LETTER

Optimal Bandwidth Reservation in Multiservice Mobile Cellular Networks with Movement Prediction

Vicent PLA[†a)], ***Member***, **José Manuel GIMÉNEZ-GUZMÁN**[†], **Jorge MARTÍNEZ**[†], ***Nonmembers***, ***and*** **Vicente CASARES-GINER**[†], ***Member***

SUMMARY We study the impact of incorporating handoff prediction information in the session admission control process in mobile cellular networks. We evaluate the performance of optimal policies obtained with and without the predictive information, while taking into account possible prediction errors. Two different approaches to compute the optimal admission policy were studied: dynamic programming and reinforcement learning. Numerical results show significant performance gains when the predictive information is used in the admission process.
key words: *cellular mobile networks, multiservice, optimization, admission control with prediction*

1. Introduction

Session Admission Control (AC) is a key aspect in the design and operation of multiservice cellular networks that provide QoS guarantees. Terminal mobility makes it very difficult to guarantee that the resources available at the time of session setup will be available in the cells visited during the session lifetime. 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 loss probabilities of both session setup and handoff requests). This paper explores the second type of issues from a novel optimization approach that exploits the availability of information related to the number of handoff requests that will be executed in the near future.

In systems that do not have predictive information available, both heuristic and optimization approaches have been proposed to improve the performance of the AC at the session level. The optimization approach in single service systems is well studied [1, 2], while the study of multiservice systems has only received attention recently [3, 4]. On the other hand, in systems that have predictive information available, most of the proposed approaches to improve performance are heuristic and only in single service scenarios, see for example [5] and references therein.

In [6] the authors determine a near-optimal policy in a single service scenario 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 [6] show that the performance gain when using this 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 handoffs obtained from the occupancy state of the neighboring cells is not sufficiently specific.

In this paper we go a step further and evaluate the performance gain that can be obtained when the AC process is provided with more specific information, both in single and multiservice scenarios. In particular, our model includes a prediction agent (PA) that provides the AC system with the information of the number of mobile terminals (MTs) that will "probably produce a handoff". We assume that only one session is active per MT.

The rest of the paper is structured as follows. In Section 2 we describe the models of both the system and the PA. The two optimization approaches are presented in Section 3. A numerical evaluation of the expected performance of the system is provided 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 resources dependent on the specific technological implementation of the radio interface. A total of N different classes of service are offered by the system. For each type of service, new and handoff session arrivals are distinguished so that there are N types of 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 (handoff) sessions of service i is λ_i^n (λ_i^h) and a request of service i consumes b_i resource units, $b_i \in \mathbb{N}$. The duration of a service i session is exponentially distributed with rate μ_i^s. The cell residence time of a service i session is exponentially distributed with rate

[†]Authors are with the Departamento de Comunicaciones, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022-Valencia, Spain. This work was supported by the *Spanish Ministry of Science and Technology* under projects TIC2001-0956-C04-04 and TIC2003-08272 and by the Universidad Politécnica de Valencia under *Programa de Incentivo a la Investigación.*
a) E-mail: vpla@dcom.upv.es

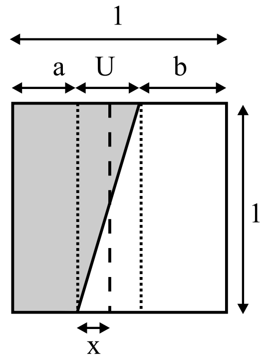

Fig. 1 Proposed model of the PA.

μ_i^r. Hence, the resource holding time for a service i session in a cell is exponentially distributed with rate $\mu_i = \mu_i^s + \mu_i^r$.

Given that the focus of our study was not the design of the PA, we used a model of it instead. An active MT entering the cell neighborhood is labeled by the PA as "probably producing a handoff" (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 handoff into the cell occurs or not (for instance because the session ends or the MT moves to another cell). The AC system is aware of the number of MTs labeled as H at any time.

