Proceedings of the 34th
Conference on Decision & Control
New Orleans, LA - December 1995

Dynamic Call Admission Control of an ATM Multiplexer with On/Off Sources

M. I. Reiman, J. Wang and D. Mitra

AT&T Bell Laboratories
Murray Hill, NJ 07974
marty@research.att.com, jie@buckaroo.att.com, mitra@research.att.com

Abstract

We consider the problem of call admission control for an ATM multiplexer by formulating it as a semi-Markov decision process. Our model consists of a single link with two types of calls. While in the system, a call sends out cells as an on/off fluid source. The objective is to maximize the long run average reward from call acceptance subject to constraints that the fraction of cells of each type that are lost must be below given levels. Our numerical results allow us to contrast two related formulations of the optimization problem: 'aggressive' and 'conservative'. In addition, we quantify the effect of using cell level control to replace the two cell loss constraints by one.

1. Introduction

ATM (Asynchronous Transfer Mode) appears to be the leading contender for the transport mechanism in emerging broadband multimedia networks. One of the primary attractions of ATM is that it provides for more efficient utilization of bandwidth, being based on cell (packet) switching rather than circuit switching. The statistical multiplexing that is at the core of ATM carries a risk, however. It is not possible to guarantee at the moment of call setup that there will be no loss of cells for a call and still achieve a multiplexing gain. A central question becomes: Given some maximum tolerable cell loss, devise a call admission control procedure that will provide acceptable service (in terms of cell loss) and maximize, in some sense, the amount of traffic carried. This paper describes the results that we have obtained by formulating the above problem as a semi-Markov decision process (SMDP).

Our model, described in Section 2, consists of a single link with two types of calls, both of which operate as on/off fluid sources while in the system. In Section 3 we outline a time scale decomposition that is used to reduce the size of the state space. The reduced state optimization problem is formulated as an SMDP in Section 4. Two related versions of the problem are considered: conservative and aggressive. In the conservative approach we require that the cell loss constraints be satisfied for every state (number of calls of each type in progress). In the aggressive approach we only require that the cell loss constraints be satisfied on a long run average basis. In Section 5 we describe how cell level control can be used to reduce the two cell loss constraints to one constraint. Numerical results are presented in Section 6. These numerical results allow us to quantify the advantages of using the aggressive rather than conservative approach, and of replacing two cell loss constraints by one.

There have been several previous studies on admission control for ATM, such as [3], [5], [6], [7], [8], [9], and [10]. Except for [8], the work in these references is in the context of static policies, which are determined by reachable sets of states. In [8], as well as in this paper, a dynamic admission control scheme is considered, so the admission decision depends on both the state to be entered and the call type awaiting admission. Our work differs from [8] in at least two substantive ways: we use an on/off source model that enables us to avoid the use of simulation, and we consider an 'aggressive' admission control scheme in addition to the 'conservative' scheme based on an acceptance region.

2. Model

We consider an ATM link with transmission capacity R and no buffer, and assume that the network has information on the statistical characteristics of the traffic at call level and burst level (in contrast to [5], where a decision-theoretic framework is used to deal with unknown parameters). There are two types of calls arriving at the link according to Poisson processes; type i calls arrive with rate λ_i, $i = 1, 2$. After a call is admitted, it alternates between on (bursting) and off as a two state Markov process. A type i call has an average on period of length β_i^{-1} and an average off period of length α_i^{-1}. In the long run it spends a fraction $p_i = \alpha_i/(\alpha_i + \beta_i)$ in the on state. When turning off, a call departs with probability q_i. When a type i call is on, it generates cells as a fluid at rate

ν_i. When cells are generated at a rate exceeding the link transmission capacity, those cells that cannot be transmitted are lost.

Two types of Quality of Service (QOS) are of interest: the cell level QOS and the call level QOS. At the cell level, we want the cell loss ratios for the two types to be smaller than f_1 and f_2, respectively. A commonly used bound is $f_1 = f_2 = 10^{-9}$. The call level QOS is reflected in call blocking probabilities.

Each admitted type i call pays the network a reward w_i. The optimization problem we consider is to maximize the (long run average) reward rate subject to satisfying the cell level QOS constraints.

