Abstract — Providing the desired call blocking probability to not only new but also existing calls has been a challenge for wireless mobile network service providers. To satisfy different requirements for new and handoff call blocking probabilities, several call admission control (CAC) schemes have been proposed in the literature. Exact analysis of these schemes using two dimensional Markov chain is computationally intensive. Therefore under specific assumptions computationally efficient methods to analyze these systems using one dimensional Markov chain models have been considered. The “traditional” approach assumes that channel holding time for new and handoff calls have equal mean values. While the “normalized” approach relaxes this assumption, it is accurate only for the new call bounding CAC scheme. In this paper, we reevaluate the analytical methods for computing call blocking probabilities for several widely known call admission control schemes under more general assumptions by providing an easy to implement method. The numerical results show that when the average values of channel holding times for new and handoff calls are different, the proposed approach gives more accurate results when compared with the traditional and normalized methods based on one dimensional Markov chain modeling, while keeping the computational complexity low.

Keywords - call admission control (CAC); call blocking probability; computational complexity; performance evaluation; priority based CAC schemes; quality of service (QoS); resource allocation; wireless mobile networks

I. INTRODUCTION

As the demand to support non-voice and multimedia services increases, radio resource management to satisfy diverse quality of service (QoS) requirements becomes increasingly important. To guarantee acceptable QoS in multi-service mobile environments, network planners need to consider certain constraints that provide upper limits for the blocking probability of different service or call types. Many call admission control (CAC) schemes have been proposed to enable the network to provide the desired QoS requirements by limiting the number of admitted calls to that network in order to reduce congestion and avoid call dropping. In wireless networks, other aspects of CAC need to be considered due to user mobility. An accepted call may be dropped before it is terminated as a result of the mobile user moving from its current cell to another, during a handoff, if the cellular network is unable to assign a new channel to the call in the new cell to continue its service due to lack of resources. Since dropping an on-going call is generally more objectionable to a mobile user than blocking a new call request, a higher priority is normally assigned for handoff calls over the new calls in order to minimize the call dropping probability. On the other hand, reducing handoff call blocking by channel reservation or other means could increase blocking for new calls. There is therefore a trade off between these two QoS measures [1]. The problem of maintaining the service continuity and QoS guarantees to the multimedia applications during handoffs is exacerbated by the increasing use of microcells and picocells in contemporary wireless networks.

CAC for highspeed networks and wireless networks has been intensively studied in the past [2] and many handoff priority-based CAC schemes have been proposed [1], [3]–[8]. These can be classified into two broad categories:

1) Guard Channel (GC) Schemes: A set of guard channels are reserved for handoff calls. There are four different schemes.
   (a) The cutoff priority scheme blocks a new call if the number of free channels is less than the number of guard channels reserved for handoff calls [3][7][8].
   (b) The fractional guard channel scheme admits a new call with certain probability that depends on the number of busy channels in the cell. It is first proposed by Ramjee et al. [1].
   (c) The new call bounding scheme limits the number of new calls admitted to the cell to some number less than the total number of available channels.
   (d) The rigid division-based scheme divides all channels available in a cell into two groups: one for common use and the other only for handoff calls [9].

2) Queuing Priority (QP) Schemes: Calls are accepted whenever there are free channels; otherwise they are queued with certain rearrangements in the queue.

Although various combinations of the above schemes are possible, we concentrate only on GC schemes as we limit our discussion to the problems of call acceptance and dropping.
In the literature (e.g., [4][5]), for reduced computational complexity, a one dimensional Markov chain is commonly used to obtain the blocking probabilities for new and handoff calls under the assumption that the channel holding times for both types of calls are identically distributed with the same parameters. This assumption may not be appropriate as new and handoff calls may have different average channel holding times. Analysis in [10] accounts for the more general case, but the approximation employed still leads to significant discrepancies with the exact solutions. In [11], a product form solution is obtained by modeling a multi-cell wireless network employing a hybrid GC/QP scheme with transfer of unsuccessful requests to neighboring cells as a network of queues. The solution is however restrictive to the protocol considered. For analysis of GC schemes, it appears that using multidimensional Markov chain models directly is the only means to obtain accurate solutions. Nevertheless, this method suffers from the curse of dimensionality that results in very high computation cost for large systems. Therefore it is still desirable to come up with approximate solutions that have high accuracy.

