A Learning Approach for Call Admission Control with Prioritized Handoff in Mobile Multimedia Networks

El-Sayed El-Alfy, Yu-Dong Yao, Harry Heffes

Wireless Systems Lab, Dept. of Electrical and Computer Eng., Stevens Inst. of Tech., Hoboken, NJ 07030-5991
{eabdelka, yyao, hheffes}@stevens-tech.edu

Abstract

In this paper we propose an alternative approach for improving the quality of service in cellular mobile multimedia networks while prioritizing handoff call requests over new call requests. The goal is to reduce the handoff failures while still making efficient use of the network resources. A performance measure is formed as a weighted linear function of new call and handoff call blocking probabilities for each service type. This problem is formulated as a semi-Markov decision process with average cost criterion. A simulation-based learning algorithm is adopted to determine a near-optimal control policy from direct interaction with the network, and without a priori knowledge of the network dynamics or traffic. Extensive simulations are provided to assess the effectiveness of the algorithm under a variety of traffic conditions. Comparisons with some well-known policies are also presented.

1. Introduction

Next generation cellular networks are expected to support multimedia traffic and wide user mobility anytime and everywhere. To increase the allocation efficiency of the limited radio resources, there is a trend toward reducing the cell size to micro-cells and even pico-cells. As a result there will be a substantial increase in the handoff rate in each cell resulting from calls migrating from one cell to another. However due to the scarcity of the available resources, the ongoing call may be prematurely terminated due to handoff failure. The potential performance measures in cellular networks are the new call blocking probability and the handoff failure rate. Since there is a trade off between the handoff failure rate and the new call blocking probability, a strict control of the available resources is essential to maintain the network throughput and protect the grade of service (GoS) for both active and new calls (or multimedia sessions). The GoS is considered to be a function of blocked calls and handoff failures.

Efficient use of the cellular network resources is affected by the adopted channel assignment scheme among cells as well as the admission control strategy. Several schemes have been proposed in the literature for channel assignment among cells ranging from static allocation to dynamic allocation [1], [2]. Recently, reinforcement learning (RL) [3] has been applied to a number of applications in communication systems such as dynamic channel allocation in cellular telephone systems [4]. In these schemes, a call is admitted if a channel is available such that the channel reuse constraint is not violated without differentiation between new calls and handoffs. However, from the user point of view, dropping an ongoing call is more undesirable than blocking a new call. As a result, much research has been devoted to call admission control and channel assignment with prioritized handoff requests [2, 5-11].

One well-known method that prioritizes handoff over new call requests is the guard channel threshold policy. The basic idea is to a priori reserve certain number of channels in each cell to handle handoff requests. A single threshold for all traffic classes has been considered in [5]. The performance analysis of a hybrid threshold scheme for multiple classes has been considered in [6] using Stochastic Petri Nets (SPN). However the selection of the threshold for each class of traffic is not considered since, as mentioned there, it is a non-trivial problem and the threshold chosen for one type of traffic clearly affects the performance of the other traffic classes. Reference [7] suggested a predictive scheme for determining a single guard threshold for all traffic classes. Other approaches based on Markov decision processes and dynamic

0-7803-6728-6/01/$10.00 ©2001 IEEE.

programming theory have been considered for a single traffic class in cellular networks [8]. However, for dynamic networks supporting multiple traffic classes with diverse characteristics, the computational complexity of these approaches becomes too high and exact solutions become intractable.

In this paper we propose an alternative approach, based on a reinforcement learning methodology, to approximate the solution of the optimal admission control problem in cellular networks supporting several traffic classes. We study the application of average cost reinforcement learning to prioritize the admission of handoff requests over new calls and compare the performance of the algorithm with the guard channel and complete sharing policies. The algorithm under study has a number of advantages: reduced state space, simple implementation, model-free, and self-adjusting to different traffic conditions.

The remaining of this paper is organized as follows. The next section describes the problem and defines the performance measures used in this paper. In Section 3 we formulate the problem as an average cost semi-Markov decision process and then we present a class of reinforcement learning schemes for solving the problem. Simulations and numerical results are given in Section 4. Comparisons with complete sharing and guard channel policies, both analytical and empirical, are also given. Finally, in Section 5 we present conclusions and future work.

