Optimal Admission Control Using Handover Prediction in Mobile Cellular Networks[*]

Vicent Pla, José Manuel Giménez-Guzmán,[†] Jorge Martínez and Vicente Casares-Giner
Department of Communications
Universidad Politécnica de Valencia (UPV)
ETSIT Camí de Vera s/n, 46022 Valencia, Spain.
Telephone: +34 963879733, Fax: +34 963877309
E-mail: `(vpla,jmartinez,vcasares)@dcom.upv.es`
`jogiguz@doctor.upv.es`

Abstract

In this paper we study the impact of incorporating handover prediction information into the session admission control process in mobile cellular networks. The objective is to compare the performance of optimal policies obtained with and without the predictive information. A prediction agent classifies mobile users in the neighborhood of a cell into two classes, those that will probably be handed over into the cell and those that probably will not. We consider the classification error by modeling the false-positive and non-detection probabilities. Two different approaches to compute the optimal admission policy were studied: *dynamic programming* and *reinforcement learning*. Results show significant performance gains when the predictive information is used in the admission process.

1 Introduction

Future mobile communication systems are expected to support broadband multimedia services with diverse Quality of Service (QoS) requirements. The cellular architecture is used in wireless networks to utilize the radio spectrum efficiently. Since mobile users may change cells a number of times during the lifetime of their sessions, availability of wireless network resources at the session setup time does not necessarily guarantee that will be available throughout the lifetime of a session. Thus users may experience a performance degradation due to their mobility. This problem is magnified by the current trend to reduce the cell size to accommodate more mobile users in a given area as handover events will occur at a much higher rate [1].

Session Admission Control (AC) is a key aspect in the design and operation of multiservice cellular networks that provide QoS guarantees. The design of the AC system must take into account not only packet level issues (like delay, jitter or losses) but also session level issues (like blocking probabilities of both session setup and handover requests) [2]. This paper explores the second type of issues from a novel optimization perspective.

AC in single service cellular systems has been thoroughly studied, see for instance the seminal work by Hong and Rappaport [3] or more recent papers like [4–6] and references therein. While most of these papers provide intuitive reservation schemes for AC a more insightful approach is adopted in [7] and [8], where AC in single service scenarios is regarded as an optimization problem. Admission control in the presence of mobility and multiple services is not that well studied although some contributions in this direction can be found in the literature [9–12].

Most of the proposed AC policies take the admission decision using only state information local to the cell, such as the number of active sessions per service. However, mobile cellular networks permit to have some anticipated knowledge about forthcoming requests, and more importantly, this predictive information concerns the most sensitive requests, namely, the handover attempts. Following that observation, several mobility prediction schemes and associated AC policies have appeared for single service scenarios, see for example [2, 13–16] and

references therein. A common feature of these studies is the proposal of heuristic AC policies which exploit the specific information provided by each mobility prediction scheme.

In this paper we study AC policies that make use of predictive information from an optimization perspective, in both single service and multiservice scenarios. Our goal is to obtain the optimal policy for a given amount of information provided by the mobility prediction scheme. We consider that such approach has not been sufficiently explored. The deployment of classical optimization techniques applied to this type of scenario provide results that help to define theoretical limits for the gain that can be expected, which could not be set by simply deploying heuristic approaches.

For a single service scenario, in [17] the authors determine a near-optimal policy by means of a genetic algorithm that takes into account not only the cell state but also the state of neighboring cells. However, results in [17] show that the performance gain obtained when using these additional information is rather insignificant. We reached the same conclusion using a different optimization method. These disappointing results suggest that the prediction of possible forthcoming handovers obtained from the occupancy state of the neighboring cells is not sufficiently specific.

Here we go a step further and evaluate the performance gain that can be obtained when the AC process is provided with more specific information. In our study the total population of active terminals in the surroundings of a cell is divided into two classes: those that will handover into the cell with high probability and those that will not. Our model of the prediction agent (PA) does not provide information about the time instant at which the handover will occur. We postpone the study of this scenario for a future work. Obviously, the more information is provided by the PA the better the performance of the AC policy will be. Unfortunately, the complexity of the PA and the optimization process increases as more information is provided.

The rest of the paper is structured as follows. In Section 2 we describe the model of the system and of the PA. The optimization approaches, both in single service and multiservice scenarios, are presented in Section 3. The numerical evaluation of the proposed model is introduced in Section 4. Finally, a summary of the paper and some concluding remarks are given in Section 5.

2   Model Description

We consider a single cell system and its neighborhood, where the cell has a total of C resource units, being the physical meaning of a unit of resource dependent on the specific technological implementation of the radio interface. A total of N different services are offered by the system. For each service new and handover session arrivals are distinguished so that there are N services and $2N$ types of arrivals.

For the sake of mathematical tractability we make the common assumptions of Poisson arrival processes and exponentially distributed random variables for cell residence time and session duration. The arrival rate for new (handover) sessions of service i is λ_i^n (λ_i^h) and a request of service i consumes b_i resource units, $b_i \in \mathbb{N}$. For service i, the session duration and cell residence rates are μ_i^s and μ_i^r respectively. The resource holding time in a cell for service i is also exponentially distributed with rate $\mu_i = \mu_i^s + \mu_i^r$.

2.1   Prediction Agent

Two main types of prediction systems have been studied in the literature [18]: history-based and positioning-based. Schemes of the first group compute movement patterns to determine movement predictions statistically, like for example estimation of the handover arrival rate to a cell. Given that mobile terminals (MTs) having a similar movement history are more likely to have common movement patterns, measurement data can be aggregated into groups, improving in this way the performance of the prediction system [2]. For schemes of the second group [19], the probability of reserving resources for a handover session increases as the MT approaches the cell. A further enhancement can be achieved by estimating the direction and speed of the MT and extrapolate this information to determine future movements [16]. It is clear that both methods can be combined to improve performance even further .