This paper is organized as follows. In the next section, we examine three of the widely known CAC schemes by evaluating their performances using the traditional and the normalized analytical methods proposed in the literature under the assumption that the new and handoff calls have different average channel holding times. In Section III we present a new analytical method that yields more accurate approximations than the traditional or normalized methods. Numerical results obtained using the traditional, normalized, proposed, and the direct methods are presented and compared in Section IV. We conclude the paper in Section V.

II. EXISTING METHODS TO ANALYSE PERFORMANCE OF CALL ADMISSION CONTROL SCHEMES

Let $\lambda_n$ and $\lambda_h$ respectively denote the arrival rates for new and handoff calls, and $1/\mu_n$ and $1/\mu_h$ respectively denote the average channel holding time for new and handoff calls. Let $C$ denote the total number of channels in a cell. We assume that the arrival process for new and handoff calls are Poisson, and the channel holding times for new calls and handoff calls are exponentially distributed.

Assuming that both new and handoff calls have the same channel holding time distributions and average values, the system model is approximated by a one dimensional Markov chain with a fixed average channel holding time for the total cell traffic. We refer to this method to derive analytical results for blocking performance as the traditional approach. To improve the inaccurate results obtained when the above assumption is no longer valid, Fang and Zhang [10] proposed normalizing the average service time (by equalizing it to unity) for each new and handoff traffic so that it becomes identical for both streams. Although this approximation, which we call the normalized approach, seems to give better accuracy than the traditional approach, it is still inaccurate especially for schemes like cutoff priority and fractional guard channel as shown in [10].

A. New Call Bounding Scheme

The idea behind this scheme is to accept fewer new calls instead of dropping the ongoing calls in the future. Analytical results for the blocking probabilities of new calls, $p_{nb}$, and handoff calls, $p_{hb}$ of this scheme have been derived in [10], where an approximation is developed by normalizing the average service time for new and handoff calls. This normalization allows the arriving traffic for each type of calls to be scaled appropriately and the blocking probabilities to be related to the traffic intensities.

As shown in [10], the traditional approach can yield significantly inaccurate results when the channel holding times for new and handoff calls have different average values. The normalized approach suggested in [10] overcome this inaccuracy by exploiting the symmetric nature and the product form of the detailed balance equations that characterizes the new call bounding scheme. We will show in the following sections that both approximations mentioned above are not good enough for obtaining new and handoff call blocking probabilities at an acceptable accuracy for other CAC schemes.

B. Cutoff Priority Scheme

This scheme works as follows. Let $m$ denote the channel occupancy threshold, upon a new arrival. If the total number of busy channels is less than $m$ when a new call arrives, the call is accepted; otherwise, the new call is blocked. A handoff call is always accepted unless no channel is available upon its arrival. This scheme has been extensively studied using one dimensional Markov chain modeling under the assumption that the average channel holding times of new and handoff calls are equal. However, since the state flows are no longer symmetric as is the case for the new call bounding scheme, the one dimensional Markov chain model provides inaccurate results when the assumption of equal channel holding time mean values for new and handoff calls does not hold true.

The blocking probabilities for new calls, $p_{nb}^n$, and handoff calls, $p_{hb}^n$, are given below for the cutoff priority scheme as derived in [10] using the normalized approach.

$$p_{nb}^n = \sum_{j=0}^{\infty} \frac{(\rho_a + \rho_h)^n \cdot \rho_h^{j-m}}{j!}$$

$$p_{hb}^n = \sum_{j=0}^{\infty} \frac{(\rho_a + \rho_h)^n \cdot \rho_a^{j-m}}{j!}$$

where $\rho_a = C! \sum_{j=0}^{\infty} \frac{(\rho_a + \rho_h)^n \cdot \rho_a^{j-m}}{j!}$ and $\rho_h = C! \sum_{j=0}^{\infty} \frac{(\rho_a + \rho_h)^n \cdot \rho_h^{j-m}}{j!}$.
On the other hand, the corresponding results for the traditional approach are as follows:

\[
p'_{nb} = \frac{\sum_{j=0}^{C} \frac{C}{j!} \left( \frac{\lambda_n + \lambda_h}{\mu_{av}} \right)^j \mu_{av}^{j-m} \lambda_m^{j-m}}{\sum_{j=0}^{m} \frac{C}{j!} \left( \frac{\lambda_n + \lambda_h}{\mu_{av}} \right)^j + \sum_{j=m+1}^{C} \frac{C}{j!} \left( \frac{\lambda_n + \lambda_h}{\mu_{av}} \right)^j \mu_{av}^{j-m} \lambda_m^{j-m}}
\]

