Analysis of a Cutoff Priority Cellular Radio System with Finite Queueing and Reneging/Dropping

Chung-Ju Chang, *Senior Member, IEEE*, Tian-Tsair Su, and Yueh-Yiing Chiang

Abstract— Queueing of new or handoff calls can minimize blocking probabilities or increase total carried traffic. This paper investigates a new cutoff priority cellular radio system that allows *finite* queueing of both new and handoff calls. We consider the *reneging* from the system of queued new calls due to caller impatience and the *dropping* of queued handoff calls by the system as they move out of a handoff area before being accomplished successfully. We use signal-flow graphs and Mason's formula to obtain the blocking probabilities of new and handoff calls and the average waiting times. Moreover, an optimal cutoff parameter and appropriate queue sizes for new and handoff calls are numerically determined so that a proposed overall blocking probability is minimized.

I. INTRODUCTION

IN A CELLULAR radio system, the blocking probabilities of new and handoff calls should be depressed as much as possible so as to improve the perceived service quality or increase the carried traffic load. In several recent papers [5]-[7], a system with a cutoff priority channel allocation strategy involving queueing of new calls has been proposed to minimize the blocking probability of handoff calls and increase the total carried traffic; a system that provides guard channels and a waiting queue for handoff calls to achieve a higher probability of successful handoffs has also been studied.

In [5], Guèrin presented a novel approach to the study of a multichannel cutoff priority cellular radio system, in which the queue size for *new* calls is *infinite* and the queued calls never *renege*. He considered two Poisson arrival streams with distinct arrival rates and the same exponential service time distributions for new and handoff calls, and obtained simple closed-form expressions for state probabilities, in which the signal-flow graph approach and then Mason's formula were utilized. This analytical method can be applied to a system with a finite queue. In the case of a system with a finite queue, however, no simple closed-form expressions for state probabilities can be found; instead, algorithmic numerical methods must be used to handle the corresponding computational problem. Moreover, the computational complexity increases with the capacity of the system buffer. In [6], [7], Hong and Rappaport described appropriate analytical models and

derived performance measures for a cellular mobile telephone system with *infinite* queueing of *handoff* calls; the performance measures included blocking probability, forced termination probability, and fraction of incomplete new calls.

Queueing of both new call and handoff calls can increase total carried traffic as well as minimize blocking probabilities. Therefore, as an alternative to the systems proposed in [5] and [6], [7], in this paper we investigate a new cutoff priority cellular radio system with *finite* queueing of both *new* and *handoff* calls. In addition, we also take into account the *reneging* of queued new calls due to caller impatience [2,4] and the *dropping* of queued handoff calls as they move out of the handoff area before being accomplished successfully [6], [9]. Such a cellular radio system is practical because finite buffering is more realistic than infinite buffering and because the related call-control packets are usually carried out on a separate control channel [5].

Our analysis is via a two-dimensional Markov chain approach. The state probabilities can be obtained computationally without any problem since the system possesses a quasi birth-death Markovian property [14]. We derive blocking probabilities for new and handoff calls, which are defined to contain their corresponding reneging and dropping probabilities, via the application of *signal-flow graphs* and *Mason's formula* [1], [8]; we also obtain average waiting times for new and handoff calls. Moreover, we heuristically define a cost function to investigate the optimal cutoff parameter and the suitable queue sizes for new and handoff calls.

This paper is organized as follows. The assumptions upon which our analysis rests are presented in Section II. In Section III, we derive the blocking probabilities of new and handoff calls by using signal-flow graphs and Mason's formula and obtain the average waiting times for new and handoff calls. We also provide details on how Mason's formula is numerically carried out in our problem and address the tractability of the numerical computation. In Section IV, some numerical examples are discussed; and overall blocking probability is proposed as a cost function for determining an optimal priority cutoff parameter and suitable queue sizes for new and handoff calls. Finally, concluding remarks are given in Section V.

II. SYSTEM MODEL

A conceptual model of the new cutoff priority cellular mobile radio system is shown in Fig. 1. The model follows those described in [5], [6], except that it considers finite queueing of both new and handoff calls and reneging/dropping of waiting calls. The assumptions involved in this model are
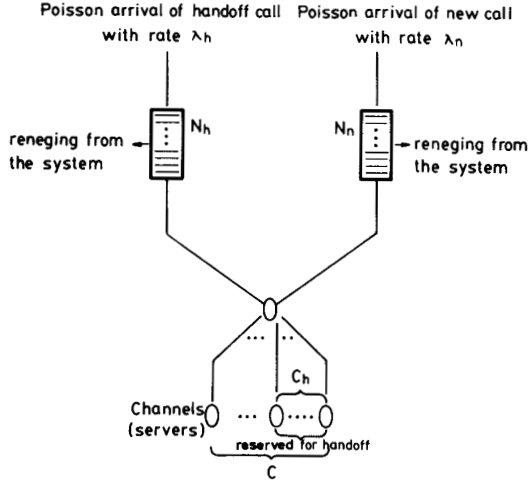
Fig. 1. The conceptual model of the new cellular mobile telephone system.

stated below.

1) The system has inputs of new and handoff calls generated according to a Poisson distribution with mean rates of λ_n and λ_h, respectively.

2) The unencumbered conversation time of a call, denoted by T_m, is assumed to be exponentially distributed with mean $1/\mu_m$.

3) The time spent in a cell by the mobile associated with a successful new (handoff) call, denoted by T_n (T_h), is approximately assumed to be exponentially distributed with mean $1/\mu_n$ $(1/\mu_h)$.

4) There are C channels available in the system. In order to protect handoff calls, the system assumes that in accessing the channels the handoff calls have priority over the new calls and a number of channels among C are reserved exclusively for handoff calls. We call the number of guard channels the *cutoff parameter* C_h. Thus, when a new call is originated, it can be successfully served only if the number of idle channels is greater than C_h. Otherwise, it will be put in the queue or blocked due to buffer overflow. The queued new call reneges from the queue unless it can be successfully served within its patience time [2], [4]; similarly, the queued handoff call is dropped from the queue by the system as it moves out of the handoff area before being accomplished successfully [6]. Here, the time a mobile spends in the handoff area will be called the dwell time of the handoff call.

5) The system provides a finite queue with capacity N_n for new calls during call setup and a finite queue with capacity N_h for handoff calls in the handoff area.

6) The patience (dwell) time of the waiting new (handoff) call is denoted by $T_{nq}(T_{hq})$ and is approximately assumed to be exponentially distributed with mean $1/\mu_{nq}(1/\mu_{hq})$.

7) The channel holding time of a call in a cell, denoted by T_{Ha}, is approximately assumed to be exponentially distributed with mean $1/\mu_{c_h}$.

Notice that λ_h and μ_{c_h} are correlated with other parameters and can be determined from them. Interested readers are referred to [7] for details. We shall address these correlations in the section below entitled Numerical Examples and Discussion.

