

Population-Based Call Admission Control in Wireless Cellular Networks

Manyoo Han and Arne A. Nilsson
Center for Advanced Computing and Communication
Department of Electrical and Computer Engineering
Box 7914
North Carolina State University
Raleigh, NC 27695-7914

Abstract

The wide spread use of mobile terminals has brought a rapid growth of cellular networks. This fast growth generates new requirements for the current wireless cellular system to provide quality-of-service guarantees for wireless traffic. To accommodate the increasing number of mobile terminals, wireless systems have been designed as micro/picocellular architectures in order to provide higher capacity. This reduced coverage area of a cell has caused a higher rate of hand-off events, and providing continuous services to wireless traffic after hand-offs has been important issues. A simple and efficient bandwidth reservation scheme is proposed to provide prioritized handling for hand-off calls. This scheme dynamically adjusts the amount of bandwidth in a cell reserved for hand-off calls according to the amount of cellular traffic in its neighboring cells. The performance of the proposed scheme is evaluated by both analytical method and simulation and is compared with the guard channel scheme which is widely used in current cellular networks.

1 Introduction

The wide spread use of mobile terminals such as mobile phones or portable computers has brought a rapid growth of cellular networks. To accommodate the increasing number of mobile terminals in the limited radio spectrum, wireless systems have been designed as micro/picocellular architectures in order to provide higher capacity [7].

Since most of traffic sources in wireless networks are considered mobile, this reduced coverage area of a cell has caused a higher rate of hand-off events [1].

Unlike in the wired networks, traffic sources or destinations in the wireless environment move freely and their future moving paths are not known to networks. Even though a mobile successfully makes a connection in its current location, there is no guarantee that it would be able to maintain its connection when moving into another area, since the cell into which the mobile moves may run out of available bandwidth and drop this connection. This characteristic of wireless networks makes it a very complex problem to provide quality-of-service (QOS) guarantees to wireless and mobile connections throughout their lifetime.

As QOS parameters in wireless/mobile networks, we can consider new-call blocking probability, hand-off call dropping probability, system overload probability/period, system utilization and so on [1]. Efficient utilization of the scarce spectrum is certainly an important issue in wireless networks. In [13], various channel assignment strategies are summarized in the sense of how to utilize the spectrum efficiently. The schemes of borrowing channels from neighboring cells or maintaining flexible channels in a cluster controlled by a Mobile Switching Center (MSC) makes the role of MSC's very complex, since it has to keep track of the channels lent to other cells or assigned from the flexible channel pool. It should also supervise borrowing procedure to prevent these channels from interfering with existing calls in neighbors. Accordingly, for practical reasons, the channel assignment is usually performed in a static way [10].

In some channel assignment strategies, hand-off calls are handled in the same manner as newly generated calls. In this case, the dropping probability due to unsuccessful hand-off attempts equals the blocking probability for the newly generated calls. This discontinuation of service is a very undesirable situation from the viewpoint of mobile stations with in-progress calls. Generally, it is considered more desirable

to keep supporting in-progress connections as long as possible. To this end, hand-off call dropping probability should be maintained below certain specified level, by handling hand-off requests with priority over new connection requests. Accordingly, many of the channel assignment strategies employ some kind of prioritized handling to hand-off calls at the expense of a tolerable increase in call blocking probability [13].

The most common method to provide prioritized handling is employing resource reservation, where the resource in this context implies bandwidth or radio channels to be allocated for wireless connections. The guard channel concept introduced in [5] reduces forced terminations of hand-off calls by simply reserving a fixed number of channels exclusively for hand-off calls at the expense of the reduction of total carried traffic, since newly generated calls have fewer channels to be assigned with [13]. This scheme is very simple to implement but does not adapt effectively to the varying traffic conditions.

For the effective and adaptive bandwidth reservation, some mechanisms have been proposed to make bandwidth reservation by predicting the mobile user's movement [9], [8], [2]. However, the prediction of the future moving direction of a mobile station requires a lot of overhead of a base station to maintain and process data to keep track of the past history of a mobile station.

In this paper, a simple bandwidth reservation scheme is proposed to provide prioritized handling for hand-off calls by dynamically adjusting the amount of reserved bandwidth of a cell according to the amount of cellular traffic in its neighboring cells. The performance of the proposed scheme is evaluated by both analytical method and simulation and is compared with the guard channel scheme which is widely used in current cellular networks.

The rest of the paper is organized as follows. In Section 2, the concept of the proposed scheme is introduced with the detailed description of operations. The defi-

definition of the performance measures used for the evaluation is described in Section 3. Section 4 describes the analytical and the simulation modeling to evaluate the performance of the reservation schemes. Numerical results are presented and discussed in Section 5. This paper concludes with suggestions for future research and concluding remarks in Section 6.

2 Bandwidth Reservation Scheme

In urban area, microcellular architecture is often employed to increase the bandwidth capacity. The radius of a cell in this architecture is usually less than 500 meters and each cell tends to be full of streets stretched vertically and horizontally and crowded with mobile stations such as automobiles or pedestrians equipped with wireless communication terminals. Due to the small coverage area of a cell, frequent hand-offs of calls between cells happen. Wireless networks try to keep these in-progress calls alive by making bandwidth reservations into the neighboring cells.

It would be ideal if we could reserve the exact amount of bandwidth required by a call into the exact cell into which a mobile station would move. Currently, however, it is not possible for wireless networks to exactly know the future direction of a mobile station. As alternatives, some reservation schemes have been proposed to make bandwidth reservations by predicting the direction of a mobile station [9], [8], [2]. Predicting moving directions requires a lot of overhead from wireless networks in terms of processing time and memory usage in base stations or large amount of control messages between base stations and/or mobile stations due to the information exchange about moving patterns of mobile stations.

According to the research in [6], drivers are very sensitive to traffic conditions and information, especially in urban area. They change their route choices depending on these factors. In this situation, it is very difficult to make an accurate prediction for

the moving direction of a mobile station. It is also mentioned in [6] that traffic signal control is one of the reasons for the traffic delay and congestion. These factors could reduce the number of hand-offs that a wireless connection of a mobile station will experience throughout its lifetime. Assuming the average call duration is 3 minutes and mobile stations move at 30 km/h on average, they move 1.5 km during their call connections. This distance corresponds to one or two hand-offs in a microcellular architecture whose cell is 1 km wide in diameter. Furthermore, when a call is first initiated in a cell, there is not any information about the mobile station with which a wireless network can predict its moving direction. There are also significant amount of slow mobile stations such as pedestrians in wireless networks, and each of them may experience at most one hand-off until it completes its call. Accordingly, the reservation schemes by predicting moving directions is considered too costly compared to its effectiveness in a real situation.

Although the users of mobile stations move through paths that can be predicted to some extent [8], from the viewpoint of a wireless network and base stations, they can be seen as moving randomly over the entire coverage area. In other words, if mobile stations are not identified individually, a base station could observe random movements of mobile stations as a whole.

Based on these observations, we propose a simple and efficient bandwidth reservation scheme, which does not require any overhead to keep track of motion of each mobile to predict the destination cell. In this scheme, since the motions of mobile stations are seen from the point of a base station, it can be assumed that they move toward random directions and spread all over the cellular coverage area. Since mobiles move randomly over a microcellular architecture in rural area, the population in each cell can be assumed to be similar to each other, and it is rare that some cells are extraordinarily crowded while others are not.

