. It is expected that the performance can be carried out as for the new call bounding scheme. Let $p(n_1, n_2)$ denote the stationary probability distribution; then

$$p(n_1, n_2) = \frac{\left[\prod_{i=0}^{n_1-1} \alpha_i\right] \rho^{n_1}}{n_1!} \cdot \frac{\rho_h^{n_2}}{n_2!} \cdot p(0,0), \qquad n_1 + n_2 \leq C$$

where

$$p(0,0) = \left[\sum_{n_1+n_2 \leq C} \frac{\left[\prod_{i=0}^{n_1-1} \alpha_i\right] \rho^{n_1}}{n_1!} \cdot \frac{\rho_h^{n_2}}{n_2!} \right]^{-1}.$$

Thus, the new call blocking probability and the handoff blocking probability are given by

$$p_{\mathrm{nb}} = \frac{\sum_{i=0}^{C}(1-\alpha_i)\left(\prod_{k=0}^{i-1}\alpha_k\right)\frac{\rho^i}{i!}\left(\sum_{j=0}^{C-i}\frac{\rho_h^j}{j!}\right)}{\sum_{i=0}^{C}\left(\prod_{k=0}^{i-1}\alpha_k\right)\frac{\rho^i}{i!}\left(\sum_{j=0}^{C-i}\frac{\rho_h^j}{j!}\right)}$$
$$p_{\mathrm{hb}} = \frac{\sum_{i=0}^{C}\left(\prod_{k=0}^{i-1}\alpha_k\right)\frac{\rho^i}{i!}\cdot\frac{\rho_h^{C-i}}{(C-i)!}}{\sum_{i=0}^{C}\left(\prod_{k=0}^{i-1}\alpha_k\right)\frac{\rho^i}{i!}\left(\sum_{j=0}^{C-i}\frac{\rho_h^j}{j!}\right)}.$$

When $\alpha_0 = \cdots = \alpha_{K-1} = 1$ and $\alpha_K = \cdots = \alpha_C = 0$, this result is reduced to (1) and (2). Applying the traditional approach, we can also obtain some similar approximate results for call blocking probabilities. We will omit the formulas here.

We remark that the thinning schemes can be generalized to handle the call admission control problem in wireless multimedia networks with different prioritized services. For example, we can classify multimedia services into different priority levels according to the QoS requirements, then apply multiple thresholds for call admission control or choose different admission probabilities for different priority levels to reflect the QoS. We will present such studies in a separate paper.

$$p_{\mathrm{nb}}^t = \frac{\sum_{j=m}^{C} \frac{1}{j!}\left(\frac{\lambda+\lambda_h}{\mu_{\mathrm{av}}}\right)^m \left(\frac{\lambda_h}{\mu_{\mathrm{av}}}\right)^{j-m}}{\sum_{j=0}^{m} \frac{1}{j!}\left(\frac{\lambda+\lambda_h}{\mu_{\mathrm{av}}}\right)^j + \sum_{j=m+1}^{C} \frac{1}{j!}\left(\frac{\lambda+\lambda_h}{\mu_{\mathrm{av}}}\right)^m \left(\frac{\lambda_h}{\mu_{\mathrm{av}}}\right)^{j-m}}$$
$$p_{\mathrm{hb}}^t = \frac{\left(\frac{\lambda+\lambda_h}{\mu_{\mathrm{av}}}\right)^m \left(\frac{\lambda_h}{\mu_{\mathrm{av}}}\right)^{C-m}}{\sum_{j=0}^{m} \frac{1}{j!}\left(\frac{\lambda+\lambda_h}{\mu_{\mathrm{av}}}\right)^j + \sum_{j=m+1}^{C} \frac{1}{j!}\left(\frac{\lambda+\lambda_h}{\mu_{\mathrm{av}}}\right)^m \left(\frac{\lambda_h}{\mu_{\mathrm{av}}}\right)^{j-m}}.$$