3. Problem Reduction Using Time Scale Decomposition

The optimization problem described above can be formulated as a Semi-Markov Decision Process (SMDP) with constraints. Let k_i denote the number of type i calls in progress, and let n_i denote the number of type i calls in the 'on' state. Under a stationary admission control strategy (k_1, k_2, n_1, n_2) is a state descriptor for a Markov process. Thus the state space of the SMDP is 4-dimensional, which will lead to computational problems. The notion of Nearly Completely Decomposable (NCD) Markov chains can be used to reduce this to a 2-dimensional problem. (Although the discussion presented here is heuristic, the problem reduction is rigorously justified by Theorem 11 of [1].) During a call's 'lifetime' it goes through many on/off cycles. Thus, the on/off cycles have a short duration relative to that of a call. Intuitively, when making a call admission decision, the number of calls of each type in progress is important, but the number of calls of each type in the on state is not, because these quantities oscillate too rapidly.

We handle the above idea mathematically as follows. Consider a family of systems indexed by $\epsilon > 0$. Let

$$\lambda_i(\epsilon) = \epsilon\lambda_i \quad \text{and} \quad q_i(\epsilon) = \epsilon q_i \; .$$

These scalings correspond (for ϵ small) to calls that consist of many short on/off periods. The average duration of a type i call is

$$\frac{\alpha_i^{-1} + \beta_i^{-1}}{q_i(\epsilon)} = \frac{\alpha_i + \beta_i}{\epsilon q_i \alpha_i \beta_i} = (\epsilon\mu_i)^{-1} \; .$$

Thus both arrival and service rates are proportional to ϵ. As $\epsilon \to 0$, the (n_1, n_2) component of the state becomes noise on the time scale where call arrival and departure rates are $O(1)$, and can be ignored for admission control purposes. This part of the state *does* affect the loss rate, so it must be 'averaged' properly. For ϵ small, the (n_1, n_2) process reaches equilibrium between changes in the (k_1, k_2) process. The equilibrium corresponds to fixed (k_1, k_2) and is given by the binomial distribution:

$$\psi(\mathbf{k}, \mathbf{n}) = \prod_{i=1}^{2} \binom{k_i}{n_i} p_i^{n_i} (1 - p_i)^{k_i - n_i} \; .$$

When the total arrival rate is $n_1\nu_1 + n_2\nu_2$, the loss rate is $[n_1\nu_1 + n_2\nu_2 - R]^+$. The average type i loss rate with (k_1, k_2) is thus

$$b_i(\mathbf{k}) = \sum_{n_1=0}^{k_1} \sum_{n_2=0}^{k_2} \psi(\mathbf{k}, \mathbf{n}) \times [n_1\nu_1 + n_2\nu_2 - R]^+ \frac{n_i\nu_i}{n_1\nu_1 + n_2\nu_2} \; . \tag{1}$$

This reflects the fact that there is no priority — the loss rate of a type is proportional to its input rate. It seems intuitively clear that the exponential distribution for on and off times plays no essential role in (1): any distribution would lead to the same result, as long as the calls are independent and have stationary probabilities p_i (for type i) of being on. Indeed, the results of [1] apply for on and off times having any phase type distributions.

4. Formulation as a Semi Markov Decision Process

States, Actions, and Transition Probabilities

We now formulate the limit control problem as an SMDP. Admission decisions are made upon call arrival. We augment the state to include the call type at an arrival epoch. (There are other ways to go about this.) The state space, which we denote by I, is the union of two sets, corresponding to states associated with call arrivals and states associated with call departures. To make this a finite state problem we may need to place an a priori bound on the number of calls that can be accepted in the system. This is taken care of below. Arrival states take the form (k_1, k_2, j), with $k_1, k_2 \geq 0$, and $j = 1, 2$. The state (k_1, k_2, j) corresponds to an arrival of type j call when there are k_1 type 1 calls and k_2 type 2 calls in progress. These are the only states in which decisions need to be made. The set of actions available is $A(k_1, k_2, j) = \{0, 1\}$, where 0 denotes rejection and 1 denotes acceptance. The state (k_1, k_2) is a departure state where the departing call leaves k_1 type 1 calls and k_2 type 2 calls behind. For states of the form (k_1, k_2) where no decision needs to be made we set $A(k_1, k_2) = \{0\}$.

