Call Admission Control: Solution of a General Decision Model with State Related Hand-off Rate

Novella Bartolini

Università di Roma "Tor Vergata", Italy
E-mail: novella@uniroma2.it

Imrich Chlamtac

University of Texas at Dallas
E-mail: chlamtac@utdallas.edu

This paper studies call admission policies for access control in cellular networks by means of a Markov Decision Process (MDP). This approach allows us to study a wide class of policies, including well known pure stationary as well as randomized policies, in a way that explicitly incorporates the dependency between the hand-off rate and the system state, assuming that the hand-off rate arriving to a cell is proportional to the occupancy level of the adjacent cells. In particular, we propose and analyze a nonpreemptive prioritization scheme, we term the cutoff priority policy. This policy consists of reserving a number of channels for the high priority requests stream. Using our analytical approach, we prove the proposed scheme to be optimal within the analyzed class.

1. Introduction

The current trend in cellular networks is exhibited in a reduced cell size to accommodate more mobile users in a given geographical area. The reduction leads to increased spectrum-utilization efficiency, but also results in more frequent hand-offs and makes guaranteed connection level QoS more difficult to achieve. Since it is impractical to completely eliminate hand-off drops, the next best alternative is a *probabilistic* guarantee on quality.

The rate of hand-off calls in a given cell depends on the number of calls in progress in the adjacent cells. Assuming that the subscriber mobility remains unchanged, an increase in the number of calls in the adjacent cells is naturally likely to increase the rate at which calls are handed off to the given cell, or in other words the hand-off rate is a function of the system state. On the other hand, the outgoing hand-off rate from a cell varies as a linear function of the number of calls in progress in the cell.

Several call admission policies have already been proposed and analytical formulas for the most important QoS parameters have been given. A comparison between the behavior of few different schemes has occasionally been introduced by means of simulative results [2,8], while analytical comparisons are made only in very few works [4] and with a very narrow class of call admission schemes.

In this paper we want to introduce a general decision model as an instrument to analytically compare the behavior of such schemes.

One of the novelties of this model is that it explicitly incorporates the dependency between the hand-off rate and the system state (in terms of number of calls in progress), and therefore can be expected to be more accurate than models based on average behavior.

Further, this model allows the analysis of call admis-

0-7803-6596-8/00/$10.00 © 2000 IEEE

sion policies that enable queueing of hand-off requests when there is not any available channel.

By means of this decision model, we search for an access control policy that gives high priority service to the hand-off requests without running the risk of compromising the whole traffic because of an insufficient consideration of initial attempts of connections.

Most of the recently proposed call admission control schemes can be studied through the decision model introduced in this paper.

An optimization analysis, using an objective function in the form of a linear combination of the loss probabilities of the two streams of arriving requests, is carried out.

This methodology is used in this paper to analytically prove the optimality of a *cutoff priority policy* (CPP) when the objective function gives higher priority to the hand-off stream and queueing of requests is not allowed.

Under CPP [2–6], priority to hand-off calls is ensured by reserving a certain number of channels, also known as *guard channels.* According to CPP, an initial attempt request is accepted only if the total number of calls in progress, regardless of their type, is below a cutoff value and a free channel is available.

This result has an immediate practical application as the optimal cutoff value can be easily computed once basic statistic parameters defining the traffic of requests are known. These parameters, like those characterizing the distributions of arrivals, cell residence time and call holding time, are used to formulate the analytical models that can be solved by means of very commonly used methods of operations research.

2. Analytical Model

Since, in the most common admission policies considered in literature works, the decision whether to accept or refuse a certain call is based on the number of ongoing calls in the given cell, it seems natural that the state of our model represents this measure of the occupancy level. Our traffic model consists of a Markov decision process in which a single cell is modeled as a service center with C servers corresponding to the available frequency channels. Arriving users, representing requests of connection to the base station, belong to two priority classes: high priority for hand-off calls and low priority for initial access requests.

Knowing the number of calls in the neighbor cells gives us some idea of how many calls we can expect will be handed off in the next unit of time. On the other hand, keeping track of this information can significantly increase the size of the state space. If we assume some uniformity in the system, for what mainly concern the geographic environment of the cells and the mobility of the subscribers, then the number of calls in the current cell gives a good indication of the number of calls in neighbor cells.