Given that the focus of our study was not the design of the PA we used a generic model of it instead. The PA informs the AC system about the number of active terminals in the neighborhood that are forecasted to produce a handover into the cell. The amount of time elapsed since an active MT is deemed as "probably producing a handover" until the handover actually occurs is not predicted by the PA and we model it by an

(a) Basic operation of the prediction scheme   (b) Proposed model of the PA
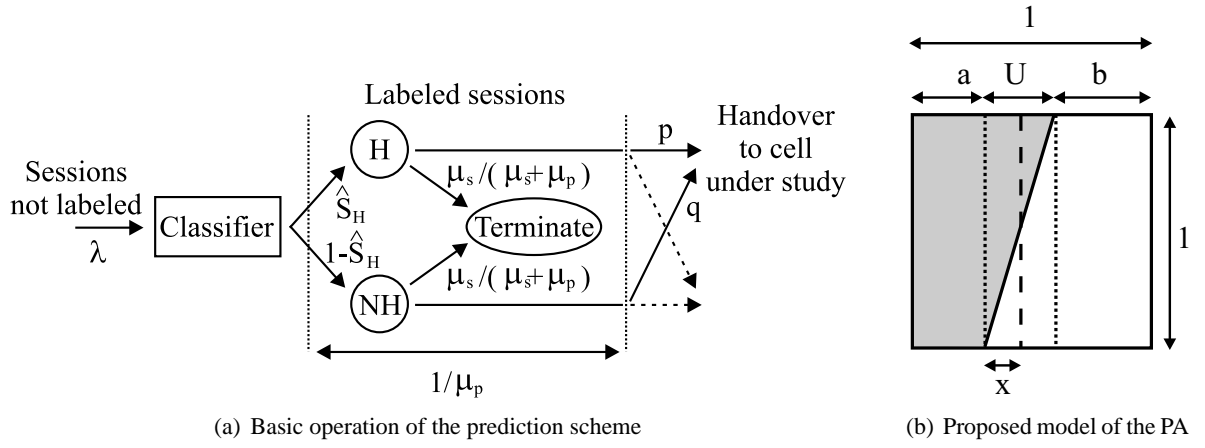
Figure 1: Model of the PA.

exponential random variable, which is an approximation widely used in the literature. The exponential assumption has been considered a good approximation for the cell dwell times [20], where in the worst case indicates general performance trends, for the time in the handover area [21] and for the inter-arrival time of handover requests [22].

An active MT entering the cell neighborhood is labeled by the PA as "probably producing a handover" (H) or the opposite (NH), according to some of its characteristics (position, trajectory, velocity, historic profile, . . .) and/or some other information (road map, hour of the day, . . .). After an exponentially distributed time, the actual destiny of the MT becomes definitive and either a handover into the cell occurs or not (for instance because the session ends or the MT moves to another cell). The AC system is aware at any time of the number of MTs labeled as H. In general, the classification into H or NH is not completely accurate and therefore our model incorporates the probabilities of non-detection and false-positive. We assume that only one session is active per MT.

Our model of the PA is characterized by three parameters: the average sojourn time of the MT in the predicted stage μ_p^{-1}, the probability p of producing a handover if labeled as H and the probability q of producing a handover if labeled as NH. Note that in general $q \neq 1 - p$. The basic operation of the prediction model is shown in Fig. 1(a). The values of p and q relate to each other through the specific model of the PA, which is shown in Fig. 1(b). In the figure there is a square (with a surface equal to one) representing the population of MTs that is going to be classified by the PA. The shaded area represents the fraction of active MTs that will ultimately move into the cell, while the white area represents the rest of active MTs. It should be pointed out that those active MTs that will ultimately move into the cell might do so after the session has terminated. The classifier sets a threshold, which is represented by a vertical dashed line, to discriminate between those MTs that will likely produce a handover and those that will not. MTs falling on the left side of the threshold are labeled as H and those on the right side as NH. There exists an uncertainty zone, which is represented by the slope of the line separating the shaded and white areas. Parameter x represents the relative position of the classifier threshold. This uncertainty produces classification errors: the white area on the left of the threshold and the shaded area on the right of the threshold. Although for simplicity we use a linear model for the uncertainty zone it would be rather straightforward to consider a different model. Let us introduce the following notation referring to the areas in Fig. 1(b): S_H denotes the shaded area and represents the fraction of active MTs in the cell neighborhood that will ultimately move into the cell, while the white area represents the rest of active MTs; $\hat{S}_H$ denotes the surface on the left of the threshold and represents the fraction of MTs labeled as H; $\hat{S}_H^e$ denotes the white surface on the left of the threshold and represents the fraction of MTs labeled as H that will not produce a handover; $\hat{S}_{NH}^e$ denotes the shaded surface on the right of the threshold and represents the fraction of MTs labeled as NH that will produce a handover. From Fig. 1(b) it follows that

$$1 - p = \frac{\hat{S}_H^e}{\hat{S}_H} = \frac{x^2}{2U(a + x)} \; ; \qquad q = \frac{\hat{S}_{NH}^e}{1 - \hat{S}_H} = \frac{(U - x)^2}{2U(1 - a - x)}$$

Parameters a and b can be expressed in terms of the fraction of MTs moving into the target cell S_H and the

degree of uncertainty in the prediction U,

$$a = S_H - U/2 \, ; \qquad b = 1 - S_H - U/2$$

and then

$$1 - p = \frac{\hat{S}_H^e}{\hat{S}_H} = \frac{x^2}{U(2S_H - U + 2x)} \, ; \qquad q = \frac{\hat{S}_{NH}^e}{1 - \hat{S}_H} = \frac{(U - x)^2}{U(2 - 2S_H + U - 2x)}$$

Referring to Fig. 1(a), the value of the session rate entering the classifier λ is chosen so that the system is in statistical equilibrium, i.e. the rate at which handover sessions enter a cell (λ_h^{in}) is equal to the rate at which handover sessions exit the cell (λ_h^{out}). Omitting the subscript referring to the service, we can write

$$\lambda_h^{in} = \lambda S_H \frac{\mu_p}{\mu_p + \mu_s}$$

$$\lambda_h^{out} = \frac{\mu_r}{\mu_r + \mu_s} [(1 - P_n)\lambda_n + (1 - P_h)\lambda_h^{in}] \tag{1}$$

where P_n (P_h) is the blocking probability of new (handover) requests.

