

Document downloaded from:

<http://hdl.handle.net/10251/56626>

This paper must be cited as:

Martínez Bauset, J.; Gimenez Guzman, JM.; Pla, V. (2012). Robustness of optimal channel reservation using handover prediction in multiservice wireless networks. *Wireless Networks*. 18(6):621-633. doi:10.1007/s11276-012-0423-6.



The final publication is available at

<http://dx.doi.org/10.1007/s11276-012-0423-6>

Copyright Springer Verlag (Germany)

Additional Information

Robustness of optimal channel reservation using handover prediction in multiservice wireless networks

Jorge Martinez-Bauset · Jose Manuel Gimenez-Guzman · Vicent Pla

2011

Abstract The aim of our study is to obtain theoretical limits for the gain that can be expected when using handover prediction and to determine the sensitivity of the system performance against different parameters. We apply an average-reward reinforcement learning (RL) approach based on afterstates to the design of optimal admission control policies in mobile multimedia cellular networks where predictive information related to the occurrence of future handovers is available.

We consider a type of predictor that labels active mobile terminals in the cell neighborhood a fixed amount of time before handovers are predicted to occur, which we call the anticipation time. The admission controller exploits this information to reserve resources efficiently. We show that there exists an optimum value for the anticipation time at which the highest performance gain is obtained. Although the optimum anticipation time depends on system parameters, we find that its value changes very little when the system parameters vary within a reasonable range. We also find that, in terms of system performance, deploying prediction is always advantageous when compared to a system without prediction, even when the system parameters are estimated with poor precision.

Keywords Channel reservation · Cellular network · Predictive information · Reinforcement learning

1 Introduction

Providing seamless mobility to users in heterogeneous mobile wireless networks is a key factor to guarantee certain degree of Quality of Service (QoS). As mobile terminals (MTs) roam across the network, unavoidably, the communications need to be handed over from one base station to another. Given that the current market trend is to integrate positioning systems in the MTs to provide location services, the mobile network operators can exploit this new functionality to predict the occurrence of handovers and improve the performance of the network. An integrated GPS-GLONASS-Galileo receiver will provide location services simultaneously from multiple satellites, substantially improving the accuracy, particularly in urban areas [1]. Two additional factors are creating increasing interest in handover prediction. One is the availability of databases that include layout information of roads and cities around the world. The other is the emergence of sophisticated algorithms that make use of layout and positioning information to estimate the movement of MTs with high accuracy [2].

MTs incorporate an increasing number of wireless technologies, including IEEE 802.11, GSM, UMTS, WiMAX, ..., so handover prediction is being studied for a wide number of wireless technologies [3–6]. The working scenario in this paper represents a mobile wireless network that generically covers any of the abovementioned technologies. We will consider a cell to be the area served by a base station or an access point, depending on the technology deployed.

To minimize the forced termination probability, conventional approaches deploy handover prioritization schemes [7] such as reserving a number of channels in each cell,

Jorge Martinez-Bauset
Dept. Comunicaciones, Universidad Politécnica de Valencia,
Cami de Vera s/n, 46022, Valencia, Spain.
E-mail: jmartinez@upvnet.upv.es

Jose Manuel Gimenez-Guzman
Dept. Automática, Universidad de Alcalá,
28871 Alcalá de Henares, Madrid, Spain.
Tel.: +34918856953
E-mail: josem.gimenez@uah.es

Vicent Pla
Dept. Comunicaciones, Universidad Politécnica de Valencia,
Cami de Vera s/n, 46022, Valencia, Spain.
E-mail: vpla@dcom.upv.es

named guard channels, only for arriving handovers. Given that the carried traffic diminishes as more guard channels are reserved, it is crucial to dimension its number appropriately. In this sense, schemes that deploy a dynamic number of guard channels, which depend on the momentary conditions of the network, are preferred to static ones. Many schemes proposed in the literature adjust the number of guard channels as a function of the predicted occurrence of handovers. See for example [2, 8–10] and references therein. Most of these studies propose a prediction system and a companion admission control (AC) scheme that makes use of the information provided by the former in a heuristic way. Besides, these studies assume that an admission policy based on guard channels is the best possible policy. In systems that do not have predictive information available, both heuristic and optimization approaches have been proposed to improve the performance of admission policies at the session level. Optimization approaches not using predictive information have been studied in [11–16].

We apply a novel optimization approach to the design of AC policies that exploits the availability of handover prediction information [17]. In particular, in this paper our study focuses on the robustness of the obtained policies to system parameter uncertainties. Our work is different from previous proposals in two distinctive aspects. One is that we study a multiservice scenario instead of a single service one. The other is that we apply a novel optimization approach, based on the formalism of Markov Decision Processes (MDPs), to search for the optimal policy instead of common heuristic approaches. As the large cardinality of the state space of the MDP that models our system makes it unfeasible to find numerical solutions, we use Reinforcement Learning (RL) as the solution method. Additionally, and more importantly, we provide the admission controller with much more specific information of the future occurrence of handovers, by considering the surface of a small neighborhood surrounding the perimeter of the cell under study and by classifying MTs with ongoing sessions in that surface according to whether they are expected to produce a handover or not. The policies we obtain show a significant performance improvement and establish theoretical limits for the gain that can be expected when handover prediction is used, which could not be established by deploying heuristic approaches.

RL [18, 19] is a simulation-based optimization technique in which an agent learns an optimal policy by interacting with an environment that rewards the agent for each executed action. In afterstates RL (ARL), which was suggested in [18], decisions are taken based on the resulting state after the action is performed rather than on the current state. Compared to conventional RL (CRL), ARL achieves better solutions and does it with higher precision. Additionally, ARL is better suited than CLR for multiservice scenarios as it produces a state space with lower cardinality [17]. ARL

has been previously used in the context of Q-learning [20] to achieve a compact representation of the value function, see for example [21]. More recently, it has also been used in model-based RL [22], in relational RL [23] and in learning general games [24].

We consider a type of prediction system that is able to label active MTs T time units before the handover is predicted to occur. We establish that an optimal T exists at which the maximum system performance is achieved. Note that a trade-off exists between the anticipation time with which the controller reserves the resources and the system performance. If the resources are reserved too early, more blocking will be experienced by new arrivals. On the other hand, if the resources are reserved too late, more forced terminations will be experienced by handover requests. Note also that, as the considered service disciplines are not preemptive, when the controller decides to reserve resources, there might not be enough free resources available in the system. One of the contributions of the paper is the study of the sensitivity of the optimal *anticipation time* T against different system parameters. We show that its optimal value is almost insensitive to system parameters like arrival and mobility rates, when their values vary within a reasonable range. This invariability is of crucial interest to network operators that use handover prediction, because once the optimum T is known an optimal policy can be computed in a short time using a conventional personal computer.

One way to deploy RL in real operating networks is to estimate system parameters like arrival rates, session durations and channel holding times periodically and then execute the optimization process to obtain the optimal policy. As estimations can incorporate errors, another contribution of the paper is the study of the robustness of the performance of optimal policies when the system parameter values provided to the optimization process are not exact. We show that deploying prediction is always advantageous in terms of system performance (with respect to the performance of a system not deploying it), even when poor estimators for the system parameters are used.

The rest of the paper is structured as follows. The next section defines the models of the system and of the prediction agent. Section 3 presents the mathematical framework of the optimization procedure. The evaluation of the sensitivity of the optimum anticipation time with different system parameters is presented in Section 4. In Section 5 we study the robustness of the performance of optimal policies when the system parameter values provided to the optimization process are not exact. Finally, a summary of the paper and some concluding remarks are given in Section 6.