To complete the specification of the SMDP we need to provide transition probabilities, mean sojourn times, rewards, and costs for each state-action pair. Let $\tau(i, a)$ denote the average time spent in state i (until the next decision epoch) if action $a \in A(i)$ is chosen. Then

$$\tau((k_1, k_2), 0) = (\lambda_1 + \lambda_2 + k_1\mu_1 + k_2\mu_2)^{-1}$$

and

$$\tau((k_1, k_2, j), a) = \begin{cases} (\lambda_1 + \lambda_2 + k_1\mu_1 + k_2\mu_2)^{-1}, & a = 0 \\ (\lambda_1 + \lambda_2 + k_1\mu_1 + k_2\mu_2 + \mu_j)^{-1}, & a = 1 \ . \end{cases}$$

Let $p(i, i', a)$ denote the transition probability from state i to i' if action a is chosen. Let

$$\overline{p}((k_1, k_2), i) = \begin{cases} k_1\mu_1/(\lambda_1 + \lambda_2 + k_1\mu_1 + k_2\mu_2) & \text{if } i = (k_1 - 1, k_2) \\ k_2\mu_2/(\lambda_1 + \lambda_2 + k_1\mu_1 + k_2\mu_2) & \text{if } i = (k_1, k_2 - 1) \\ \lambda_1/(\lambda_1 + \lambda_2 + k_1\mu_1 + k_2\mu_2) & \text{if } i = (k_1, k_2, 1) \\ \lambda_2/(\lambda_1 + \lambda_2 + k_1\mu_1 + k_2\mu_2) & \text{if } i = (k_1, k_2, 2), \end{cases}$$

and 0 otherwise. Then

$$p((k_1, k_2), i, 0) = \overline{p}((k_1, k_2), i) \ ,$$

and

$$p((k_1, k_2, j), i, a) = \begin{cases} \overline{p}((k_1, k_2), i), & a = 0, \\ \overline{p}((k_1 + 1, k_2), i), & a = 1, \ j = 1, \\ \overline{p}((k_1, k_2 + 1), i), & a = 1, \ j = 2. \end{cases}$$

A Conservative Approach to the Cell Loss Constraint

We formulate a conservative approach to the cell loss constraint as follows. Let

$$C = \{(k_1, k_2) : b_i(k_1, k_2) \leq f_i k_i p_i \nu_i, \quad i = 1, 2\}. \quad (2)$$

Then C is the set of (k_1, k_2) such that when there are k_1 type 1 and k_2 type 2 calls in the system *forever*, the long run average cell loss rate constraints are always satisfied. The set C uniquely determines a state space I: for $j = 1, 2, (k_1, k_2, j) \in I$ if and only if $(k_1, k_2) \in C$; and $(k_1, k_2) \in I$ if and only if either $(k_1+1, k_2) \in C$ or $(k_1, k_2 + 1) \in C$. The action sets for the states on the boundary of I require minor modification: given $(k_1, k_2) \in C$, if $(k_1 + 1, k_2) \notin C$, then $A((k_1, k_2, 1)) = \{0\}$; if $(k_1, k_2 + 1) \notin C$, then $A((k_1, k_2, 2)) = \{0\}$. In this manner C uniquely determines a SMDP. Note that given f_1, f_2 and R, the set C is always finite. The optimal policy obtained by solving this SMDP is conservative in terms of cell loss constraints because it will never go into any state for *any period of time* where the cell loss constraints will be violated if we stay there *forever*. This SMDP can be solved using value iteration.

An Aggressive Approach to the Cell Loss Constraint

We now describe an 'aggressive' approach to the cell loss constraint. In order to be able to apply standard numerical solution procedures to find the optimal policy, we need to make the state space of the SMDP finite. We achieve this by placing an a priori bound on the number of calls that can be accepted into the system. Let M be a fixed large positive number. We consider a state space $I(M)$ of the form

$$\begin{aligned} I(M) = \{&(k_1, k_2, j) : k_1, k_2 \in \mathbf{Z}^+; \ j = 1, 2; \\ &\text{and } k_1 e_1 + k_2 e_2 \leq M\} \cup \{(k_1, k_2) : \\ &k_1, k_2 \in \mathbf{Z}^+ \ \text{ and } k_1 e_1 + k_2 e_2 \leq M\} \end{aligned}$$

where e_1 and e_2 are the *effective bandwidths* of the two types of calls, which are calculated from a static model (see Section 6). For states (k_1, k_2, j) such that $k_1 e_1 + k_2 e_1 > M - e_j$, we let $A((k_1, k_2, j)) = \{0\}$.

Since M was chosen arbitrarily to make the problem finite, we need to solve a series of problems with increasing M's, until the associated optimal policy and reward stop changing. Then the optimal policy for $M = \infty$ will have been obtained.