III. ANALYSIS

We define (n_1, n_2) as the system state with probability P_{n_1,n_2}, where n_1 is the sum of the number of occupied channels and the number of handoff calls waiting in the queue, n_2 is the number of new calls waiting in the queue, and $0 \leq n_1 \leq C + N_h, 0 \leq n_2 \leq N_n$. The state-transition diagram of the system can then be obtained on the basis of assumptions 1) through 7) above. The diagram is shown in Fig. 2. From this diagram, we can obtain the state-transition equations shown below.

(i) If $n_2 = 0$, then

$$(\lambda_n + \lambda_h + n_1\mu_{ch})P_{n_1,0} = \\ (\lambda_n + \lambda_h)P_{n_1-1,0} + (n_1 + 1)\mu_{ch}P_{n_1+1,0}, \\ \text{for } 0 \leq n_1 \leq C - C_h - 1;$$

$$(\lambda_n + \lambda_h + n_1\mu_{ch})P_{n_1,0} = (\lambda_n + \lambda_h)P_{n_1-1,0} \\ + (n_1 + 1)\mu_{ch}P_{n_1+1,0} + (n_1\mu_{ch} + \mu_{nq})P_{n_1,1}, \\ \text{for } n_1 = C - C_h;$$

$$(\lambda_n + \lambda_h + n_1\mu_{ch})P_{n_1,0} = \\ \lambda_h P_{n_1-1,0} + (n_1 + 1)\mu_{ch}P_{n_1+1,0} + \mu_{nq}P_{n_1,1}, \\ \text{for } C - C_h + 1 \leq n_1 \leq C - 1;$$

$$[\lambda_n + \lambda_h + C\mu_{ch} + (n_1 - C)\mu_{hq}]P_{n_1,0} = \lambda_h P_{n_1-1,0} \\ + [C\mu_{ch} + (n_1 + 1 - C)\mu_{hq}]P_{n_1+1,0} + \mu_{nq}P_{n_1,1}, \\ \text{for } \leq n_1 \leq C + N_h - 1;$$

$$(\lambda_h + C\mu_{ch} + N_h\mu_{hq})P_{n_1,0} = \\ \lambda_h P_{n_1-1,0} + \mu_{nq}P_{n_1,1}, \qquad \text{for } n_1 = C + N_h.$$

(ii) If $1 \leq n_2 \leq N_n - 1$, then

$$(\lambda_n + \lambda_h + n_1\mu_{ch} + n_2\mu_{nq})P_{n_1,n_2} = \\ \lambda_n P_{n_1,n_2-1} + (n_1 + 1)\mu_{ch}P_{n_1+1,n_2} \\ + [n_1\mu_{ch} + (n_2 + 1)\mu_{nq}]P_{n_1,n_2+1}, \\ \text{for } n_1 = C - C_h;$$

$$(\lambda_n + \lambda_h + n_1\mu_{ch} + n_2\mu_{nq})P_{n_1,n_2} = \\ \lambda_h P_{n_1-1,n_2} + \lambda_n P_{n_1,n_2-1} + (n_1 + 1)\mu_{ch}P_{n_1+1,n_2} \\ + (n_2 + 1)\mu_{nq}P_{n_1,n_2+1}, \\ \text{for } C - C_h + 1 \leq n_1 \leq C - 1;$$

$$[\lambda_n + \lambda_h + C\mu_{ch} + (n_1 - C)\mu_{hq} + n_2\mu_{nq}]P_{n_1,n_2} = \\ \lambda_h P_{n_1-1,n_2} + \lambda_n P_{n_1,n_2-1} \\ + [C\mu_{ch} + (n_1 + 1 - C)\mu_{hq}]P_{n_1+1,n_2} \\ + (n_2 + 1)\mu_{nq}P_{n_1,n_2+1}, \\ \text{for } C \leq n_1 \leq C + N_h - 1;$$

$$(\lambda_n + C\mu_{ch} + N_h\mu_{hq} + n_2\mu_{nq})P_{n_1,n_2} = \\ \lambda_h P_{n_1-1,n_2} + \lambda_n P_{n_1,n_2-1} + (n_1+1)\mu_{nq}P_{n_1,n_2+1}, \\ \text{for } n_1 = C + N_h.$$
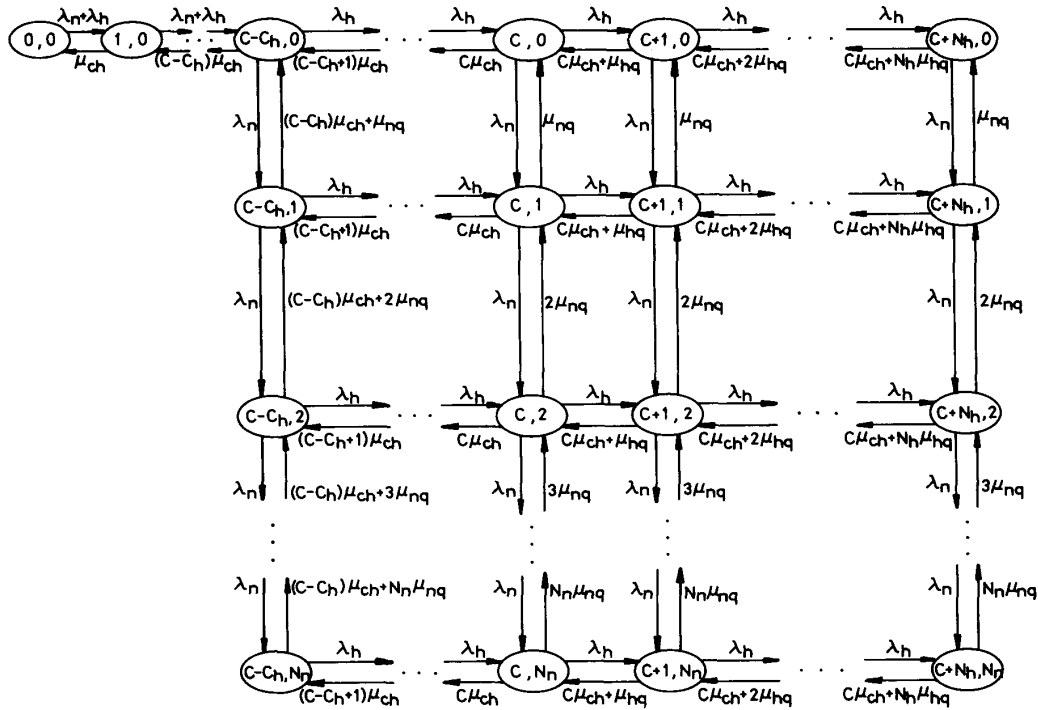
Fig. 2. The state transition diagram.