Making $\lambda_h^{in} = \lambda_h^{out}$, substituting P_h by

$$P_h = \frac{P_{ft}}{1 - P_{ft}} \cdot \frac{\mu_s}{\mu_r}$$

where P_{ft} is the probability of forced termination of a successfully initiated session, and after some algebra we get

$$\lambda = (1 - P_n)(1 - P_{ft})\lambda_n(\mu_r/\mu_s + \mu_r/\mu_p)(1/S_H)$$

3 Optimization of the Admission Policy

The information provided by the PA and the state of the cell (number of active sessions) is used to find the optimal admission policy and its performance. The generic definition of the system state space is

$$\mathcal{S} := \big\{ \boldsymbol{x} = (x_1, \ldots, x_N, x_{N+1}, \ldots, x_{2N}) \big\}.$$

where x_i is the number of ongoing sessions of service i, $1 \leq i \leq N$, in the cell under study, x_{i+N} is the number of ongoing sessions of service i in the cell neighborhood which are labeled as H.

We make use of the theory of *Markov decision processes* (MDPs) [23] to find a policy that minimizes the average expected cost rate. A MDP can be viewed as a stochastic automaton in which an agent's actions influence the transitions between states, and costs are imputed depending on the states visited by an agent. Formally, a MDP can be defined as a tuple $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{C}\}$, where $\mathcal{S}$ is a finite set of states, $\mathcal{A}$ is a finite set of actions, $\mathcal{P}$ is a state transition function and $\mathcal{C}$ is a cost function. The agent can control the state of the system by choosing actions a from $\mathcal{A}$, influencing in this way the state transitions. The results of an action are stochastic in that the actual transition cannot be predicted with certainty. The transition function $\mathcal{P} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$ specifies the effect of taking an action at a given state. We denote by $p_{\boldsymbol{xy}}(a)$ the transition probability from state $\boldsymbol{x}$ to state $\boldsymbol{y}$ when action a is taken at state $\boldsymbol{x}$, and require that $0 \leq p_{\boldsymbol{xy}}(a) \leq 1 \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}, \forall a \in \mathcal{A}$ and $\sum_{\boldsymbol{y} \in S} p_{\boldsymbol{xy}}(a) = 1$.

The agent knows the state of the system $\boldsymbol{x}$ at any time and it chooses actions based only on the current state. We consider deterministic stationary Markovian policies, $\pi : \mathcal{S} \rightarrow \mathcal{A}$, which defines the next action of the agent based only on the current state $\boldsymbol{x}$, i.e. an agent adopting this policy performs action $\pi(\boldsymbol{x})$ in state $\boldsymbol{x}$. For the problems we consider, optimal stationary Markovian policies always exist.

We assume a bounded, integer-valued cost function $\mathcal{C} : \mathcal{S} \rightarrow \mathbb{N}$, and denote by $c(\boldsymbol{x}, a)$ the finite cost for executing action a in state $\boldsymbol{x}$. Different optimality criteria can be adopted to measure the cost of a policy π, all measuring in some way the cost accumulated by the agent as it follows policy π. In this work we focus on the average cost criterion because is more appropriate for the problem under study than other discounted cost approaches [24].

When the agent minimizes a discounted cumulative sum of costs, and we suppose that starting from state x_0 and using policy π the system evolves through states $\{x_0, x_1, \ldots, x_t\}$ in interval $[0, t]$, then the discounted cost of policy π is defined as

$$v^\pi(x_0) = \lim_{t \to \infty} E\Big(\sum_{m=0}^{t} \gamma^m c\big(x_m, \pi(x_m)\big) \Big)$$

where $\gamma \leq 1$ is the discount factor. This allows simpler computational methods to be used, as discounted total reward will be finite. Note that an agent minimizing $v^\pi(x)$ will prefer actions that generate an immediate cost reduction instead of those ones generating the same cost reduction some steps into the future, due to the discounted factor. In many situations, discounted methods can be justified by the nature of the problem, like in economics or when the tasks terminate. Notwithstanding, a more natural measure of optimality exists for infinite-horizon tasks like the one studied in this work, based on minimizing the average cost per action. We define the total cost accumulated in the interval $[0, t]$ as

$$w^\pi(x_0, t) = \sum_{m=0}^{t} c\big(x_m, \pi(x_m)\big)$$

If the environment is stochastic then $w^\pi(x_0, t)$ is a random variable. Under the average cost criterion we seek to minimize the average expected cost rate over time t, as $t \to \infty$. When the system starts at state x and follows policy π, the average expected cost rate is denoted by $g^\pi(x)$ and is defined as:

$$g^\pi(x) = \lim_{t \to \infty} \frac{1}{t} E\left[w^\pi(x, t)\right]$$

In this work we minimize a weighted sum of loss rates and therefore the average cost criterion is more appropriate for the problem under study than other discounted cost approaches. In a system like ours, it is not difficult to see that for every deterministic stationary policy the embedded Markov chain has a unichain transition probability matrix, and therefore the average expected cost rate does not vary with the initial state [25]. We call it the "cost rate" of the policy π, denote it by g^π and consider the problem of finding the policy π^* that minimizes g^π, which we name the optimal policy.

It can be shown that for our system

$$g^\pi = \sum_{i=1}^{N} (\beta_i^n P_i^n \lambda_i^n + \beta_i^h P_i^h \lambda_i^h)$$

where β_i^n (β_i^h) is the relative weight associated to the blocking of a new (handover) request and P_i^n (P_i^h) is the blocking probability of new (handover) requests, both of service i. In general, $\beta_i^n < \beta_i^h$ to account for the fact that the blocking of a handover request is less desirable than the blocking of a new session request.

Two different optimization approaches have been used to find the optimal policy. The first approach is based on *dynamic programming* (DP) [23], specifically we used a policy improvement method [25]. This approach is applied to a single service scenario. The second is an automatic learning approach based on the theory of *reinforcement learning* (RL) [26], more specifically we used the average reward reinforcement learning algorithm proposed in [27]. This approach is applied to a multiservice scenario. DP gives an exact solution and allows to evaluate the theoretical limits of incorporating movement prediction in the AC problem, whereas RL tackles more efficiently the curse of dimensionality and offers the important advantage of being a model-free method, i.e. transition probabilities and average costs are not needed by the method. As a consequence, in the RL approach, neither the numerical values of the PA parameters (p, q, μ_p) nor the arrival rates and holding times need to be known beforehand. Moreover, the learning algorithm can adapt to variations of those parameters.