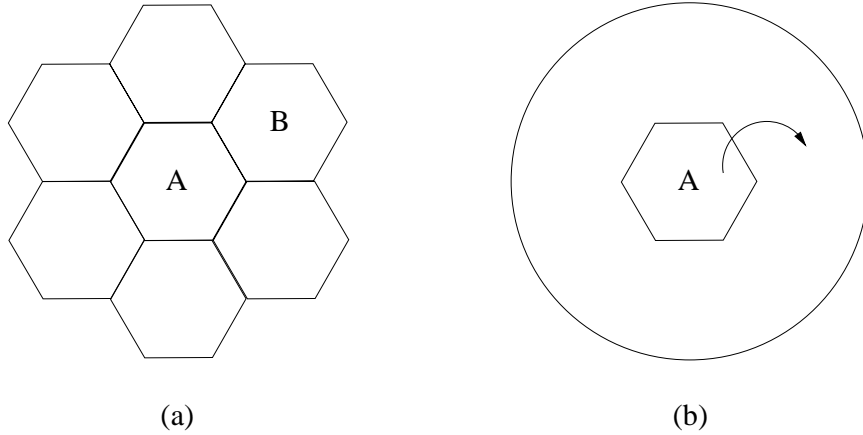


Figure 1: Seven-cell cluster model and its approximation

For clarity, three bandwidth types are defined. Reserved bandwidth is the one reserved for hand-off calls, and not to be assigned to new calls. Allocated bandwidth is currently occupied one by either new calls or hand-off calls. Available bandwidth is the one neither reserved nor allocated.

Assuming the seven-cell cluster model as in Figure 1(a) as a wireless cellular network, a mobile station with an in-progress call in cell *A* is going to move into only one of its six neighboring cells if it does not complete its call while staying inside the cell. If, however, we consider the simplified model consisting of a cell and its one big neighbor as shown in Figure 1(b), the mobile station in cell *A* would definitely move into the big neighbor if it needs to be handed over.

If bandwidth reservations could be managed at this big neighborhood level, not at the individual neighboring cell level, higher bandwidth utilization could be achieved as well as lower dropping probability since the prediction for future moving direction would always be correct, and only the exact amount of bandwidth required for a call would be reserved. It is, however, a very difficult and complex task to manage bandwidth at a neighborhood level, since each and every cell in a cellular network has its own unique neighborhood and also belongs to someone else's neighborhood.

Under the assumption that each mobile moves at random direction from the point of a base station, an active call will be handed off into any one of the neighboring cells with an equal probability. For example, in seven-cell cluster model as in Figure 1(a), a mobile station in cell A will move into cell B with a probability of $1/6$. Statistically speaking, cell B needs to reserve only one sixths of required bandwidth for this call. Since the cell B is also surrounded by its own six neighboring cells and if there are some active calls in its neighborhood at most of time, the base station of cell B can collect the fractional bandwidth reservations from each of these calls and manage the reserved bandwidth in the form of a pool, so that any hand-off calls are to be served by this reservation pool if the pool is not empty.

Accordingly, the basic idea of this scheme is that each base station adjusts the amount of bandwidth reserved for hand-off calls according to the amount of current traffic in its neighboring cells, from which the amount of potential hand-offs can be predicted. The amount of traffic may be represented by the amount of allocated bandwidth.

To describe the detailed procedure to operate the proposed bandwidth reservation scheme, which is referred to as Population-Based Bandwidth Reservation (PBR), cell i is supposed to have n_i neighboring cells. For example, if a seven-cell cluster model is assumed for a cellular network, n_i is equal to 6 for any cell i in a network. In real circumstances, however, cells may have the different number of neighboring cells, depending on the geographical characteristics. In extreme case such as a highway area, the base station that covers that area may assume that it has only two neighboring cells, since most of mobile stations move along the highway path. Thus, n_i of cell i may not be equal to n_j of cell j .

Whenever a call which consumes b bandwidth units is admitted into a cell j as either a newly generated call or a hand-off call, the base station of the cell requests a

fractional bandwidth reservation for the amount of b/n_j to each of its n_j neighboring cells. Whenever this call is leaving the cell either by call completion or by hand-off into its neighbor, the base station requests a fractional bandwidth release for the same amount as requested for the reservation to each of its n_j neighboring cells, even to the cell into which this call is handed over. This step is to inform the neighborhood of appearance and disappearance of a potential hand-off.

Each base station in a cellular network maintains a counter that records transactions for fractional bandwidth reservation or release requests from its neighboring cells. Everytime it receives a fractional bandwidth reservation request or a release request, it increments or decrements the counter by the requested amounts, respectively. Eventually, this counter shows the current amount of allocated bandwidth in its neighborhood, which reflects the amount of potential hand-offs in the future. Denoting the set of the neighboring cells of cell i by N_i , the counter value of cell i , q_i , is equal to:

$$q_i = \sum_{j \in N_i} \frac{C_{oj}}{n_j},$$

where C_{oj} is the number of allocated bandwidth units in cell j .

Among q_i of potential hand-off traffic, however, some of them may terminate their calls before initiating hand-offs. Even every call that is to be handed over into cell i will not request a hand-off at the same time. Thus, it is not necessary to reserve all q_i bandwidth at the same instance. Reflecting this fact, q_i is multiplied by a fractional parameter, f_i , which ranges between 0 and 1, and implies how much actual hand-off traffic will occur to cell i within the next certain period. The value of this parameter can be determined statically or adaptively. Currently, f_i is set statically, and an adaptive method is for further study.

Since bandwidth is utilized in bandwidth units, a base station tries to reserve $\lceil q_i \cdot f_i \rceil$ bandwidth units exclusively for hand-off traffic. It is, however, not always

possible to reserve this amount because of already allocated bandwidth. Accordingly, this reservation can be considered as a virtual reservation, rather than a physical reservation, and used for a call admission control criteria.

Another factor in determining the amount of bandwidth reserved for hand-off calls is the current load level or crowdedness of the cell in which those bandwidth are to be reserved. It is true that, as a cell becomes more crowded, it becomes more difficult to satisfy the reservation requests. Accordingly, it is necessary to limit admitting new calls, depending on the current load level of a cell, before it becomes more crowded. For this purpose, we multiply a weight, $(1 + \frac{C_{oi}}{C_i})$, to the amount requested for bandwidth reservation.

The admission of a call which demands b bandwidth units in cell i is decided as follows; a newly generated call is admitted only if

$$C_{oi} + b \leq C_i - \lceil q_i f_i (1 + \frac{C_{oi}}{C_i}) \rceil, \quad (1)$$

and a hand-off call is admitted only if

$$C_{oi} + b \leq C_i, \quad (2)$$

where C_i is a capacity of cell i in bandwidth units.

The admission criteria for a newly generated call in Equation 1 implies couple of things. First, the admission of a newly generated call in the proposed PBR scheme is determined by the traffic level in the neighborhood as well as that in the current cell itself. Secondly, hand-off calls are handled with much higher priority over newly generated calls than in physical reservation schemes.

As indicated in [2], if a newly generated call were admitted into a cell only according to the current cell's traffic information such as the number of allocated or physically reserved bandwidth, these admitted calls might be handed over to the already crowded neighboring cell and dropped. The virtual reservation level, $\lceil q_i f_i (1 + \frac{C_{oi}}{C_i}) \rceil$,

of PBR scheme indicates the degree of crowdedness in the neighborhood, since q_i is proportional to the number of allocated bandwidth in the neighborhood. It indicates not the current level of physical reservation, but the amount of bandwidth necessary to handle incoming traffic during the next certain period.

When a call leaves the cell in which it has been residing, due to either call completion or hand-off, it releases the bandwidth it has occupied. Suppose that the capacity of a cell is fully utilized and there is a newly generated call waiting to be served at this moment. The released bandwidth would be immediately allocated to this new call, if a physical reservation were considered for an admission control, even though a hand-off request from a neighboring cell is expected soon. The hand-off call cannot help being dropped, and this is a very undesirable situation.