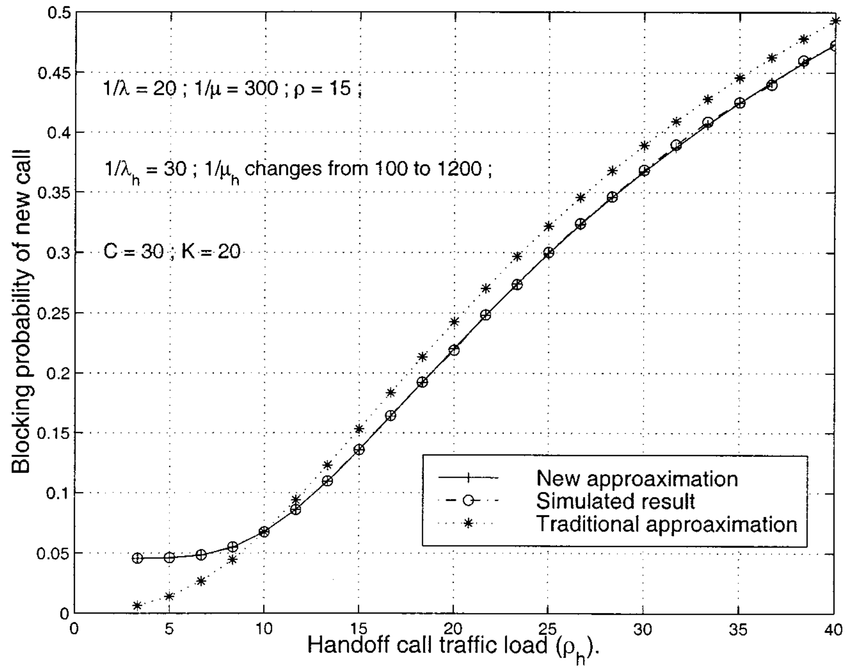


Fig. 6. New call blocking probability in the new call bounding scheme.

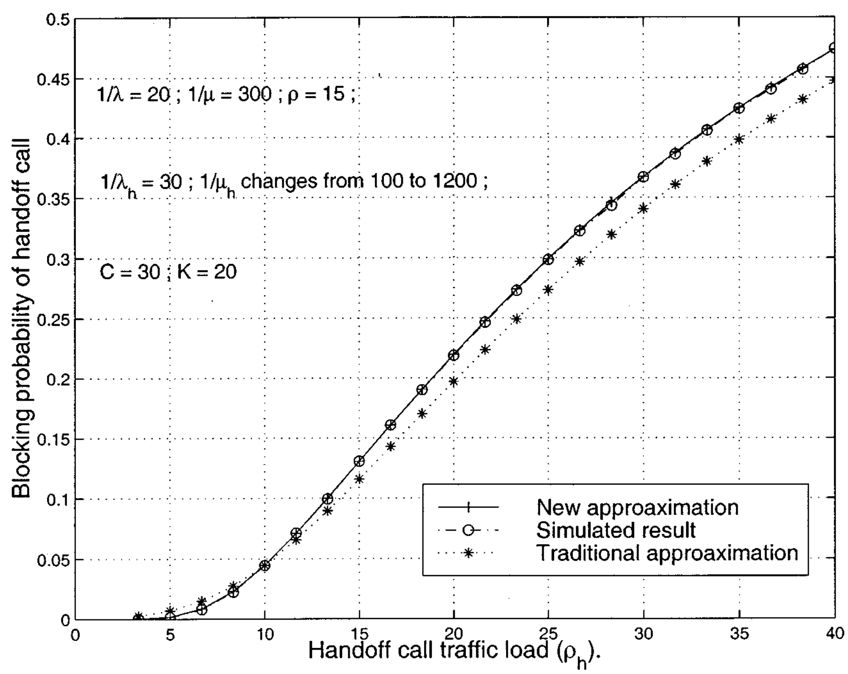


Fig. 7. Handoff call blocking probability in the new call bounding scheme.