The model of the PA is shown in Fig. 1 where the square (with a surface equal to one) represents the population of active MTs to be classified by the PA. 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. Notice that part of the MTs that will move into the cell can finish their active sessions before doing so. The classifier sets a threshold, which is represented by a vertical dashed line, to discriminate between those MTs that will likely produce a handoff 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, 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 ($\hat{S}_H^e$) and the shaded area on the right of the threshold ($\hat{S}_{NH}^e$). The 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 handoff if labeled as H and the probability q of producing a handoff if labeled as NH. Note that in general $q \neq 1-p$. It can be shown that $1-p = \hat{S}_H^e/\hat{S}_H = x^2/\left(U(2S_H - U + 2x)\right)$ and that $q = \hat{S}_{NH}^e/(1-\hat{S}_H) = (U-x)^2/\left(U(2-2S_H+U-2x)\right)$.

3. Optimizing the AC Policy

We make use of the theory of Markov decision process to find a policy that minimizes the average expected cost rate. When the system stars at state $\boldsymbol{x}$ and follows policy π then the average expected cost rate over time t, as $t \to \infty$, is denoted by $g^\pi(\boldsymbol{x})$ and defined as: $g^\pi(\boldsymbol{x}) = \lim_{t\to\infty} \frac{1}{t} E\left[w^\pi(\boldsymbol{x},t)\right]$, where $w^\pi(\boldsymbol{x},t)$ is a random variable that expresses the total cost incurred in interval $[0,t]$. For the systems we are considering, 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 [7]. We call it the "cost" 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 optimum policy.

It can be shown for our systems that $g^\pi = \sum_{i=1}^{N}(\omega_i^n P_i^n \lambda_i^n + \omega_i^h P_i^h \lambda_i^h)$, where ω_i^n (ω_i^h) is the relative weight associated to the loss of a new (handoff) request and P_i^n (P_i^h) is the loss probability of new (handoff) requests, both of service i. In general, $\omega_i^n < \omega_i^h$ to account for the fact that the loss of a handoff request is less desirable than the loss of a new session setup request.

Two different optimization approaches have been used to find the optimal policy. The first approach is based on dynamic programming, specifically we used a policy iteration method [7]. This approach is applied to a single service scenario. The second is an automatic learning approach based on the theory of reinforcement learning, more specifically we used the average reward reinforcement learning algorithm proposed in [8]. This approach is applied to a multiservice scenario. Dynamic programming gives an exact solution and allows to evaluate the theoretical limits of incorporating handoff prediction in the AC system, whereas reinforcement learning tackles more efficiently the curse of dimensionality.

3.1 Single Service

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 labeled as H in the cell neighborhood. The set of possible states of the system is $S := \{\boldsymbol{x} = (i,j) : 0 \le i \le C;\ 0 \le j \le C_p\}$, where C_p is the maximum number of MT that can be labeled as H at a given time. We use a large value for C_p so that it has no practical impact in our results. At each state $(i,j), i < C$, the set of possible actions is defined by $A := \{a : a = 0, 1\}$, being $a = 0$ the action that rejects an incoming new session and $a = 1$ the action that accepts an incoming new session. Handoff 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.

For this system, loss rates are given by

$$P_1^n \lambda_1^n = \sum_{\substack{\boldsymbol{x} \\ \pi(\boldsymbol{x})=0}} \lambda_1^n p(\boldsymbol{x}) \qquad P_1^h \lambda_1^h = \sum_{\substack{\boldsymbol{x}=(C,j) \\ 0 \le j \le C_p}} \big(\lambda_h' + jp\mu_p\big) p(\boldsymbol{x})$$

where $p(\boldsymbol{x})$ is the stationary probability of state $\boldsymbol{x}$, $\lambda_h' = q\lambda(1 - \hat{S}_H)\mu_p/(\mu_p + \mu_1^s)$ is the average arrival rate of handoffs that have not been predicted and λ is the input rate to the PA. The optimization problem pursues to find the policy (π^*) that minimizes g^π. Starting from any policy, for example the complete sharing policy, and deploying the *Policy Iteration* method [7, Section 8.6] the optimal policy can be found in a finite — and typically small — number of iterations.