With the proposed scheme, however, just released bandwidth is not assigned to a newly generated call, but reserved for a hand-off call. While the amount of virtual reservation, $\lceil q_i f_i (1 + \frac{C_{oi}}{C_i}) \rceil$, is larger than $C_i - C_{oi}$, it is implied that the base station should have reserved more bandwidth units, but could not, since there had not remained any available bandwidth units in the cell. With at least one active call in the neighborhood, $\lceil q_i f_i (1 + \frac{C_{oi}}{C_i}) \rceil$ is always greater than or equal to 1, and, in this example, C_{oi} is equal to C_i just before the bandwidth is released. Thus, the system is in need of a reservation, and the released bandwidth is used to compensate the deficit in reservation.

3 Performance Measures

When a newly generated call is not admitted into a cell due to the lack of resources such as bandwidth, it is called blocked. Thus, the blocking probability denotes the ratio of blocked calls over all newly generated calls. When a hand-off attempt fails due to the same reason as blocking, it is called dropped, and the dropping probability

denotes the ratio of dropped hand-off attempts over all hand-off attempts. The QOS in cellular networks is mainly determined by these two quantities [10].

In [1], wireless connections are differentiated into real-time and non-real-time connections. When a cell is under congested situations, real-time connections are to be dropped, while non-real-time connections can wait for being allocated bandwidth. Thus, the latter are to experience much higher delay than under normal conditions. To control the delay constraints, the overload probability is defined, which is the probability of being in an overload state, and the duration of such a state is considered as another performance measure.

In this paper, however, we do not consider the delayed service in studying the performance of bandwidth reservation schemes. The calls which are not allocated bandwidth when moving into a new cell is to be dropped, instead of being delayed. In this sense, the probability or the duration of overload state is not considered.

The radio spectrum is very precious and limited resource in wireless networks, and it is extremely important how well it is utilized. In this sense, the system utilization is also considered, which is defined as a ratio of average number of bandwidth occupied over a cell capacity.

4 Performance Analysis

To evaluate the performance of the proposed algorithm, we analyze the cellular network under this scheme as well as under the guard channel scheme for the comparison purpose. The analysis is performed in both analytical and simulation method. The seven-cell cluster model where each cell is shaped as a hexagon and surrounded by six neighboring cells is considered as a basic cellular network model, shown in Figure 2, although a simplified version of this model is hired for analytical modeling.

In terms of bandwidth or channel assignment strategies, various strategies can be

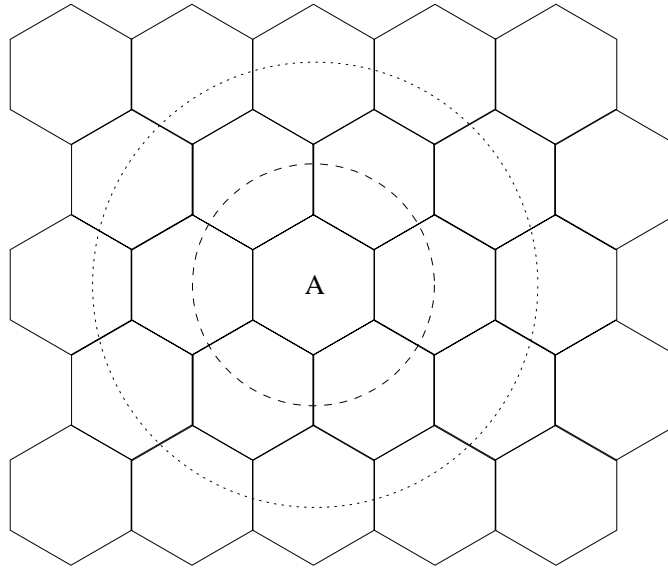


Figure 2: Cellular network for performance analysis modeling

applied to this analysis such as dynamic channel assignment [3] or flexible channel assignment [12]. For simplicity of implementation, however, the basic fixed channel assignment strategy is employed. Here, a set of C bandwidth units are permanently assigned to each cell, and a call attempt in a cell is served only by the idle bandwidth units from this predetermined set of bandwidth units dedicated to this cell.

Throughout the analysis, it is assumed that new calls are generated at random locations all over a wireless coverage area and according to a Poisson process with mean λ . Newly generated calls cannot utilize the reserved bandwidth for hand-off calls. Accordingly, they are admitted into the cell only when there remain available bandwidth units in a cell.

Generally, the call duration could be assumed to be exponentially distributed, if it were generated in wired networks. The call duration means the length of a lifetime of a call, during which the call may experience hand-offs in wireless cellular networks. The channel holding time of a call in a cell is the duration between the time at which a call is assigned a channel in a cell and the time at which the channel is released due

to either a hand-off into another cell or termination of a call. This channel holding time is actually dependent on the location in a cell where a mobile station is assigned a channel, its moving direction and its speed.

In [5], channel holding time in a cell is investigated, and its distribution is approximated by an exponential distribution for the analysis. This approximation is also adopted for our analysis and channel holding time is assumed exponentially distributed with mean $1/\mu_h$. The call duration is also assumed exponentially distributed with $1/\mu_c$, and, by nature, $1/\mu_c \geq 1/\mu_h$.

4.1 Analytical Modeling

In this section, an analytical model is developed to evaluate the performance of the proposed reservation scheme. A similar model is also used to analyze the guard channel scheme proposed in [5], where it was modeled by one-dimensional markov chain with the state of a cell defined by the number of calls allocated in the cell. In the analysis in [5], both the average new call arrival rate and the average hand-off attempt rate per cell were assumed constant for all states. In fact, however, the hand-off attempt rate changes, according to various conditions. For example, as the number of active calls in the neighborhood of a cell increases, the cell will observe more hand-off attempts.

In our analytical model, we take into consideration the call population not only in a cell, but also in its neighboring cells to incorporate the varying hand-off call rate, and investigate the interactions between them in terms of change in the amount of allocated bandwidth units. For an exact analysis, interactions between a cell and each of its neighboring cells should be investigated individually. Since, however, it would increase the complexity of this analysis tremendously, a simplified cellular network model and approximating assumptions are made for the sake of analytical tractability.

4.1.1 Traffic Modeling

In the cellular network as seen in Figure 2, cell A is surrounded by six immediate neighboring cells referred to as a first-tier neighborhood, circled with a dashed line, and another twelve farther neighboring cells, a second-tier neighborhood, circled with a dotted line. The cellular network surrounding cell A can be simplified by considering two tiers of neighboring cells as two big layered neighboring rings as shown in Figure 3.

The state of cell A is represented with a pair of two state variables, (m, n) , m denoting the number of allocated bandwidth units in the cell and n that in the first-tier neighborhood. This state is affected by various events occurring in its own cell or its neighborhood such as new call arrivals or call terminations or hand-offs between cells. We analyze the proposed scheme by building a Markov model for the state transitions due to such events.

Since the analysis is to be concentrated on the interactions between cell A and its immediate neighboring cells, the number of allocated bandwidth units in the second-tier neighborhood may be approximated to reduce the complexity of the Markov model. As mentioned previously in Section 2, it is assumed that mobile stations move randomly and spread all over a cellular coverage area, and the second-tier neighborhood contains twice as many cells as the first-tier neighborhood. Thus, the second-tier neighborhood can be assumed to have twice as many active calls as the first-tier neighborhood. We also consider a single class of calls in the analytical modeling for the ease of analysis. Assuming each call consumes one bandwidth unit, the number of allocated bandwidth units in the second-tier neighborhood can be approximated twice the number of allocated units in the first-tier.