2 Model Description and Prediction System

We consider a single cell system and its neighborhood, where the cell has a total of C resource units and the neighborhood C_p 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 service classes are supported by the system and we refer to them as services. It is usually accepted that it is more disturbing for a subscriber in a cellular network to have an ongoing session dropped than the blocking of a new session setup. Then, for each service, new and handover session arrivals are distinguished, so that there are N services and $2N$ arrival types. This distinction allows the admission controller to provide differentiated treatment to different arrival types.

For the sake of mathematical tractability, we make the common assumptions of Poisson arrival processes and exponentially distributed random variables for the cell residence time and session duration. However, we also study scenarios in which the cell residence time is modeled by different distributions. The exponential assumption is considered a good approximation for the inter-arrival time of new and handover requests, particularly when the performance indices of interest are blocking probabilities [25]. The arrival rate for new (handover) sessions of service i to the cell under study is λ_i^{nc} (λ_i^{hc}) and a request of service i consumes b_i resource units, $b_i \in \mathbb{N}$, when accepted. For a packet based air interface, b_i represents the effective number of resources required by the session [26,27]. As shown later, the admission control problem is formulated as a linear function of the number of active sessions of each service. In this function, each accepted session of each service class contributes with a constant amount, possibly different for different service categories. A new session is accepted if the value of the linear function (that describes the amount of some kind of resource consumed in the system at some time instant, like channels, power, etc.) is lower than a constant limit (that describes the total amount of the resource). See for example [28,29] for CDMA systems, [30] for OFDM systems and [31] for WLANs. For service i , the session duration and cell residence rates are μ_i^s and μ_i^r respectively. The resource holding time for a service i session in a cell is also exponentially distributed with rate $\mu_i = \mu_i^s + \mu_i^r$, and the mean number of handovers per session is $N_i^h = \mu_i^r / \mu_i^s$ when the number of resource units in the network is infinite. Without loss of generality, we consider that only one session is active per MT.

Given that the focus of our study was not the design of the prediction system, we used a model of it instead. A classifier labels an active MT in the vicinity of a cell as *probably producing a handover* (H) or the opposite (NH). We denote this system as classifier for incoming handovers (CIH). After some time the actual destiny of the MT becomes defini-

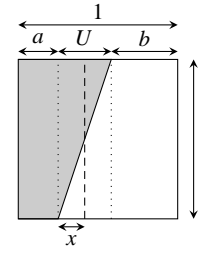


Fig. 1 Basic parameters of the classifier.

tive. Two outcomes are possible: either a handover into the cell occurs or not (for instance because the session ends or the MT moves to another cell). We envisage the CIH system to be located at the base station or access point and to receive positioning updates from the MTs, possibly at a higher rate as MTs approach the cell neighborhood.

A wide range of techniques have been proposed to perform handover predictions. They can be classified into five categories according to the information used for predictions [3]: i) RSS (Received Signal Strength) [32, 33, 6]; ii) movement extrapolation [34, 35]; iii) history data [8, 36, 37]; iv) mobility pattern [38]; and v) distance between MTs and BSs [39]. These schemes are based on the estimation of one or several of the next parameters: position, RSS, direction and velocity. It is also remarkable that to improve the predictions accuracy some studies also use city or road maps [2, 40], or consider multifaceted user behavior using also the group, time-of-day and duration characteristics of mobile users [5]. As it can be observed, most prediction methods are based on the estimation of parameters that are relatively easy to measure. The most important and widely used parameter is the position of MTs, which typically obtained using a GPS receiver [2].

In the next two subsections we study two important aspects of the CIH. First, its model, which allows us to determine the probabilities of false positives and non-detections, and second, the labeling instant. Note that the admission controller reserves resources for future arrivals based on the number and importance of the predicted handovers. However, the predictions made by the CIH are only visible to the controller once the MTs have been labeled.

2.1 Model of the CIH

The labeling process performed by the classifier produces classification errors of two types: false-positives and non-detections. The model of the classifier, shown in Fig. 1, has been designed to take into account these classification errors. The model can be depicted by a square with a surface equal to one (1×1), which represents the population of active MTs to be classified. The shaded area represents the fraction of MTs (S_H) that will ultimately move into the cell, while the white area represents the rest of active MTs. No-

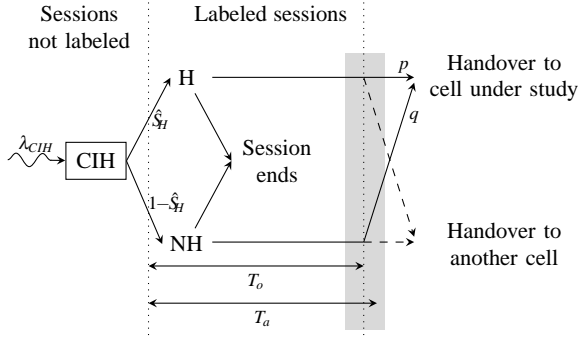


Fig. 2 Temporal evolution of sessions labeled by the CIH.

tice that part of the MTs that will move into the cell can finish their active sessions before doing so. The classifier sets a threshold (represented by a vertical dashed line) to discriminate between those MTs that will likely produce a handover and those that will not. The fraction of MTs falling on the left side of the threshold (\hat{S}_H) are labeled as H and those on the right side as NH. There exists an uncertainty zone, of width U , which accounts for classification errors: the white area on the left of the threshold (\hat{S}_H^e) and the shaded area on the right of the threshold (\hat{S}_{NH}^e). The parameter x represents the relative position of the classifier threshold within the uncertainty zone.

The model of the CIH can be alternatively characterized by two parameters: 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 $1 - p$ and q model the false-positive and non-detection probabilities respectively, and in general $q \neq 1 - p$. It can be shown that

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

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

2.2 CIH labeling instant

The time elapsed since an active MT is labeled until its destiny becomes definitive will be referred to as the labeling *anticipation time*. We assume that residence time of MTs in the cell neighborhood follow an exponential distribution. Consider a seven cell cluster with one cell at the center. It can be easily shown that if the residence time in a cell is exponentially distributed, then the residence time in the ring of cells surrounding the central cell is also exponentially distributed. Assuming hexagonal shaped cells, we denote by $f_g(t)$ the probability density function of the time an MT resides in the ring of cells surrounding a given cell, by $f_r(s)$ the probability density function of the time an MT resides in

a cell and by $f_g^*(s)$ and $f_r^*(s)$ their respective Laplace transforms. It is clear that $f_g^*(s) = \sum_{k=0}^{\infty} f_r^*(s)(2/6)^k(4/6)$. When the cell residence time $f_r(s)$ is exponentially distributed with rate μ^r then $f_r^*(s) = \mu^r/(\mu^r + s)$ and we obtain that $f_g^*(s) = (4\mu^r/6)/((4\mu^r/6) + s)$. Therefore $f_g(t)$ is also exponentially distributed with rate $(4\mu^r/6)$. In our model, the so-called neighborhood is only a narrow area surrounding the perimeter of the cell under study instead of a six-cell ring, but by analogy the residence time of an MT in the neighborhood is assumed to be exponentially distributed. However, we also study scenarios in which the cell neighborhood residence time is modeled by different distributions. Note that this model might account for the fact that actual cell boundaries are fuzzy and irregularly shaped.

We also assume that the CIH will always label an MT (session) at the required anticipation time. However, note that the prediction accuracy might decrease when the time elapsed since the MT arrives into the neighborhood and the labeling instant is too short. This might happen for short residence times. Although, as expected, the performance gain decreases with the accuracy of predictions, the gain is substantially above 1 even for high values of U (uncertainty zone of the predictor). This means that the performance of a system with prediction is substantially better than the performance of a system without prediction, even when deploying low precision predictors [17].