3.1 Single Service

In this section we describe the optimization approach based on DP. Since there is only one service through this section we simplify notation by omitting the subscript referring to the service, i.e. $\lambda_n = \lambda_1^n$, $\lambda_h = \lambda_1^h$, $\mu_s = \mu_1^s$, $\mu_r = \mu_1^r$, $\mu = \mu_1$. Without loss of generality, we assume that $b_1 = 1$.
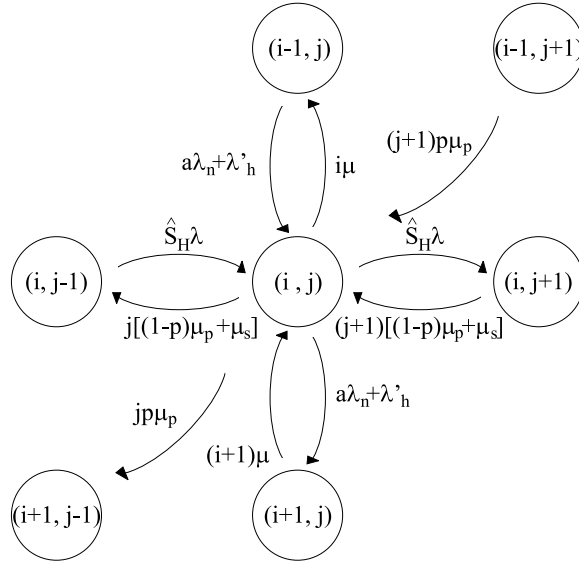
Figure 2: Transition rates.

Let us represent the system state by (i, j) where i is the number of active sessions in the cell and j is the number of MTs in the cell neighborhood labeled as H. The set of possible states of the system is

$$\mathcal{S} := \left\{ \boldsymbol{x} = (i, j) : 0 \le i \le C; \ 0 \le j \le C_p \right\}$$

where C_p represents the maximum number of MT that can be labeled as H at a given time. We use a large value for C_p so it has no practical impact on our results. At each state $(i, j), i < C$, the set of possible actions is defined by $\mathcal{A} := \{a : a = 0, 1\}$, being $a = 0$ the action that rejects an incoming new session and $a = 1$ the action that accepts an incoming new session. Handover sessions have priority over new sessions and they are accepted as long as resources are available ($i < C$). At state (C, j) only the action $a = 0$ is possible.

Figure 2 shows the transition rates from and to state (i, j). Note that some of the transition rates depend on the action $a = 0, 1$. In the figure we introduced λ'_h which is the average arrival rate of handovers that have not been predicted, and it is given by

$$\lambda'_h = (1 - \hat{S}_H) \frac{\mu_p}{\mu_p + \mu_s} q\lambda$$

where λ is the input rate to the PA.

The model described is a continuous-time Markov chain, which we convert to a *discrete time Markov chain* (DTMC) by applying uniformization (see [28, Section 4.7]). It can be shown that $\Gamma = C_p(\mu_p + \mu_s) + C(\mu_r + \mu_s) + \lambda + \lambda_n$ is an uniform upper-bound for the outgoing rate of all the states. If $r_{\boldsymbol{xy}}(a)$ denotes the transition rate from state $\boldsymbol{x}$ to state $\boldsymbol{y}$ when action a is taken at state $\boldsymbol{x}$, then the transition probabilities of the resulting DTMC are given by

$$p_{\boldsymbol{xy}}(a) = \frac{r_{\boldsymbol{xy}}(a)}{\Gamma} \quad \text{if } \boldsymbol{y} \ne \boldsymbol{x} \qquad \text{and} \qquad p_{\boldsymbol{xx}}(a) = 1 - \sum_{\boldsymbol{y} \in S} p_{\boldsymbol{xy}}(a)$$

We define the incurred cost rate at state $\boldsymbol{x}$ when action a is taken by

$$c(\boldsymbol{x}, a) = \begin{cases} 0, & i < C, \ a = 1 \\ \lambda_n, & i < C, \ a = 0 \\ \lambda_n + \beta(\lambda'_h + jp\mu_p), & i = C, \ a = 0 \end{cases}$$

The weighting factor β (typically) accounts for the fact that blocking a handover request is less desirable than blocking of a new session request. Costs are defined so that the average expected cost rate g^π equals a weighted sum of the average loss rate of new sessions ($P_n\lambda_n$) and handover attempts ($P_h\lambda_h$), i.e.

$$g^\pi = \lim_{n \to \infty} E\left[\frac{1}{n+1} \sum_{t=0}^{n} c\big(\boldsymbol{x}(t), \pi(\boldsymbol{x}(t))\big) \right] = P_n\lambda_n + \beta P_h\lambda_h$$

where $\boldsymbol{x}(t)$ is the state visited at time t when policy π is deployed. Loss rates are given by

$$P_n \lambda_n = \sum_{\boldsymbol{x}: \pi(\boldsymbol{x})=0} \lambda_n p(\boldsymbol{x}) ; \qquad P_h \lambda_h = \sum_{\substack{\boldsymbol{x}=(C,j) \\ 0 \leq j \leq C_p}} (\lambda_h' + jp\mu_p)p(\boldsymbol{x})$$

where $p(\boldsymbol{x})$ is the stationary probability of state $\boldsymbol{x}$. Thus, the optimization problem pursues to find the policy π^* that minimizes g^π. Since state $(0,0)$ can be reached from any other state regardless of the policy deployed, by virtue of the *Corollary 6.20* and the subsequent remark, both in [23], we know that an optimal stationary policy exists.