The events that can initiate state transitions, that is, increase or decrease the number of allocated bandwidth units in cell A or its first-tier neighborhood in the simplified cellular network are depicted in Figure 3, and the descriptions for each

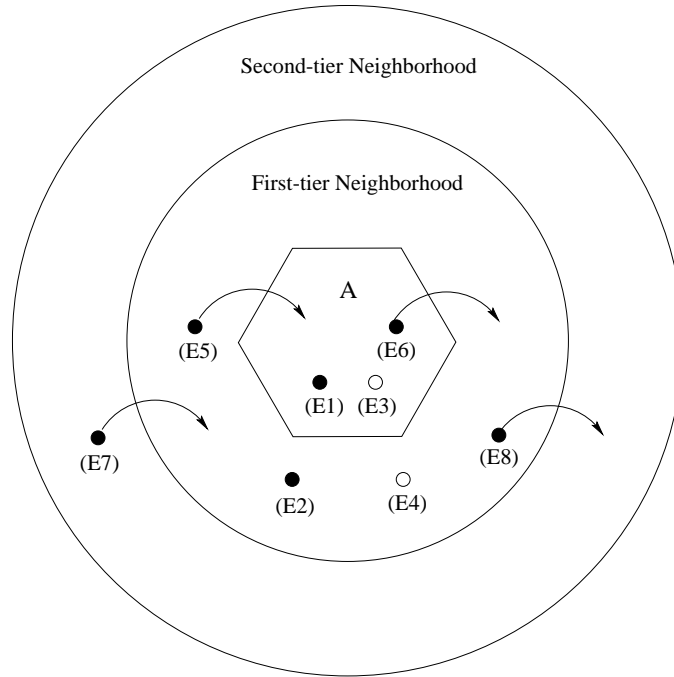


Figure 3: Events in simplified seven-cell cluster model

event are listed below:

- (E1) New call arrival in cell A :
- (E2) New call arrival in the first-tier neighborhood
- (E3) Call termination in cell A
- (E4) Call termination in the first-tier neighborhood
- (E5) Call hand-off from the first-tier neighborhood into cell A
- (E6) Call hand-off from cell A into the first-tier neighborhood
- (E7) Call hand-off from the second-tier neighborhood into the first-tier neighborhood
- (E8) Call hand-off from the first-tier neighborhood into the second-tier neighborhood

The set of events listed above does not include the hand-offs between the cells within the same tier of neighborhood, which might decrease the number of allocated bandwidth units in the tier. For example, if all of the bandwidth units of one of six cells in the first-tier were fully allocated, the hand-off attempt into this cell from its neighboring cells in the same tier would be dropped due to no available bandwidth units, and the total number of allocated bandwidth units in this tier would be decremented. In the simplified model, however, we do not take into consideration the call hand-off inside a tier and the call is assumed to remain inside the same tier.

An active call in a cell eventually either terminates due to completion of communication or attempts a hand-off into a neighboring cell when crossing a cell boundary. The call hand-off rate out of a cell is related to several factors such as population of active calls in the cell, their positional distribution from the edge of the cell, moving speeds of mobile stations associated with calls, and so on. For simplicity, however, it is assumed that call hand-off attempt rate depends only on the population of active calls. Accordingly, if a cell holds m active calls, the call hand-off rate out of the cell is computed as $m \mu_h$, and the call termination rate as $m \mu_c$, where $1/\mu_h$ is average channel holding time and $1/\mu_c$ average call duration.

In Figure 4, the possible call hand-off directions from cells in a first-tier or a second-tier neighborhood are depicted. As seen in this figure, assuming that a mobile station's moving direction is distributed uniformly over a cellular coverage area, a call in cell B which is located in a first-tier neighborhood of cell A will either be handed off into cell A with a probability of $1/6$, or into a second-tier with $3/6$ or remain in the first-tier with $2/6$, if it is to be handed off. Since this is common to all six cells in the first-tier neighborhood, calls in this tier are handed off into cell A with a rate $\frac{1}{6}n \mu_h$ with n active calls in this tier, and into a second-tier neighborhood with a rate $\frac{1}{2}n \mu_h$.

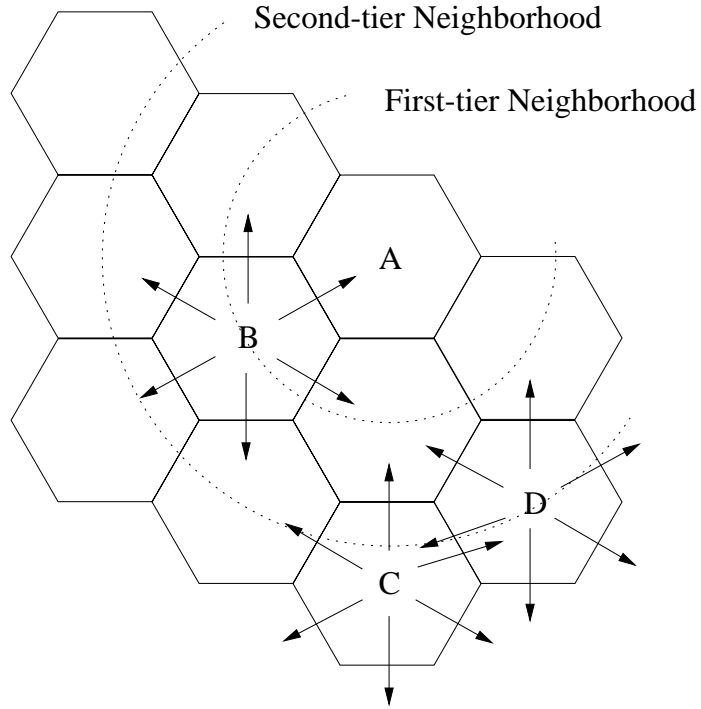


Figure 4: Possible hand-off directions

In the same way, the hand-off attempt rate from a second-tier neighborhood can be determined. As seen in Figure 4, considering cells C and D as a pair, a call in this pair of cells can move into a first-tier neighborhood through three different paths out of twelve, and this is common to all six pairs of cells in the second-tier. The hand-off attempt rate into a first-tier neighborhood becomes $\frac{1}{4}l\mu_h$ with l active calls in a second-tier neighborhood. Since, however, the call population in a second-tier is approximated to twice the population in a first-tier previously, this rate becomes $\frac{1}{2}n\mu_h$ with n active calls in a first-tier neighborhood.

4.1.2 Markov Model for PBR scheme

Based on the above discussion, we build a two-dimensional state transition rate matrix, \mathbf{R} , for the proposed bandwidth reservation scheme, in which each state is specified by (m, n) -tuple, where m is for the number of allocated bandwidth units in cell

A , and n for that in the first-tier neighborhood of cell A . As noted previously, in single traffic class analysis, the number of allocated bandwidth units is equal to the number of active calls. It is assumed that m and n are integers which are bounded as follows:

$$0 \leq m \leq C,$$

$$0 \leq n \leq 6C,$$

where C is the capacity of a cell in bandwidth units.

The state transition diagram for the rate matrix \mathbf{R} is shown in Figure 5. Due to the large scope of the actual diagram, one small diagram is devised, which only shows state transitions involving state (m, n) . All possible state transitions from or into state (m, n) are shown and transitions between other states are omitted for simplicity in this diagram. Depending on the values of m and n , some of the transitions may occur and some others may not. To complete each row of the rate matrix \mathbf{R} , it is enough to know the rates of the transitions originating at the state which corresponds to the row of the rate matrix \mathbf{R} . In this sense, only the transitions which originate at (m, n) are labeled in the diagram, and their rates are discussed under various boundary conditions of m and n .

Before the transition rates are discussed in detail, the new call admission condition in the first-tier neighborhood of cell A needs to be stated. As seen in Figure 2, a cell located in the first-tier neighborhood of cell A is surrounded by three cells from the second-tier, two cells from the same first-tier and cell A . With n calls in the first-tier, each cell in both the first-tier and the second-tier can be assumed to have $n/6$ calls on average, and, accordingly, a cell in the first-tier to have $(\frac{5}{6}n + m)$ calls on average in its own neighborhood, assuming m calls in cell A . Thus, the new call admission in the first-tier neighborhood of cell A is determined based on this assumption.