Given that our interest is to find theoretical limits for the gain that can be expected when using handover prediction, we use a type of CIH that provides the admission controller with more specific information about the future occurrence of handovers. More precisely, at any time instant, the admission controller is able to know the number and importance of the handovers that will take place in future time window of fixed size. A similar approach is used in [2], where the system predicts the incoming and outgoing handovers that will take place in a time window of fixed size. To achieve this, the labeling anticipation time of an active MT in the cell neighborhood is $\hat{t} = \min\{T, t^r\}$, where T is a constant and t^r is the predicted residence time in the neighborhood. In practical scenarios we expect that $T < E[t^r]$, and therefore for most MTs, $\hat{t} = T$. For that reason we refer to T as the anticipation time.

As it will be shown in Section 4, there exists an optimum value for T that maximizes the system performance, which in addition is almost insensitive to small perturbations of common system parameters, like load, mobility, etc. We also study the impact that deviations from the ideal constant elapsed time since labeling to handover execution have on the system performance. That is, even if we configure the anticipation to be a constant time T , the time elapsed since the CIH labels an MT until its destiny becomes definitive will be a random variable. See Fig. 2, where T_o and T_a are the optimum value for T and the actual anticipation time,

and the shaded surface represents an interval for the likely values of T_a . More details are provided in Section 4.

Finally, and to summarize, the basic operation of the proposed system is: i) for each active session in the cell neighborhood, the CIH is in charge of estimating if the session (MT) will generate a handover requests or not, and the time instant of its occurrence; ii) at some time instant (the labeling instant) the CIH communicates to the admission controller the future occurrence of a handover. The labeling instant is the minimum of the residence time in the neighborhood and the anticipation time T ; iii) the admission controller, learning from its past experience, takes the optimal actions with respect to the number of resources to reserve and when the reservation takes place.

2.3 Relations between system parameters

Instead of defining the arrival rate of new sessions to the cell neighborhood (ng) and the handover rate to the cell neighborhood from the outside of the system (so) as independent parameters of the model, it is more realistic to relate them to other system parameters. To do so, we assume a circular-shaped cell of radius r and a holed-disk-shaped neighborhood with inner (outer) radius $1.0r$ ($1.5r$). Note that these values should be understood as idealized average values, because, as mentioned before, cell boundaries are fuzzy and irregularly shaped.

Then, 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 rates to the cell neighborhood from the outside of the system (so) and from the cell (hc) is made equal to the ratio of their perimeters, $\lambda_i^{so} = 1.5\lambda_i^{hc}$.

The classifier operates on sessions arriving from the outside to the neighborhood as well as on new sessions initiated in the neighborhood. Therefore, in Fig. 2 we have that $\lambda_{CIH} = \sum_{i=0}^N (\lambda_i^{so} + \lambda_i^{ng})$. Note that in our numerical experiments the arrival rates are chosen to achieve realistic operating values for P_i^n ($\approx 10^{-2}$) and P_i^{ft} ($\approx 10^{-3}$), where P_i^n is the blocking probability of new requests and P_i^{ft} is the forced termination probability, both of service i .

3 Optimization by Reinforcement Learning

The information provided by the CIH and the state of the cell (number of occupied resources) are used to find the optimal admission policy. We formulate the optimization problem as an infinite-horizon finite-state *Semi-Markov Decision Process* (SMDP) under the average cost criterion. We focus on the average cost criterion instead of using a discounted approach as it is more appropriate for this type of continuous-time systems with long-term objectives [41]. SMDPs are a

special kind of MDPs appropriate for modeling continuous-time systems in which the time between decision epochs is not constant.

We consider deterministic stationary Markovian policies, $\pi : S \rightarrow A$, which define the next action of the agent based only on the current state x , i.e. an agent adopting this policy performs action $\pi(x)$ in state x . For the problems we consider, optimal stationary Markovian policies always exist.

When the system starts at state x and follows policy π , the average expected cost rate, denoted by $\gamma^\pi(x)$, is defined as

$$\gamma^\pi(x) = \lim_{t \rightarrow \infty} \frac{1}{t} E [K^\pi(x, t)]$$

where $K^\pi(x, t)$ is the total cost accumulated in the interval $[0, t]$ when the system starts in state x . Clearly, if the environment is stochastic then $K^\pi(x, t)$ is a random variable.

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 [42]. We call it the *cost rate* of the policy π , denote it by γ^π and consider the problem of finding the policy π^* that minimizes γ^π , which we name the optimal policy.

At decision epochs an action is selected from the set of possible actions $A := \{0 = \text{reject}, 1 = \text{admit}\}$. Given that no actions are taken at session departures, the time instants at which these occur are not considered decision epochs. Furthermore one of the $2N$ arrival types is selected as the highest priority one, being its requests always admitted while free resources are available. Therefore, only one action is possible at those decision epochs corresponding to arrivals of the highest priority stream.

The cost function is defined as follows. At any decision epoch, the cost incurred by accepting any type of arrival is zero and by rejecting a new (handover) request of service i is ω_i^n (ω_i^h). A further accrual of cost between two decision epochs occurs when the system has to reject requests of the highest priority. It can be shown that for the defined cost function, the cost rate represents a weighted sum of the loss rates

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

where P_i^h is the blocking probability of handover requests. In general, $\omega_i^n < \omega_i^h$ since the loss of a handover request is less desirable than the loss of a new session setup request.

3.1 The Afterstates Approach

Intuitively, ARL is based on the idea that what is relevant in a decision process like ours is the state reached immediately after the action is taken. More specifically, all states

at decision epochs in which the immediate actions drive the system to the same afterstate, would accumulate the same future cost if the same future actions are taken.

As we learn over the state reached immediately after the action is taken, the arrival type is not needed in the learning process. Therefore, in the ARL approach the cardinality of the state space is independent of the number of services involved. This characteristic is specially important in systems with a high number of services where ARL tackles more efficiently than CRL the curse of dimensionality. Besides, as any RL optimization method, ARL offers the important advantage of being a model-free method, i.e. transition probabilities and average costs are not needed in advance.

To solve the optimization problem we deploy a modified version of the SMART algorithm [43] that follows an ARL approach using a temporal difference method (TD(0)). For the pseudo code of the proposed algorithm, please refer to Fig. 3. Note that at the m^{th} decision epoch an exploratory action is taken with probability p_m , which is decayed to zero by using the following rule $p_m = p_0/(1 + u_m)$, where $u_m = m^2/(\phi + m)$, $p_0 = 0.3$ and $\phi = 10^{13}$ [44]. The exploration of the state space is a common RL technique used to avoid being trapped at local minima. Finally, due to the simulation-based nature of RL, each point in the figures presented along the paper represents the average of 10 different simulation runs initialized with different seeds.