III. Numerical Results

In this section, we present the simulation results for comparison purposes. They will show how much discrepancy may be caused by using our approximate model and the traditional approach (which does not distinguish between new calls and handoff calls). On the other hand, comparison will also show how much accuracy our new approach can achieve.

First, we investigate the *new call bounding scheme*. We choose the following set of parameters: $C = 30, K = 15, \lambda = \lambda_h = 1/30, \mu_h = 1/450$, and μ is varying from 1/600 to 1/200. Figs. 4 and 5 depict the new call blocking probability and handoff call blocking probability, respectively, under different new call traffic load. In Figs. 4 and 5, handoff call traffic load is given as $\rho_h = 15$. It is observed that when the handoff call traffic load is higher than the new call traffic load (i.e., $\rho_h > \rho$), the traditional approach will overestimate the new call blocking probability (see Fig. 4), while it will underestimate the handoff call blocking probability (see Fig. 5). On the other hand, when the handoff call traffic load is lower than the new call traffic load (i.e., $\rho_h < \rho$), the traditional approach will underestimate the new call blocking probability and overestimate the handoff call blocking probability.

A similar conclusion can be drawn from Figs. 6 and 7, which show those blocking probabilities under different handoff call traffic load ($\lambda = 1/20, \lambda_h = 1/30, K = 20, \mu = 1/300$ while

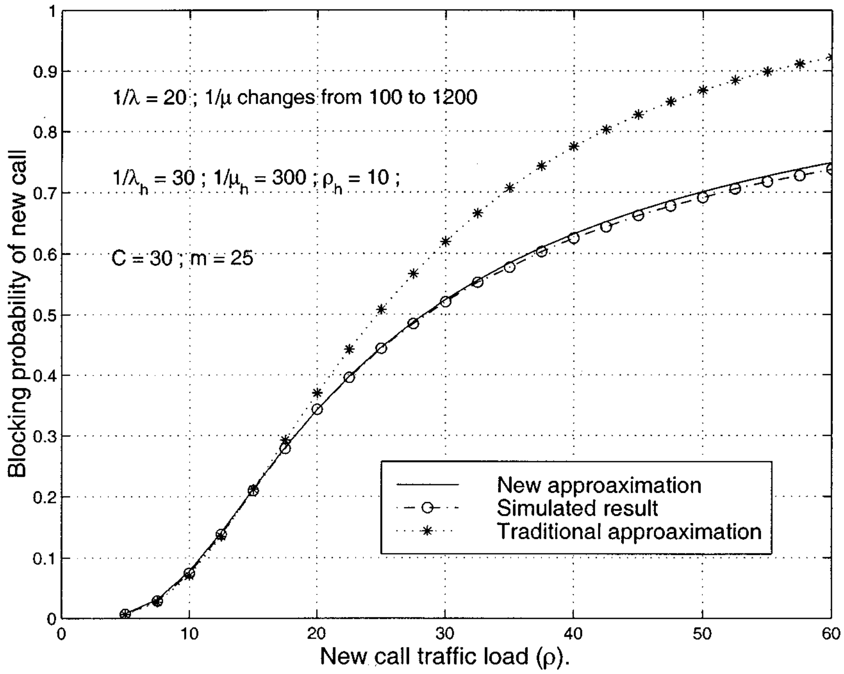


Fig. 8. New call blocking probability in the cutoff priority scheme.