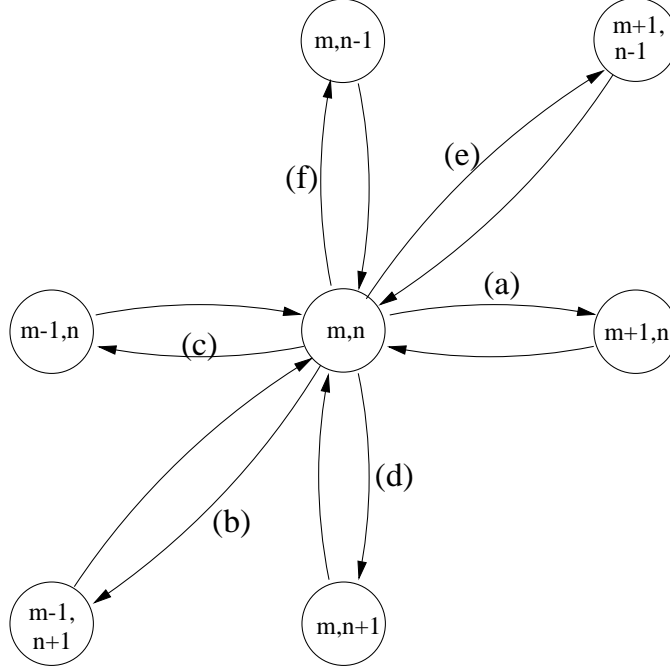


Figure 5: Markov state-transition diagram for bandwidth reservation scheme

- (a) This transition is caused by a new call generated inside cell A , (event **E1**), which is admitted into cell A . The transition rate, r_a , is defined as follows:

$$r_a = \begin{cases} \lambda & \text{if } C - m > \lceil \frac{1}{6}nf(1 + \frac{m}{C}) \rceil \\ 0 & \text{otherwise} \end{cases}$$

where f is the fractional parameter.

- (b) This transition is caused by a call hand-off from cell A into its first-tier neighborhood (**E6**), which is admitted into the tier. The transition rate, r_b , is defined as follows:

$$r_b = \begin{cases} m\mu_h & \text{if } m > 0, n < 6C \\ 0 & \text{otherwise} \end{cases}$$

- (c) This transition is caused by either a call termination inside cell A (**E3**) or a call hand-off from cell A into its first-tier neighborhood (**E6**), which is failed since all the bandwidth units of the first-tier neighborhood are fully allocated. The

transition rate, r_c , is defined as follows:

$$r_c = \begin{cases} m\mu_c & \text{if } m > 0, l < 6C \\ m\mu_c + m\mu_h & \text{if } m > 0, l = 6C \\ 0 & \text{otherwise} \end{cases}$$

- (d) This transition is caused by either a new call generated inside the first-tier neighborhood of cell A (**E2**), which is admitted into the tier, or a call hand-off from the second-tier into the first-tier neighborhood of cell A (**E7**), which is admitted into the first-tier. The transition rate, r_d , is defined as follows:

$$r_d = \begin{cases} 6\lambda + \frac{1}{2}n\mu_h & \text{if } 6C - n > 6\lceil \frac{1}{6}(\frac{5}{6}n + m)f(1 + \frac{m}{C}) \rceil \\ \frac{1}{2}n\mu_h & \text{if } 0 < 6C - n \leq 6\lceil \frac{1}{6}(\frac{5}{6}n + m)f(1 + \frac{m}{C}) \rceil \\ 0 & \text{otherwise} \end{cases}$$

where f is the fractional parameter.

- (e) This transition is caused by a call hand-off into cell A from its first-tier neighborhood (**E5**), which is admitted into the cell. The transition rate, r_e , is defined as follows:

$$r_e = \begin{cases} \frac{1}{6}n\mu_h & \text{if } m < C, n > 0 \\ 0 & \text{otherwise} \end{cases}$$

- (f) This transition is caused by either a call termination inside the first-tier neighborhood of cell A (**E4**), or a call hand-off from the first-tier either into the second-tier neighborhood of cell A (**E8**) or into cell A (**E5**), which is failed since all the bandwidth units of cell A are fully allocated. The transition rate, r_f , is defined as follows:

$$r_f = \begin{cases} n\mu_c + \frac{1}{2}n\mu_h & \text{if } m < C, n > 0 \\ n\mu_c + \frac{2}{3}n\mu_h & \text{if } m = C, n > 0 \\ 0 & \text{otherwise} \end{cases}$$

Letting $\pi_{m,n}$ denote the steady-state probability for state (m, n) of cell A , we have a steady-state probability vector π that satisfies the followings:

$$\pi \mathbf{R} = \mathbf{0}, \tag{3}$$

$$\pi \mathbf{e} = 1,$$

where \mathbf{e} is a column vector of 1's. Due to the complexity of this Markov process, it is not feasible to find a closed form solution for π . Thus, π should be computed by a numerical method such as Gaussian elimination [11].

With π , we can compute performance measures. Since new calls are generated according to a Poisson process with the mean rate which is constant for all states, the blocking probability for new calls, P_b , is simply a sum of the steady-state probabilities of states in the blocking region, which corresponds to any pairs of m and n that satisfy the following inequality,

$$C - m \leq \lceil \frac{1}{6}nf(1 + \frac{m}{C}) \rceil.$$

Thus,

$$P_b = \sum_{n=0}^{6C} \sum_{0 \leq m \leq C, C-m \leq \lceil \frac{1}{6}nf(1 + \frac{m}{C}) \rceil} \pi_{m,n}. \quad (4)$$

On the other hand, hand-off attempts into a cell are dropped only if the capacity of the cell is fully utilized, that is, they are dropped only at the states (C, n) for any n . The mean rate of hand-off attempts, however, is not constant as in the case of new call arrivals. It varies depending on the population of a neighborhood. Letting $h_{m,n}$ denote the hand-off attempt rate into cell A when the system is at state (m, n) , the dropping probability for hand-off calls, P_d , is computed as:

$$\begin{aligned} P_d &= \frac{\sum_{n=0}^{6C} h_{C,n} \pi_{C,n}}{\sum_{m=0}^C \sum_{n=0}^{6C} h_{m,n} \pi_{m,n}} \\ &= \frac{\sum_{n=0}^{6C} n \pi_{C,n}}{\sum_{m=0}^C \sum_{n=0}^{6C} n \pi_{m,n}}, \end{aligned} \quad (5)$$

since $h_{m,n} = \frac{1}{6}n\mu_h$.

System utilization, ρ , is:

$$\rho = \frac{1}{C} \sum_{m=0}^C \sum_{n=0}^{6C} m \pi_{m,n}. \quad (6)$$

4.1.3 Markov Model for Guard Channel Scheme

The only difference in between modeling guard channel scheme and PBR scheme is the new call admission conditions for cell A and its first-tier neighborhood. In guard channel scheme, predefined amount of bandwidth units, C_h , are reserved exclusively for hand-off calls. Thus, new calls are admitted into a cell only if the current number of allocated bandwidth units in the cell is smaller than $C - C_h$, where C is the capacity of the cell.