3.2 Representation of the state of the system

In [17] we explored different representations of the state of the cell and the neighborhood and found that the weighted representation defined in (1) was particularly convenient, as it provides a good compromise between cardinality and precision of the results. Clearly, the larger the state space, the larger the exploration required and the higher the complexity of the learning procedure. When deploying the weighted representation, the state space is defined as:

$$S := \{x = (x_T, x_\omega^{ng}) : x_T \leq C; x_\omega^{ng} \leq \kappa C_p\} \quad (1)$$

where we denote by x_T the total number of resource units occupied in the cell and by x_ω^{ng} the weighted number of resource units occupied by sessions labeled as H in the neighborhood (i.e. the total weighted number of resources units required by the forecasted handover sessions). We define $\kappa = \omega_H^h/\omega_L^h$, being H (L) the highest (lowest) priority service. For example, for one of the systems we study the number of services is $N = 2$, having the handover arrivals of the second service the highest priority, i.e. its requests are always admitted while free resources are available. The weights associated to the blocking of new and handover requests are $\omega_1^h = 1$, $\omega_2^h = 2$, $\omega_1^l = 20$ and $\omega_2^l = 40$. Then, $\kappa = \omega_2^h/\omega_1^h = 2$ and $x_\omega^{ng} = x_1^{ng} + \kappa x_2^{ng}$, where x_i^{ng} is the number of resource

```

1: Initialize  $h(x), \forall x \in S$ , arbitrarily (usually zeros).
2: Initialize  $\gamma$  arbitrarily (usually zeros)
3: Initialize  $N(x) = 0, W_T = 0$  and  $T_T = 0$ 
4: Repeat forever:
    We denote by  $a$  the action taken in the current state  $y$ , by
     $y'_{reject}$  ( $y'_{accept}$ ) the afterstate when the reject (accept) action is
    taken and by  $\omega_{reject}$  the immediate cost when the request is re-
    jected.
5:   Take action  $a$ :
6:     Exploration: random action
7:     Greedy: action selected from
           if  $(\omega_{reject} + h(y'_{reject})) < h(y'_{accept})$  then
              $a = reject$ 
           else
              $a = accept$ 
8:      $\alpha = 1/(1 + N(x'))$ 
           being  $\alpha$  the learning rate,  $x'$  the previous afterstate and  $N(x')$ 
           the number of times the afterstate  $x'$  has been updated:
9:      $h(x') \leftarrow (1 - \alpha)h(x') + \alpha[w_c(x', y) + w(y, a) + h(y') - \gamma\tau]$ 
            $N(x') \leftarrow N(x') + 1$ 
           being  $w_c(x', y)$  the accrued cost when the system evolves from
            $x'$  to  $y$ ,  $w(y, a)$  the immediate cost of taking action  $a$  in state  $y$  and
            $\tau$  the time elapsed between decision epochs  $m$  and  $m + 1$ .
10:    if  $a$  is greedy:
11:       $W_T \leftarrow W_T + w_c(x', y) + w(y, a)$ 
12:       $T_T \leftarrow T_T + \tau$ 
13:       $\gamma \leftarrow W_T/T_T$ 
14:       $x' \leftarrow y'$ 

```

Fig. 3 SMART algorithm with afterstates.

units occupied by sessions of service i labeled as H. The rationale behind the weighted prediction is to provide information to the optimization process about the *importance* of the forecasted handovers. Recall that the predictive information described by x_ω^{ng} refers only to those ongoing sessions that will probably be handed over in less than T time units.

Note that the admission controller does not know the exact time instant at which individual handovers will occur, nor has an explicit knowledge of the exact duration of the time window over which those handovers will occur. The admission controller is not informed by the CIH about the value of anticipation time T since, owing to its reduced complexity, it would not know how to use that information. Instead, the admission controller learns by experience and adapts its response to the temporal immediacy of the information received from the CIH, which is expressed as the weighted number of handovers that will occur in the near future.

In [17] we compared the performance of our predictive admission control scheme to the performance of the scheme proposed in [8]. In addition, for comparative purposes, we also determined the performance of an optimal fractional guard channel policy (that does not exploit predictive information). There, we showed that our scheme clearly outper-

formed the other two. As discussed above, many schemes proposed in the literature adjust the number of guard channels as a function of the predicted occurrence of handovers. Then it is implicitly assumed that an admission policy based on guard channels is the best possible policy. In our comparative study we also showed that, in general, this is not true.

3.3 Reference scenarios

For the numerical evaluations in this paper we use two different reference scenarios, one with $N = 2$ and the other with $N = 4$. In both scenarios, $C = 50$ and $C_p = 100$ resource units. Note that the value of C_p is chosen big enough to guarantee that a session is never blocked when handed over to the neighborhood. For the classifier we set $S_H = 0.4$, $x = U/2$ and $U = 0.2$. With these values, the probability of false-positive is $1 - p = 6.25 \cdot 10^{-2}$ and the probability of non-detection is $q = 4.166 \cdot 10^{-2}$.

For the scenario with $N = 2$, $b_1 = 1$ and $b_2 = 2$. We also set $\mu_1 = 1$, $\mu_2 = 3$, $N_i^h = 1$ and $\mu_i^r/\mu_i^p = 0.5$, $i = 1, 2$, where μ_i^p is the residence rate in the cell neighborhood for a service i session. The arrival rates of new sessions to the cell are: $\lambda_1^{nc} = f_1 \lambda_T$, $\lambda_2^{nc} = f_2 \lambda_T$, and $\lambda_T = \sum \lambda_i^{nc} = 20$. We call f_i the penetration factor of service i , and clearly $\sum_{i=1}^N f_i = 1$. We set $f_1 = 0.8$ and $f_2 = 0.2$. The relative weights associated to the blocking of new and handover requests were defined in the previous subsection.

For the scenario with $N = 4$, $b_i = \{1, 2, 4, 6\}$, $N_i^h = 1$, $\mu_i = 1$ and $\mu_i^r/\mu_i^p = 0.5$. The different weights associated to the blocking of new sessions and handovers are $\omega_i^n = \{1, 2, 4, 8\}$ and $\omega_i^h = \{20, 40, 80, 160\}$, and, therefore, $x_\omega^{ng} = x_1^{ng} + 2x_2^{ng} + 4x_3^{ng} + 8x_4^{ng}$. Given that in multiservice wireless networks the bandwidth required by different sessions is quite different, the rate charged per minute is also quite different, and this in turn makes the arrival patterns quite different. To achieve this arrival rate differentiation, we set $\lambda_i^{nc} = f_i \lambda_T$, being $f_i = \varphi_i / \sum \varphi_j$, $\varphi_i = \chi^{i-1}$ and $\chi = 0.2$, i.e. for a given service, its arrival rate of new sessions is set to 20% of the arrival rate of the service with the next lower index. The value of λ_T has been chosen to achieve that the load per resource unit offered by the new sessions ρ is the same as the one offered in the previous scenario where $N = 2$:

$$\rho = \frac{1}{C} \sum_{i=1}^N \frac{\lambda_i^{nc}}{\mu_i} b_i = \frac{\lambda_T}{C} \sum_{i=1}^N \frac{f_i}{\mu_i} b_i.$$

For the sake of clarity and unless otherwise specified, the numerical examples will be evaluated for the simpler scenario with $N = 2$ services. However, we also include numerical examples for the scenario with $N = 4$ services to show that the main contributions of the paper are also valid for

scenarios with higher number of services. Following common practice, the units of the rate parameters are not specified. As an example, if the rates are expressed in events per minute, then the units of time parameters like the anticipation time T are expressed in minutes. Also, it is common practice to assume that all rates of the Markov model are expressed in the same units.

4 Sensitivity of the optimum anticipation time against system parameters

In this section we study the sensitivity of the optimum value of T , at which the maximum performance gain is obtained, and the sensitivity of the maximum performance gain against different system parameters like: number of resource units of the cell (C), system load (λ_T), distribution of the residence time and mobility of terminals. We measure the performance gain by the ratio $\gamma_{wp}^\pi / \gamma_p^\pi$, where γ_p^π (γ_{wp}^π) is the cost rate of the optimal policy in a system with (without) prediction. Note that an optimal policy in a system without prediction is obtained when the optimization process is provided with no information about the future occurrence of handovers.

The study is motivated by the fact that when deploying the CIH the operator must configure it with the optimum anticipation time to maximize system performance. As it will be shown, the optimum value of T depends on the system parameters but we find that its value changes very little when the system parameters vary within a reasonable range. This invariability is of crucial interest to network operators that use handover prediction systems, because the optimum T can be precomputed for a given set of network conditions. Once the optimum T is known, an optimal policy can be computed quite quickly, even using a large exploration, as only the optimal policy at the optimum value of T , which maximizes the system performance gain, needs to be obtained. For example, when an Intel Pentium IV HT 3GHz personal computer and a large exploration characterized by $\phi = 10^{13}$ is used, a good policy is computed in less than 2 minutes.