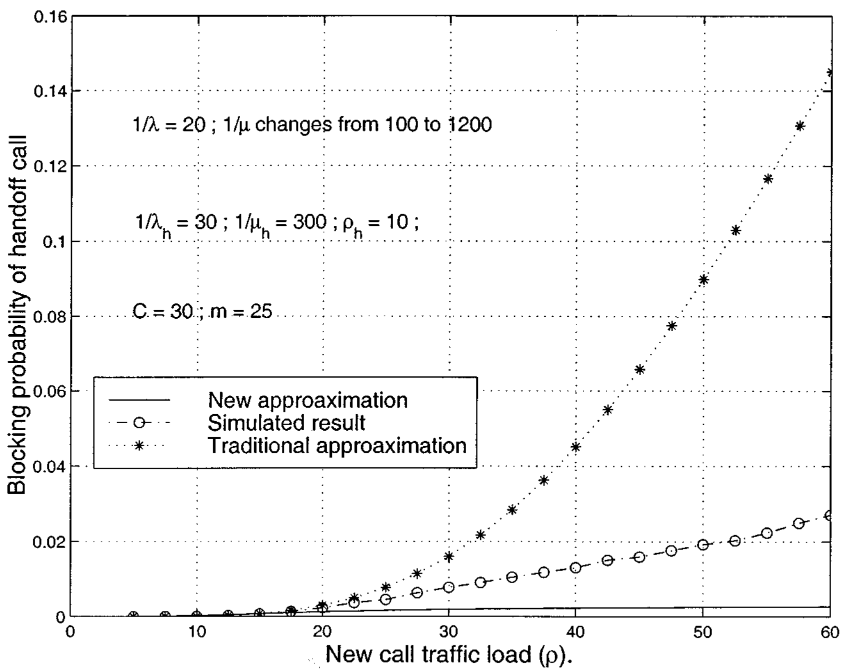


Fig. 9. Handoff call call blocking probability in the cutoff priority scheme.

μ_h varies from 1/1200 to 1/100). In addition, we observe that when the handoff traffic load (ρ_h) increases, new call blocking probability obtained from Fig. 6 and handoff call blocking probability achieved from Fig. 7 tend to be the same value. That makes sense because the number of new calls may not be able to reach the bound if handoff calls contribute a heavy traffic load. Since the new call bound scheme will not have an impact on the new calls, the new calls and the handoff calls will sustain the same blocking probability. However, the traditional approach does not yield similar results on blocking probabilities of new calls and handoff calls.

In summary, the traditional approach may either underestimate or overestimate the blocking probabilities on a new call or handoff call. It may lead in practice to either overdimensioning the network or not meeting the design requirement. However, comparisons from the above figures show that our approach can easily overcome such inaccuracy.

Next, we compare the two approaches under the *cutoff priority scheme*. A special case of our new call thinning scheme is applied: set $\beta_0 = \cdots = \beta_{m-1} = 1$ and $\beta_m = \cdots = \beta_C = 0$. Then the new call thinning scheme becomes the cutoff priority scheme. We choose the following parameters: $C = 30, m = 25, \lambda = 1/20$, $\lambda_h = 1/30, \mu_h = 1/300$, while μ varies from 1/1200 to 1/100.

Figs. 8 and 9 show the blocking probabilities for new calls and for handoff calls versus the new call traffic load, respectively.