If we denote by $h(\boldsymbol{x})$ the relative cost rate of state $\boldsymbol{x}$ under policy π, then we can write

$$h(\boldsymbol{x}) = c(\boldsymbol{x}, \pi(\boldsymbol{x})) - g^\pi + \sum_{\boldsymbol{y}} p_{\boldsymbol{xy}}(\pi(\boldsymbol{x}))h(\boldsymbol{y}) \qquad \forall \boldsymbol{x} \qquad (2)$$

from which we can obtain the average cost and the relative costs $h(\boldsymbol{x})$ up to an undetermined constant. Thus we arbitrarily set $h(0,0) = 0$ and then solve the linear system of equations (2) to obtain g^π and $h(\boldsymbol{x})$, $\forall \boldsymbol{x}$. Having obtained the average and relative costs under policy π an improved policy π' can be calculated as

$$\pi'(\boldsymbol{x}) = \arg\min_{a=0,1} \left\{ c(\boldsymbol{x}, a) - g^\pi + \sum_{\boldsymbol{y}} p_{\boldsymbol{xy}}(a)h(\boldsymbol{y}) \right\}$$

so that the following relation holds $g^{\pi'} \leq g^\pi$. Moreover, if the equality holds then $\pi' = \pi = \pi^*$, where π^* denotes the optimal policy, i.e. $g^{\pi^*} \leq g^\pi \; \forall \pi$.

If we repeat iteratively the solution of system (2) and the policy improvement until we obtain a policy which does not change after improvement. This process is called *Policy Iteration* [25, Section 8.6] and it leads to the average optimal policy in a finite — and typically small — number of iterations.

3.2 Multiservice

In this section we study a scenario in which the information available to the AC system is also the state of the cell and the state of the cell neighborhood. The state could be conceptually represented by a vector with $2N$ elements, each of them being the number of ongoing sessions initiated either as new or handover requests. However, we adopt a more compact representation of the state space, including only the number of resource units occupied in the cell and in its neighborhood. This is motivated by the fact that reducing the state space helps the RL algorithm to find better solutions and by the conclusions of previous studies [12], which show that the performance of policies which base their decisions only on the number of resource units occupied (trunk reservation policies), are close to the performance of the optimum policy.

We formulate the optimization problem as an infinite-horizon finite-state semi-Markov decision process (SMDP) under the average cost criterion. It is evident that we search for policies that minimize g^π. Decision epochs correspond to time instants at which arrivals occur. Given that no actions are taken at session departures, then only the arrival events are relevant to the optimization process. At each decision epoch the system has to select one action from the set of possible actions $\mathcal{A} := \{0 = \text{reject}, 1 = \text{admit}\}$.

The state space is defined as

$$\mathcal{S} := \{\boldsymbol{x} = (x_0, x_1, k) : x_0, x_1, k \in \mathbb{N}; x_0 \leq C; x_1 \leq C_p, 1 \leq k \leq (2N-1)\}$$

where x_0 is the number of resource units occupied in the cell under study, x_1 is the number of resource units occupied in the cell neighborhood by ongoing sessions labeled as H by the PA and k, $1 \leq k \leq (2N-1)$, is the arrival type. We select one of the $2N$ arrival types as the highest priority one, being its requests always admitted while free resources are available, and therefore no decisions are taken for them.

The cost structure is defined as follows. At any decision epoch, the cost incurred by accepting any arrival type is zero and by rejecting a new (handover) request of service i is β_i^n (β_i^h). With this framework, further accrual of cost occurs when the system has to reject requests of the highest priority arrival type between two decision epochs.

We denote by $h(\boldsymbol{x})$ the relative cost rate of state $\boldsymbol{x}$, which can be interpreted as the expected long-term advantage in total cost for starting in state $\boldsymbol{x}$ in addition to $t \cdot g^\pi$, the expected total cost at time t on the average. The Bellman optimality recurrence equations for an SMDP under the average cost criterion have the form

$$h^*(\boldsymbol{x}) = \min_{a \in A_x} \left\{ c(\boldsymbol{x}, a) - g^* \tau(\boldsymbol{x}, a) + \sum_{\boldsymbol{x} \in S} p_{\boldsymbol{x}\boldsymbol{y}}(a) h^*(\boldsymbol{y}) \right\}$$

where $h^*(\boldsymbol{x})$ is an optimal state dependent relative value function and $c(\boldsymbol{x}, a)$ and $\tau(\boldsymbol{x}, a)$ are the average cost and the average sojourn time when taking decision a in state $\boldsymbol{x}$. The greedy policy π^* defined by selecting actions that minimize the right-hand side of the above equation is gain-optimal [27].

If the parameters of the model can be derived, then the solution to the Bellman equations can be obtained through dynamic or linear programming techniques. In multiservice scenarios, where the number of states can be large, the derivation of the model parameters can be complex and make the problem intractable (course of dimensionality). We propose an alternative approach based on a reinforcement learning algorithm named Semi-Markov Average Reward Technique (SMART) [27].

The Bellman equations can be rewritten as

$$h^*(\boldsymbol{x}, a) = \min_{a \in A_x} \left\{ c(\boldsymbol{x}, a) - g^* \tau(\boldsymbol{x}, a) + \sum_{\boldsymbol{x} \in S} p_{\boldsymbol{x}\boldsymbol{y}}(a) \min_{a' \in A_y} h^*(\boldsymbol{y}, a') \right\}$$

where $h^*(\boldsymbol{x}, a)$ is the average expected relative value of taking the optimal action a in state $\boldsymbol{x}$ and then continuing indefinitely by choosing actions optimally. Then, the optimal policy is

$$\pi^*(\boldsymbol{x}) = \arg\min_{a \in A_x} h^*(\boldsymbol{x}, a)$$

The SMART algorithm estimates $h^*(\boldsymbol{x}, a)$ by simulation, using a temporal difference method (TD(0)). If at the $(m-1)^{th}$ decision epoch the system is in state $\boldsymbol{x}$, action a is taken and the system is found in state $\boldsymbol{y}$ at the mth decision epoch, then we update the relative state-action values as follows:

$$h_{new}(\boldsymbol{x}, a) = (1 - \alpha_m) h_{old}(\boldsymbol{x}, a) + \alpha_m \left\{ c_m(\boldsymbol{x}, a, \boldsymbol{y}) - g_m \tau_m(\boldsymbol{x}, a, \boldsymbol{y}) + \min_{a' \in A_y} h_{old}(\boldsymbol{y}, a') \right\}$$

where $c_m(\boldsymbol{x}, a, \boldsymbol{y})$ is the actual cumulative cost incurred between the two successive decision epochs, $\tau_m(\boldsymbol{x}, a, \boldsymbol{y})$ is the actual sojourn time between the decision epochs, α_m is the learning rate parameter at the mth decision epoch, and g_m is the average cost rate estimated as:

$$g_m = \frac{\sum_{k=1}^{m} c_k \left(\boldsymbol{x}_{(k)}, a_{(k)}, \boldsymbol{y}_{(k)} \right)}{\sum_{k=1}^{m} \tau_k \left(\boldsymbol{x}_{(k)}, a_{(k)}, \boldsymbol{y}_{(k)} \right)}$$

4 Numerical Evaluation

We evaluated the performance gain when introducing prediction by the ratio g_{wp}^π / g_p^π, where g_p^π (g_{wp}^π) is the expected average cost rate of a policy that is optimal in a system with (without) prediction. We assume a circular-shaped cell of radio r and a dick-shaped neighborhood with inner (outer) radio $1.0r$ ($1.5r$).