(iii) If $n_2 = N_n$, then

$$(\lambda_h + n_1\mu_{ch} + n_2\mu_{nq})P_{n_1,n_2} =$$
$$\lambda_n P_{n_1,n_2-1} + (n_1 + 1)\mu_{ch}P_{n_1+1,n_2},$$
$$\text{for } n_1 = C - C_h;$$

$$(\lambda_h + n_1\mu_{ch} + n_2\mu_{nq})P_{n_1,n_2} =$$
$$\lambda_h P_{n_1-1,n_2} + \lambda_n P_{n_1,n_2-1} + (n_1+1)\mu_{ch}P_{n_1+1,n_2},$$
$$\text{for } C - C_h + 1 \le n_1 \le C - 1;$$

$$[\lambda_h + C\mu_{ch} + (n_1 - C)\mu_{hq} + n_2\mu_{nq}]P_{n_1,n_2} = \lambda_h P_{n_1-1,n_2}$$
$$+ \lambda_n P_{n_1,n_2-1} + [C\mu_{ch} + (n_1 + 1 - C)\mu_{hq}]P_{n_1+1,n_2},$$
$$\text{for } C \le n_1 \le C + N_h - 1;$$

$$(C\mu_{ch} + N_h\mu_{hq} + n_2\mu_{nq})P_{n_1,n_2} = \lambda_h P_{n_1-1,n_2}$$
$$+ \lambda_n P_{n_1,n_2-1}, \text{for } n_1 = C + N_h. \quad (1)$$

The above state-transitions belong to a class of Markov chains: the quasi-birth-death (QBD) process [14]. The structure of the transition matrix of the QBD process is in *block tri-diagonal form*. When a larger system is considered, the matrix can be computationally solved by a so-called *folding algorithm*. This type of algorithm, exploited in [14], begins with a forward reduction phase and then executes a backward expansion phase to find the solution. Thus, for any size system the state probabilities can be obtained without any problem. In the following, several performance measures will be derived, including the blocking probability of a new call and the probability of a call being forced into termination during conversation.

A. The Average Blocking Probabilities

Blocking of a new call may occur for two reasons. One is that as a new call originates, the number of available idle channels is less than or equal to C_h and there are no free buffers left in the waiting queue. The other is that although a new call has been accepted and is waiting in the queue, it fails to access a free channel within its patience time and so reneges from the system. The reneging probability of a waiting new call can be easily obtained by comparing the reneging rate with the effective arrival rate, as in [2], [6], [7]. Nevertheless, we here propose an alternative approach to obtaining the reneging (or blocking) probability by considering an arbitrarily selected new call (or, say, new call of interest). We denote the blocking probability of an arbitrarily selected new call by P_B^N. P_B^N can be obtained by

$$P_B^N = \sum_{n_1=C-C_h}^{C+N_h} P_{n_1,N_n} + \sum_{n_1=C-C_h}^{C+N_h} \sum_{n_2=0}^{N_n-1} P_{n_1,n_2} R_n(n_1, n_2) \quad (2)$$

where $R_n(n_1, n_2)$ is the reneging probability of an arbitrarily selected new call given that the system state is (n_1, n_2) just at the instant when the call is accepted and put in the waiting queue. Clearly, $C - C_h \le n_1 \le C + N_h$ and $0 \le n_2 \le N_n - 1$. The derivation of $R_n(n_1, n_2)$ is more complicated than the method used in [2], [6], [7], but it paves the way for obtaining the average waiting times later.

We find $(1 - R_n(n_1, n_2))$ instead of $R_n(n_1, n_2)$. $(1 - R_n(n_1, n_2))$ is the probability that the arbitrarily selected new

call can finally get a free channel within its patience time, given that the system state is (n_1, n_2) at the instant the call is accepted by the system and begins waiting in the queue. When the arbitrarily selected waiting new call successfully accesses a free channel within its dwell time, the quasi-system state is at $(C - C_h - 1, 0)$, where the quasi-system state is defined as the system state observed by the arbitrarily selected new call, excluding those waiting new calls coming after the call of interest. In deriving $(1 - R_n(n_1, n_2))$, we use the signal-flow graph shown in Fig. 3 to portray the transitions of quasi-system states from the input node y_{in} of state (n_1, n_2) to the output node y_{out} of state $(C - C_h - 1, 0)$ and the respective branch gains (the probabilities of transitions). Any intermediate quasi-system state (m_1, m_2) in the graph may have three possibilities of transition: (m_1, m_2) to (m_1+1, m_2), to $(m_1 - 1, m_2)$, or to $(m_1, m_2 - 1)$. The possibility of transition from (m_1, m_2) to $(m_1, m_2 + 1)$ is not included. In this graph, the transition probability from y_{in} to (n_1, n_2) is 1 because the system state is given at (n_1, n_2) as the arbitrarily selected waiting new call is just accepted by the system.

The transition from (m_1, m_2) to (m_1+1, m_2) for $C - C_h \le m_1 \le C, 0 \le m_2 \le n_2 \le N_n - 1$ indicates the arrival of an acceptable handoff call. We denote this transition probability by $P_{m_1+1,m_2|m_1,m_2}$. If $C - C_h \le m_1 \le C, 0 \le m_2 \le n_2$, the transition occurs when the remaining interarrival time of the handoff call, denoted by T_ξ, is smaller than the remaining channel holding time of any of the m_1 calls in progress, T_{H_a}, the remaining patience time of any of the m_2 new calls waiting in the queue, and the remaining patience time of the accepted waiting new call of interest, T_{nq}. If $C + 1 \le m_1 \le C + N_h - 1, 0 \le m_2 \le n_2$, the transition occurs when T_ξ is smaller than the remaining channel holding time of any of the C calls in progress, T_{H_a}, the remaining dwell time of any of the $(m_1 - C)$ handoff calls waiting in the queue, T_{hq}, the remaining patience time of any of the m_2 new calls waiting in the queue, and the remaining patience time of the accepted waiting new call of interest, T_{nq}. Since $T_\xi, T_{H_a}, T_{nq}, and\ T_{hq}$ are mutually independent and are all assumed to be exponentially distributed, $P_{m_1+1,m_2|m_1,m_2}$ can

be obtained by

$$
P_{m_1+1,m_2|m_1,m_2} = \begin{cases} \frac{\lambda_h}{m_1\mu_{ch}+(m_2+1)\mu_{nq}+\lambda_h}, \\ \quad \text{for } C - C_h \le m_1 \le C, 0 \le m_2 \le n_2. \\ \\ \frac{\lambda_h}{C\mu_{ch}+(m_1-C)\mu_{hq}+(m_2+1)\mu_{nq}+\lambda_h}, \\ \quad \text{for } C + 1 \le m_1 \le C + N_h - 1, 0 \le m_2 \le n_2. \end{cases}
$$
(3)