Let $r(k_1, k_2, j, a)$ denote the reward earned in state (k_1, k_2, j) when action a is chosen. (The reward earned in states of the form (k_1, k_2) is 0.) Then

$$r(k_1, k_2, j, a) = \begin{cases} 0, & a = 0 \\ w_j, & a = 1 \ . \end{cases}$$

We also define 'costs' which will play a role in the constraint on loss probability. Let

$$\overline{c}_\ell(k_1, k_2) = b_\ell(k_1, k_2)\tau((k_1, k_2), 0), \quad \ell = 1, 2 \ .$$

We define

$$c_\ell((k_1, k_2), 0) = \overline{c}_\ell(k_1, k_2) \ ,$$

and

$$c_\ell((k_1, k_2, j), a) = \begin{cases} \overline{c}_\ell(k_1, k_2), & a = 0 \\ \overline{c}_\ell(k_1 + 1, k_2), & a = 1, \ j = 1 \\ \overline{c}_\ell(k_1, k_2 + 1), & a = 1, \ j = 2 \ . \end{cases}$$

As defined above, $c_\ell(i, a)$ is the expected amount of type ℓ traffic lost until the next decision epoch. Let

$$\overline{c}_{2+\ell}(k_1, k_2) = p_\ell k_\ell \alpha_\ell \tau((k_1, k_2), 0) \ ,$$
$$c_{2+\ell}((k_1, k_2), 0) = \overline{c}_{2+\ell}(k_1, k_2) \ ,$$

and

$$c_{2+\ell}((k_1, k_2, j), a) = \begin{cases} \overline{c}_{2+\ell}(k_1, k_2), & a = 0 \\ \overline{c}_{2+\ell}(k_1 + 1, k_2), & a = 1, \ j = 1 \\ \overline{c}_{2+\ell}(k_1, k_2 + 1), & a = 1, \ j = 2. \end{cases}$$

Then $c_{2+\ell}(i,a)$ is the expected amount of type ℓ traffic to arrive until the next decision epoch.

Let $\phi(k_1,k_2,j)$ denote the probability that action 1 (accept) is chosen in state (k_1,k_2,j). From results of Feinberg [4], we know that it suffices to consider randomized stationary policies for this problem, and the optimal solution can be obtained from the linear program:

$$\text{maximize} \sum_{i\in I}\sum_{a\in A(i)} r(i,a)z_{ia} \tag{3}$$

subject to

$$\sum_{a\in A(j)} z_{ja} - \sum_{i\in I}\sum_{a\in A(i)} p(i,j,a)z_{ia} = 0, \quad j\in I, \tag{4}$$

$$\sum_{i\in I}\sum_{a\in A(i)} [c_\ell(i,a) - f_\ell c_{\ell+2}(i,a)]z_{ia} \le 0, \quad \ell = 1,2, \tag{5}$$

$$\sum_{i\in I}\sum_{a\in A(i)} \tau(i,a)z_{ia} = 1, \tag{6}$$

$$z_{ia} \ge 0\ , \qquad i\in I,\ a\in A(i). \tag{7}$$

The above LP is clearly feasible, because $z_{(0,0),0} = [\tau((0,0),0)]^{-1} = \lambda_1+\lambda_2$, and $z_{ia}=0$ otherwise is a feasible solution (corresponding to never allowing any calls to enter). Given an optimal solution, z, of the LP, we obtain an optimal randomized stationary policy as

$$\phi(k_1,k_2,j) = \frac{z_{(k_1,k_2,j),1}}{z_{(k_1,k_2,j),0} + z_{(k_1,k_2,j),1}}$$

if $z_{(k_1,k_2,j),0} + z_{(k_1,k_2,j),1} > 0$, and $\phi(k_1,k_2,j) = 0$ otherwise. (The quantity $z_{ia}\tau(i,a)$ corresponds to the fraction of time spent in state i with action a chosen.)

Comparison of the Two Approaches

There are a couple of advantages in terms of solution procedure for formulating cell loss constraints with the conservative approach. First, because the state space of the SMDP is finite we do not have to place any a priori bound to make it finite as we do for the aggressive approach, hence we do not have to choose M. Secondly, because the cell loss constraints are already reflected in the state space, we have a SMDP without constraints, which allows us to use the more efficient value iteration method to find the optimal policy. The most noticeable disadvantage of this conservative approach is the sacrifice at the call level QOS: the call blocking probabilities are higher. However, the conservative approach is more consistent with our assumption that calls and bursts are on two very different time scales. When we choose one of the two approaches, we are trading cell level QOS against call level QOS.