The state transition diagram for guard channel scheme is as same as that for PBR scheme, shown in Figure 5, and rates for transitions are as same as those of PBR scheme except r_a and r_d , which are the rates for the transitions that can be caused by new call arrivals in cell A and its first-tier neighborhood, respectively. r_a and r_d for guard channel scheme are defined as follows:

$$r_a = \begin{cases} \lambda & \text{if } m < C - C_h \\ 0 & \text{otherwise} \end{cases}$$

$$r_d = \begin{cases} 6\lambda + \frac{1}{2}n\mu_h & \text{if } n < 6(C - C_h) \\ \frac{1}{2}n\mu_h & \text{if } 6(C - C_h) \geq n < 6C \\ 0 & \text{otherwise} \end{cases}$$

The blocking region for new calls in guard channel scheme is simply where the current number of allocated bandwidth units is equal to or greater than $C - C_h$. Thus, the blocking probability for this scheme is computed as:

$$P_b = \sum_{n=0}^{6C} \sum_{m=C-C_h}^C \pi_{m,n}.$$

The dropping probability and system utilization are computed in the same way as for PBR scheme.

4.2 Simulation Modeling

In simulation modeling, we consider a cellular network which is composed of N by N cells as shown in Figure 2. New calls are generated at individual cell level. We also

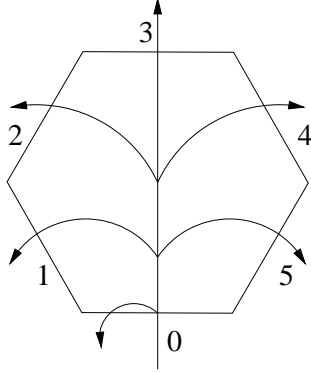


Figure 6: Mobile station's moving directions

consider multiple classes of traffic. The mean new call arrival rate per class per cell is adjusted so that each class has the same bandwidth unit demand rate, that is, for any class i and j ,

$$\lambda_i b_i = \lambda_j b_j,$$

where λ_i is a mean new call arrival rate per cell for class i , and b_i is the number of bandwidth units required by class i call.

It is assumed that each class of a call has the same call duration distribution with mean $1/\mu_c$ as previously mentioned. Thus, the offered load per cell, Λ , in Erlangs/cell, is defined as follows:

$$\Lambda = \frac{1}{\mu_c} \sum_i \lambda_i b_i. \quad (7)$$

Hand-off of a call is simulated by transferring an active call to one of its neighboring cells. Each admitted call stays inside the cell for a certain duration, which is distributed exponentially with mean $1/(\mu_h + \mu_c)$. After that, it either moves with probability of p_h into one of six neighboring cells or terminates with probability of $(1 - p_h)$. The hand-off probability, p_h , is computed as follows:

$$p_h = \frac{\mu_h}{\mu_c + \mu_h}. \quad (8)$$

In Figure 6, the sides of a hexagonal cell are numbered 0 through 5 and possible

moving directions are depicted, where the side 0 is the one through which a call is handed over into a cell. Assuming a mobile station tends to move straight, it is most probable that it will move toward the side 3 and least likely that it will go back toward the side 0, although it may change depending on the pattern of streets in a cell. Letting h_i denote the probability that a mobile station which comes in through the side 0 moves toward the side i , the following is assumed throughout the simulation,

$$h_3 > h_2 = h_4 > h_1 = h_5 > h_0, \sum_{i=0}^5 h_i = 1.$$

The neighboring cell adjacent to the side i is selected as a destination cell with a probability of h_i if the call is to be handed over. For the call initiated inside a cell, however, the destination cell for hand-off is selected arbitrarily with a probability of $1/6$ for each neighboring cell, since no information about its past is available.

The calls handed off outside the cellular coverage area, which may occur in the cells located at the edge cells in Figure 2, are fed back into the other end of the coverage area to offer traffic fairly to all of cells. This can be achieved by handling a cellular coverage area as if sets of N by N cells shown in Figure 2 are tiled.

During the simulation, the number of hand-off attempts and those admitted into the cell located at the middle of the cellular coverage area are counted to compute dropping probability for hand-off calls, and the number of occurrences of newly generated calls and those admitted into the same cell are counted for blocking probability for newly generated calls. The system utilization is measured by reading the number of allocated bandwidth in the same cell at random interval during the simulation.

For the comparison of the performance of the proposed reservation scheme, PBR, the same simulation is performed with guard channel scheme [5], in which C_h bandwidth units are assigned exclusively for hand-off calls out of C units of a cell, and the remaining $C - C_h$ units are shared by both new calls and hand-off calls. A newly

Parameter	Value	Description
N	5	cellular network dimension
μ_h	0.5	1/average channel holding time
μ_c	0.333	1/average call duration
p_h	0.6	hand-off probability
q_3	0.4	probability of moving toward side 3
q_2, q_4	0.2	probability of moving toward side 2 or 4
q_1, q_5	0.09	probability of moving toward side 1 or 5
q_0	0.02	probability of moving toward side 0

Table 1: Simulation parameters

generated call which demands b bandwidth units is blocked if $C_o + b$ is greater than C_h when the call is initiated, where C_o is the number of currently allocated bandwidth units, while a hand-off call is dropped only if $C_o + b$ is greater than C .

5 Numerical Results

Table 1 shows the values of common parameters used in the simulation. Unless specified otherwise, these values are assumed throughout this paper. In most cases, we consider a cell whose diameter is 1 km and average speed of a mobile station is assumed 30 km/h. We also assume that average call duration is 3 minutes per call if it completes without being dropped during hand-off. Since, at this speed, it takes 2 minutes to move 1 km, the average channel holding time in a cell is set to 2 minutes, even though it varies depending on the location at which a call is initiated.

We have used a simplified cellular network model and approximating assumptions to develop the analytical models for bandwidth reservation schemes in Section 4.1. These assumptions are for the ease of analysis, but should be reasonable enough to well represent the actual model. To verify the correctness of our analytical model, we compare the results from numerical solutions with those from simulations.

In Figure 7, both of results from numerical solution and simulation for the pro-

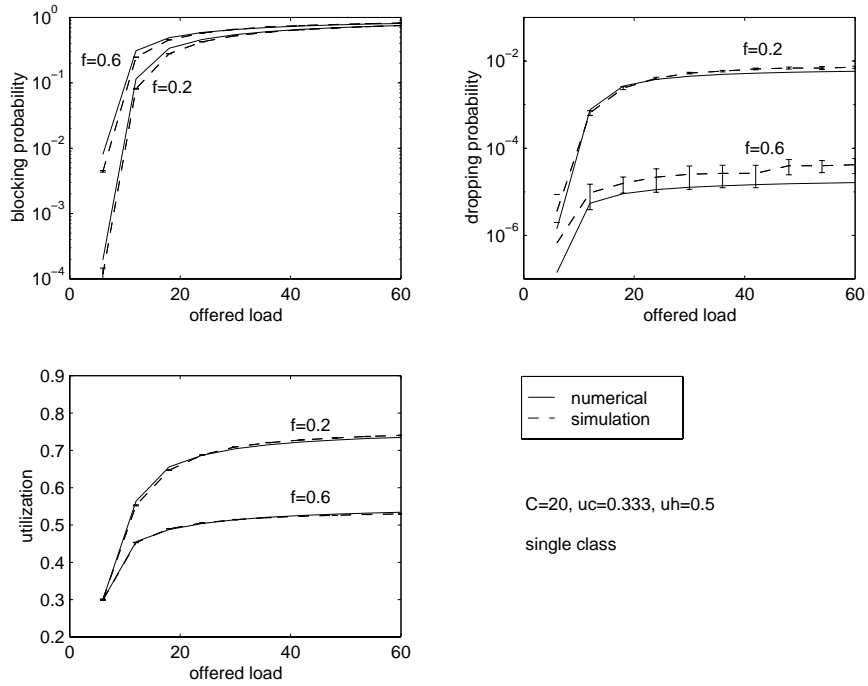


Figure 7: Verifying analytical model for PBR

posed PBR scheme are plotted at cell capacity of 20 bandwidth units. The offered load in the figure is defined as Equation 7. The results for guard channel scheme are also shown in Figure 8.