We denote the probability of transition from (m_1, m_2) to $(m_1 - 1, m_2)$ by $P_{m_1-1,m_2|m_1,m_2}$ for $m_1 = C - C_h, m_2 = 0$ or $C - C_h + 1 \le m_1 \le C + N_h, 0 \le m_2 \le n_2 \le N_n - 1$. If $m_1 = C - C_h, m_2 = 0$ or $C - C_h + 1 \le m_1 \le C, 0 \le m_2 \le n_2, P_{m_1-1,m_2|m_1,m_2}$ is contributed by the probability that the m_1 channels in use are reduced by 1 due to completion of a conversation. If $C + 1 \le m_1 \le C + N_h, 0 \le m_2 \le n_2, P_{m_1-1,m_2|m_1,m_2}$ is contributed by the probability that (i) the C channels in use are reduced by 1 due to completion of a conversation of (ii) the number of handoff calls waiting in the queue is reduced by 1 due to dropping of a call. In a manner similar to that used to derive $P_{m_1+1,m_2|m_1,m_2}$ in (3), $P_{m_1-1,m_2|m_1,m_2}$ can be obtained by (4) shown at the bottom of the page, where T'_{H_a} is the channel holding time of the other $(m_1 - 1)$ calls in progress and T'_{hq} is the dwell time of the other waiting handoff calls. T'_{H_a} has the same distribution as T_{H_a}, and T'_{hq} has the same distribution as T_{hq}.

We denote the probability of transition from (m_1, m_2) to $(m_1, m_2 - 1)$ by $P_{m_1,m_2-1|m_1,m_2}$ for $C - C_h \le m_1 \le C + N_h, 1 \le m_2 \le n_2 \le N_n - 1$. If $m_1 = C - C_h, 1 \le m_2 \le n_2, P_{m_1,m_2-1|m_1,m_2}$ is contributed by the probability that (i) a call among m_1 now in progress will complete its conversation or (ii) there is a waiting new call among m_2 reneging from the system. If $C - C_h + 1 \le m_1 \le C + N_h, 1 \le m_2 \le n_2, P_{m_1,m_2-1|m_1,m_2}$ is contributed by the probability that there is a waiting new call reneging from the system. Accordingly, $P_{m_1,m_2-1|m_1,m_2}$ is given by (see (5) shown below), where T'_{nq}, the patience time of the other waiting new calls, has the same distribution as T_{nq}.

Based on the established signal-flow graph shown in Fig. 3 and the branch gains obtained in (3)–(5), $(1 - R_n(n_1, n_2))$, the probability that a new call attempt will succeed, can be

$$
P_{m_1-1,m_2|m_1,m_2} = \begin{cases} \frac{m_1\mu_{ch}}{m_1\mu_{ch}+(m_2+1)\mu_{nq}+\lambda_h}, & \text{for } m_1 = C - C_h, m_2 = 0 \quad \text{or } C - C_h + 1 \le m_1 \le C,\ 0 \le m_2 \le n_2 \\ \\ \frac{C\mu_{ch}+(m_1-C)\mu_{hq}}{C\mu_{ch}+(m_1-C)\mu_{hq}+(m_2+1)\mu_{nq}+\lambda_h}, & \text{for } C + 1 \le m_1 \le C + N_h, 0 \le m_2 \le n_2. \end{cases}
$$
(4)

$$
P_{m_1,m_2-1|m_1,m_2} = \begin{cases} \frac{m_2\mu_{nq}}{m_1\mu_{ch}+(m_2+1)\mu_{nq}+\lambda_h}, & \text{for } C - C_h + 1 \le m_1 \le C, 1 \le m_2 \le n_2 \\ \\ \frac{m_2\mu_{nq}}{C\mu_{ch}+(m_1-C)\mu_{hq}+(m_2+1)\mu_{nq}+\lambda_h}, & \text{for } C + 1 \le m_1 \le C + N_h, 1 \le m_2 \le n_2. \end{cases}
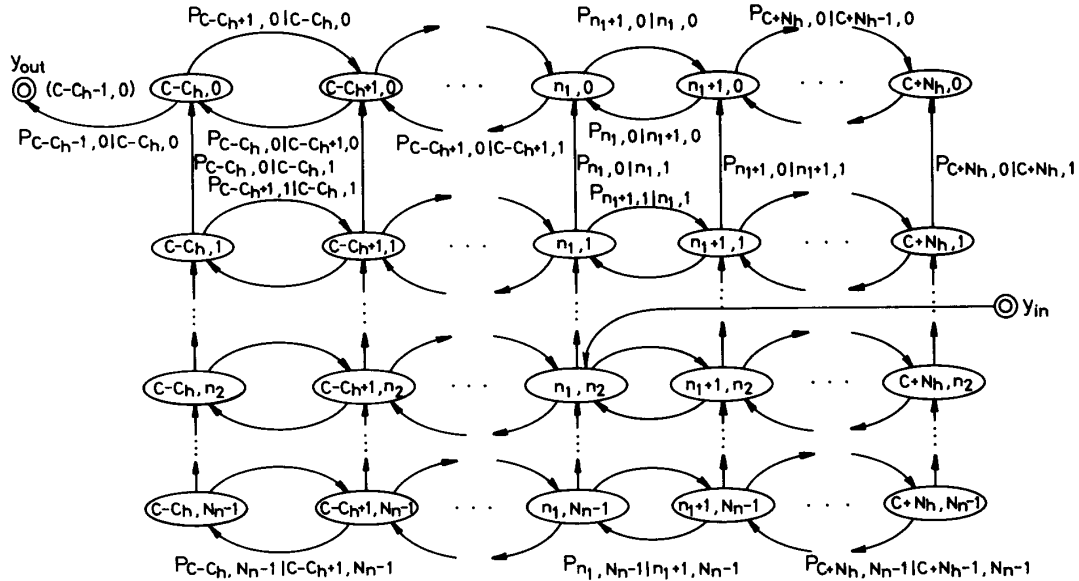$$
(5)

Fig. 3.   The signal-flow graph for obtaining $R_n(n_1, n_2)$.

obtained by using the general gain formula (Mason's rule) [8], which is given by

$$1 - R_n(n_1, n_2) = \sum_{k=1}^{N} \frac{M_k \Delta_k}{\Delta}, \qquad (6)$$

where N is the total number of *forward paths*, which are defined to be paths from the input node (n_1, n_2) to the output node $(C - C_h - 1, 0)$, M_k is the k-th forward path gain, which is the product of the branch gains encountered in traveling the k-th forward path, $\Delta = 1-$ (sum of the gains of all individual loops) + (sum of products of gains of all possible combinations of two nontouching loops) $-$ (sum of products of gains of all possible combinations of three non-touching loops) $+...$, and $\Delta_k =$ the Δ for that part of the signal-flow graph that is nontouching with the k-th forward path. Note that loops are called *nontouching* if they do not share a common node.