5. Cell Level Control

Both the aggressive and conservative approaches involve two cell level constraints, one for each type. It is possible in both cases to transform the problem to one involving one constraint using cell level control, which consists of deciding how many cells of each type are lost. This transformation will result in improved performance of the associated optimal control. Equation (1) presents the cell loss rate under a 'no priority' assumption. Not surprisingly, typically only one of the two cell level constraints is tight in the solution. Thus, by giving priority to the type whose constraint is tight it seems clear that we can improve the solution. Related results and discussion are contained in Bean [2].

Consider the conservative approach, for which the 'acceptance region', C, is given by (2). Let

$$b(\mathbf{k}) = \sum_{n_1=0}^{k_1}\sum_{n_2=0}^{k_2} \psi(\mathbf{k},\mathbf{n})[n_1\nu_1 + n_2\nu_2 - R]^+\ . \tag{8}$$

There is a cell level control (it will depend on k_1,k_2) that can achieve

$$b_i(k_1,k_2) \le f_i k_i p_i \nu_i\ , \quad i=1,2 \tag{9}$$

if

$$b(k_1,k_2) \le f_1k_1p_1\nu_1 + f_2k_2p_2\nu_2\ , \tag{10a}$$

$$b(k_1,0) \le f_1k_1p_1\nu_1\ , \tag{10b}$$

$$\text{and} \quad b(0,k_2) \le f_2k_2p_2\nu_2\ . \tag{10c}$$

If $f_1 = f_2$, then (10a) implies (10b) and (10c), and the two constraints of (9) are replaced by the one constraint (10a). In this case, the acceptance region becomes

$$C = \{(k_1,k_2) : b(k_1,k_2) \le f[k_1p_1\nu_1 + k_2p_2\nu_2]\}\ , \tag{11}$$

where $f = f_1 = f_2$.

The aggressive case is a bit more complicated because the constraints involve the stationary distribution over all states. Consider the constraint

$$\sum_{i\in I}\sum_{a\in A(i)} [c_1(i,a) + c_2(i,a) - f_1c_3(i,a) - f_2c_4(i,a)]\, z_{ia} \le 0\ . \tag{12}$$

For the case $f_1 = f_2$, if we solve the LP (3), (4), (6), (7), (12), we can find a cell level control such that the constraints (5) are satisfied. A related result holds with $f_1 \ne f_2$.

6. Numerical Results

We now describe our numerical results. All the numerical results described here are based on a system with $R = 45$. The data for the two call types is given in Table 1. In addition to the data in Table 1 we need to specify rewards w_1 and w_2. We choose $w_i = e_i/\mu_i$, where e_i is an empirically obtained 'effective bandwidth'. We obtain e_1 and e_2 as follows. Consider the acceptance region defined in equation (11) with parameters as in Table 1 (μ_1 and μ_2 have no effect), which is the line labelled "conservative QOS" in Figure 1. We want a linearly constrained region which is contained in the acceptance region. There is more than one choice, but once we insist that the line intersect the point $(14, 0)$, there is a unique 'maximal' linear constraint, which is labelled "effective bandwidth" in Figure 1. This line, which passes through $(3, 77)$ in addition to $(14, 0)$ yields $e_1 = 3.2143$ and $e_2 = .459$. Based on this we have $w_1 = 32.143$ and $w_2 = .459$.

Type i	μ_i	ν_i	p_i	f_i
1	0.1	6.0	0.025	10^{-9}
2	1.0	1.5	0.100	10^{-9}

Table 1: Traffic Data

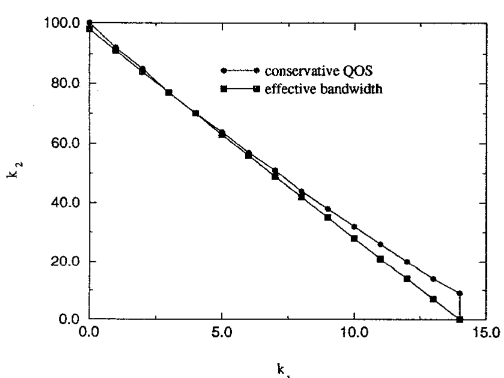

Figure 1: Effective Bandwidth

It becomes clear in the course of doing numerical calculations that we cannot use the same 'load' for both the conservative and aggressive approaches and obtain interesting results. Figures 2 and 3 present call blocking probabilities as a function of λ_1 under constant 'load' for the conservative and aggressive schemes respectively. Constant load means that for each value of λ_1 we choose λ_2 such that