As seen in the figures, the results from numerical solutions are very close to those from simulations for each of three performance measures in both PBR scheme and guard channel scheme. The differences between two results are considered partly due to simulation error and partly due to the call admission condition at a first-tier neighborhood in an analytical model.

In analytical model shown in Figure 3, the first-tier neighborhood of cell A admits more calls than it does in real circumstances. When we determine the admittance of new calls in the first-tier neighborhood, it is assumed that the population in this tier is equally divided into each of six cells of this tier, and the call admission condition is tested at individual cell level. In a real cellular network, however, the populations of

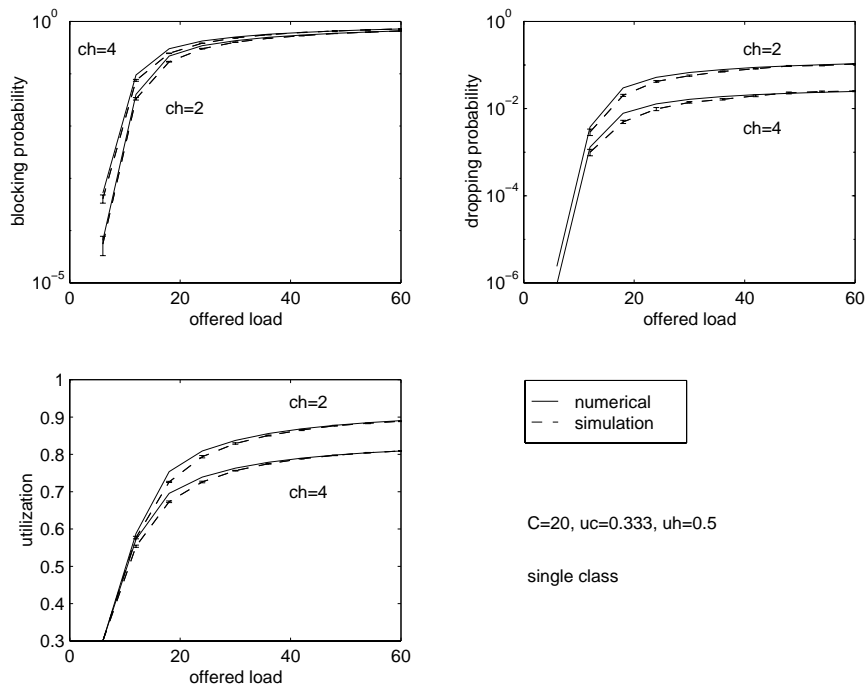


Figure 8: Verifying analytical model for guard channel

individual cells may not be equal at most of time. Accordingly, with certain first-tier population with which the analytical model would not observe any new call blocking, there might be some blockings in a real situation. In the same way, the analytical model admits more hand-off calls during hand-off attempts between cell A and its first-tier or between first and second tiers.

The way of handling hand-off attempts within a first-tier neighborhood can also contribute to the causes of the difference in two results. In an analytical model, hand-offs between the same tiered cells are regarded as staying inside the tier, which could be dropped in a real network, especially in a heavily loaded situation. Since, however, as seen in the results, the hand-off dropping probability is maintained relatively very low, it can be said that hand-offs inside a first-tier does not affect the results.

The performances of PBR scheme and guard channel scheme are compared in Figure 9 which was obtained from analytical model at cell capacity of 20 bandwidth

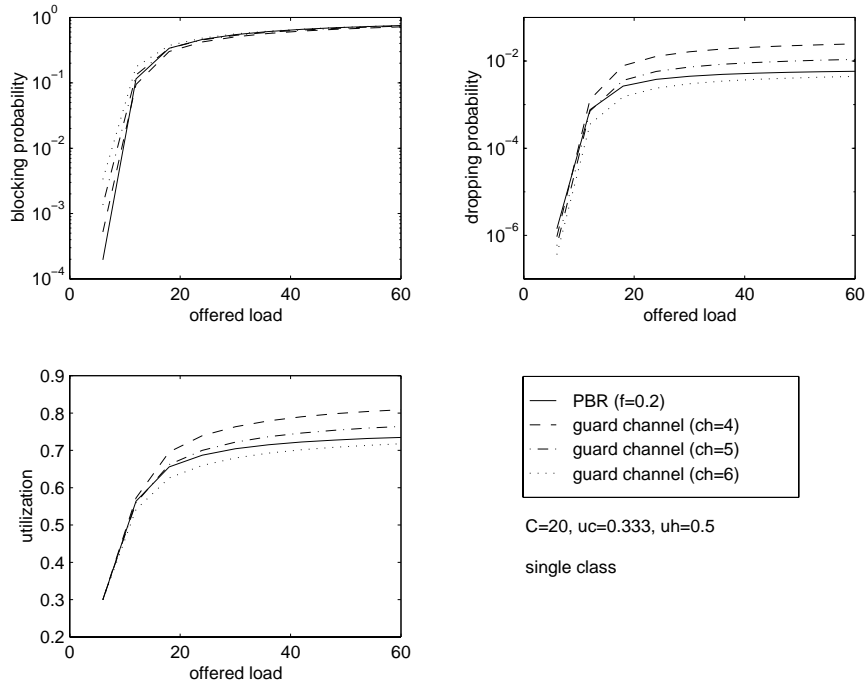


Figure 9: Performance comparison between PBR and guard channel scheme

units. Three different values for the number of guard channels are selected to compare with PBR scheme with $f = 0.2$. For the case of dropping probability, guard channel scheme with $C_h = 6$ performs most similarly to the selected PBR. It, however, shows much higher blocking probability than PBR, when offered load is light. On the other hand, guard channel scheme with $C_h = 4$ works very closely to the selected PBR in terms of blocking probability, while it shows higher dropping probability than PBR as offered load increases. As seen in this figure, it is proved that PBR scheme adjusts itself fairly well to the different traffic conditions. When a cellular network is not crowded, it admits more new calls by acting like guard channel scheme with $C_h = 4$, while it achieves better dropping probability at heavy offered load by acting like guard channel scheme with $C_h = 6$.

Three performance measures of PBR scheme are plotted as a function of f in Figure 10, for four different levels of offered load. As f increases, the amount of

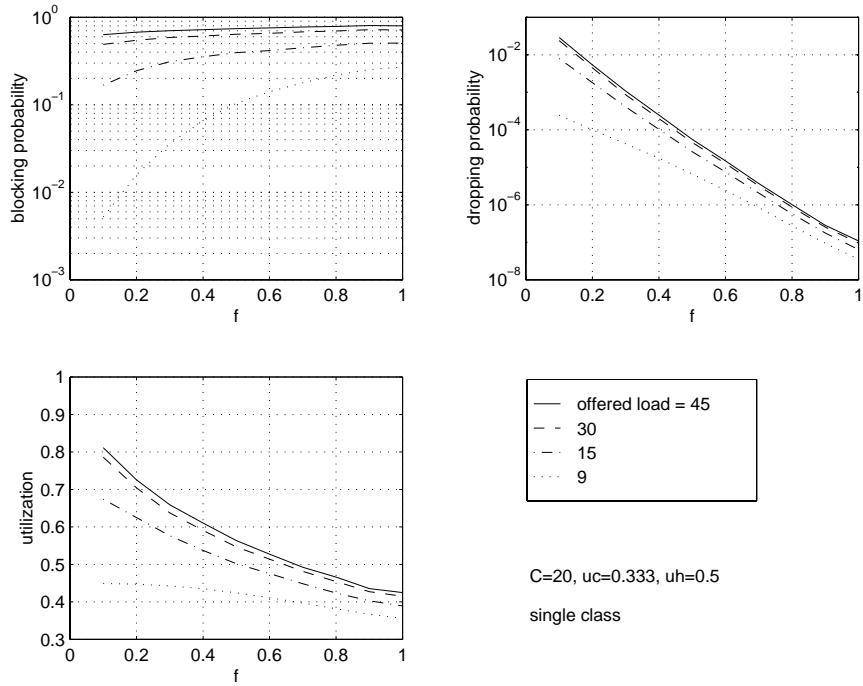


Figure 10: Performance of PBR vs. f

bandwidth units requested for reservation also increases, assuming constant call population. Naturally, blocking probability becomes higher and dropping probability and bandwidth utilization becomes lower.