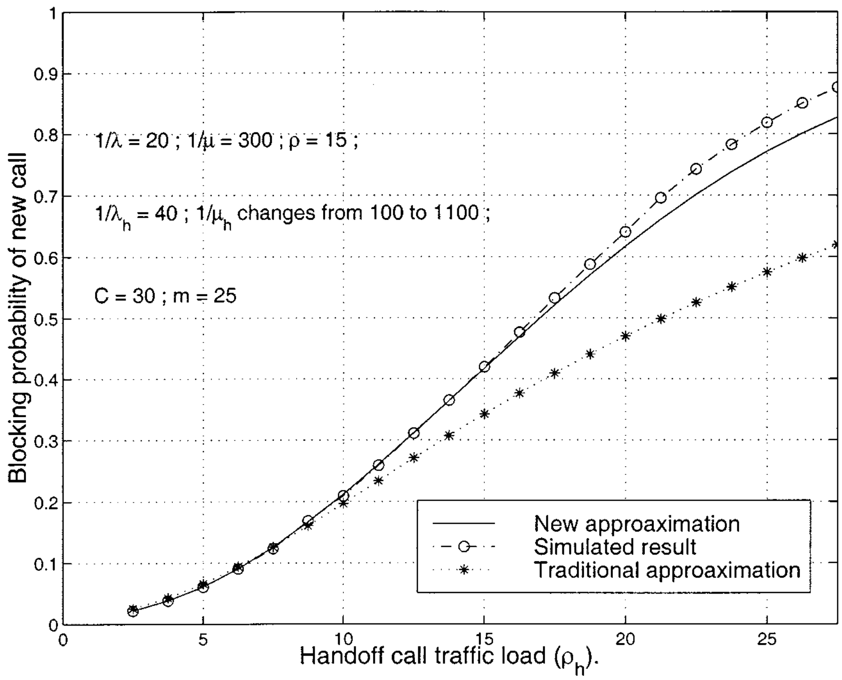


Fig. 10. New call blocking probability in the cutoff priority scheme.

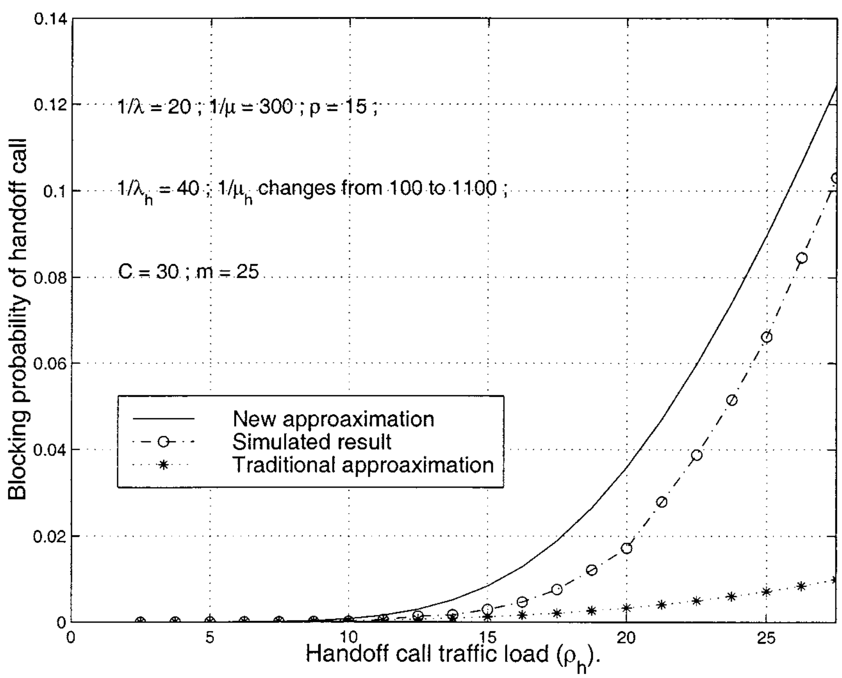


Fig. 11. Handoff call call blocking probability in the cutoff priority scheme.

We observe that when the new and handoff calls have significant different average values, the traditional approach for new call blocking probability gives significant discrepancy, while the result obtained from our approximate approach matches the simulated result very well. We also note that the traditional approximation overestimates the handoff call blocking probability while the new approximation underestimates it.

Figs. 10 and 11 show the blocking probabilities for new calls and for handoff calls versus the handoff call traffic load, respectively. We observe that our new approximate curve is much closer to the simulated result than the traditional approximate one for the new call blocking probability, especially in the range of interest (lower than 40%). We also obtain a much better result for the handoff call blocking probability.

In Figs. 12 and 13, we make another comparison: we change the new call arrival rate instead of the channel holding time. We choose the parameters as follows: $C = 30, m = 25, \mu = 1/300, \lambda_h = 1/30, \mu_h = 1/450$, and λ varies from 1/60 to 1/12. They show that we can obtain very accurate results for the new call blocking probability if our approximation approach is deployed. However, we also observe that both approaches are not good enough for the handoff call blocking probability.

