Asymptotically Achievable Performance in ATM Networks

Sem Borst, Debasis Mitra

Bell Laboratories, Lucent Technologies

P.O. Box 636, Murray Hill, NJ 07974-0636

USA

Abstract

The primary objective in the present paper is to gain fundamental understanding of the achievable performance of ATM networks as a function of the various system characteristics. We derive limit theorems that characterize the achievable performance in terms of the offered traffic, the admissible region, and the revenue measure. The insights obtained allow for substantial simplifications in the design of real-time connection admission control algorithms. In particular, we describe how the boundaries of admissible regions with convex complements may be linearized – thus reducing the admissible region – so as to obtain a convenient loss network representation. The asymptotic results for the achievable performance suggest that the potential reduction in revenue is immaterial in high-capacity networks. Numerical experiments confirm that the actual reduction is typically negligible, even in networks of moderate capacity.

AMS 1991 subject classifications: primary 60K30; secondary 60K20, 68M20, 90B15, 90B22, 93E20.

Keywords & phrases: achievable performance, ATM networks, connection admission control, Markov decision processes, revenue maximization.

1 Introduction

Emerging ATM networks critically rely on connection admission control for maximizing bandwidth utilization, while satisfying certain quality of service guarantees. Major aspects of the quality of service requirements relate to performance criteria at the cell level, like delay characteristics or loss probabilities. Bandwidth utilization is a crucial yet not exclusive measure of global network performance. For the purpose of design and analysis, it is usually convenient to adopt a single proxy for network performance, called revenue, which may include several components.

An admission control algorithm operates at the connection or call level. The algorithm determines whether to accept or reject offered calls, depending on the number of calls of various classes in progress. The latter may conveniently be thought of as the network occupancy. In case of acceptance, the algorithm may also have to select a suitable route through the network. Clearly, the decision should comply with the quality of service guarantees, but is not based on that condition alone. To maximize revenues, the decision should also weigh the immediate earnings from accepting calls against the potential risk of lost earnings due to rejecting future calls.

Now suppose we could translate the quality of service requirements into an *admissible region*, representing the admissible network occupancies. A network occupancy is considered to be admissible, in case the quality of service requirements would be satisfied if the network maintained that occupancy permanently. This corresponds to the *conservative* approach in Reiman et al. [14], as opposed to the *aggressive* approach, in which the quality of service guarantees may be temporarily violated, as long as they remain satisfied over an infinite-time period. Also, assume we could express the revenue as a function of the network occupancy and the decisions taken by the admission control algorithm. Under exponentiality assumptions, the problem of determining an optimal algorithm may then be formulated in the framework of Markov decision processes. The Markov decision problem may, in principle, be solved numerically, using value iteration, policy iteration, or linear programming. Computationally, value iteration is considered to be the most efficient approach, especially when enhanced with features like dynamic relaxation factors to speed up the convergence.

Even then, however, the computation of an optimal policy is prohibitively demanding for all but the smallest networks.

The value iteration approach may, in principle, also be used for analyzing the structure of an optimal policy. In the case of a single link, which is offered calls with similar characteristics but possibly different revenues, the optimal strategy has a simple threshold structure, referred to as *trunk reservation*, see Lippman [11], Miller [12]. Usually, however, the optimal policy has a rather complicated structure. Even in the case of a single link, which is offered calls with different characteristics, the optimal policy may lack simple monotonicity properties, see Ross & Tsang [16]. In a network environment, the optimal strategy may clash with the necessity for a decentralized control mechanism. So, even if the computation of an optimal strategy was feasible, the implementation would be most impractical.

Considering the above, we prefer policies that are not necessarily theoretically optimal, but easy to compute and simple to implement. In constructing such policies, knowledge of the achievable performance is of critical importance. The present paper provides an explicit characterization of the achievable performance in terms of the offered traffic, the admissible region, and the revenue measure. The primary motivations are twofold: (i) understanding of what is *definitely not* achievable may provide a useful measure of the quality of heuristic strategies; and (ii) insight in what *is* achievable, and *how*, may guide the construction of efficient strategies. Our goal in the present paper is *not* to derive numerical approximations for the achievable performance, *nor* to make a methodological contribution to the field of probability theory. Our primary focus is rather to gain fundamental understanding of the achievable performance of ATM networks as a function of the various system characteristics, and to use those insights to obtain substantial simplifications in the complex problem of designing efficient real-time connection admission control algorithms.