2. Problem Description

Consider a cellular network with a limited number of bandwidth units or channels. Here the concept of a channel is generic. It can be a time slot, a frequency carrier or a spreading code depending on the access technology used whether FDMA, TDMA, or CDMA. Here multimedia traffic is defined in terms of their traffic characteristics and resource requirements, e.g., the effective bandwidth requirements. We consider K different classes indexed by k=1, 2, ..., K with each class characterized by a set of traffic parameters such as mean arrival rate, bit rate or bandwidth requirements, mean service time and a weighting factor indicating the relative importance of each class or its priority level. Under a fixed channel assignment scheme and the assumption of spatially uniform traffic conditions, the cellular network model can be studied through a single cell [9] as shown in Fig 1.a. Figure 1.b. shows the corresponding traffic model. We make the commonly used assumptions that the new call and handoff traffic arrivals are according to mutually independent Poisson processes with average arrival rates λ_{nk}, λ_{hk} respectively. The call session duration is exponentially distributed with mean $1/\mu_{sk}$. The cell dwelling time or the inter-handoff time is exponential with mean $1/\mu_{hk}$. The channel holding time, which is the minimum of the call session time and the cell dwelling time, is also exponential with mean $1/\mu_k = 1/(\mu_{sk}+\mu_{hk})$. The bandwidth requested by type-k calls is b_k.

2.1. The nonprioritized scheme

The nonpriotized or complete sharing scheme does not differentiate between new call and handoff requests. Therefore both handoffs and new calls have the same blocking probability for each traffic class, i.e., $B_{nk}=B_{hk}$ for all k. The total arrival rate for class-k traffic is $\lambda_k=\lambda_{nk}+\lambda_{hk}$. Under complete sharing policy the evolution of the system state can be modeled as a multidimensional Markov chain. Let $\boldsymbol{n}=(n_1, n_2, \ldots, n_K)$ be the system state where n_k is the number of ongoing calls of type-k. The state space of the system, S, is a finite set given by

$$S = \{\boldsymbol{n} \mid \boldsymbol{n}\cdot\boldsymbol{b}^T \le C\}, \tag{1}$$

where $\boldsymbol{b}^T$ is the transpose of the required bandwidth vector $\boldsymbol{b}$, and $\boldsymbol{n}\cdot\boldsymbol{b}^T$ is the dot product. Following [12], the equilibrium state distribution has product form solution and is expressed as

$$P(\boldsymbol{n}) = \prod_{k=1}^{K} \frac{\rho_k^{n_k}}{n_k!} G^{-1}(S), \tag{2}$$

where $G(S) = \sum_{\boldsymbol{n}\in S}\left(\prod_{k=1}^{K}\frac{\rho_k^{n_k}}{n_k!}\right)$ and $\rho_k = \dfrac{\lambda_k}{\mu_k}$.

The new call and handoff blocking probabilities are given by

$$B_{nk} = B_{hk} = \sum_{\boldsymbol{n}\in S \wedge \boldsymbol{n}_k^+ \notin S} P(\boldsymbol{n}), \tag{3}$$

where $\boldsymbol{n}_k^+ = (n_1, n_2, \ldots, n_k+1, \ldots, n_K)$.

2.2. The guard channel scheme

One approach to prioritize handoff calls over new calls is called guard channel approach [10]. The idea is to assign a higher capacity limit for handoff calls rather than new calls for each traffic class. New calls of type-k will be blocked if the number of occupied channels is G_k or greater but handoff calls will have access up to C_k. The difference between the two limits, C_k-G_k, is known as guard channels. Guard channels are exclusively reserved for handling handoffs after the network carried load reaches a certain value.

2.3. The optimization problem

The objective of this study is to find a channel allocation policy that attempt to minimize a weighted linear function of new call and handoff blocking probabilities of each type as defined by,

$$P=\sum_{k=1}^{K}\left\{w_{nk}\frac{\lambda_{nk}}{\lambda_{nk}+\lambda_{hk}}B_{nk}+w_{hk}\frac{\lambda_{hk}}{\lambda_{nk}+\lambda_{hk}}B_{hk}\right\}, \quad (4)$$

where w_{nk} and w_{hk} represent the relative weights of type-k traffic with $w_{hk} > w_{nk}$ reflecting the fact that rejecting a handoff request is more undesirable than blocking a new call attempt. In the next section we present a class of RL-based channel allocation schemes.