As often happens in literature works, arrivals are assumed to be generated according to Poisson processes.

The arrival rate of new requests of connection, that will be served with low priority, will be λ_L while we can assume a hand-off rate proportional to the number i of busy channels i.e. $i \cdot \lambda_H$ that will be treated with high priority, where λ_H is a measure proportional to the hand-off rate per ongoing call from an adjacent cell to the considered cell.

Blocked initial requests are lost, while a blocked hand-off call can wait in the hand-off queue for a channel of the new cell by continuing to use a channel of the previous cell. The queueing scheme is briefly described as follows. No initial access request is granted a channel before the hand-off requests in the queue are served. When a MS reaches the overlapping region between two adjacent cells, also called *hand-off region* (HR), and no free channels are available in the destination cell, the call remains queued until either an available channel in the new cell is found, or the MS abandons the HR before a channel becomes available, thus causing the forced termination of the hand-off call and its departure from the queue. In the case of high demand for hand-off, hand-off calls will be denied queueing due to the limited size of the hand-off queue. The queueing device has a finite number of places M_H.

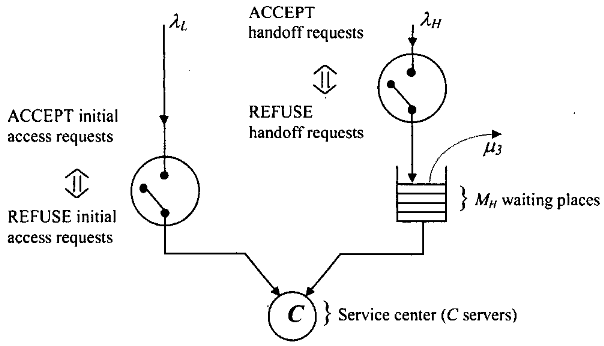

Figure 1. System configuration

In this model a call may exit from the control of the base station in different ways:

1. the conversation is completed (it may happen even with a queued hand-off request, which thus abandons the queue);
2. the MS goes out of cell;
3. a waiting hand-off call is terminated because it is not served before passing the HR, thus it abandons the queue.

The distribution of these events is supposed to be exponential with parameters μ_1, μ_2 and μ_3 respectively.

Fig. 1 shows our model configuration.

The switches represent the two actions (accept or refuse) that can be chosen by the access control policy when a call arrives. A refused call is definitively lost, regardless of its priority class.

A call generated within a cell can be accepted or not, according to a certain control policy. If accepted, it can be completed before the MS goes out of the cell, otherwise the hand-off procedure is initiated. The base station of the destination cell may decide to refuse the hand-off request, to provide it with a new channel, if available, or to put it into the hand-off queue while waiting for a new available channel. While in the hand-off queue, the call continue to use a channel of the old cell. During the time the call spends in the queue, the user may exit from the HR before obtaining a new channel, terminating the hand-off procedure unsuccessfully, or may also decide to conclude the call.

2.1. A general decision model

The memoryless property of all probability distributions in a Markov process makes impossible to represent policies for which the behavior of the system strictly depends on its past history, unless we use several different states to represent the same occupancy level. Each state $\mathbf{s}$, belonging to the finite state space E of the Markov decision process, can be defined through a couple of indexes (i,t), where i represents the number of busy servers, while t is a *state tag*, with $t \in \{1, 2, ..., n\}$ introduced to allow different decisions in correspondence with the same occupancy level i. Let us consider the following set of possible actions that can be undertaken at each state of the process:

a_1 : accept requests belonging to both streams
a_2 : deny access to initial attempts
a_3 : deny access to hand-off calls
a_4 : deny access to both streams of requests.

We define the function $n(\mathbf{s})$ as follows. Given $\mathbf{s} \in E$, $n(\mathbf{s})$ is the occupancy level characterizing the state $\mathbf{s}$. Thus $n(\mathbf{s})$ is the sum of the number of busy channels and the number of busy places in the queueing device. If $\mathbf{s} = (i, t)$, then $n(\mathbf{s}) = i$. If $C \leq n(\mathbf{s}) < C + M_H$, the set of feasible actions reduces to a_2, a_4 because we have no queueing device for initial access requests and to a_4 if $n(\mathbf{s}) = C + M_H$.