The values of the parameters that define the scenario are: $C = 10$ and $C_p = 60$ resource units, $N_h = \mu_i^r / \mu_i^s = 1$, $\mu_i^r / \mu_i^p = 0.5$, $S_H = 0.4$, $x = U/2$. For the single service scenario ($N = 1$) we use $b_1 = 1$, $\lambda_1^n = 1$, $\mu_1 = \mu_1^s + \mu_1^r = 1$, $\beta_1^n = 1$, and $\beta_1^h = \beta = 20$. As mentioned in Section 2.1, the value of λ is chosen so that the system is in statistical equilibrium, i.e. the rate at which handover sessions enter a cell equals the rate at which handover sessions exit the cell. For small values of P_1^n ($\approx 10^{-2}$) and P_1^{ft} ($\approx 10^{-3}$), we make the approximation $\lambda = 0.989 \lambda_1^n (N_h + \mu_1^r / \mu_1^p)(1/S_H)$. For the multiservice scenario we use $N = 2$ services, $b_1 = 1$ and $b_2 = 2$ resource units. The arrival rates of new sessions to the cell are $\lambda_1^{nc} = 0.8 \lambda_T$, $\lambda_2^{nc} = 0.2 \lambda_T$, where $\lambda_T = 2$. The ratio of arrival rates of new sessions to the cell neighborhood (ng) and to the cell (nc) is made equal to the ratio of their surfaces, $\lambda_i^{ng} = 1.25 \lambda_i^{nc}$. The ratio of handover arrival rates to

the cell neighborhood from the outside of the system (ho) and from the cell (hc) is made equal to the ratio of their perimeters, $\lambda_i^{ho} = 1.5\lambda_i^{hc}$. From equation (1) it follows that $\lambda_i^{hc} = (1 - P_i^n)(1 - P_i^{ft})N_h\lambda_i^{nc}$, which we approximate by $\lambda_i^{hc} = 0.989N_h\lambda_i^{nc}$. We also set $\mu_1 = \mu_1^s + \mu_1^r = 1$, $\mu_2 = \mu_2^s + \mu_2^r = 3$, $\beta_1^n = 1$, $\beta_2^n = 20$, $\beta_1^h = 10$ and $\beta_2^h = 200$.

In regards to the reinforcement learning algorithm, we use a constant learning rate $\alpha_m = 0.01$ but the exploration rate p_m is decayed to zero by using the following rule $p_m = p_0/(1 + u)$, where $u = m^2/(\gamma + m)$. We used $\gamma = 1.0 \cdot 10^{11}$ and the algorithm starts with an exploration rate $p_0 = 0.1$.

4.1 Single Service

When no predictive information is used in the single service scenario, the optimization is carried without considering the second component of the system state, i.e. the number of MT labeled as H, and the optimal policy results to be of the *guard channel* type [7].

The curves in Fig. 3 represent the quotient between the average expected cost rate of the optimal policy when no prediction is deployed and the optimal policy deploying prediction. As expected, using prediction induces a gain in all cases and that gain decreases as prediction uncertainty (U) increases. In Fig. 3(a) we varied the average number of handovers per session. In Fig. 3(b) we varied the weighting factor β which quantifies the priority of handover requests over new sessions: the higher the value of β the lower the blocking probability of handover sessions compared to the blocking probability of new sessions. It is observed that higher values of β lead to higher performance gains. The position of the decision threshold within the uncertainty zone is evaluated in Fig. 3(c), the curves indicate that a threshold in the middle of the uncertainty zone is the best choice. Finally, Fig. 3(d) shows the effect of the elapse time since an MT is classified as H until it is handed over into the target cell or it moves to another cell. Both, short and long prediction periods, have a negative effect on the performance gain.

In all the cases that we examined the optimal policy when prediction is deployed had a *dynamic guard channel* structure, in which the number of reserved channels increases with the number of MTs labeled as H. More formally, let $p(i, j)$ be the probability of accepting a new session when the system is at state (i, j), then

$$p(i,j) = \begin{cases} 1, & \text{if } i \le i_{th}(j) \\ 0, & \text{if } i > i_{th}(j) \end{cases} \quad \text{and} \quad i_{th}(j) \le i_{th}(j') \quad \text{if} \quad j > j'.$$

where $i_{th}(j)$ is the threshold for a given j.

4.2 Multiservice

When no predictive information is used in the multiservice scenario, the optimization is carried without considering the second component of the system state, i.e. the number of resources occupied in the neighborhood by sessions labeled as H.

Figure 4(a) displays the variation of the ratio g_{wp}^π/g_p^π with different values of the uncertainty U in multiservice scenarios. As a reference, the same figure also shows the variation of the ratio g_{wp}^π/g_p^π for single service scenarios. In the multiservice scenario, for each value of U we run 10 simulations with different seeds and we display the averages. As expected, using prediction induces a gain in all cases and that gain decreases as the prediction uncertainty (U) increases.

Finally it is worth noting that the main challenge in the design of efficient bandwidth reservation techniques for mobile cellular networks is to balance two conflicting requirements: reserving enough resources to achieve a low forced termination probability and keeping the resource utilization high by not blocking too many new setup requests. Figure 4(b), which shows the variation of the utilization gain, i.e. the ratio $utilization_{wp}/utilization_p$, for different values of U, justifies the efficiency of our optimization approach.