So far, we have not made any assumptions on the admissible region. In many important cases, the admissible region S may be fairly accurately described by a set of linear constraints, i.e.,

$$\sum_m b_{lm} x_m \leq C_l, \qquad l = 1, \ldots, L, \tag{1}$$

with x_m denoting the number of class-m calls. (The index set $l = 1, \ldots, L$ may or may not correspond to the set of links in the network, see Remark 1.1 below.) For example, single ATM nodes endowed with large multiplexing buffers give rise to admissible regions with linear boundaries. See, for instance, Gibbens & Hunt [7] and Elwalid & Mitra [3]. If the l-th constraint in (1) is associated with a single link of capacity C_l, the coefficient b_{lm} may be interpreted as the *effective bandwidth* per class-m call on the l-th link.

Remark 1.1

In general, there need not be a one-to-one correspondence between links and constraints. Elwalid & Mitra [4] analyze a multi-service multiplexing system, in which separate buffers are dedicated to each service, and these have prioritized access to an outgoing channel. For example, one class may have stringent delay requirements, and hence have small buffers and high priority access to the channel, while another class with much less stringent delay requirements may be given large buffers and lower priority. A related priority system without buffers is studied by Bean [1]. In those cases, a single link may generate several constraints, as illustrated in Figure 1. There are also cases conceivable, in which several links in combination may induce a single constraint.

$\square$

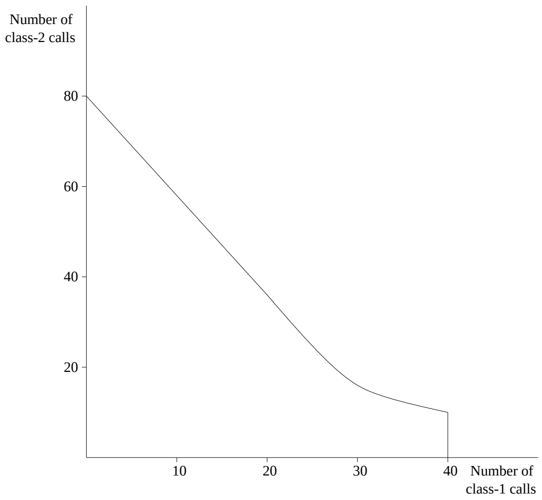

Figure 1: Admissible region for priority system.

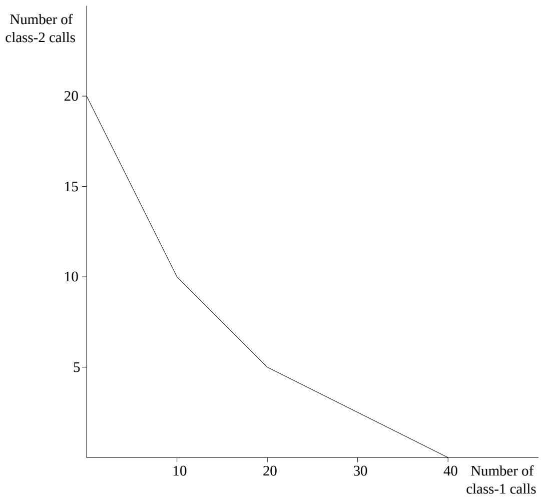

Figure 2: Admissible region with piece-wise linear boundary.

Note that, without any essential loss of generality, we may assume the coefficients C_l and b_{lm} in (1) to be integers. Thus, interpreting b_{lm} as the number of circuits required per class-m call on a link l comprising C_l circuits, we obtain a multi-service loss network representation. The conceptual and computational advantages of that framework for ATM network design have been recognized and are well established. See, for instance, Faragó et al. [6], Hui [8], Kelly [9], Mitra et al. [13], and Ross [15].