3. The Proposed Approach

In this section we develop an alternative approach based on reinforcement learning for call admission control with prioritized handoff. We first formulate the optimization problem as an ergodic average cost semi-Markov decision problem (AC-SMDP) then we present an RL-based scheme for solving the AC-SMDP online.

3.1. Problem formulation as AC-SMDP

When a new (handoff) call arrives, a decision is required to be made whether to admit or reject the call. If the new (handoff) call of type-k is rejected the controller incurs an immediate cost equals w_{nk} (w_{hk}). The controller observes the system state at the occurrence of each event $e_k \in$ {new arrival, handoff arrival, call departure}. When a call departs, there is no real decision to be made and no cost incurred. The system state, $\boldsymbol{n}$, at the decision epochs evolves over time and the system dynamics can be modeled as a controlled Markov chain (S, A, P). Here S is a finite set of all states given by (1), A is a finite set of all feasible actions, and P is the controlled state transition probability matrix with components $p_{xy}(a|e_k)$ denoting the probability that the system state at the next decision epoch is $y\in S$ given that the system is currently at state $x\in S$, an event e_k occurs, and executing action $a\in A$. A deterministic stationary policy is a mapping from state-event pairs to actions. For ergodic Markov decision processes, the long-term average cost rate or policy gain is independent of the initial state and given by

$$g^{\pi}=\lim_{n\to\infty}E\{\sum_{t=1}^{n}c_t\}\Big/E\{\sum_{t=1}^{n}\tau_t\}, \quad (5)$$

where c_t is the immediate cost incurred at the decision epoch t and τ_t is the time duration until the next decision epoch. Let Π denote the set of all feasible policies. The controller objective is to determine a policy, $\pi^*\in\Pi$, that minimizes the average cost rate, i.e. $g^{\pi^*}=g^*=\min_{\pi\in\Pi}g^{\pi}$. The optimality equations for the AC-SMDP have the following recursive form

$$h^*(x)+g^*\tau(x)=E_e\left[\min_{a\in A_{x,e}}\{c(x,e,a)+h^*(y)\}\right], \forall x\in S, \quad (6)$$

where $h^*(\boldsymbol{x})$ is an optimal average-adjusted state value function $h^*: S\to\Re$, $c(\boldsymbol{x}, e, a)$ is the expected immediate cost incurred when being in state $\boldsymbol{x}$, event e occurs and action a is selected, and $\tau(\boldsymbol{x})$ is the average sojourn time until the next decision epoch when being in state $\boldsymbol{x}$. By observing the system state at each event, the next state y is deterministically determined as a function of the current state, event type and selected action.

Other formulations are possible. For example, the decision epochs can be defined only at time points when a real decision is needed. To simulate the single cutoff scheme for all traffic classes, the decision epochs correspond to the arrival of new calls. Similarly for a hybrid cutoff scheme, the decision epochs correspond to the arrival of new calls or handoff requests. Under such conditions the system state at the next decision epoch is a random variable and the above equation is modified by replacing $h^*(y)$ by the average value and this requires the knowledge of the transition probability. Another technique is to use the action value functions or Q-values [3] instead. Dynamic programming (DP) techniques can be used to solve this optimization problem. However DP techniques are off-line planning policies which presume the priori knowledge of a perfect system model. Also they fail to scale well for large state space, complex dynamic systems. In the following subsection we present a stochastic simulation-based approximation of the DP methodology for learning an admission policy on-line from direct interaction with the network as depicted in Fig. 2. This policy autonomously adjusts to the traffic conditions.

3.2. RL-based scheme

The controller uses the sample information to incrementally update the state or action value functions, to solve the system of equations (6) using temporal difference [3] as follows

$$h_{new}(x) = h_{old}(x) + \alpha\left\{\left[c_{im} - g^{*}\tau_{im} + h_{old}(y)\right] - h_{old}(x)\right\} \tag{7}$$

where $\alpha \in [0, 1]$ is a learning rate or step size parameter. The setting for α may be fixed or diminishing over time. When a call of type-k arrives when the system state is $\boldsymbol{x}$ the decision maker selects whether to admit or reject as follows

$$\pi(x) = \begin{cases} \text{admit} & x_k^{+} \in S \wedge h(x_k^{+}) - h(x) \le w_{mk} \\ \text{reject} & \text{otherwise} \end{cases}, \tag{8}$$

for $m \in \{n, h\}$. This selection rule is known as the greedy selection policy and based on the assumption of certainty equivalence. Another approach, called ε-greedy, is to select the greedy action with a high probability and with a small probability, ε, uniformly select among other actions. The later policy has an advantage over the greedy policy with respect to the rate of convergence since it allows the state values to be updated for non-greedy actions and maintains a balance between exploiting the available information and exploring the system. For large state spaces, the evaluation functions can be represented as a parameterized function using function approximation methods such as neural networks [3].