For the signal-flow graph shown in Fig. 3, we find that the graph excluding the output node $(C - C_h - 1, 0)$ has a rectangular structure with $(n_2 + 1)$ rows and $(C_h + N_h + 1)$ columns. It has $(C_h + N_h + 1)^{n_2}$ forward paths and $(C_h + N_h) \times (n_2 + 1)$ individual loops. Loops belonging to different rows or belonging to the same row but not adjacent to each other are non-touching. The gain of each loop is simply the product of the gains of two branches. On the basis of the specific features of the signal-flow graph, we can numerically compute Δ and Δ_k using a recursive algorithm. However, the large number of $(C_h + N_h + 1)^{n_2}$ forward paths prevents our analytical method from applying to all cases. Fortunately, N_n and N_h need not be large in real applications, due to call reneging and dropping. We shall examine this characteristic in the numerical examples discussed in the next section. In summary, the analytical method presented here is

computationally tractable. Via the general gain formula in (6), we can numerically obtain $R_n(n_1, n_2)$ and in turn the blocking probability of an arbitrarily selected new call P_B^N in (2).

We also derive the blocking probability of a handoff call by considering an arbitrarily selected handoff call (or, say, handoff call of interest). Blocking of an arbitrarily selected handoff call occurs in two situations. The first is that there are no free channels and no free buffers available as the call moves into a handoff area. The second is that, although the handoff call has been accepted by the system and is waiting in the queue, the call cannot access a free channel within its dwell time in the handoff area and so is dropped from the queue by the system. The blocking probability of the arbitrarily selected handoff call, denoted by P_B^H, can be similarly obtained by

$$P_B^H = \sum_{n_2=0}^{N_n} P_{C+N_h, n_2} + \sum_{n_1=C}^{C+N_h-1} \sum_{n_2=0}^{N_n} P_{n_1, n_2} R_h(n_1, n_2) \qquad (7)$$

where $R_h(n_1, n_2)$ is the dropping probability of the arbitrarily selected handoff call given that the system state is (n_1, n_2) just at the instant when the call is accepted by the system and waits in the queue.

In this case, we again find the probability $(1 - R_h(n_1, n_2))$ instead of $R_h(n_1, n_2)$. In obtaining $(1 - R_h(n_1, n_2))$ for $C \leq n_1 \leq C + N_h - 1, 0 \leq n_2 \leq N_n$, we once again construct a signal-flow graph and find the respective branch gain. When the arbitrarily selected waiting handoff call successfully accesses a free channel within its dwell time, the quasi-system state is at $(C - 1, n_2)$, where the quasi-system state is defined as the system state observed by the arbitrarily selected waiting handoff call, excluding the new and handoff calls coming after the call of interest. Fig. 4 shows a signal-flow graph that portrays the transitions of quasi-system states from the input node y_{in} of state
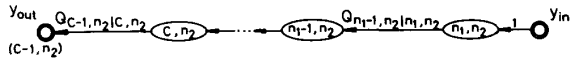
Fig. 4. The signal-flow graph for obtaining $R_h(n_1, n_2)$.

(n_1, n_2) to the output node y_{out} of state $(C - 1, n_2)$. For any intermediate quasi-system state (m_1, n_2) in the signal-flow graph, $C \leq m_1 \leq n_1$, its only transition is from (m_1, m_2) to $(m_1 - 1, n_2)$.

We denote the probability of transition from (m_1, n_2) to $(m_1 - 1, n_2)$ by $Q_{m_1-1,n_2|m_1,n_2}$. $Q_{m_1-1,n_2|m_1,n_2}$ is contributed by two probabilities. The first is the probability that the remaining channel holding time of any of the C calls in progress, T_{H_a}, is smaller than the remaining channel holding time of any of the other $(C - 1)$ calls in progress T'_{H_a}, the dwell time of any of the $(m_1 - C)$ waiting handoff calls, and the dwell time of the waiting handoff call of interest, T_{hq}. The second is the probability that the remaining dwell time of any of the $(m_1 - C)$ handoff calls waiting in the queue, T_{hq}, is smaller than the channel holding time of any of the C calls in progress, T_{H_a}, the dwell time of any of the other $(m_1 - C - 1)$ waiting handoff calls, and the dwell time of the waiting handoff call of interest, T'_{hq}. Thus $Q_{m_1-1,n_2|m_1,n_2}$ can be obtained by

$$Q_{m_1-1,n_2|m_1,n_2} = \frac{C\mu_{ch} + (m_1 - C)\mu_{hq}}{C\mu_{ch} + (m_1 - C + 1)\mu_{hq}},$$
$$\text{for } C \leq m_1 \leq n_1. \quad (8)$$

Fortunately, because there are no individual loops in the signal-flow graph and only one possible path between y_{in} and y_{out}, we can obtain a closed-form solution for $(1 - R_h(n_1, n_2))$ via the general gain formula. The solution is given by

$$1 - R_h(n_1, n_2) = \prod_{m_1=C}^{n_1} \frac{C\mu_{ch} + (m_1 - C)\mu_{hq}}{C\mu_{ch} + (m_1 - C + 1)\mu_{hq}}. \quad (9)$$

From (7) and (9), P_B^H can be obtained.

However, the probability that a successfully call is forced into termination during conversation due to handoff blocking could be more significant than P_B^H [7]. A call may succeed in the first $(\kappa - 1)$ handoff attempts but be forced into termination at the κ-th handoff due to blocking. We denote this probability by P_F and express it as

$$P_F = \sum_{\kappa=1}^{\infty} \Psi_N \cdot [(1 - P_B^H) \cdot \Psi_H]^{\kappa-1} \cdot P_B^H = \frac{\Psi_N \cdot P_B^H}{1 - (1 - P_B^H)\Psi_H} \quad (10)$$

where Ψ_N and Ψ_H are the handoff requirement probabilities for a new and a handoff call, respectively. They can be obtained by

$$\Psi_N = Prob\{T_m > T_n\} = \frac{\mu_n}{\mu_m + \mu_n} \quad (11)$$

and

$$\Psi_H = Prob\{T_m > T_h\} = \frac{\mu_h}{\mu_m + \mu_h}. \quad (12)$$

B. The Average Waiting Times

We first derive the average waiting time for queued new calls. The waiting time of a queued new call is here defined as the time that an arbitrarily selected waiting new call spends from the time it is accepted by the system to the time it successfully accesses a free channel. We denote $W_n(n_1, n_2)$ the waiting time of a queued new call given that the system state is at (n_1, n_2) when the call just arrives at the system and waits in the queue. Clearly, $C - C_h \leq n_1 \leq C + N_h$, $0 \leq n_2 \leq N_n - 1$. The waiting time $W_n(n_1, n_2)$ can be obtained via the probability $(1 - R_n(n_1, n_2))$, which can be expressed as

$$1 - R_n(n_1, n_2) = Prob\{T_{nq} > W_n(n_1, n_2)\}. \quad (13)$$

Since T_{nq} is assumed to be exponentially distributed, $W_n(n_1, n_2)$ can be obtained by

$$W_n(n_1, n_2) = -\frac{1}{\mu_{nq}} \cdot In(1 - R_n(n_1, n_2)), \quad (14)$$

where "In" is the natural logarithmic function. Consequently, the average waiting time of a queued new call, denoted by $\overline{W}_n$, can be obtained by

$$\overline{W}_n = \sum_{n_1=C-C_h}^{C+N_h} \sum_{n_2=0}^{N_n-1} P_{n_1,n_2} \cdot W_n(n_1, n_2)$$
$$/ \sum_{n_1=C-C_h}^{C+N_h} \sum_{n_2=0}^{N_n-1} P_{n_1,n_2}. \quad (15)$$