5 Conclusion

In this paper we analyzed the performance gain that can be obtained when handover prediction information is considered in order to optimize the admission control policy in a mobile cellular network. Predictive information is provided by a prediction agent that labels the active mobile terminals in the neighborhood of the cell
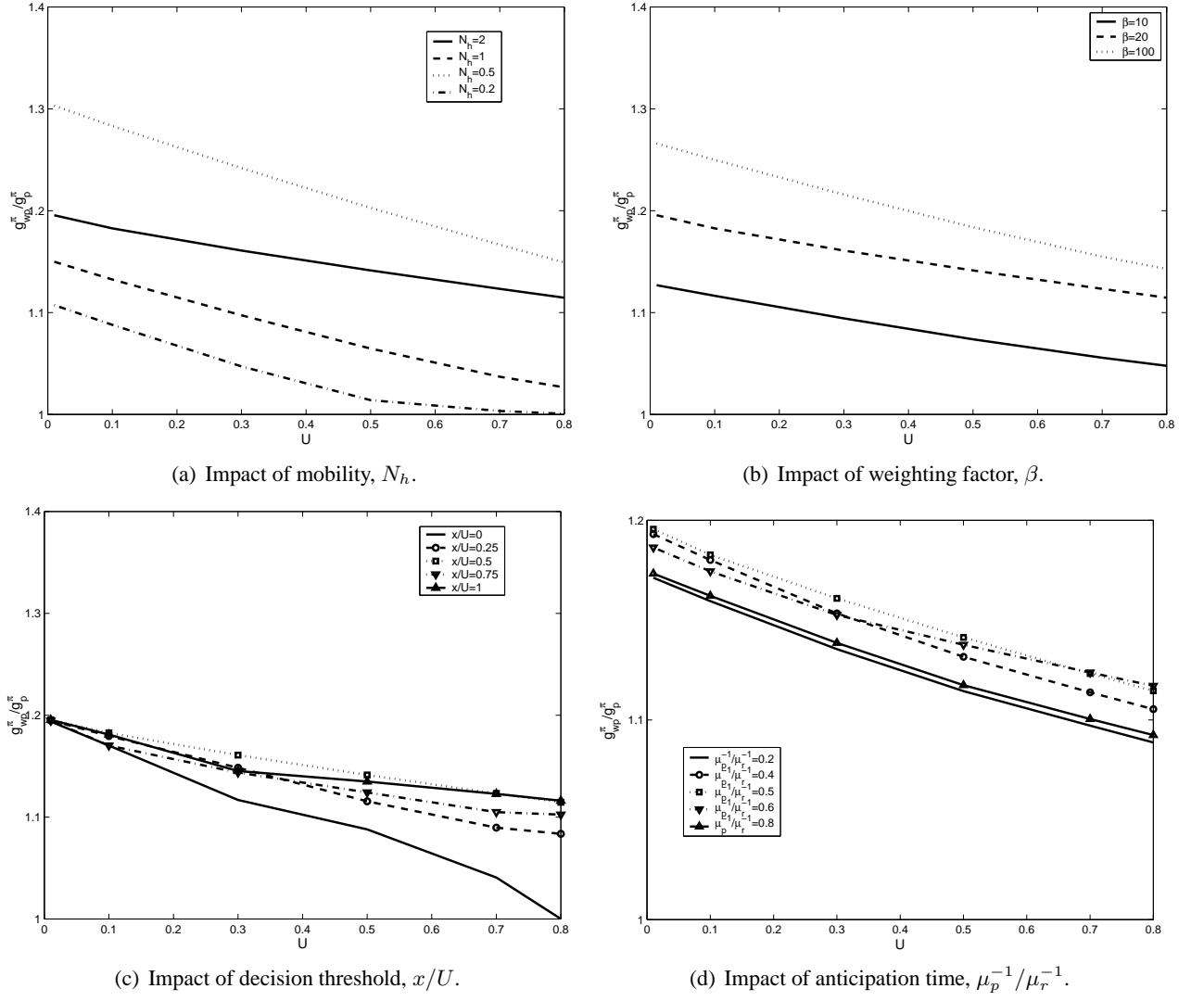
(a) Impact of mobility, N_h.

(b) Impact of weighting factor, β.

(c) Impact of decision threshold, x/U.

(d) Impact of anticipation time, μ_p^{-1}/μ_r^{-1}.

Figure 3: Performance comparison in the single service scenario.

(a) Performance gain when introducing handover prediction.

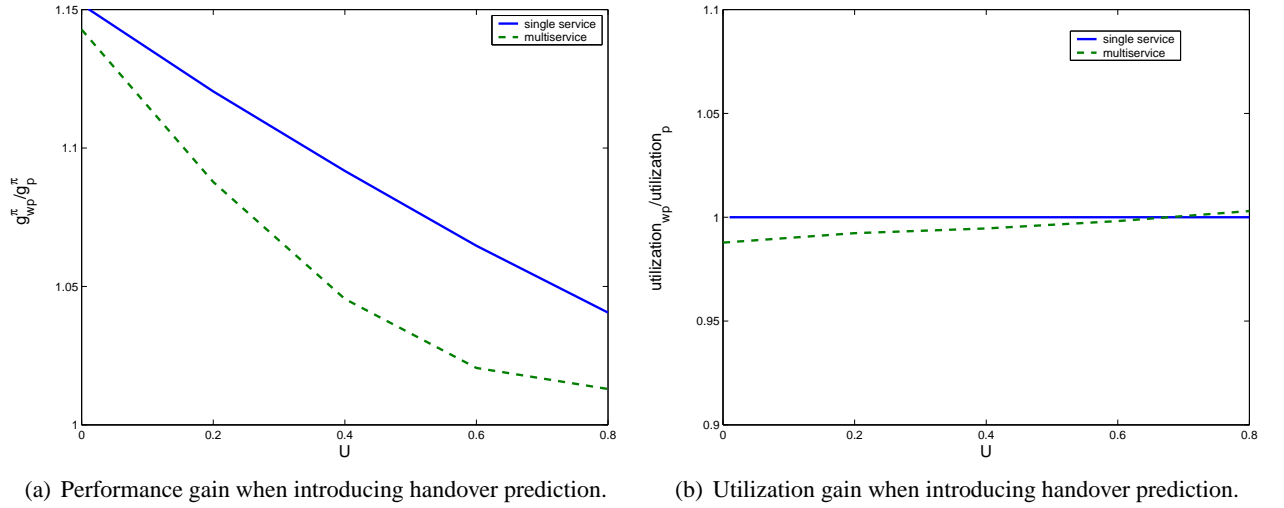(b) Utilization gain when introducing handover prediction.