Therefore, deploying RL in real operating networks is possible either using historical information or estimating the system parameters (like arrival and residence rates) periodically and feeding them to the simulation program. With this second type of operation, as it is not expected that new system parameter estimates differ substantially from the previous ones, the time required to compute a new optimal policy from a close one is much shorter than computing it from an arbitrary initialization.

We follow an experimental approach based on defining two test scenarios with realistic operating conditions and comparing the cost rate of optimal policies obtained in sys-

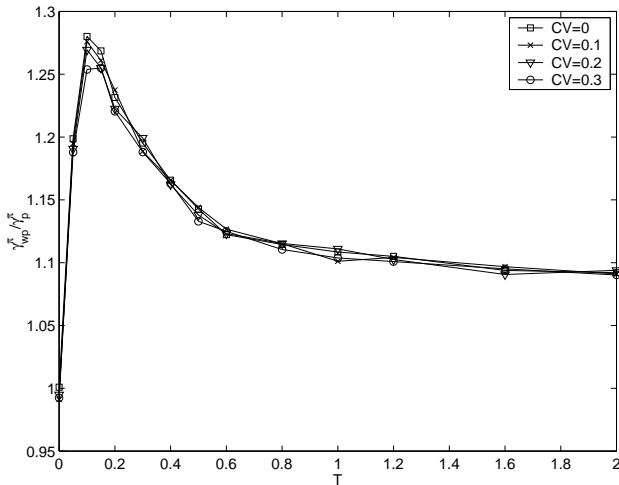


Fig. 4 Impact of the accuracy of the predicted anticipation time on the performance gain.

tems with different characteristics. The reference scenarios were defined in Section 3.3.

When describing the CIH in Section 2.2 it was discussed that even if we configure the anticipation to be T time units, the time elapsed since the CIH labels an MT until the handover actually occurs will be different from T and we model it as a random variable. We are interested in evaluating the impact that the deviation of the real anticipation time from an ideal constant one has on the system performance. For that, we assume an ideal scenario where the CIH achieves a 100% accuracy when labeling an MT and determine optimal policies for different values of T . Then, we evaluate the performance gain of these policies in a real scenario where the time elapsed since the CIH labels an MT until the handover actually occurs is modeled by a normal distribution of mean T and standard deviation $T \cdot CV$, being CV the coefficient of variation¹. Note that for a normal distribution $N(\mu, \sigma^2)$ the fraction of samples in the interval $[\mu - \sigma, \mu + \sigma]$ is approximately 0.68.

As observed in Fig. 4, an optimum value for the anticipation time T exists at which the performance gain attains its maximum. Note also that even when the actual anticipation time is not constant, the optimum value of T remains approximately the same. However, as it could be expected, the maximum gain diminishes as the CV (error in the labeling instant) increases. For the numerical results in this section we are deploying the reference scenario with $N = 2$ services in which we consider a certain degree of classification errors by configuring $U = 0.2$. As observed in Fig. 4, even for a relative error up to 20% ($CV = 0.2$) in the labeling instant, the performance of the obtained policy is not significantly worse than the one obtained for an exact CIH ($CV = 0$). For

¹ The coefficient of variation of random variable X is the ratio of its standard deviation to its mean, $CV_X = \sigma_X / E[X]$.

that reason, in the rest of the paper we assume an exact anticipation time.

To determine the performance gain (cost) of the policies obtained by our optimization approach in all the results presented, we first solve the MDP using RL and obtain the optimal policy for each system configuration. Then, the performance gain of the policy is determined using a discrete-event simulation model. That is, for each scenario of study, we first determine the optimal policy and then, we measure by simulation the blocking probabilities of new and handover requests perceived by the different service classes when the obtained policy is enforced.

4.1 Sensitivity against the number of resource units of the cell

Maintaining the load per resource unit constant, Fig. 5 shows the performance gain $\gamma_{wp}^\pi / \gamma_p^\pi$ as a function of T for three different values of the number of resource units of the cell, $C = \{30, 40, 50\}$, when $N = 2$. As observed, the optimum value of T decreases slightly as the number of resources increases. This might be due to the fact that in a system with more resource units, resources are released at higher rate and, consequently, the admission controller is able to reserve resources with less anticipation time. Also note that the maximum gain increases as the number of resource units of the cell increases. This is explained by the trade-off between the anticipation time with which the admission controller reserves the resources and the system performance, which was described in Section 1. The shorter the time required to enforce reservations, the shorter the time elapsed since the resources are reserved until they are needed and, then, the better the system performance. This trade-off also explains why providing the admission controller with the information of future handover arrivals too early (large T) leads to poor performance. Although the optimum value of T changes slightly with C , this parameter is only expected to vary in the design phase of the network.

4.2 Sensitivity against the system load

Maintaining the rest of the parameters constant, Fig. 6 shows the performance gain $\gamma_{wp}^\pi / \gamma_p^\pi$ as a function of T for a system with $N = 2$ services for three different system loads $\lambda_T = \{18, 20, 22\}$. The different loads have been selected to produce realistic blocking probabilities in real operating networks. As observed, the optimum value of T is almost insensitive to system load and the maximum $\gamma_{wp}^\pi / \gamma_p^\pi$ decreases as load increases. Note that for the study, the load region of interest has been chosen high enough to obtain an advantage by deploying prediction, but at the same time complying with the QoS (blocking) objectives (see Subsection 2.3).

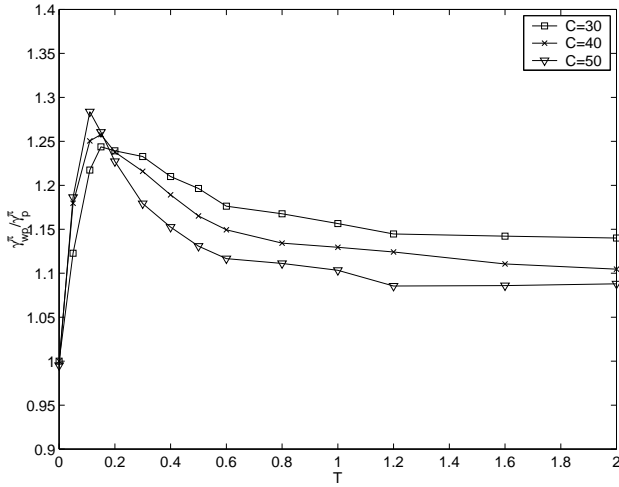


Fig. 5 Sensitivity of the optimum T and the maximum performance gain against the number of resources units of the cell.

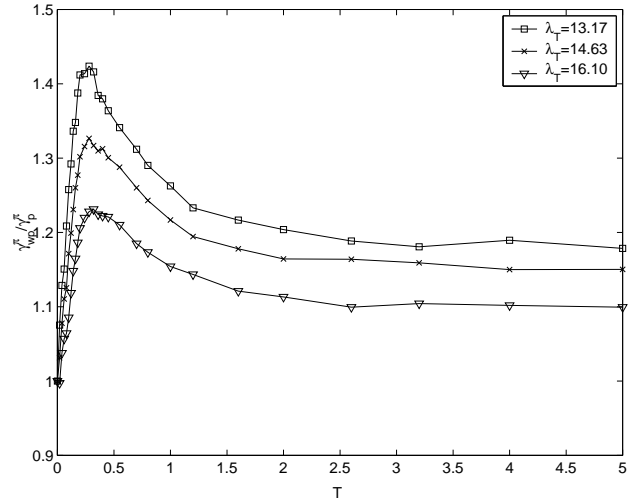


Fig. 7 Sensitivity of the optimum T and the maximum performance gain against the system load in a system with $N = 4$ services.