Similarly, we define the waiting time of queued handoff calls as the time that an arbitrarily selected handoff call takes from the time it is accepted and begin waiting in the queue to the time it successfully accesses a free channel. We denote the waiting time of a queued handoff call given that the system is at (n_1, n_2) when it arrives at the system and waits in the queue by $W_h(n_1, n_2)$. Clearly, $C \leq n_1 \leq C + N_h - 1$, $0 \leq n_2 \leq N_n$. $W_h(n_1, n_2)$ can be through a formula analogous to (14) by

$$W_h(n_1, n_2) = -\frac{1}{\mu_{hq}} \cdot In(1 - R_h(n_1, n_2)). \quad (16)$$

And the average waiting time of a queued handoff call, denoted by $\overline{W}_h$, can be obtained by

$$\overline{W}_h = \sum_{n_1=C}^{C+N_h-1} \sum_{n_2=0}^{N_n} P_{n_1,n_2} \cdot W_h(n_1, n_2)$$
$$/ \sum_{n_1=C}^{C+N_h-1} \sum_{n_2=0}^{N_n} P_{n_1,n_2}. \quad (17)$$

IV. NUMERICAL EXAMPLES AND DISCUSSION

As mentioned previously, the arrival rate of handoff call λ_h is correlated with other parameters. From [7,(17)], λ_h can be obtained by

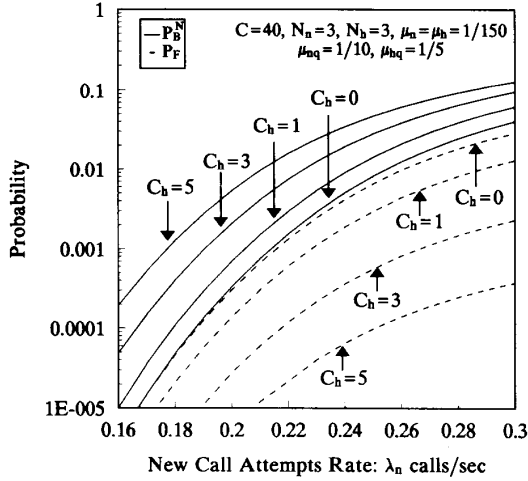$$\lambda_h = \frac{\lambda_n(1 - P_B^N)\Psi_N}{1 - (1 - P_F)\Psi_H}. \quad (18)$$

Fig. 5. The probabilities of P_B^N and P_F versus λ_n for various C_h's.

Fig. 6. The probabilities of P_B^N and P_F versus λ_n for various N_h's.

As in the derivation in [7, eqs. (4)-(7)], we obtain the cumulative distribution function (CDF) of the channel holding time of a call in a cell T_{H_a}, denoted by $F_{H_a}(t)$, by

$$F_{H_a}(t) = 1 - \frac{\lambda_n(1 - P_B^N)}{\lambda_n(1 - P_B^F) + \lambda_h(1 - P_F)} e^{-(\mu_m + \mu_n)t}$$
$$- \frac{\lambda_h(1 - P_F)}{\lambda_n(1 - P_B^F) + \lambda_h(1 - P_F)} e^{-(\mu_m + \mu_h)t}. \quad (19)$$

If the time spent in a cell by the mobile associated with a successful new call $1/\mu_n$ is assumed to be equal to the time spent in a cell by the mobile associated with a successful handoff call $1/\mu_h$, T_{H_a} should be exponentially distributed with mean $\mu_{ch} = \mu_m + \mu_n$. If $1/\mu_n \neq 1/\mu_h$, T_{H_a} can be closely approximated by an exponential distribution as a result of [6, Table I]. In the following examples, we assume $1/\mu_h = 1/\mu_n$ for convenience. We also assume the following parameters: $C = 40, 1/\mu_m = 120, 1/\mu_h = 1/\mu_n = 150, 1/\mu_{nq} = 10$, and $1/\mu_{hq} = 5$. We use an iterative method to numerically compute the solutions. With the above parameters and an initial guess for λ_h, we obtain temporary performance measures of P_B^N and P_F; we then substitute the temporary P_B^N and P_F into (18) to obtain a new value of λ_h and execute the computation process again. The entire computation process is repeated until there are no further changes within four significant figures for P_B^N and P_F.

Fig. 5 shows the probabilities of P_B^N and P_F versus the new call arrival rate λ_n for different cutoff parameters C_h, where we assume the queue size for new calls $N_n = 3$ and the queue size for handoff calls $N_h = 3$. The effects of C_h on P_B^N and P_F can be seen from the figure. As C_h increases, P_B^N increases and P_F decreases for all traffic loads. This is intuitively reasonable. If there is only one channel reserved for handoff calls $(C_h = 1), P_F$ can be at least five times smaller than P_B^N in this example. Also, notice that if there is no channel reserved for handoff calls, P_F is still smaller than P_B^N. This is because the handoff call in the waiting queue has a higher priority to access channels than the new call does.
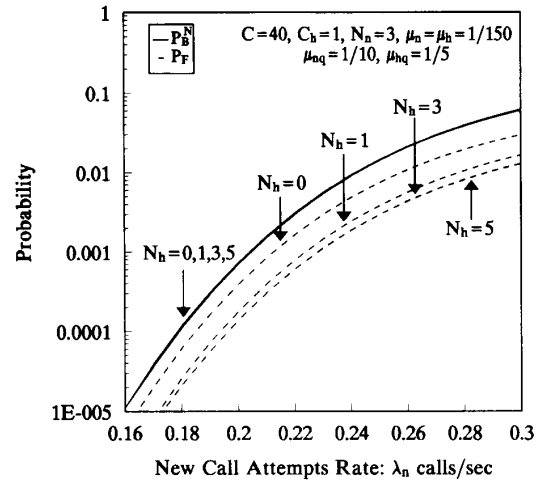

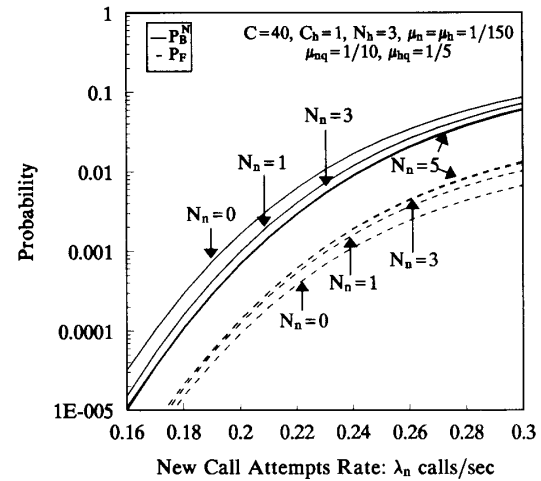Fig. 7. The probabilities of P_B^N and P_F versus λ_n for various N_n's.