For cases where some of the constraints may be concave, i.e.,

$$d_l(x) \leq C_l, \qquad l = 1, \ldots, L, \tag{2}$$

with $d_l(\cdot)$ denoting some concave function, Hui [8] and Kelly [9] propose the procedure of bounding the admissible region by a tangent hyperplane passing through an operating point on the boundary. There are important cases, however, where the boundary may be highly non-linear, so that it is not immediately clear if linearizing the boundary is not overly conservative. For example, in the situation of Figure 1, the admissible region may be substantially diminished if the curve in the boundary (see Weiss [17]) is to be eliminated. Highly non-linear boundaries may also arise as a result of multiplexing calls with widely different characteristics. Recently, Elwalid et al. [5] have developed the concept of a *critical* link capacity. Depending on whether the link capacity exceeds a certain class-dependent threshold or not, calls may behave as statistically multiplexable or non-multiplexable, see also Boorstyn et al. [2]. When calls from statistically multiplexable and non-multiplexable classes are being intermixed on a single link, the boundary of the admissible region is no longer approximately linear. Elwalid et al. [5] show that the boundary is at best approximately *piece-wise* linear, with a convex complement, as exemplified in Figure 2.

To examine the impact of the above, we focus in the final section of the paper on the case that the admissible region for the network may be characterized as the intersection of several regions, each with a convex complement, the boundaries possibly being highly non-linear. We describe how the boundaries may be linearized – thus reducing the admissible region – so as to obtain a circuit-switched loss network representation. Asymptotic results for the achievable performance suggest that the potential reduction in revenue is immaterial in high-capacity networks. Numerical experiments confirm that the actual reduction is typically negligible, even in networks of moderate capacity.

The remainder of the paper is organized as follows. In Section 2, we present a detailed model description, and then proceed to derive an upper bound on the achievable performance. Section 3 characterizes the achievable performance in a sequence of models, in which the offered traffic and the admissible region both grow large, in proportion to one another. Section 4 describes how the boundaries of admissible regions with convex complements may be linearized without any significant reduction in the potential revenue. We present the results of the numerical experiments in Section 5. In Section 6, we make some concluding remarks.

2 Model description

We consider a network of arbitrary topology, which is offered calls from K classes. Class-k calls arrive as a Poisson process of rate λ_k, and have independent exponentially distributed holding times with mean $1/\mu_k$. Subject to admissibility of the network occupancy, class-k calls may be carried on any route from a given set $\mathcal{R}_k$. For conciseness, a class-route pair (k,m), $m \in \mathcal{R}_k$, will be called a *stream*. Denote by $\rho_k := \lambda_k/\mu_k$ the offered traffic from class-k calls. Define the 'offered-traffic' region as $T := \{x \in \mathbb{R}_+^{\mathcal{I}} : \sum\limits_{m \in \mathcal{R}_k} x_{(k,m)} \leq \rho_k, k = 1, \dots, K\}$, with the set $\mathcal{I}$ indexing the streams. The admissible region $P \subseteq \mathbb{N}^{\mathcal{I}}$ represents the set of admissible network occupancies. When the network occupancy is $x \in P$, revenue is generated at rate $r(x)$.

Remark 2.1

Note that revenue is accrued gradually over time rather than earned instantaneously upon call arrival or departure. Later, we will focus on the case that $r(\cdot)$ is linear, i.e., gradual rewards at rate $\sum\limits_{(k,m) \in \mathcal{I}} c_{(k,m)} x_{(k,m)}$ when the network occupancy is $x \in P$. Little's law shows that – as far as the mean reward rate is concerned – that is equivalent to instantaneous rewards $c_{(k,m)}/\mu_k$ for each class-k call carried on route m. In general, there need not be such a correspondence between gradual and instantaneous revenues.

$\square$

The admission control policy determines whether to accept or reject offered calls, depending on the network occupancy. In case of acceptance, the policy also selects a suitable route through the network. Call acceptance is always subject to admissibility of the network occupancy. Once admitted, a call is carried for the complete duration of the holding time. Define Π as the set of feasible policies. Denote by $\mathbf{X}^\pi$ a stochastic vector representing the network occupancy for policy $\pi \in \Pi$. Denote by $r_{\max} := \max_{\pi \in \Pi} \mathrm{E}r(\mathbf{X}^\pi)$ the maximum achievable mean reward rate. Denote by $\mathbf{Y}^\pi$ a stochastic vector representing the *aggregated* network occupancy for policy $\pi \in \Pi$, i.e., $\Pr\{\mathbf{Y}^\pi = y\} = \Pr\{\sum_{m \in \mathcal{R}_k} \mathbf{X}^\pi_{(k,m)} = y_k, k = 1, \ldots, K\}$.