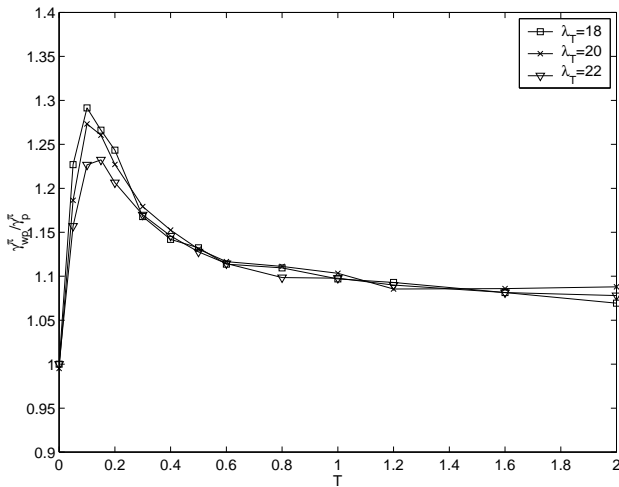


Fig. 6 Sensitivity of the optimum T and the maximum performance gain against the system load in a system with $N = 2$ services.

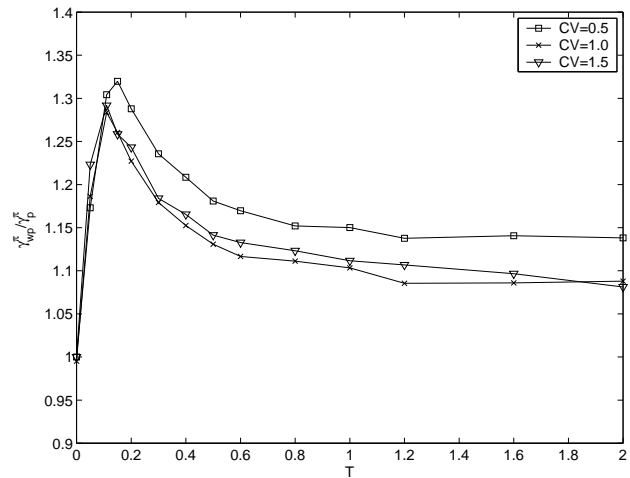


Fig. 8 Sensitivity of the optimum T and the maximum performance gain against the CV of the distributions of the residence time in the cell and in the neighborhood.

Figure 7 shows the performance gain as a function of T for a system with $N = 4$ services. Here, with $\lambda_T = 14.6332$ we get the same load per resource unit as the one obtained in a system with $N = 2$ services and $\lambda_T = 20$. Note that the optimum value of T is almost insensitive to the system load as well.

4.3 Sensitivity against the distribution of the residence time in the cell and the neighborhood

In addition to the exponential distribution, for which we have a $CV = 1$, we also consider the hyperexponential and Erlang distributions for the residence time. The CV of the Erlang distribution can be adjusted in the range $0 < CV < 1$, while the CV of the hyperexponential distribution can be set to $CV > 1$. Using these distributions, we evaluate the variation

of the performance gain $\gamma_{wp}^{\pi}/\gamma_p^{\pi}$ with T for three different values of the CV of the distribution of the residence time, $CV = \{0.5, 1.0, 1.5\}$, when $N = 2$.

Figure 8 shows the impact on performance of the CV of the residence time in the cell and in the neighborhood. Note that the mean residence times, i.e. $1/\mu_i^c$ and $1/\mu_i^p$, are kept constant and equal to the values deployed in the rest of experiments. Clearly, the optimum value for T remains approximately constant. We have also studied the scenarios in which the residence time distribution in the cell or in the neighborhood, but not both, is different from an exponential distribution, reaching the same conclusion. Note that when the resource holding time is not exponentially distributed, the policies obtained might not be optimal, as the underlying model is not an MDP and the corresponding theoretical framework cannot be applied.

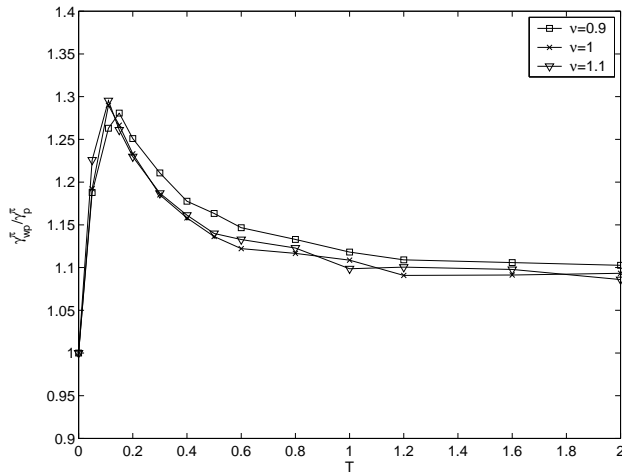


Fig. 9 Sensitivity of the optimum T and the maximum performance gain against the mobility of terminals.

4.4 Sensitivity against the mobility of terminals

In this section we study the evolution of the performance gain $\gamma_{wp}^\pi / \gamma_p^\pi$ as a function of T for different values of the mobility when $N = 2$. We express the mobility of terminals as $\mu_i = v\mu_i^*$, where μ_i^* is the channel holding time rate of service i in the reference scenario, v is a multiplicative factor and we keep constant the mean session duration $1/\mu_i^s$. Although the arrival rate of sessions from outside of the system (λ_i^{so}) depends on the mobility of MTs, we keep λ_i^{so} constant and equal to the value defined for the reference scenario, i.e. with $v = 1$. Results are depicted in Fig. 9. Again, the optimum value of T remains almost invariant with the mobility of MTs. Note that for the values of v used in Fig. 9, the cell residence rate μ_i^l changes by $\pm 20\%$ or more, with respect to its nominal value.

5 Robustness of the obtained policies

As described in the previous section, one way to deploy RL in operating networks is to estimate the system parameters periodically and then run a simulation program to obtain a good policy. In this section we explore the robustness of the obtained policies when the estimation of the system parameters is done with low precision. As in previous sections, we deploy the reference scenario with $N = 2$ services when obtaining numerical results. We assume that the operator has already determined the optimum anticipation time ($T = 0.1$) for the system under study. We also assume that the parameters that characterize the classifier ($S_H = 0.4$, $x = U/2$ and $U = 0.2$) are known and that the classifier estimates the anticipation time exactly.

In a first study we want to determine if it is still possible to obtain a performance gain in a system with prediction,

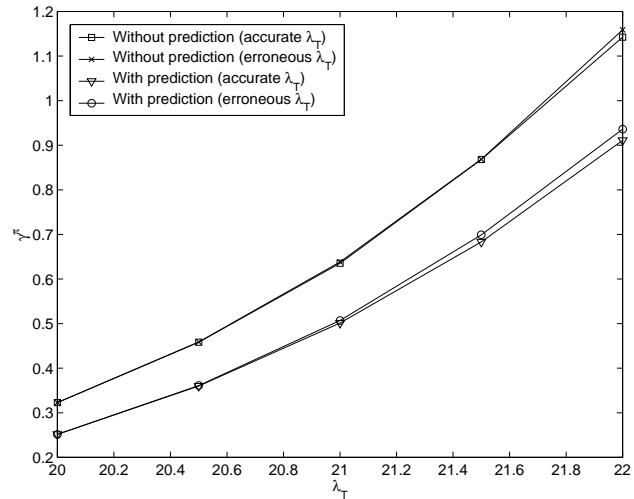


Fig. 10 Robustness of optimal policies when the arrival rate λ_T is estimated with poor precision when $N = 2$ services.

when compared to another without prediction, when the estimation of the aggregated arrival rate λ_T is done with poor precision. We compare the performance of optimal policies obtained in a system with prediction and other without prediction. The performance of an optimal policy is determined by computing its cost rate γ^π . For each system we show two different curves. One is obtained by determining the optimal policies for exact estimations of λ_T . The other is obtained by determining the optimal policy for $\lambda_T = 20$ and using this policy for other values of $\lambda_T \in [20.0, 22.0]$, i.e. we are considering the worst case where the value of λ_T is being underestimated. The interval for λ_T has been chosen to achieve realistic operating values for P_1^n and P_1^{ft} . As an example, we provide values for $(P_1^n, P_1^{ft}, P_2^n, P_2^{ft})$, achieved by optimal policies at $\lambda_T = 20.0$ and $\lambda_T = 22.0$: (0.00530, 0.02007, 0.00016, 0.00043), (0.00801, 0.02841, 0.00058, 0.00142).