$$\frac{\lambda_1}{\mu_1} e_1 + \frac{\lambda_2}{\mu_2} e_2 = \rho R,$$

where $\rho = .7$ for the conservative scheme and $\rho = .9$ for the aggressive scheme. Using $\rho = .7$ for the aggressive scheme would produce call blocking probabilities indistinguishable from zero on the scale of Figure 2, while using $\rho = .9$ for the conservative scheme would produce call blocking probabilities as high as .1 to .4 (too large to be of practical interest). These figures clearly show the advantage of using cell level control.

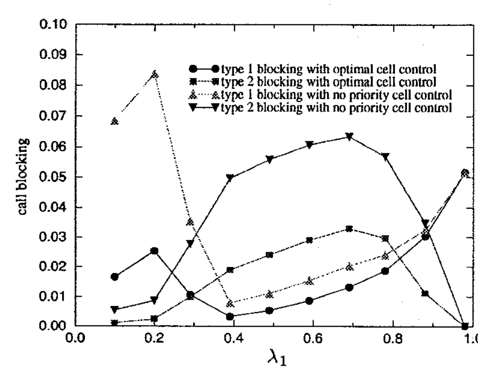

Figure 2: Call Blocking for Conservative Scheme

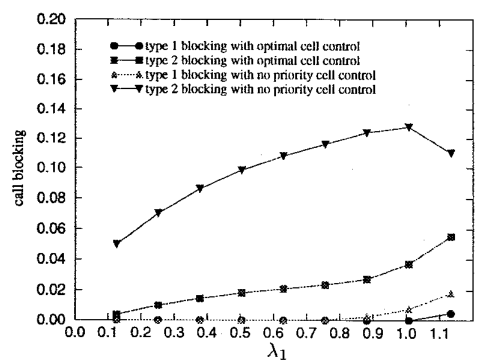

Figure 3: Call Blocking for Aggressive Scheme

Figures 2 and 3 do not provide a direct comparison of the conservative and aggressive schemes. It is clear that the aggressive scheme has lower call blocking probabilities. We investigate how much more traffic can be handled with the aggressive scheme for a given constraint on the call blocking probability. In Figure 4 we plot the boundaries of the 'feasible regions' in (λ_1, λ_2) space for both the conservative and aggressive schemes. If a point (λ_1, λ_2) is inside the set bounded by the two coordinate axes and the aggressive boundary, then there is some call admission policy that yields feasible cell level behavior (with $f_1 = f_2 = 10^{-9}$) and call blocking probability of at most 0.01 for both types. If a point (λ_1, λ_2) is outside of the above set then no such policy exists. The same holds true for the conservative approach, with its associated boundary.

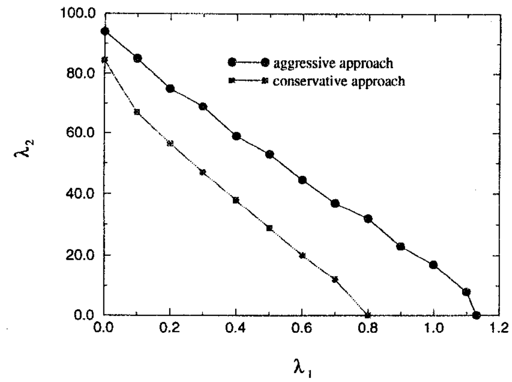

Figure 4: Feasible Regions (1%)

The boundaries of the feasible regions look roughly linear. To explore this further, in Figure 5 we plot the boundaries of the feasible regions for the aggressive approach corresponding to 0%, 1%, and 5% blocking. We also plot the straight line corresponding to 90% load ($\rho = .9$). If we accept linearity of the boundary of the feasible region as a reasonable approximation, we can calculate the boundary via its end points. These ends points are the solutions of one dimensional problems (only one call type) that are substantially simpler than the two dimensional problems we have been dealing with.