(3)

\[
p'_{hb} = \frac{\sum_{j=0}^{C} \frac{C}{j!} \left( \frac{\lambda_n + \lambda_h}{\mu_{av}} \right)^j \mu_{av}^{j-m} \lambda_m^{j-m}}{\sum_{j=0}^{m} \frac{C}{j!} \left( \frac{\lambda_n + \lambda_h}{\mu_{av}} \right)^j + \sum_{j=m+1}^{C} \frac{C}{j!} \left( \frac{\lambda_n + \lambda_h}{\mu_{av}} \right)^j \mu_{av}^{j-m} \lambda_m^{j-m}}
\]

(4)

where

\[
\frac{1}{\mu_{av}} = \frac{\lambda_n}{\lambda_n + \lambda_h} \cdot \frac{1}{\mu_n} + \frac{\lambda_h}{\lambda_n + \lambda_h} \cdot \frac{1}{\mu_h} = \frac{\rho_n + \rho_h}{\lambda_n + \lambda_h}
\]

(5)

C. Fractional Guard Channel Scheme

This scheme admits new calls with certain probabilities, \( \beta_n \), that depend on the number of busy channels denoted by \( i \). The new call stream is smoothly throttled as the network traffic is building up. An arriving handoff call will always be admitted unless there are no channels available, in which case all calls will be blocked. Obviously, when \( \beta_0 = \ldots = \beta_{m-1} = 1 \) and \( \beta_m = \ldots = \beta_C = 0 \), it becomes the cutoff priority scheme.

Let \( q(c) \) denote the equilibrium channel occupancy probability when exactly \( c \) channels are occupied in a cell. Using a corollary given in [12], we can obtain the occupancy probabilities \( q(c) \), \( c = 0, \ldots, C \), which satisfy the following recursive equation.

\[
\frac{\lambda_n \cdot \beta_{c+1}}{\mu_n} + \frac{\lambda_h}{\mu_h} \cdot q(c-1) = c \cdot q(c), c = 1, \ldots, C
\]

(6)

Yet, as given in [10], normalizing the average service time for both types of calls can transform (6) to

\[
(\rho_n \cdot \beta_{c+1} + \rho_h) \cdot q(c-1) = c \cdot q(c), c = 1, \ldots, C
\]

(7)

giving us a general expression for all GC schemes. Solving for \( q(0) \) in the equation \( \sum_{j=0}^{C} q(j) = 1 \), we obtain

\[
q(j) = \frac{\prod_{k=0}^{j-1} (\rho_n \cdot \beta_k + \rho_s)}{j!} \cdot q(0), 1 \leq j \leq C
\]

(8)

where

\[
q(0) = \left[ 1 + \sum_{j=1}^{C} \frac{\prod_{k=0}^{j-1} (\rho_n \cdot \beta_k + \rho_s)}{j!} \right]^{-1}
\]

(9)

Following (8) and (9) the blocking probabilities for new and handoff calls are obtained as follows:

\[
p_{nb} = \sum_{j=0}^{C} \left( 1 - \beta_j \right) \cdot q(j), \beta_C = 0
\]

(10)

\[
p_{hb} = q(C)
\]

(11)

III. THE PROPOSED PERFORMANCE EVALUATION METHOD

Although the normalized approach [10] provides better approximation than the traditional one, it is still inaccurate for many GC schemes. Therefore, in order to provide more accurate results while keeping the computational complexity low, we present the following novel performance evaluation method for GC schemes.