As observed in Fig. 10, in the system with prediction, even underestimating λ_T by 10%, the performance of the obtained policy is considerably better than the performance of optimal policies obtained by an exact estimation of λ_T in a system without prediction. An interesting additional finding is that the optimal policies obtained by the ARL approach are quite robust, i.e. the performance obtained with exact and inexact parameters differ very little.

Figure 11 compares the performance of optimal policies obtained in a system with prediction and another without prediction when $N = 4$ services. Clearly, the robustness is as good as the one obtained in the scenario with $N = 2$ services.

Figure 12 shows the result of a second study related to the estimation of the mobility of terminals. As in Subsection 4.4, it is expressed by $\mu_i = v\mu_i^*$, where μ_i^* is the channel holding time rate of service i in the reference scenario, v is a multiplicative factor and we keep constant the mean ses-

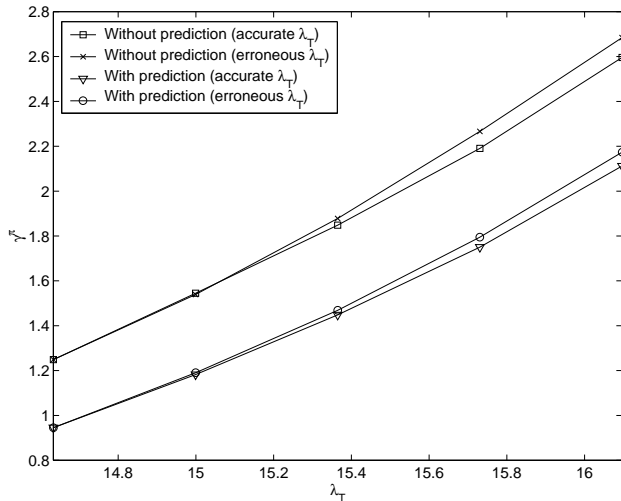


Fig. 11 Robustness of optimal policies when the arrival rate λ_T is estimated with poor precision when $N = 4$ services.

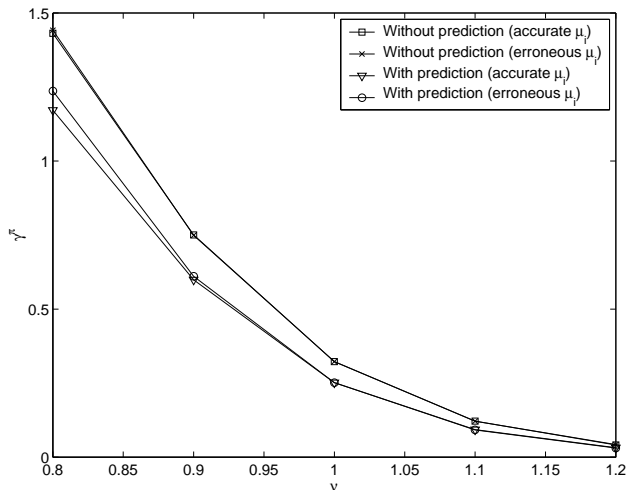


Fig. 12 Robustness of optimal policies when the mobility is estimated with error.

sion duration $1/\mu_i^s$. For each system we show two different curves. One is obtained by determining the optimal policies for exact estimations of the mean channel holding time $1/\mu_i$. The other is obtained by determining the optimal policy for $v = 1$ and using this policy for other values of $v \in [0.8, 1.2]$, i.e. we are considering in this case both underestimations and overestimations of the channel holding time value. As it can be observed, the performance of policies obtained in a system with prediction and with a poor estimation of the mobility are better than the performance of optimal policies obtained in a system without prediction and an exact estimation of the mobility. Note that for the values of v used in Fig. 12, the error in the estimation of the cell residence rate μ_i^r is as big as $\pm 40\%$ or more, with respect to its correct value.

6 Conclusions

In this paper we have studied the robustness of an admission controller that exploits the availability of predictive information. A system that we denoted as Classifier for Incoming Handovers is able to estimate the future trajectory of an MT and, therefore, to estimate the time instant at which an MT with an ongoing session is expected to produce a handover. The Classifier for Incoming Handovers labels active MTs in the neighborhood T time units before handovers actually take place. The admission controller makes use of the information provided by the classifier to optimize its admission decisions. We showed that there exists an optimum value for the anticipation time T at which the highest performance gain is obtained.

The problem of determining optimal admission policies that make use of the predictive information has been formulated as a semi-Markov decision process, using a novel reinforcement learning algorithm based on the concept of afterstates as solving methodology.

We evaluated the sensitivity of the optimum value of the anticipation time T against different system parameters. We concluded that the optimum value of T depends on system parameters but we find that its value changes very little when the system parameters vary within a reasonable range. This feature can be exploited by the network operator to determine optimal policies with less computational cost.

Finally, we studied the robustness of optimal policies to errors in the estimation of system parameters like arrival and cell residence time (mobility) rates. We showed that in a system deploying prediction, the performance of optimal policies obtained when the system parameters are estimated with low accuracy are considerably better than the performance of optimal policies obtained with exact estimations in a system without prediction, even when estimation errors are high. This robustness is important for network operators as it guarantees that deploying prediction is always advantageous in terms of system performance, even when the system parameters are estimated with poor precision.

Acknowledgements The authors would like to thank the reviewers for their valuable comments that helped to improve the quality of the paper. This work has been supported by the Spanish Ministry of Education and Science and European Commission (30% PGE, 70% FEDER) under projects TIN2008-06739-C04-02 and TIN2010-21378-C02-02 and by Comunidad de Madrid through project S-2009/TIC-1468.

References

1. Ji, S., Chen, W., Ding, X., Chen, Y., Zhao, C., & Hu, C. (2010). Potential benefits of GPS/GLONASS/GALILEO integration in an urban canyon—Hong Kong. *Journal of Navigation*, 63(4), 681–693.
2. Soh, W., & Kim, H. (2006). A predictive bandwidth reservation scheme using mobile positioning and road topology information. *IEEE/ACM Transactions on Networking*, 14(5), 1078–1091.