Consider now a partition of the set E into classes E_i with the following properties: $E = \bigcup_{i=0}^{C} E_i$, where $E_i = \{\mathbf{s} \in E, n(\mathbf{s}) = i\}$. From any state $\mathbf{s} \in E_i$, a new request acceptance leads the system to any state $\mathbf{q}$ of the class E_{i+1}, denoted by $Succ(\mathbf{s})$. The choice of the next state among the members of this class follows a certain probability distribution $\pi^+_{\mathbf{sq}}$, where $\mathbf{q} \in Succ(\mathbf{s})$, with $\sum_{\mathbf{q} \in Succ(\mathbf{s})} \pi^+_{\mathbf{sq}} = 1$.

The transition rate from $\mathbf{s}$ to any state $\mathbf{q}$ of the class $Succ(\mathbf{s})$ is $\lambda(\mathbf{s}, a) \cdot \pi^+_{\mathbf{sq}}$, where

$$\lambda(\mathbf{s}, a) = \begin{cases} n(\mathbf{s}) \cdot \lambda_H + \lambda_L & \text{if } a = a_1, \\ n(\mathbf{s}) \cdot \lambda_H & \text{if } a = a_2, \\ \lambda_L & \text{if } a = a_3, \\ 0 & \text{if } a = a_4. \end{cases} \quad (1)$$

On the other hand, from any state $\mathbf{s} \in E_i$, the termination of a service, either due to call completion or to the MS movements outside the cell, brings the system to any state $\mathbf{k}$ of the class E_{i-1}, denoted by $Prec(\mathbf{s})$ with rate $n(\mathbf{s})(\mu_1 + \mu_2)\pi^-_{\mathbf{sk}}$, if $n(\mathbf{s}) \leq C$ and $\{n(\mathbf{s})\mu_1 + C\mu_2 + [n(\mathbf{s}) - C]\mu_3\}\pi^-_{\mathbf{sk}}$ if $n(\mathbf{s}) > C$, where $\sum_{\mathbf{k} \in Prec(\mathbf{s})} \pi^-_{\mathbf{sk}} = 1$.

The transition diagram of the process is represented in Fig. 2, which illustrates all the possible outgoing transitions from one state.

The parameters $\pi^+_{\mathbf{sq}}$, $\pi^-_{\mathbf{sq}}$ and the stationary state-decision associations can be adequately set to turn our general decision model into the models of CPP and of most of the well known policies.

CPP can be obtained by selecting $\pi^+_{\mathbf{sq}}$ and $\pi^-_{\mathbf{sq}}$ with $\mathbf{s} = (i_s, j_s)$ and $\mathbf{k} = (i_k, j_k)$, in the following way: $\pi^+_{\mathbf{sq}} = \pi^-_{\mathbf{sq}} = 1$ if $j_s = j_k = fixed\_tag$, for any $fixed\_tag$ else $\pi^+_{\mathbf{sq}} = \pi^-_{\mathbf{sq}} = 0$, and taking the decision a_1 for all the state $\mathbf{s}$ with i_s lower than the cutoff value T, the decision a_2 if $T \leq i_s < C$ and the decision a_4 if $i_s = C$.

The related model is shown in Fig. 3.

In [1] we show how common policies proposed by other authors [2,4,8,9] can be viewed as particular instances of the general decision model we are analyzing.

2.2. Optimization within the general class

In order to give higher priority to the hand-off stream, rather than to the initial access stream, we introduce a cost function which assigns different penalties to the loss of the two kinds of requests. The system is forced to pay a high penalty H if a hand-off call is refused or if it is firstly queued but no channel is assigned before the MS exits from the HR. If service is denied to an initial attempt of access, the system pays a lower penalty $L < H$.

The optimization procedure can be summed up as follows, while all the details can be found in [1].

- The continuous-time process introduced in the previous section is uniformized and discretized in order to apply discrete-time optimization methods. The objective function is introduced with direct application to this discrete-time model [10,12,11,13].

Then the analysis of the discretized model follows.

- We analytically prove that it exists an optimal deterministic stationary policy, i.e. not randomized, for which the decision chosen in correspondence to each state is always the same, independently of the particular instant of time.
- Moreover, we introduce a dynamic programming formulation to prove the existence of an optimal policy for which the optimal decision does not depend on the state tag, but on its occupancy level only.
- The optimality of CPP is proved through the analysis of the structural properties of the optimal cost function.