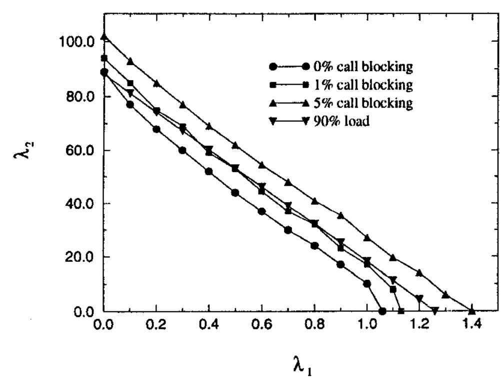

Figure 5: Feasible Regions for Aggressive

We briefly digress to describe the one dimensional problems. Let $s(k)$ denote the mean cell generation rate with k calls in progress ($s(k) = k\nu p$), and let $b(k)$ denote the mean cell loss rate with k calls in progress. It is intuitively clear that the control in the one dimensional problem has a simple threshold (possibly with randomization). Let κ denote a nonrandomized threshold. The cell loss ratio, $L(\kappa)$, is given by

$$L(\kappa) = \sum_{k=0}^{\kappa} \left[(\lambda/\mu)^k/k!\right] b(k) \Big/ \sum_{k=0}^{\kappa} \left[(\lambda/\mu)^k/k!\right] s(k).$$

The call blocking probability is

$$B(\kappa) = (\lambda/\mu)^\kappa/\kappa! \Big/ \sum_{k=0}^{\kappa} (\lambda/\mu)^k/k!.$$

We determine an endpoint associated with cell level constraint f and call level constraint q by finding the maximum λ for which we can find a κ yielding $B(\kappa) \leq q$ and $L(\kappa) \leq f$.

The charging rates w_1 and w_2 played no role in Figures 4 and 5, since feasibility and not optimality was being explored in those figures. In Figures 2 and 3, where the charging rates do play a role, we used rates based on effective bandwidths as determined from Figure 1. Our model allows us to investigate the effect of using other charging rates. Note that, because we are not modeling the effect of price (charging rate) on demand (through a demand curve), charging rates really need to be viewed as internal signals to the system. Viewed in this way, charging rates can be used to get the system to operate in some 'desired' state. Two possible desired states are equal call blocking probabilities and maximum utilization. It is not clear how to set w_1 and w_2 to equalize blocking probabilities. In Figure 6 we plot call blocking probabilities as a function of the ratio w_1/w_2 for the aggressive scheme, using the data in Table 1. In Table 2

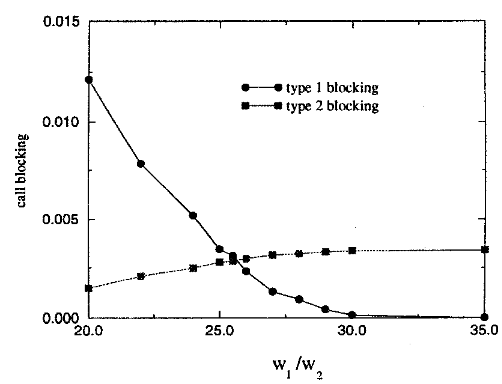

Figure 6: Effect of Charging Rate on Call Blocking

we present the value of the ratio w_1/w_2 that results in equal call blocking probabilities for several cases. Case 1 corresponds to the data in Table 1. In Case 2 $\nu_2 = 2.0$, and the other data is as in Case 1. In Case 3 $\nu_1 = 4.0$ and $p_1 = 0.05$, with the other data as in Case 1. The common call blocking probability is denoted by B. Maximizing utilization corresponds to charging per cell, so that $w_i = p_i\nu_i/\mu_i$. This would give lower call blocking probability to the bursty traffic than under the charging rate $w_i = e_i/\mu_i$. This effect is seen in Table 3. The call blocking probability of type i is denoted by B_i.

Figures 2 and 3 provided an indication of the advan-

Case	(λ_1, λ_2)	$(\frac{w_1}{w_2})^*$	B	$\frac{e_1/\mu_1}{e_2/\mu_2}$
1	Agg.:(0.38, 57.4)	25.5	0.003	70.02
	Con.:(0.29, 44.6)	66.0	0.006	
2	Agg.:(0.38, 37.7)	20.5	0.017	42.74
	Con.:(0.29, 29.3)	49.0	0.014	
3	Agg.:(0.69, 61.1)	18.0	0.005	38.00
	Con.:(0.54, 47.5)	35.5	0.003	