3. Kwon, H., Yang, M., Park, A., & Venkatesan, S. (2008). Handover prediction strategy for 3G-WLAN overlay networks. In *Proceedings: IEEE Network Operations and Management Symposium (NOMS)* (pp. 819–822).
4. Huang, C., Shen, H., & Chuang, Y. (2010). An adaptive bandwidth reservation scheme for 4G cellular networks using flexible 2-tier cell structure. *Expert Systems with Applications*, 37(9), 6414–6420.
5. Wanalertlak, W., Lee, B., Yu, C., Kim, M., Park, S., & Kim, W. (2011). Behavior-based mobility prediction for seamless handoffs in mobile wireless networks. *Wireless Networks*, 17(3), 645–658.
6. Becvar, Z., Mach, P., & Simak, B. (2011). Improvement of handover prediction in mobile WiMAX by using two thresholds. *Computer Networks*, 55, 3759–3773.
7. Sgora, A., & Vergados, D. (2009). Handoff prioritization and decision schemes in wireless cellular networks: a survey. *IEEE Communications Surveys and Tutorials*, 11(4), 57–77.
8. Choi, S., & Shin, K. G. (2002). Adaptive bandwidth reservation and admission control in QoS-sensitive cellular networks. *IEEE Transactions on Parallel and Distributed Systems*, 13(9), 882–897.
9. Ye, Z., Law, L., Krishnamurthy, S., Xu, Z., Dhirakaosal, S., Tripathi, S., & Molle, M. (2007). Predictive channel reservation for handoff prioritization in wireless cellular networks. *Computer Networks*, 51(3), 798–822.
10. Abdulova, V., & Aybay, I. (2011). Predictive mobile-oriented channel reservation schemes in wireless cellular networks. *Wireless Networks*, 17(1), 149–166.
11. Ramjee, R., Nagarajan, R., & Towsley, D. (1997). On optimal call admission control in cellular networks. *Wireless Networks*, 3(1), 29–41.
12. Bartolini, N. (2001). Handoff and optimal channel assignment in wireless networks. *Mobile Networks and Applications*, 6(6), 511–524.
13. Bartolini, N., & Chlamtac, I. (2002). Call admission control in wireless multimedia networks. In *Proceedings: Personal, Indoor and Mobile Radio Communications (PIMRC)* (pp. 285–289).
14. Pla, V., & Casares-Giner, V. (2003). Optimal admission control policies in multiservice cellular networks. In *Proceedings of the International Network Optimization Conference (INOC)* (pp. 466–471).
15. Chu, K., Hung, L., & Lin, F. (2009). Adaptive channel reservation for call admission control to support prioritized soft handoff calls in a cellular CDMA system. *Annals of Telecommunications*, 64(11), 777–791.
16. El-Alfy, E., & Yao, Y. (2011). Comparing a class of dynamic model-based reinforcement learning schemes for handoff prioritization in mobile communication networks. *Expert Systems With Applications*, 38(7), 8730–8737.
17. Gimenez-Guzman, J. M., Martinez-Bauset, J., & Pla, V. (2007). A reinforcement learning approach for admission control in mobile multimedia networks with predictive information. *IEICE Transactions on Communications*, E-90B(7), 1663–1673.
18. Sutton, R., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*, Cambridge, Massachusetts: The MIT press.
19. Busoniu, L., Babuska, R., De Schutter, B., & Ernst, D. (2010). *Reinforcement learning and dynamic programming using function approximators*, CRC Press.
20. Watkins, C., & Dayan, P. (1992). Q-learning. *Machine learning*, 8(3–4), 279–292.
21. Brown, T. (2001). Switch packet arbitration via queue-learning. *Advances in Neural Information Processing Systems*, 14, 1337–1344.
22. Proper, S., & Tadepalli, P. (2006). Scaling model-based average-reward reinforcement learning for product delivery. In *Proceedings 17th European Conference on Machine Learning* (pp. 735–742).
23. Driessens, K., Ramon, J., & Gärtner, T. (2006). Graph kernels and Gaussian processes for relational reinforcement learning. *Machine Learning*, 64(1), 91–119.
24. Banerjee, B., & Stone, P. (2007). General game learning using knowledge transfer. In *Proceedings 20th International Joint Conference on Artificial Intelligence* (pp. 672–677).
25. Martinez-Bauset, J., Pla, V., Garcia-Roger, D., Domenech-Benlloch, M. J., & Gimenez-Guzman, J. M. (2008). Designing admission control policies to minimize blocking/forced-termination. In *Advances in Wireless Networks: Performance Modelling, Analysis and Enhancement* (pp. 359–390). Nova Science Pub Inc.
26. Biswas, S., & Sengupta, B. (1997). Call admissibility for multi-rate traffic in wireless ATM networks. In *Proceedings IEEE INFOCOM*, 2 (pp. 649–657).
27. Evans, J. S., & Everitt, D. (1999). Effective bandwidth-based admission control for multiservice CDMA cellular networks. *IEEE Transactions on Vehicular Technology*, 48(1), 36–46.
28. Gilhousen, K., Jacobs, I., Padovani, R., Viterbi, A., Weaver, L.A., J., & Wheatley, C.E., III. (1991). On the capacity of a cellular CDMA system. *IEEE Transactions on Vehicular Technology*, 40(2), 303–312.
29. Hegde, N., & Altman, E. (2006). Capacity of multiservice WCDMA networks with variable GoS. *Wireless Networks*, 12, 241–253.
30. Ben-Shimol, Y., Kitroser, I., & Dinitz, Y. (2006). Two-dimensional mapping for wireless OFDMA systems. *IEEE Transactions on Broadcasting*, 52(3), 388–396.
31. Gao, D., Cai, J., & Ngan, K. N. (2005). Admission control in IEEE 802.11e wireless LANs. *IEEE Network*, 19(4), 6–13.
32. Liu, T., Bahl, P., & Chlamtac, I. (1998). Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks. *IEEE Journal on Selected Areas in Communications*, 16(6), 922–936.
33. Hu, F., & Sharma, N. (2004). Priority-determined multiclass hand-off scheme with guaranteed mobile qos in wireless multimedia networks. *IEEE Transactions on Vehicular Technology*, 53(1), 118–135.
34. Chan, J., & Seneviratne, A. (1999). A practical user mobility prediction algorithm for supporting adaptive QoS in wireless networks. In *Proceedings IEEE International Conference on Networks (ICON)* (pp. 104–111).
35. Jayasuriya, A., & Asenstorfer, J. (2002). Mobility prediction model for cellular networks based on the observed traffic patterns. In *Proceedings of Wireless and Optical Communications (WOC)*, 2.
36. Diederich, J., & Zitterbart, M. (2005). A simple and scalable handoff prioritization scheme. *Computer Communications*, 28(7), 773–789.
37. Rashad, S., Kantardzic, M., & Kumar, A. (2006). User mobility oriented predictive call admission control and resource reservation for next-generation mobile networks. *Journal of Parallel and Distributed Computing*, 66(7), 971–988.
38. Soh, W.-S., & Kim, H. (2003). QoS provisioning in cellular networks based on mobility prediction techniques. *IEEE Communications Magazine*, 41(1), 86–92.
39. Lott, M., Siebert, M., Bonjour, S., vonHugo, D., & Weckerle, M. (2004). Interworking of WLAN and 3G systems. In *Proceedings IEE Communications*, 151(5), 507–513.
40. Sanabani, M., Shamala, S., Othman, M., & Zukarnain, Z. (2007). An enhanced bandwidth reservation scheme based on road topology information for QoS sensitive multimedia wireless cellular networks. In *Proceedings of the 2007 International Conference on Computational Science and its Applications - Part II (ICCSA)* (pp. 261–274).
41. Mahadevan, S. (1996). Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Machine Learning*, 22(1–3), 159–196.

42. Puterman, M. L. (1994). *Markov Decision Processes : Discrete Stochastic Dynamic Programming*. New York: John Wiley & Sons.
43. Das, T. K., Gosavi, A., Mahadevan, S., & Marchallick, N. (1999). Solving semi-markov decision problems using average reward reinforcement learning. *Management Science*, 45(4), 560–574.
44. Darken, C., Chang, J., & Moody, J. (1992). Learning rate schedules for faster stochastic gradient search. In *Proceedings of the IEEE-SP Workshop on Neural Networks for Signal Processing II* (pp. 3–12).