In fact, the optimal value of f should be determined as a function of speed of mobile stations, average call duration, cell size, and so on, since f indicates the amount of bandwidth units required for hand-off calls that will occur for the next certain period. This kind of plot, however, may provide assistance in selecting f which satisfies the given QOS. Although f ranges between 0 and 1, the values of f which are larger than the hand-off probability, p_h , defined as Equation 8, may not be considered, since mobile stations will finish their call with probability of $(1 - p_h)$, instead of hand-off. The adaptive method to choose f is for further study.

Figure 11 shows instantaneous dropping probabilities of PBR scheme and guard channel scheme under varying traffic load. The traffic load is let to change as seen

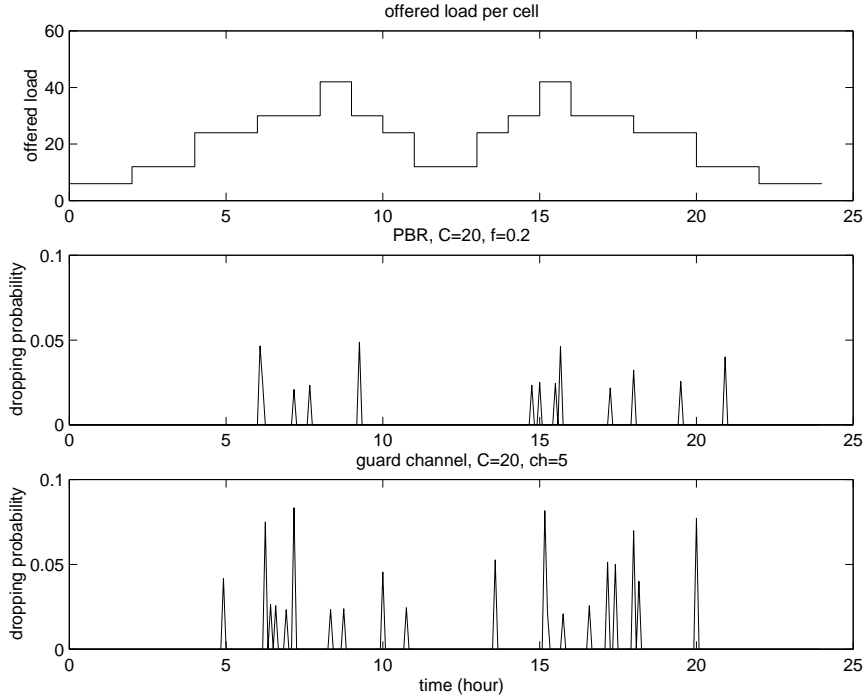


Figure 11: Instantaneous dropping probability with varying traffic load

in the first graph in the figure for 24 hours, and dropping probability during every 5 minute interval is computed. The number of guard channels is set to 5 bandwidth units to compare with PBR scheme with $f = 0.2$, since the average load for 24 hours given for this example is 21 Erlangs, and both performs similarly at this level of offered load as seen in Figure 9. According to this figure, PBR scheme shows much lower peak value and less fluctuations for dropping probability than guard channel scheme does. It implies that PBR scheme adjusts to changes in traffic load more effectively and promptly.

6 Conclusion

To guarantee continuous services to cellular traffic after hand-off, bandwidth reservation method is employed to provide prioritized handling of hand-off calls. For the effective and adaptive bandwidth reservation, some mechanisms have been proposed

to make bandwidth reservation by predicting mobile user's movement, which requires a lot of overhead of base stations. The population-based reservation scheme, proposed in this paper, however, is very simple to implement and requires almost no overhead of base stations. It dynamically adjusts the amount of bandwidth reservation according to changing traffic condition, which is represented by current call population in cellular networks.

An analytical model for cellular networks is developed to evaluate the performance of proposed scheme and guard channel scheme. For analytical tractability, approximating assumptions are used in building the analytical model, and verified through simulation. Using analytical model and simulation, it is shown that the proposed scheme effectively adapts to changing traffic conditions. It shows good dropping performance at heavy traffic load, while maintaining low blocking probability at light traffic load. It also shows less fluctuations of dropping probability against varying traffic load, compared to guard channel scheme.

In this work, we investigate the performance of the proposed PBR scheme at arbitrary values of fractional parameter, f . The value of this parameter, however, should be determined based on the given QoS requirement, and, to meet the requirement, the value of f must be able to be adjusted dynamically according to the changing traffic conditions. The adaptive method to adjust f will be studied and we will continue investigating PBR scheme under various traffic conditions.

References

- [1] A. S. Acampora and M. Naghshineh, "Control and quality-of-service provisioning in high-speed microcellular networks," *IEEE Personal Communications*, Second Quarter 1994, pp. 36–42.
- [2] S. Choi and K. G. Shin, "Predictive and adaptive bandwidth reservation for hand-offs in QoS-sensitive cellular networks," in *Proc. ACM SIGCOMM '98*, Vancouver, Canada, Sep. 1998, pp. 155–166.

- [3] D. C. Cox and D. O. Reudnik, "Increasing channel occupancy in large-scale mobile radio systems: dynamic channel assignments," *IEEE Trans. on Vehicular Technology*, Vol. VT-22, Nov. 1973, pp. 218–222.
- [4] A. Gersht and K. J. Lee, "Virtual-circuit load control in fast packet-switched broadband networks," in *Proc. Globecom '88*, pp. 214–220.
- [5] 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. on Vehicular Technology*, Vol. VT-35, No. 3, Aug. 1986, pp. 77–92.
- [6] S. Katsuki and E. Hato, "A study of drivers' behavior and traffic management," in *Proc. 1994 Vehicle Navigation & Information Systems Conference*, Yokohama, Japan, 1994, pp. 255–258.
- [7] W. C. Y. Lee, "Smaller cells for greater performance," *IEEE Communications Magazine*, Nov. 1991, pp. 19–23.
- [8] D. A. Levine, I. F. Akyildiz, and M. Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," *IEEE/ACM Trans. on Networking*, Vol. 5, No. 1, Feb. 1997, pp. 1–12.
- [9] C. Oliveira, J. B. Kim and T. Suda, "Quality-of-service guarantee in high-speed multimedia wireless networks," in *Proc. ICC '96*, pp. 728–733.
- [10] M. Sidi and D. Starobinski, "New call blocking versus handoff blocking in cellular networks," *Wireless Networks*, Vol. 3, No. 1, 1997, pp. 15–27.
- [11] W. J. Stewart, *Introduction to the numerical solution of Markov chains*, Princeton University Press, 1994.
- [12] J. Tajima and K. Imamura, "A strategy for flexible channel assignment in mobile communication systems," *IEEE Trans. on Vehicular Technology*, Vol. 37, May 1988, pp. 92–103.
- [13] S. Tekinay and B. Jabbari, "Handover and channel assignment in mobile cellular networks," *IEEE Communications Magazine*, Nov. 1991, pp. 42–46.
- [14] P. Tran-Gia and F. Hubner, "An analysis of trunk reservation and grade of service balancing mechanisms in multiservice broadband networks," *IFIP Transactions C-15: Modelling and Performance Evaluation of ATM Technology*, North-Holland, Amsterdam, 1993, pp. 83–97.