In sum, our approximation approach can achieve much better results than the traditional one, especially for the new call

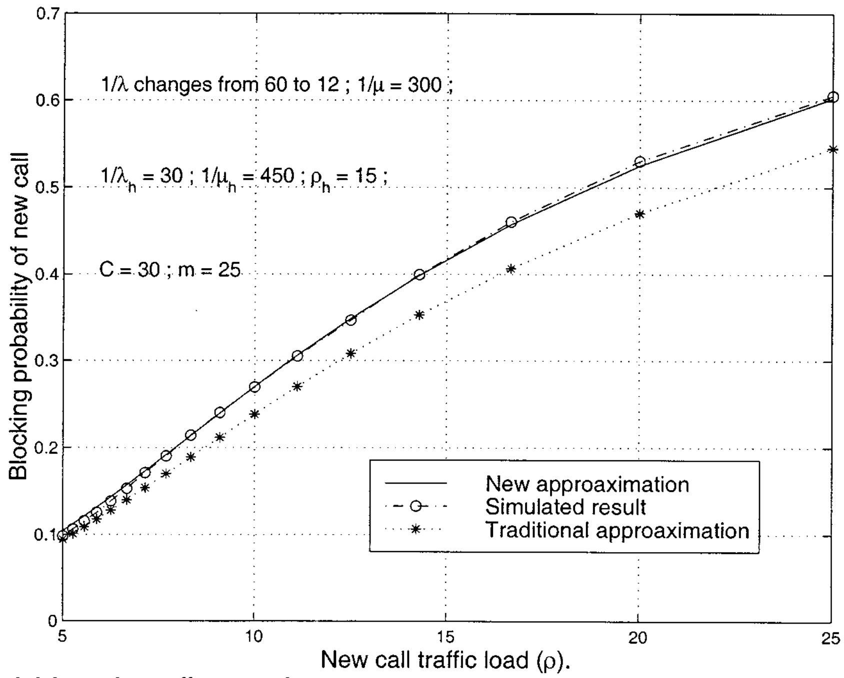


Fig. 12. New call blocking probability in the cutoff priority scheme.

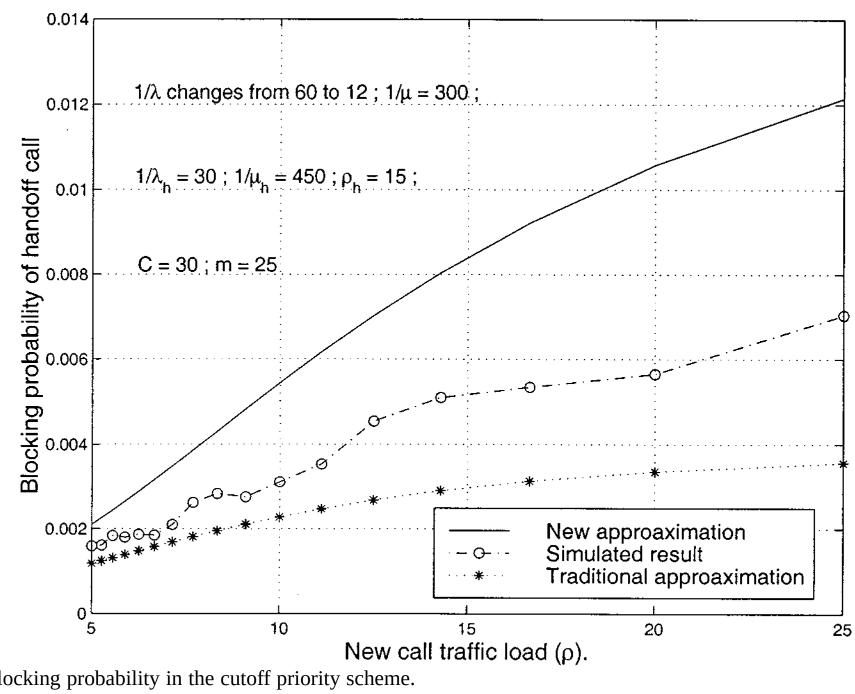


Fig. 13. Handoff call call blocking probability in the cutoff priority scheme.

blocking probability. This paper calls again for the necessity of reexamining the classical analytical results in traffic theory, which are used for the analysis and design of wireless mobile networks.