We now derive an upper bound for $r_{\max}$. A rather crude bound is

$$r_{\max} \leq \max_{x \in P} r(x),$$

saying that we cannot earn anything more than if we were able to permanently maintain the most rewarding admissible network occupancy.
To obtain a tighter bound, we need not only consider what is admissible, but also what is achievable. Denote by $Q := \bigcup_{\pi \in \Pi} \{\mathrm{E}\mathbf{X}^\pi\}$ the set of achievable mean network occupancies. For conciseness, Q is referred to as the achievable region. For any set $A \subseteq \mathbb{R}^{\mathcal{I}}$, denote by $\mathrm{conv}(A)$ the convex hull of A.

Lemma 2.1

Suppose there is a concave function $\bar{r}(\cdot)$, defined on $\mathrm{conv}(P)$, such that $r(x) = \bar{r}(x)$ for all $x \in P$. Then

$$r_{\max} \leq \max_{x \in Q} \bar{r}(x),$$

with equality if $\bar{r}(\cdot)$ is linear.

Proof

By concavity of $\bar{r}(\cdot)$,

$$\mathrm{E}r(\mathbf{X}^\pi) = \sum_{x \in P} \bar{r}(x) \Pr\{\mathbf{X}^\pi = x\} \leq \bar{r}(\sum_{x \in P} x \Pr\{\mathbf{X}^\pi = x\}) = \bar{r}(\mathrm{E}\mathbf{X}^\pi),$$

for any $\pi \in \Pi$. In particular,

$$r_{\max} = \max_{\pi \in \Pi} \mathrm{E} r(\mathbf{X}^\pi) \leq \max_{\pi \in \Pi} \bar{r}(\mathrm{E}\mathbf{X}^\pi) = \max_{x \in Q} \bar{r}(x).$$

□

The above lemma says that we cannot earn anything more than if we were able to permanently maintain the most rewarding achievable mean link occupancy. We now proceed to determine an upper bound for the achievable region Q, so as to establish an upper bound for $r_{\max}$.

Lemma 2.2

$$Q \subseteq \mathrm{conv}(P) \cap T.$$

Proof

Assume $z \in Q$. By definition, there is a policy $\pi \in \Pi$ such that $\mathrm{E}\mathbf{X}^\pi = z$. Note that $\mathrm{E}\mathbf{X}^\pi = \sum_{x \in P} x \Pr\{\mathbf{X}^\pi = x\}$, so $z \in \mathrm{conv}(P)$. Also, $\sum_{m \in \mathcal{R}_k} z_{(k,m)} = \mathrm{E}\mathbf{Y}_k^\pi$ for all $k = 1, \ldots, K$. By Little's law, $\mathrm{E}\mathbf{Y}_k^\pi = \alpha_k \rho_k$, with α_k the fraction of class-k calls accepted, so $z \in T$. Hence, $z \in \mathrm{conv}(P) \cap T$.

□

Combining Lemmas 2.1 and 2.2 leads to the following corollary.

Corollary 2.1

Suppose there is a concave function $\bar{r}(\cdot)$, defined on $\mathrm{conv}(P)$, such that $r(x) = \bar{r}(x)$ for all $x \in P$. Then

$$r_{\max} \leq \max_{x \in \mathrm{conv}(P) \cap T} \bar{r}(x).$$

Remark 2.2

To illustrate the above corollary, consider the case where the reward function $r(x) = \sum_{(k,m) \in \mathcal{I}} c_{(k,m)} x_{(k,m)}$ is linear, and the admissible region $P = \{x \in \mathbb{N}^{\mathcal{I}} : \sum_{(k,m) \in \mathcal{I}} b_{l,(k,m)} x_{(k,m)} \leq$