Fig. 6 shows the probabilities of P_B^N and P_F versus λ_n for various N_h's, where we assume $C_h = 1$ and $N_n = 3$. We find that as the queue capacity for handoff calls N_h increases, P_B^N deteriorates very little (almost no change) but P_F improves greatly for all traffic loads. For example, when $\lambda_n = 0.26$, P_F is reduced from 0.0108 at $N_h = 0$ to 0.0044 at $N_h = 3$. A similar conclusion is also presented in [9, p.280]. This result is because N_h has a direct effect on the performance of handoff calls but an indirect effect on that of new calls. We also observe that the improvement in P_F becomes saturated as N_h becomes larger. This is because of the dropping effect of the handoff calls, and this tells us that it is not necessary to provide a large queue size for handoff calls for a given dropping rate of handoff calls. In this example $N_h = 3$ is sufficient.

Fig. 7 shows the probabilities of P_B^N and P_F versus λ_n for various N_n's, where we assume $C_h = 1$ and $N_h = 3$. When the queue capacity for new calls N_n increases, P_B^N decreases but P_F increases. The improvement in P_B^N is significant but
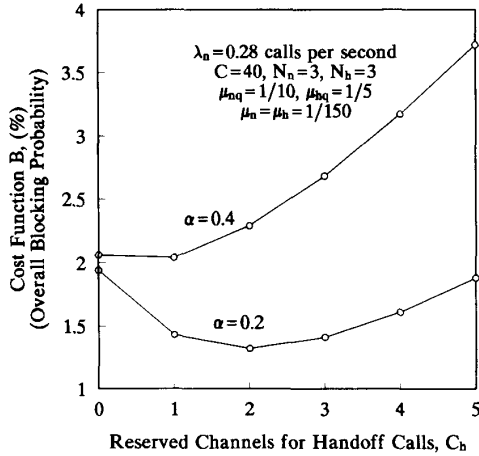
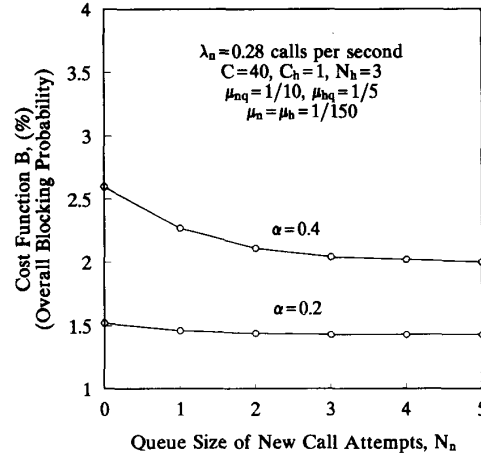Fig. 8. The cost function B versus C_h for various α's.

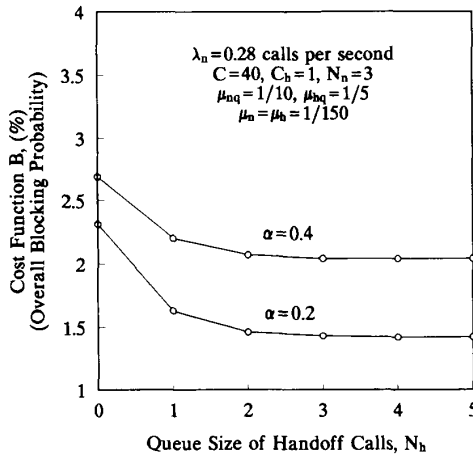Fig. 10. The cost function B versus N_n for various α's.

Fig. 9. The cost function B versus N_h for various α's.
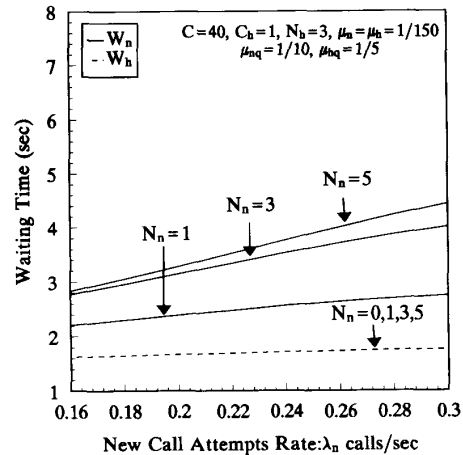
Fig. 11. The average waiting times $\overline{W}_n$ and $\overline{W}_h$ versus λ_n for various N_n's.

the deterioration in P_F is insignificant, because a very small value of P_F does not result in a significant deterioration in service for customers. For example, when $\lambda_n = 0.26$, P_B^N is reduced from 0.0344 at $N_n = 0$ to 0.0208 at $N_n = 3$, while P_F is increased from 0.0024 at $N_n = 0$ to 0.0044 at $N_n = 3$. The improvement in P_B^N becomes saturated as N_n becomes larger because of the reneging effect of new calls. This also tells us that it is not necessary to provide a large queue size for new calls for a given reneging rate of new calls. In this example $N_n = 3$ is sufficient.

There is a tradeoff between the two performance measures, P_B^N and P_F, for different cutoff parameters of C_h, as can be seen in Fig. 5. There should exist an optimal cutoff parameter C_h. Here, to determine an optimal C_h we define an overall blocking probability (or cost function), denoted by B, similar to that used in [7]. B is given by

$$B = \alpha \cdot P_B^N + (1 - \alpha) \cdot P_F, \qquad (20)$$

where α is a weighting factor, $0 \leq \alpha \leq 1$. The value of α depends on the stress laid on the quality-of-service

requirements for P_B^N and P_F. Usually, interruptions of handoff calls upset customers much more than blocking of new calls. In other words, P_F is more significant than P_B^N. Therefore the value of α should be confined to be less than 0.5. Fig 8 shows the overall blocking probability B versus the cutoff parameter C_h for $\alpha = 0.2$ and 0.4, given that $\lambda_n = 0.28$, $N_n = 3$, and $N_h = 3$. We observe that in this particular example the optimal value of C_h is 2 if $\alpha = 0.2$ and the optimal value of C_h is 1 if $\alpha = 0.4$. Figs. 9 and 10 also show the overall blocking probability B versus N_h and N_n, respectively, for $\alpha = 2$ and 0.4. We find that the overall blocking probability B decreases as $N_h(N_n)$ increases, asymptotically approaching a bounded value at around $N_h = 3(N_n = 3)$.The results in Figs. 9 and 10 justify our earlier statements that queueing of new and handoff calls can minimize the overall blocking probability and that the queue sizes for new and handoff calls need not be large in real applications.