3.2 Multiservice

We formulate the optimization problem as an infinite-horizon finite-state semi-Markov decision process (SMDP) under the average cost criterion, which is more appropriate for the problem under study than other discounted cost approaches. The decision epochs correspond to the time instants at which an arrival occurs. Given that no actions are taken at session departures, then only the arrival events are relevant to the optimization process. Additionally, 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 generic definition of the state space is $S := \{\boldsymbol{x} = (x_0, x_1, k) : x_0, x_1, k \in \mathbb{N}; x_0 \le C; x_1 \le C_p, 1 \le k \le (2N-1)\}$, where x_0 is the number of resource units occupied in the cell under study, x_1 is the number of resource units occupied by sessions labeled as H and k is the arrival type. At each decision epoch the system has to select an action from the set of possible actions $A := \{0 = \text{reject}, 1 = \text{admit}\}$.

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 (handoff) request of service i is ω_i^n (ω_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.

The Bellman optimality recurrence equations for an SMDP under the average cost criterion can be written as $h^*(\boldsymbol{x}, a) = \min_{a \in A_x} \{w(\boldsymbol{x}, a) - g^* \tau(\boldsymbol{x}, a) + \sum_{\boldsymbol{x} \in S} p_{\boldsymbol{xy}}(a) \min_{a' \in A_{\boldsymbol{y}}} h^*(\boldsymbol{y}, a')\}$, 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, $w(\boldsymbol{x}, a)$ is the average cost of taking action a in state $\boldsymbol{x}$, $\tau(\boldsymbol{x}, a)$ is the average sojourn time in state $\boldsymbol{x}$ under action a and $p_{\boldsymbol{xy}}(a)$ is the probability of moving from state $\boldsymbol{x}$ to state $\boldsymbol{y}$ under action $a = \pi(\boldsymbol{x})$. The greedy policy π^* defined by selecting actions that minimize the right-hand side of the above equation is gain-optimal [8].

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 systems where the number of states can be large, like multiservice scenarios, reinforcement learning 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 in advance.

We deploy a reinforcement learning algorithm named Semi-Markov Average Reward Technique (SMART) [8]. 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 m^{th} decision epoch then we update the relative state-action values as follows: $h_{new}(\boldsymbol{x}, a) = (1 - \alpha_m) h_{old}(\boldsymbol{x}, a) + \alpha_m \{w_m(\boldsymbol{x}, a, \boldsymbol{y}) - g_m \tau_m(\boldsymbol{x}, a, \boldsymbol{y}) + \min_{a' \in A_{\boldsymbol{y}}} h_{old}(\boldsymbol{y}, a')\}$, where $w_m(\boldsymbol{x}, a, \boldsymbol{y})$ 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 m^{th} decision epoch and g_m is the average cost rate estimated as: $g_m = \sum_{k=1}^m w_k\big(\boldsymbol{x}_{(k)}, a_{(k)}, \boldsymbol{y}_{(k)}\big) / \sum_{k=1}^m \tau_k\big(\boldsymbol{x}_{(k)}, a_{(k)}, \boldsymbol{y}_{(k)}\big)$

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 average expected 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 holed-disk-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 $b_1 = 1$, $\lambda_1^n = 1$, $\mu_1 = \mu_1^s + \mu_1^r = 1$, $w_1^n = 1$, and $w_1^h = 20$. The value of the input rate to the PA λ is chosen so that the system is in statistical equilibrium, i.e. the rate at which handoff sessions enter a cell is equal to the rate at which handoff sessions exit the cell. It can be easily shown that for our single service scenario $\lambda = (1 - P_1^n)(1 - P_1^{ft})\lambda_n(N_h + \mu_1^r/\mu_1^p)(1/S_H)$, where $P_1^{ft} = P_1^h/(P_1^h + \mu_1^s/\mu_1^r)$ is the probability of forced termination. Note that in our numerical experiments the values of the arrival rates are chosen to achieve realistic operating values for P_i^n and P_i^{ft}. For such values of $P_1^n (\approx 10^{-2})$ and $P_1^{ft} (\approx 10^{-3})$, we make the approximation $\lambda \approx 0.989\lambda_n(N_h + \mu_1^r/\mu_1^p)(1/S_H)$.