Since it is crucial to find an approximation that overcomes the curse of dimensionality when the state dimension is large, it is inevitable to attempt reducing the two dimensional Markov chain model to a one dimensional one. However, we convert (6) to (13) by replacing the average channel holding times for both new and handoff calls with an approximation of the average channel holding time for the whole system. This idea was first proposed by Gerstb and Lee [13] for the stochastic knapsack problem. We call this approximation, the average effective channel holding time and denote it by \( 1/\mu_{eff} \). Inspired by the well known Little’s theorem, \( \mu = \lambda / N \), they defined \( \mu_{eff} \) as the ratio of expected number of both types of call arrivals to the expected number of occupied channels, which is given by

\[
\mu_{eff} = \frac{\sum_{c=1}^{C} \left( \lambda_n \cdot \beta_j \cdot q(j) \right) + \sum_{c=1}^{C} \left( \lambda_h \cdot q(j) \right)}{\sum_{j=0}^{C} q(j)}
\]

(12)

We now simply approximate the occupancy probabilities by setting \( q(c) = \hat{q}(c) \), \( c = 0, \ldots, C \) and using the updated recursive formula given below.

\[
(\lambda_n \cdot \beta_{c+1} + \lambda_h) \cdot \hat{q}(c-1) = c \cdot \mu_{eff} \cdot \hat{q}(c), c = 1, \ldots, C
\]

(13)

Solving for \( \hat{q}(0) \) in the equation \( \sum_{j=0}^{C} \hat{q}(j) = 1 \), we obtain

\[
\hat{q}(j) = \frac{\prod_{k=0}^{j-1} (\lambda_n \cdot \beta_k + \lambda_h)}{\mu_{eff} \cdot j!} \cdot \hat{q}(0), 1 \leq j \leq C
\]

(14)

where

\[
\hat{q}(0) = \left[ 1 + \sum_{j=1}^{C} \frac{\prod_{k=0}^{j-1} (\lambda_n \cdot \beta_k + \lambda_s)}{j!} \right]^{-1}
\]

(15)

In their knapsack problem approach, Gerstb and Lee [13] suggested obtaining \( \mu_{eff} \) using (12) by replacing \( q(c) \) with \( \hat{q}(c) \) and updating the approximate equilibrium occupancy probabilities iteratively, using (14) and (15) until each \( \hat{q}(c) \) changes by no more than \( \varepsilon \) for all \( c = 0, \ldots, C \), where \( \varepsilon \) is small.
Even though their approach did not emphasize starting with an appropriate initial value for each approximate equilibrium occupancy probability, this becomes a problem since more than one set of probabilities can satisfy (14) and (15). What makes a set to be the unique solution depends on the values of arrival rates and average channel holding times of both types of calls. Therefore, it is important for those values to be included in the computation to obtain the approximate results. The call arrival rates for both types of calls are already taken into consideration in (12), (14) and (15); however the average channel holding time for both types of calls is not considered directly as the average channel effective time, $1/\mu_{eff}$, is used for the calculation of the same set of equations.

Therefore, we suggest setting the approximate equilibrium occupancy probabilities initially with the values obtained by the normalized approach proposed by Fang and Zhang [10] and calculating $\mu_{eff}$ using Little’s theorem to obtain the new and handoff call blocking probabilities analytically.

To sum up, here is how our proposed method looks like:

1. Initialize equilibrium occupancy probabilities ($\hat{q}(c)$ for $c = 0,...,C$) by setting them to the corresponding values obtained from the normalized approach.
2. Calculate $\mu_{eff}$ using (12) by replacing $q(c)$’s with $\hat{q}(c)$’s.
3. Calculate $\hat{q}(c)$ for $c = 0,...,C$ using (14) and (15).
4. Obtain new and handoff call blocking probabilities using $\hat{q}(c)$.

Despite Gerstl and Lee suggested an iterative approach, our method is in closed form since $\hat{q}(c)$ no longer changes once Little’s formula is applied. After the approximate equilibrium occupancy probabilities are calculated we can obtain the new and handoff call blocking probabilities as follows:

$$
\begin{align*}
{p_{off}} &= 1 - \frac{\sum_{j=0}^{C-1} \lambda_n \cdot \beta_j \cdot \hat{q}(j)}{\sum_{j=0}^{C} \hat{q}(j)} \quad (16) \\
{p_{h}} &= 1 - \frac{\sum_{j=0}^{C} \lambda_h \cdot \hat{q}(j)}{\sum_{j=0}^{C} \hat{q}(j)} \quad (17)
\end{align*}
$$

**IV. NUMERICAL RESULTS**

In this section we present numerical results of performance evaluations using our novel approximation method presented in Section III and compare them with those obtained using the existing traditional and the normalized approaches based on one dimensional Markov chain models. We also obtain accurate results using a multidimensional Markov chain model for comparison purposes. This is accomplished using a numerical method called direct (also known as LU decomposition) to calculate the exact values of the performance metrics, as the results are labeled as “direct method”.