The knowledge that, the optimal policy is CPP, dramatically decreases the feasible region of the optimization problem which is reduced to the only search for the optimal cutoff value T. This allows a relevant reduction of the number of iterations for the solution with the most common algorithms like the simplex or the policy improvement.

3. Conclusions

This paper proposes an optimization method of call admission control which takes into account the dependency of the hand-off rates on the average occupancy level of the cells. The model is based on a cost function which gives higher priority to hand-off requests than to originating calls.

This cost function has been studied and optimized through a Markov decision model characterized by a great generality. The proposed model is shown to be able to represent both not stationary policies and randomized fractional policies. Moreover, owing to the particular shape of its transition diagram, it becomes possible to study key policies such as the threshold policy and algorithms with one or more cycles of hysteresis. It is analytically proven that if the objective function is

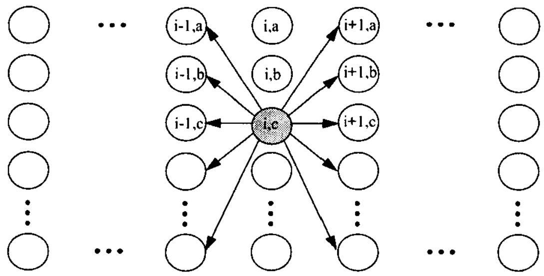

Figure 2. Possible transitions from state (i, k).

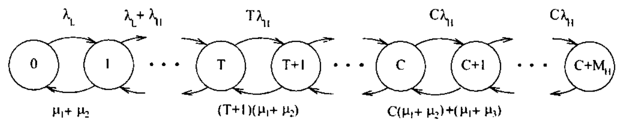

Figure 3. State diagram of CPP.

the total discounted cost function, or the average cost function applied to the infinite horizon problem, the policy CPP is optimal, when no queueing of requests is allowed. Numerical results confirm the optimality of CPP even when hand-off queueing is allowed.

References

[1] N. Bartolini, I. Chlamtac, "Call admission control: solution of a general decision model with state related hand-off rate and exponential assumption of arrivals and call holding times", *Technical Report of University of Texas at Dallas n. UTD/EE/01/00.*

[2] B. Gavish, S. Shridar, "Threshold priority policy for channels assignment in cellular networks", *IEEE Transactions on Computers, Vol. 46, No. 3, March 1997.*

[3] R. Guerin, "Queueing-blocking system with two arrival streams and guard channels", *IEEE Transactions on Communications, Vol. 36, No. 2, February 1988.*

[4] R. Ramjeee, R. Nagarajan, D. Towsley, "On optimal call admission control in cellular networks", *Proceedings of IEEE INFOCOM '96 Conference, March 1996.*

[5] D. Hong, S. S. Rappaport, "Priority oriented channel access for cellular systems serving vehicular and portable radio telephones", *IEEE Proceedings, Vol. 136, No. 5, October 1989.*

[6] Q. Zeng, K. Mukumoto, A. Fukuda, "Performance analysis of mobile cellular radio systems with priority reservation hand-off procedures", *IEEE 0-7803-1927-3/94.*

[7] S. S. Lam, M. Reiser, "Congestion control of store-and-forward networks by input buffer limits - an analysis", *IEEE Transactions on Communications, Vol. 27, pp. 127-133, January 1979.*

[8] D. McMillan, "Delay analysis of a cellular mobile priority queueing system", *IEEE/ACM Transactions on Networking, Vol. 3, No. 3, June 1995.*

[9] R. G. Scherer, "On a cutoff priority queueing system with hysteresis and unlimited waiting room", *Computer Networks and ISDN Systems, No.20, 1990.*

[10] A. T. Bharucha-Reid, *Elements of the theory of Markov processes and their applications.* McGraw-Hill, 1960.

[11] J. Keilson, *Markov chain models. Rarity and exponentiality.* Springer-Verlag, New York, 1979.

[12] D. P. Heyman, M. J. Sobel, *Stochastic models in operations research.* Vol. 1 and Vol.2, McGraw-Hill, 1984.

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