For the multiservice scenario we use $N = 2$ services, $b_1 = 1$ and $b_2 = 2$ resource units. The ar-

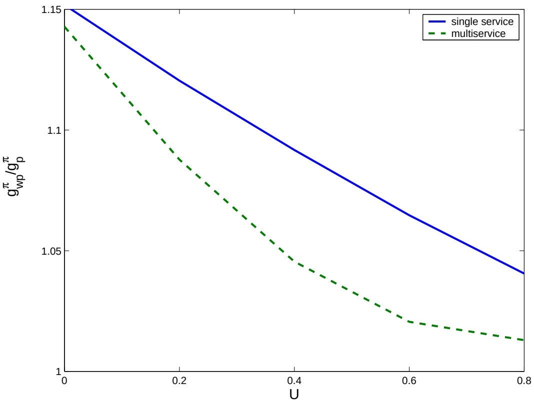

Fig. 2 Performance gain of introducing handoff prediction.

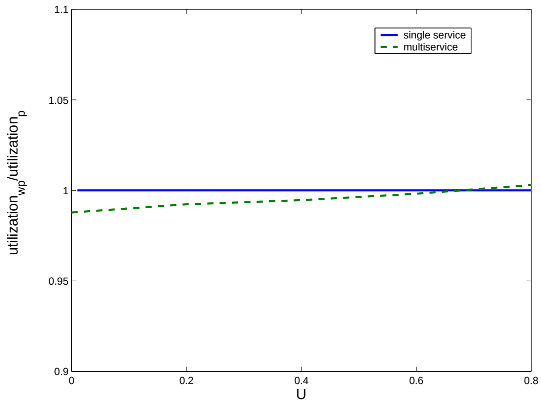

Fig. 3 Impact on utilization of introducing handoff prediction.

rival 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 and to the cell is made equal to the ratio of their surfaces, $\lambda_i^{ng} = 1.25\lambda_i^{nc}$. The ratio of handoff arrival rates to the cell neighborhood from the outside of the system and from the cell is made equal to ratio of their perimeters, $\lambda_i^{ho} = 1.5\lambda_i^{hc}$. Using the flow equilibrium property, we can write $\lambda_i^{hc} = (1-P_i^n)(1-P_i^{ft})(\mu_i^r/\mu_i^s)\lambda_i^{nc}$ which we approximate by $\lambda_i^{hc} \approx 0.989(\mu_i^r/\mu_i^s)\lambda_i^{nc}$. We also set $\mu_1 = \mu_1^s + \mu_1^r = 1$, $\mu_2 = \mu_2^s + \mu_2^r = 3$, $w_1^n = 1$, $w_2^n = 20$, $w_1^h = 10$ and $w_2^h = 200$.

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 $\gamma = 1.0 \cdot 10^{11}$ to obtain $p_m = 1 \cdot 10^{-3} p_0$ when $m = 1 \cdot 10^7$. We start with an exploration rate $p_0 = 0.1$.

Figure 2 shows the gain when introducing prediction for different values of the uncertainty U. In the multiservice scenario, for each value of U we run 10 simulations with different seeds and 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 3, which shows the utilization gain for different values of U, justifies the efficiency of our optimization approach.

5. Conclusion

In this paper we analyze the performance gain that can be obtained when handoff prediction information is considered in order to optimize the admission control policy in a mobile cellular network. 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 algorithm. 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 a future work we will consider a more sophisticated model of the prediction agent including, for instance, a more precise estimation of the time instants at which handoffs will occur and not only the incoming handoffs to the cell but also the outgoing handoffs. This last consideration could be relevant to increase performance as taking into account only incoming handoffs could lead the AC system to reserve too many resources.

References

[1] 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.

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

[3] N. Bartolini and I. Chlamtac, “Call admission control in wireless multimedia networks,” Proceedings of IEEE PIMRC, 2002.

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

[5] W.S. Soh and H.S. Kim, “Dynamic bandwidth reservation in cellular networks using road topology based mobility prediction,” Proceedings of IEEE INFOCOM, 2004.

[6] 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.

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

[8] 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.