IV. Conclusion

In this paper, we investigate the call admission control strategies for the wireless networks. We point out that when the average channel holding times for new calls and handoff calls are significantly different, the traditional one-dimensional Markov chain model may not be suitable; two-dimensional Markov chain theory must be applied. We also propose a new approximation approach to reduce the computational complexity. It seems that the new approximation performs much better than the traditional approach. Future work includes research on finding out how good this new approximation is analytically.

Acknowledgment

The authors would like to express their gratitude to the reviewers for their comments, which greatly enhanced the quality of this paper.

REFERENCES

[1] C. Chang, C. J. Chang, and K. R. Lo, "Analysis of a hierarchical cellular system with reneging and dropping for waiting new calls and handoff calls," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1080–1091, 1999.

[2] D. C. Cox, "Wireless personal communications: What is it?," *IEEE Personal Commun. Mag.*, pp. 20–35, Apr. 1995.

[3] E. D. Re, R. Fantacci, and G. Giambene, "Handover queueing strategies with dynamic and fixed channel allocation techniques in low earth orbit mobile satellite systems," *IEEE Trans. Commun.*, vol. 47, no. 1, pp. 89–102, 1999.

[4] D. E. Everitt, "Traffic engineering of the radio interface for cellular mobile networks," *Proc. IEEE*, vol. 82, no. 9, pp. 1371–1382, 1994.

[5] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling for PCS networks," *IEEE Trans. Commun.*, vol. 47, pp. 1062–1072, July 1999.

[6] Y. Fang, I. Chlamtac, and Y. B. Lin, "Channel occupancy times and handoff rate for mobile computing and PCS networks," *IEEE Trans. Comput.*, vol. 47, pp. 679–692, June 1998.

[7] R. A. Guerin, "Queueing-blocking system with two arrival streams and guard channels," *IEEE Trans. Commun.*, vol. 36, no. 2, pp. 153–163, 1988.

[8] D. Grillo, R. A. Skoog, S. Chia, and K. K. Leung, "Teletraffic engineering for mobile personal communications in ITU-T work: The need to match practice and theory," *IEEE Personal Commun.*, vol. 5, pp. 38–58, Dec. 1998.

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

[10] B. Jabbari, "Teletraffic aspects of evolving and next-generation wireless communication networks," *IEEE Personal Commun. Mag.*, pp. 4–9, Dec. 1996.

[11] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *IEEE Personal Commun.*, vol. 3, pp. 10–31, June 1996.

[12] L. Kleinrock, *Queueing Systems: Theory*. New York: Wiley, 1975, vol. I.

[13] M. D. Kulavaratharasah and A. H. Aghvami, "Teletraffic performance evaluation of microcellular personal communication networks (PCN's) with prioritized handoff procedures," *IEEE Trans. Veh. Technol.*, vol. 48, no. 1, pp. 137–152, 1999.

[14] V. K. N. Lau and S. V. Maric, "Mobility of queued call requests of a new call-queueing technique for cellular systems," *IEEE Trans. Veh. Technol.*, vol. 47, no. 2, pp. 480–488, 1998.

[15] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, 1994.

[16] B. Li, C. Lin, and S. Chanson, "Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks," *Wireless Networks*, vol. 4, no. 4, Aug. 1998.

[17] Y. B. Lin, S. Mohan, and A. Noerpel, "Queueing priority channel assignment strategies for handoff and initial access for a PCS network," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 704–712, 1994.

[18] P. V. Orlik and S. S. Rappaport, "A model for teletraffic performance and channel holding time characterization in wireless cellular communication with general session and dwell time distributions," *IEEE J. Select. Areas Commun.*, vol. 16, no. 5, pp. 788–803, 1998.

[19] R. Ramjee, D. Towsley, and R. Nagarajan, "On optimal call admission control in cellular networks," *Wireless Networks*, vol. 3, pp. 29–41, 1997.