Figure 4: Performance comparison in the multiservice scenario.

which will probably produce a handover into the cell. The policy optimization has been performed in a Markov or semi-Markov decision process framework and two optimization methods have been applied: policy iteration and a model free reinforcement learning methods. Our numerical results show that typical performance gains are around 10% although improvement ratios up to 30% have also been observed in some specific scenarios. In

future work we will consider a more sophisticated model of the prediction agent including, for instance, a more precise estimation of the time instant at which a handover will occur.

References

[1] W. Y. Lee, "Smaller cells for greater performance," *IEEE Communications Magazine*, vol. 29, no. 11, pp. 19–23, Nov. 1991.

[2] S. Choi and K. G. Shin, "Adaptive bandwidth reservation and admission control in QoS-sensitive cellular networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 9, pp. 882–897, Sep 2002.

[3] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Transactions on Vehicular Technology*, vol. VT-35, no. 3, pp. 77–92, Aug. 1986, see also: CEAS Technical Report No. 773, June 1, 1999, College of Engineering and Applied Sciences, State University of New York, Stony Brook, NY 11794, USA.

[4] F. Barceló, "Performance analysis of handoff resource allocation strategies through the state-dependent rejection scheme," *IEEE Transactions on Wireless*, vol. 3, no. 3, pp. 900–909, May 2004.

[5] D. J. Daley and L. D. Servi, "Loss probabilities of hand-in traffic under various protocols: II. model comparisons," *Performance Evaluation*, vol. 55, no. 3-4, pp. 231–249, Feb. 2004.

[6] V. Pla and V. Casares-Giner, "Analysis of priority channel assignment schemes in mobile cellular communication systems: a spectral theory approach," *Performance Evaluation*, vol. 59, no. 2-3, pp. 199–224, Feb 2004.

[7] R. Ramjee, R. Nagarajan, and D. Towsley, "On optimal call admission control in cellular networks," *Wireless Networks Journal (WINET)*, vol. 3, no. 1, pp. 29–41, 1997.

[8] N. Bartolini, "Handoff and optimal channel assignment in wireless networks," *Mobile Networks and Applications (MONET)*, vol. 6, no. 6, pp. 511–524, 2001.

[9] N. Bartolini and I. Chlamtac, "Call admission control in wireless multimedia networks," in *Proceedings of IEEE PIMRC*, 2002.

[10] H. Heredia-Ureta, F. A. Cruz-Pérez, and L. Ortigoza-Guerrero, "Capacity optimization in multiservice mobile wireless networks with multiple fractional channel reservation," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 6, pp. 1519 – 1539, Nov. 2003.

[11] V. Pla and V. Casares-Giner, "Optimal admission control policies in multiservice cellular networks," in *Proceedings of the International Network Optimization Conference (INOC)*, 2003, pp. 466–471.

[12] D. García, J. Martínez, and V. Pla, "Admission control policies in multiservice cellular networks: Optimum configuration and sensitivity," *Lecture Notes in Computer Science*, vol. 3427, pp. 121–135, Feb 2005.

[13] D. Levine, I. Akyildiz, and M. Naghshineh, "A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 1–12, Feb. 1997.

[14] J. Hou and Y. Fang, "Mobility-based call admission control schemes for wireless mobile networks," *Wireless Communications and Mobile Computing*, vol. 1, no. 3, pp. 269–282, 2001.

[15] F. Yu and V. Leung, "Mobility-based predictive call admission control and bandwidth reservation in wireless cellular networks," *Computer Networks*, vol. 38, no. 5, pp. 577–589, 2002.

[16] W.-S. Soh and H. S. Kim, "Dynamic bandwidth reservation in cellular networks using road topology based mobility prediction," in *Proceedings of IEEE INFOCOM*, 2004.

[17] C. Yener, A. Rose, "Genetic algorithms applied to cellular call admission: local policies," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 1, pp. 72–79, 1997.

[18] R. Zander and J. M. Karlsson, "Predictive and adaptive resource reservation (parr) for cellular networks," *International Journal of Wireless Information Networks*, vol. 11, no. 3, pp. 161–170, July 2004.

[19] Y. Zhao, "Standardization of mobile phone positioning for 3g systems," *IEEE Communications Magazine*, vol. 40, no. 7, pp. 108 –116, July 2002.

[20] F. Khan and D. Zeghlache, "Effect of cell residence time distribution on the performance of cellular mobile networks," in *Proceedings of VTC'97*.   IEEE, 1997, pp. 949 – 953.

[21] V. Pla and V. Casares-Giner, "Effect of the handoff area sojourn time distribution on the performance of cellular networks," in *Proceedings of IEEE MWCN*, Sept. 2002.

[22] P. V. Orlik and S. S. Rappaport, "On the handoff arrival process in cellular communications," *Wireless Networks Journal (WINET)*, vol. 7, no. 2, pp. 147–157, March/April 2001.

[23] S. M. Ross, *Applied probability models with optimization applications*.   Holden-Day, 1970.

[24] S. Mahadevan, "Average reward reinforcement learning: Foundations, algorithms, and empirical results," *Machine Learning, Special Issue on Reinforcement Learning (edited by Leslie Kaebling)*, vol. 22, no. 1–3, pp. 159–196, Jan./Feb./March 1996.

[25] M. L. Puterman, *Markov Decision Processes : Discrete Stochastic Dynamic Programming*.   John Wiley & Sons, 1994.

[26] R. Sutton and A. G. Barto, *Reinforcement Learning*.   Cambridge, Massachusetts: The MIT press, 1998.

[27] T. K. Das, A. Gosavi, S. Mahadevan, and N. Marchalleck, "Solving semi-markov decision problems using average reward reinforcement learning," *Management Science*, vol. 45, no. 4, pp. 560–574, 1999.

[28] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*.   Englewood Cliffs, NJ: Prentice Hall, 1989.