$C_l, l = 1, \ldots, L\}$ is a polytope. The solution y^* to the following linear programming problem then provides an upper bound for $r_{\max}$:

$$\begin{array}{lll}
\max & y = \sum_{(k,m)\in\mathcal{I}} c_{(k,m)} x_{(k,m)} & \\
\text{sub} & \sum_{(k,m)\in\mathcal{I}} b_{l,(k,m)} x_{(k,m)} \le C_l, \qquad l = 1, \ldots, L; & \\
& \sum_{m\in\mathcal{R}_k} x_{(k,m)} \le \rho_k, \qquad k = 1, \ldots, K; & \\
& x_{(k,m)} \ge 0, \qquad (k,m) \in \mathcal{I}, &
\end{array}$$

where $x_{(k,m)}$ may be interpreted as the mean number of class-k class carried on route $m \in \mathcal{R}_k$.

□

To gain further insight into the achievable mean reward rate, it helps considerably to focus on high-capacity networks. That is the topic of the next section.

3 The asymptotically achievable performance

We now consider a sequence of models, indexed by n, in which the arrival rates and the admissible region are both scaled up, while the holding times, rewards, and routing sets are kept constant. In the n-th model, the arrival rate of class-k calls is $\lambda_k^{(n)} := n\lambda_k$. The admissible region is $P^{(n)} := nS \cap \mathbb{N}^{\mathcal{I}}$ for some compact monotone set $S \subseteq \mathbb{R}^{\mathcal{I}}$, i.e., a call completion always leaves the network occupancy admissible. (Throughout the paper, we use the short-hand notation $\kappa A := \bigcup_{x\in A} \{\kappa x\}$, for any scalar $\kappa \in \mathbb{R}$ and set $A \subseteq \mathbb{R}^{\mathcal{I}}$.) The reward function is $r^{(n)}(x) := nu(x/n)$ for some continuous function $u(\cdot)$ defined on S. Define $\Pi^{(n)}$ as the set of feasible policies in the n-th model.

Define by $\mathbf{Z}^\pi := \frac{1}{n}\mathbf{X}^\pi$ a stochastic vector representing the normalized network occupancy for policy $\pi \in \Pi^{(n)}$. The normalized mean reward rate for policy $\pi \in \Pi^{(n)}$ is $\frac{1}{n}\mathrm{E}r^{(n)}(\mathbf{X}^\pi) = \mathrm{E}u(\mathbf{Z}^\pi)$. For any point $x \in \mathbb{R}^{\mathcal{I}}$, let $B_{\alpha,\epsilon}^{(n)}(x)$ be the sphere in $\mathbb{R}^{\mathcal{I}}$ with center x and radius ϵn^α. We use the short-hand notation $B_{\alpha,\epsilon}^{(n)}(A) := \bigcup_{x\in A} B_{\alpha,\epsilon}^{(n)}(x)$, for any set $A \subseteq \mathbb{R}^{\mathcal{I}}$.

We first state an auxiliary lemma.

Lemma 3.1

Let $\mathbf{Y}^{(n)}$ be the number of busy servers in an infinite-server system with a Poisson arrival process of rate $n\lambda$ and service time distribution $B(\cdot)$ with first moment β and finite second moment $\beta^{(2)}$. Then

$$\Pr\{\mathbf{Y}^{(n)} \geq n\lambda\beta + \epsilon n^{\alpha+1}\} \leq \frac{\lambda\beta + \epsilon n^{\alpha}}{\lambda\beta + \epsilon n^{\alpha} + \epsilon^2 n^{2\alpha+1}},$$

for any $\epsilon > 0$ and $n > 0$.

Proof

See Appendix A.

$\square$

Theorem 3.1

$$\min_{\pi \in \Pi^{(n)}} \Pr\{\mathbf{Z}^{\pi} \in B_{\alpha,\epsilon}^{(n)}(S \cap T)\} \geq 1 - \sum_{k=1}^{K} \frac{\rho_k + \epsilon n^{\alpha}/\sqrt{K}}{\rho_k + \epsilon n^{\alpha}/\sqrt{K} + \epsilon^2 n^{2\alpha+1}/K},$$

for any $\epsilon > 0$ and $n > 0$.