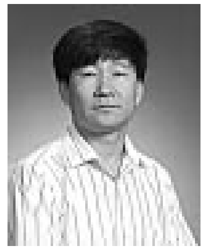
[20] S. S. Rappaport, "Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms," in *Proc. Inst. Elect. Eng. I*, vol. 140, 1993, pp. 389–401.

[21] M. Shafi, A. Hashimoto, M. Umehira, S. Ogose, and T. Murase, "Wireless communications in the twenty-first century: A perspective," *Proc. IEEE*, vol. 85, no. 10, pp. 1622–1638, 1997.

[22] W. Stallings, *High-Speed Networks: TCP/IP and ATM Design Principles*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[23] S. Tekinay and B. Jabbari, "Handover and channel assignment in mobile cellular networks," *IEEE Commun. Mag.*, pp. 42–46, Nov. 1991.

[24] C. H. Yoon and C. K. Un, "Performance of personal portable radio telephone systems with and without guard channels," *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 911–917, 1993.



Yuguang Fang (S'92–M'94–SM'99) received the B.S. and M.S. degrees in mathematics from Qufu Normal University, Qufu, Shandong, China, in 1984 and 1987, respectively, the Ph.D. degree in systems and control engineering from the Department of Systems, Control and Industrial Engineering, Case Western Reserve University, Cleveland, OH, in 1994, and the Ph.D. degree in electrical engineering from Department of Electrical and Computer Engineering, Boston University, Boston, MA, in 1997.

From 1987 to 1988, he held research and teaching positions in both the Department of Mathematics and the Institute of Automation at Qufu Normal University. From September 1989 to December 1993, he was a Teaching/Research Assistant in the Department of Systems, Control and Industrial Engineering, Case Western Reserve University, where he was a Research Associate from January 1994 to May 1994. He held a postdoctoral position in the Department of Electrical and Computer Engineering, Boston University, from June 1994 to August 1995. From September 1995 to May 1997, he was a Research Assistant in the Department of Electrical and Computer Engineering, Boston University. From June 1997 to July 1998, he was a Visiting Assistant Professor in the Department of Electrical Engineering, University of Texas at Dallas. From July 1998 to May 2000, he was an Assistant Professor in the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark. Since May 2000, he has been an Assistant Professor in the Department of Electrical and Computer Engineering, University of Florida, Gainesville. His research interests span many areas, including wireless networks, mobile computing, mobile communications, automatic control, and neural networks. He has published more than 60 papers in refereed professional journals and conferences. He has also engaged in many professional activities. He has been actively involved with many professional conferences, such as ACM MobiCom'01, IEEE INFOCOM'98, INFOCOM'00, IEEE WCNC'02, WCNC'00 (Technical Program Vice-Chair), WCNC'99, and International Conference on Computer Communications and Networking (IC3N'98) (Technical Program Vice-Chair).

Prof. Fang is a member of the ACM. He is an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, and for *ACM Wireless Networks*,. In addition, he is a Feature Editor for "Scanning the Literature" in IEEE PERSONAL COMMUNICATIONS, an Area Editor for *ACM Mobile Computing and Communications Review*, and an Associate Editor for the *Journal on Wireless Communications and Mobile Computing*. He received the National Science Foundation Faculty Early Career Award in 2001. He is listed in Marquis *Who's Who in Science and Engineering*, *Who's Who in America*, and *Who's Who in the World*.

Yi Zhang received the B.S. degree in computer communication networks from Xi'An University of Electrical Science and Technology (Xidian University), Xi'An, ShaanXi, China, in 1993 and the M.S. degree in telecommunication networking from New Jersey Institute of Technology, Newark, in 2000.

His research interests include wireless communication networks and routing and signaling protocols for multiservice broadband networks. He was a System Engineer with Siemens-China Ltd. from July 1993 to December 1995. He was with Nokia-China Inc. from January 1996 to July 1998 as a Senior Networking System Engineer and later Team Leader on GSM systems. He joined the InterNetworking System (INS) Division of Lucent Technologies, Westford, MA, in 2000, where he is a Senior Software Engineer in the routing and signaling group.