Average cost estimation. Since the controller has no information about the long-run average cost g^* it can be estimated incrementally using a similar update rule. Another way to estimate g^*, is to use sample average as

$$g_n = \sum_{t=1}^{n} c_t \Big/ \sum_{t=1}^{n} \tau_t . \tag{9}$$

The above estimates can update the value of g only when selecting a greedy action.

4. Simulation Results

In this section we use computer simulations to evaluate the performance of the learning approach. Comparisons with the complete sharing and guard channel schemes are provided. In the first example we consider a single cell within a cellular network with fixed channel assignment, $C = 30$ channels, and a single traffic class. The traffic parameters are $\lambda_n = 20$, $\lambda_h = 10$, $\mu = 1$, $w_n = 1$, $w_h = 5$ and each call requests one channel. The optimum number of reserved channels is computed analytically for various handoff arrival rates. The performance of the reinforcement learning (RL) compared with that of complete sharing (CS) and optimal guard channel (GC) policies is depicted in Fig. 3.

In the second example, we consider a cellular network supporting two classes of traffic, $K = 2$, e.g., narrowband (NB) and wideband (WB). We consider a single cell with a limited number of channels $C = 50$. New calls and handoffs of type-k are assumed to arrive according to mutually independent Poisson processes with means λ_{nk}, and λ_{hk} respectively. Each traffic class requests a fixed number of channels, b_k, during the call duration and the channel holding time is exponentially distributed with mean $1/\mu_k$. The numerical values of the traffic parameters are shown in Table 1. Simulations results depicting the blocking probabilities and average cost incurred for the learning algorithm compared to those of complete sharing are shown in Figs. 4-5 over a single simulation run. The time units used are generic time units (TU).

Table 1. Traffic parameters for example two.

	NB	WB
New call Arrival rate, λ_{nk}	20	10
Handoff arrival rate, λ_{hk}	10	5
Channel holding mean time μ^{-1}_k	1	1
Bandwidth requirements, b_k	1	2
New calls weighting factor, w_{nk}	1	2
Handoff weighting factors, w_{hk}	5	10

5. Conclusions

The call admission control in cellular mobile multimedia networks with prioritized handoffs has been formulated as an AC-SMDP and an RL-based scheme for approximating the optimal admission policy has been proposed. A key finding of this study is that the reinforcement leaning algorithm has a superior performance over complete sharing and a comparative performance to the optimal guard threshold policy. Also, we found that for multiple traffic classes the learning algorithm generalizes the concept of guard threshold to guard states. This paper is part of a major study which explores the application of the learning algorithms for prioritized resource allocation in multimedia cellular mobile networks with QoS provisioning.

References

[1] G. L. Stuber, *Principles of mobile communication*. Kluwer Academic Publishers, 1996.

[2] S. Tekinay and B. Jabbari, "Handover and channel assignment in mobile cellular networks," *IEEE Commun. Mag.*, vol. 29, pp. 42-46, 1991.

[3] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. MIT Press, 1998.

[4] S. P. Singh and D. P. Bertsekas, "Reinforcement learning for dynamic channel allocation in cellular telephone