Fig. 11 shows the average waiting times of queued new and handoff calls, $\overline{W}_n$ and $\overline{W}_h$, versus the new call arrival rate λ_n
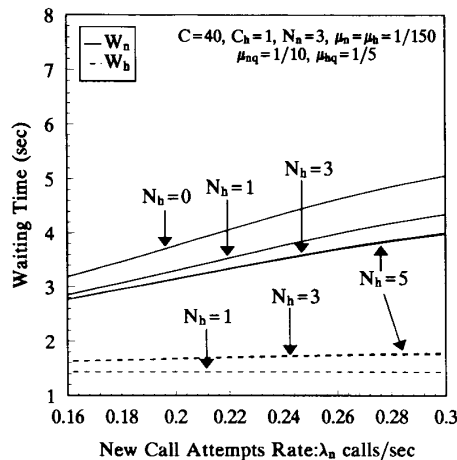
Fig. 12. The average waiting times $\overline{W}_n$ and $\overline{W}_h$ versus λ_n for various N_h's.

for various N_n's, given that $C_h = 1$ and $N_h = 3$. We can see from the figure that, as the new call arrival rate λ_n increases, the increment in $\overline{W}_n$ is more significant than the increment in $\overline{W}_h$; moreover, the larger the queue capacity for new calls N_n is, the greater the increment in $\overline{W}_n$ will be. These findings are intuitive. Fig. 12 shows the average waiting times of queued new and handoff calls versus the new call arrival rate λ_n for various N_h's, given that $C_h = 1$ and $N_n = 3$. We can see from the figure that, as the queue capacity for handoff calls N_h increases, $\overline{W}_h$ becomes larger and approaches saturation. It is because of the dropping effect of the handoff call. On the other hand, it is interesting to see that as N_h increases, $\overline{W}_n$ becomes smaller and approaches saturation. This is because those waiting new calls in the rear of the queue, which would contributed more to $\overline{W}_n$ if they could be served, are likely to renege as N_h increases.

V. CONCLUDING REMARKS

This paper studies a new cutoff priority cellular radio system, in which the ability for both types of calls to wait in a finite queue and the possibility for calls to renege or to be dropped form the queue are considered. High-priority handoff calls have access to all channels, while low-priority new calls can be served only if enough channels are idle. All arrival processes are Poisson and the service time distributions are exponential. The analysis is via a two-dimensional Markov chain approach; the blocking probability of new calls and the probability of a call being forced into termination during conversation are derived using signal-flow graphs and Mason's formula. We also obtain the average waiting time for queued new and handoff calls. We conclude that, in the new cutoff priority cellular radio system, there exists an optimal cutoff parameter and appropriate queue sizes for both new and handoff calls.

Note that the system designed here is a system with only single-call platform type. A system that supports a mixture of platform types and queueing of handoff calls was studied in [11]-[13]. An extension of the system that considers multiple-

call platform types and queueing of both new and handoff calls is now being studied.

REFERENCES

[1] E. Ayanoglu, "Signal flow graphs for path enumeration and deflection routing analysis in multihop networks," *Proc. GLOBECOM'89*, pp. 1022-1029.
[2] D. Y. Barrer, "Queueing with impatient customers and ordered services," *Oper. Res.*, vol. 5, pp. 650-656, 1957.
[3] C. J. Chang and Y. Y. Chiang, "Analysis of priority channel allocation in a cell for mobile radio telephone systems via signal-flow graphs," *IEEE TENCON'92*, Melbourne, Australia, pp. 418-422.
[4] D. Gross and C. M. Harris, *Fundamental of Queueing Theory*. New York: Wiley, 1985, pp. 122-125.
[5] R. Guérin, "Queueing-blocking system with two arrival streams and guard channels," *IEEE Trans. Commun.*, vol. 36, pp.153-163, Feb. 1988.
[6] 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.
[7] ——, "Priority oriented channel access for cellular systems serving vehicular and portable radio telephones," *IEE Proc., Part I*, vol. 136, no.5, pp. 339-346, 1989.
[8] B. C. Kuo, *Automatic Control Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1987, 5th ed.
[9] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*. New York: McGraw-Hill, 1990.
[10] S. H. Oh and D. W. Tcha, "Prioritized channel assignment in a cellular radio network," *IEEE Trans. Commun.*, vol.40, pp. 1259-1269, July 1992.
[11] C. Purzynski and S. S. Rappaport, "Traffic performance analysis for cellular communication systems with mixed platform types and queued hand-offs," *IEEE VTC'93*, pp. 172-175.
[12] S. S. Rappaport, "The multiple call hand-off problem in high capacity cellular communication systems," *IEEE Trans. Veh. Technol.*, vol. VT-40, pp. 546-557, Aug. 1991.
[13] ——, "Modeling the hand-off problem in personal communication networks," *IEEE VTC'91*, pp. 517-523.
[14] J. Ye and S.-Q. Li, "Analysis of multi-media traffic queues with finite buffer and overload control—Part I: Algorithm," *IEEE INFOCOM'91*, pp. 1464-1474.

Chung-Ju Chang (S'81–M'85–SM'94) received the B.E. and M.E. degrees in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1972 and 1976, respectively, and the Ph.D degree in electrical engineering from National Taiwan University in 1985.

From 1976 to 1988 he was with Telecommunication Laboratories, Directorate General of Telecommunications, Ministry of Transportation and Communications, China, as a Design Engineer, Supervisor, Project Manager, and then Division Director. There, he was involved in designing digital switching system, ISDN user-network interface, and ISDN service and technology trials. In the meantime, he also acted as a Science and Technical Advisor for Ministry of Transportation and Communications from 1987 to 1989. In August 1988, he joined the faculty of the Department of Communication Engineering and Center for Telecommunications Research, National Chiao Tung University, where he is currently Professor. His research interests include performance evaluation, ATM (asynchronous transfer mode) networks, and mobile radio networks.

Dr. Chang is a member of the Chinese Institute of Engineers (CIE).

Tian-Tsair Su received the B.E. and M.E. degrees in electrical engineering from National Cheng Kung University, Taiwan, in 1984 and 1986, respectively. He is currently pursuing the Ph.D degree in communication engineering in National Chiao Tung University, Hsinchu, Taiwan.

Since 1986 he has been with Telecommunication Laboratories, Directorate General of Telecommunications, Ministry of Transportation and Communications, China, where he is working on satellite and mobile radio networks in the Transmission Technology Laboratory. His research interests include performance evaluation, satellite communication system, and personal communications.

Yueh-Yiing Chiang received the B.S. degree in communication engineering from Chiao Tung University, Taiwan, in 1989 and the M.S. degree in communication engineering from Chiao Tung University, Taiwan, in 1991. His M.S. work was on the modeling and analysis of cellular mobile communication systems.

In July 1991, he joined the Air Force as a platoon leader of the artillerymen and finished his army duty in May 1993. He is currently employed by the Navy as an electrical engineer.