The results not only show that the normalized and the traditional approach can deviate significantly from the accurate results from the direct approach, but more importantly, show that our new approach can achieve results very close to the accurate direct results.

We do not give the results here for the new call bounding scheme, since the normalized approach can overcome the inaccuracy of the traditional approach by exploiting the symmetric nature of the state transitions and thus leaves little room for improvement. However, we focus on the other schemes considered above for which this property does not apply.

Since fractional guard channel is a generalization of the cutoff priority scheme, we evaluate only the cutoff priority scheme here due to space constraints as same property applies for both schemes. Before starting to examine this scheme, we should determine the range of values to be applied to parameters like $\lambda_n$, $\lambda_h$, $1/\mu_n$ and $1/\mu_h$ in the performance evaluations in order to reflect practical situations.

It is generally accepted in the literature to set the new call arrival rate, $\lambda_n$, and the average channel holding time, $1/\mu_n$, in proportion with the handoff call arrival rate, $\lambda_h$, and the average channel holding time, $1/\mu_h$, respectively. Following the scenarios that have been considered in the literature, we apply the arrival rates and average channel holding times for new and handoff calls as shown in Table I in the performance evaluations.

To put it simply, we assume both ratios to have values changing within the range of 4 and 0.5 in order to cover scenarios commonly considered in the literature. We further set the number total number of channels, $C = 30$ and the channel occupancy threshold $m = 25$. We evaluate the approaches mentioned above by grouping the possible scenarios into four different cases with the parameter values as shown in Table I.

<table>
<thead>
<tr>
<th>Cases</th>
<th>$\lambda_n$</th>
<th>$\lambda_h$</th>
<th>$1/\mu_n$ (sec)</th>
<th>$1/\mu_h$ (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1/5</td>
<td>1/20</td>
<td>800 – 100</td>
<td>200</td>
</tr>
<tr>
<td>II</td>
<td>1/10</td>
<td>1/20</td>
<td>800 – 100</td>
<td>200</td>
</tr>
<tr>
<td>III</td>
<td>1/20</td>
<td>1/20</td>
<td>800 – 100</td>
<td>200</td>
</tr>
<tr>
<td>IV</td>
<td>1/40</td>
<td>1/20</td>
<td>800 – 100</td>
<td>200</td>
</tr>
</tbody>
</table>

A different ratio of new and handoff call arrival rates is chosen for each case where average channel holding times of new calls are being varied relative to that of handoff calls within the limits of the proportions given above. The new call traffic load is varied by changing the average channel holding times of new calls to obtain handoff call blocking probabilities within the range of 0.01 and 0.1, which is the interval of interest to provide QoS guarantees for CAC schemes.
In Figs. 1 to 8, the new and handoff call blocking probabilities computed by the mentioned approaches are shown for each case given in Table I. We observe that when the new and handoff calls have significantly different average channel holding times, the traditional and the normalized approaches result in significant discrepancies compared to the direct method, especially for handoff call blocking probability which is overestimated by the former approach while it is underestimated by the latter one. However, the results obtained by the proposed approach can overcome such inaccuracy.

V. CONCLUSION

In this paper we have examined various call admission control schemes in wireless mobile networks to evaluate their performances analytically by using a one dimensional Markov chain model. When the average channel holding times for new and handoff calls are significantly different, we showed that the traditional and the normalized approaches may not be appropriate to use due to their discrepancies in comparison with the exact results. Although using a two dimensional Markov chain model could solve this problem and yield exact results, it gives rise to another problem known as the curse of dimensionality since the dimension of the state space in such a model can increase very quickly. With the objective of developing a practical solution to this problem, we have proposed a new method, which gives more accurate results when compared to the existing approximation approaches while keeping the computational complexity low.

As computational complexity plays an important role in real time applications, we believe a better approximation method with low computational complexity for performance evaluation of CAC schemes will help motivate the practical implementation of these schemes in the future. We are extending this work to include analytical performance evaluation of multi-service models with multiple channel requests, considering that the level of relative prioritization provided to different service types with different QoS requests is specified by relative blocking/dropping probabilities.

ACKNOWLEDGMENT

This work was supported by a grant from Bell Canada under the Bell University Laboratories program, and by the Canadian Natural Sciences and Engineering Research Council under grant STPGP 269872-03.
REFERENCES