systems," In *Advances in NIPS 9*, MIT Press, pp.974-980, 1997.
[5] S. S. Rappaport and C. Purzynski, "Prioritized resource assignment for mobile cellular communication systems with mixed services and platform types," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 443-457, 1996.
[6] B. Li, C. Lin, and S. T. Chanson "Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks," *Wireless Networks* 4, pp. 279-290, 1998.
[7] P. Ramanathan, K. M. Sivalingam, and P. Agrawal, "Resource allocation during handoff through dynamic schemes for mobile multimedia wireless networks," *Proc. of INFOCOM*, 1999.
[8] M. Saquib and B. Yates, "Optimal call admission to a mobile cellular network," *IEEE Veh. Tech. Conf.*, VTC-95, pp. 190-194, 1995.
[9] J. Ho and C. T. Lea, "Improving call admission policies in wireless networks," *Wireless Networks*, vol. 5, pp. 257-265, 1999.
[10] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Tech.*, vol. 35, pp. 77-92, 1986.
[11] S. Oh and D. Tcha, "Prioritized channel assignment in a cellular radio networks," *IEEE Trans. On Commun.*, vol. 40, pp. 1259-1269, 1992.
[12] J. S. Kaufman, "Blocking in a shared resource environment," *IEEE Trans. On Commun.*, vol. 29, no. 10, pp. 1474-1481, October 1981.

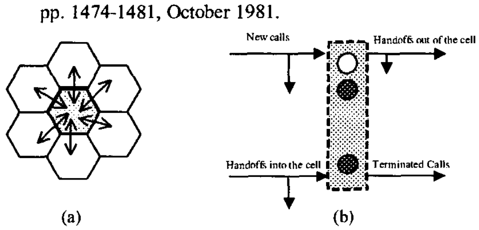

Fig. 1. Cellular system and traffic model.

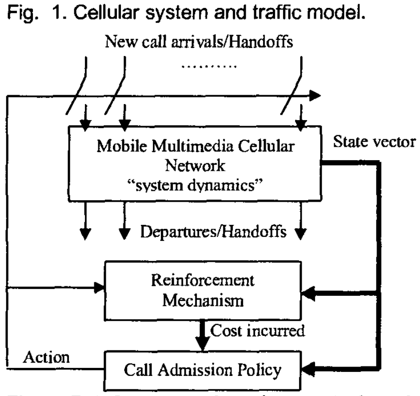

Fig. 2. Reinforcement learning control model.

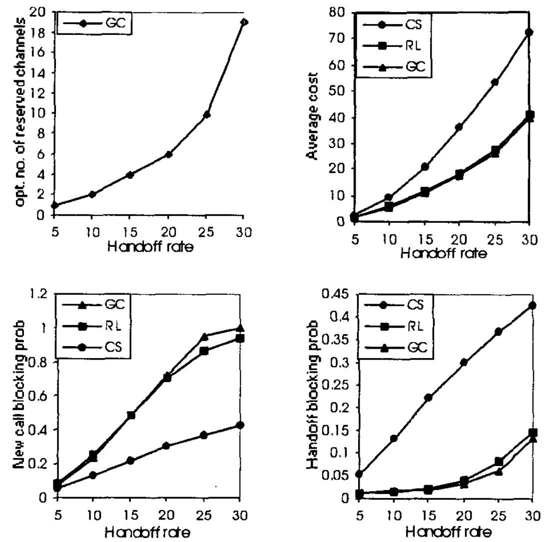

Fig. 3. Performance of CS, GC and RL vs. handoff arrival rates, C=30, λ_n=20, w_n=1, w_h=5, b=1, and μ=1. Each point is an average over 5 runs for different traffic scenarios.

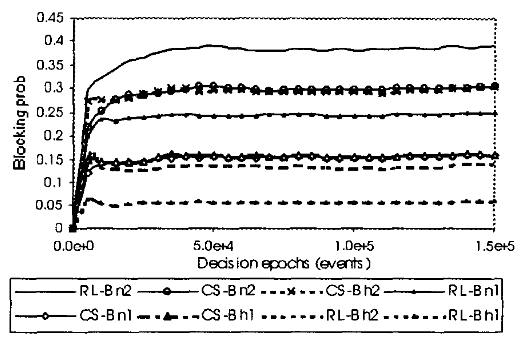

Fig. 4. Learning and complete sharing blocking probabilities for the two-traffic class example.

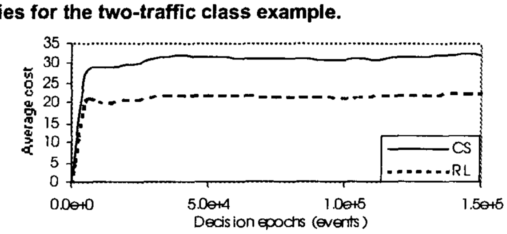

Fig. 5. Learning and complete sharing average cost incurred for the two-traffic class example.