Table 2: Equalizing Call Blocking Probabilities

Case	$\frac{w_1}{w_2}$	(λ_1, λ_2)	B_1	B_2
1	10.0	Agg.:(0.38, 57.4)	0.025	0.000
		Con.:(0.29, 44.6)	0.022	0.001
2	7.5	Agg.:(0.38, 37.7)	0.119	0.001
		Con.:(0.29, 29.3)	0.044	0.005
3	13.3	Agg.:(0.69, 61.1)	0.023	0.003
		Con.:(0.54, 47.5)	0.008	0.002

Table 3: Charging According to the Mean Rates

tage of cell level control. Another indication of this advantage can be seen by comparing the acceptance regions for the conservative scheme. Figure 7 displays the acceptance regions associated with one and two constraints, as given by equations (2) and (11). In

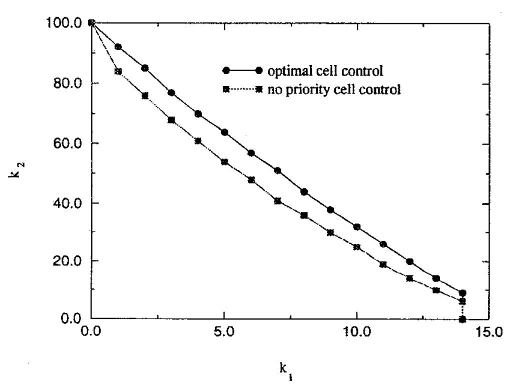

Figure 7: Acceptance Regions ($f_1 = f_2 = 10^{-9}$)

Figure 8, where we use $f_1 = 10^{-9}$ and $f_2 = 10^{-7}$, the advantage of cell level control is more dramatic.

Acknowledgement. We gratefully acknowledge the invaluable help of K. G. Ramakrishnan, in the course of many conversations, on the solution of linear programs and related numerical problems.

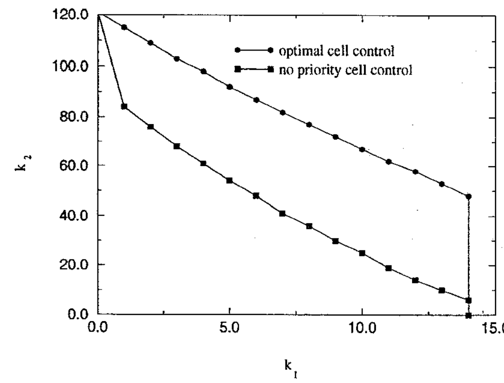

Figure 8: Acceptance Region ($f_1 = 10^{-9}, f_2 = 10^{-7}$)

References

[1] E. Altman and V. A. Gaitsogory, "Stability and Singular Perturbations in Constrained Markov Decision Problems," *IEEE Trans. Aut. Contr.*, **38**, 971–975, 1993.

[2] N. G. Bean, "Effective bandwidths with different quality of service requirements," *Integrated Broadband Communication Networks and Services*, V. B. Iverson (Ed.), IFIP, 1993.

[3] M. Beshai, R. Kositpaiboon, and J. Yan, "Interaction of Call Blocking and Cell Loss in an ATM Network," *IEEE JSAC*, **12**, 1051–1058, 1994.

[4] E. A. Feinberg, "Constrained Semi-Markov Decision Processes With Average Rewards," ZOR, to appear.

[5] R. J. Gibbens, F. P. Kelly, and P. B. Key, "A Decision-Theoretic Approach to Call Admission Control in ATM Networks," *IEEE JSAC*, **13**, 1101–1114, 1995.

[6] T. R. Griffiths, "Analysis of a Connection Acceptance Strategy for Asynchronous Transfer Mode Networks," Globecom '91, Phoenix, 1991.

[7] F. Hübner and P. Tran-Gia, "Quasi-Stationary Analysis of a Finite Capacity Asynchronous Multiplexer with Modulated Deterministic Input," ITC 13, Copenhagen, 1991.

[8] J. M. Hyman, A. A. Lazar, and G. Pacifici, "A Separation Principle between Scheduling and Admission Control for Broadband Switching," *IEEE JSAC*, **11**, 605–616, 1993.

[9] D. Mitra and J. A. Morrison, "Erlang Capacity and Uniform Approximations for Shared Unbuffered Resources," *IEEE/ACM Trans. Networking*, **2**, 558–570, 1994.

[10] H. Saito, "Call Admission Control in an ATM Network Using Upper Bound of Cell Loss Probability," *IEEE Trans. Comm.*, **40**, 1512–1521, 1992.