Proof

Note that $B_{\alpha,\epsilon}^{(n)}(S \cap T) \cap S = B_{\alpha,\epsilon}^{(n)}(T) \cap S$, so

$$\Pr\{\mathbf{Z}^{\pi} \in B_{\alpha,\epsilon}^{(n)}(S \cap T)\} = \Pr\{\mathbf{Z}^{\pi} \in B_{\alpha,\epsilon}^{(n)}(T)\}. \tag{3}$$

Define $T_{\alpha,\epsilon}^{(n)} := \prod_{k=1}^{K} [0, n\rho_k + \epsilon n^{\alpha+1}/\sqrt{K}]$. It is easily verified that if $y \in T_{\alpha,\epsilon}^{(n)}$, with $y_k = \sum_{m \in \mathcal{R}_k} x_{(k,m)}$, $x_{(k,m)} \geq 0$ for all $(k,m) \in \mathcal{I}$, then $\frac{x}{n} \in B_{\alpha,\epsilon}^{(n)}(T)$, so

$$\Pr\{\mathbf{Z}^{\pi} \in B_{\alpha,\epsilon}^{(n)}(T)\} \geq \Pr\{\mathbf{Y}^{\pi} \in T_{\alpha,\epsilon}^{(n)}\}. \tag{4}$$

Note that

$$\Pr\{\mathbf{Y}^{\pi} \in T_{\alpha,\epsilon}^{(n)}\} \geq 1 - \sum_{k=1}^{K} \Pr\{\mathbf{Y}_k^{\pi} \geq n\rho_k + \epsilon n^{\alpha+1}/\sqrt{K}\}. \tag{5}$$

Denote by $\mathbf{Y}^{(n)}$ a stochastic vector representing the aggregated network occupancy in the n-th model in the hypothetical situation in which all offered calls could be accommodated. Evidently, $\mathbf{Y}_k^{(n)}$ is stochastically larger than $\mathbf{Y}_k^{\pi}$ for all $\pi \in \Pi^{(n)}$, $k = 1, \ldots, K$, in particular,

$$\Pr\{\mathbf{Y}_k^{\pi} \geq n\rho_k + \epsilon n^{\alpha+1}/\sqrt{K}\} \leq \Pr\{\mathbf{Y}_k^{(n)} \geq n\rho_k + \epsilon n^{\alpha+1}/\sqrt{K}\}, \tag{6}$$

for all $\pi \in \Pi^{(n)}$, $k = 1, \ldots, K$.

Observe that $\mathbf{Y}_k^{(n)}$ has the distribution of the number of busy servers in an infinite-server system with a Poisson arrival process of rate $n\lambda_k$ and exponentially distributed service times with mean $1/\mu_k$. Hence, Lemma 3.1 implies

$$\Pr\{\mathbf{Y}_k^{(n)} \geq n\rho_k + \epsilon n^{\alpha+1}/\sqrt{K}\} \leq \frac{\rho_k + \epsilon n^{\alpha}/\sqrt{K}}{\rho_k + \epsilon n^{\alpha}/\sqrt{K} + \epsilon^2 n^{2\alpha+1}/K}, \tag{7}$$

for all $k = 1, \ldots, K$. Combining (3)-(7) completes the proof.

□

The next corollary captures the essential information from Theorem 3.1.

Corollary 3.1

$$\lim_{n\to\infty} \min_{\pi \in \Pi^{(n)}} \Pr\{\mathbf{Z}^{\pi} \in B_{\alpha,\epsilon}^{(n)}(S \cap T)\} = 1,$$

for any $\alpha > -1/2$ and $\epsilon > 0$.

Note that for any $\alpha < 0$ the set $B_{\alpha,\epsilon}^{(n)}(S \cap T)$ 'shrinks' to the set $S \cap T$ as $n \to \infty$. So, asymptotically, for all strategies, the normalized network occupancy vector will be in or arbitrarily close to $S \cap T$ all the time.

We now proceed with an auxiliary lemma.

Lemma 3.2

Let $\mathbf{Y}^{(n)}$ be the number of busy servers in an $C^{(n)}$-server loss system with a Poisson arrival process of rate $n\lambda$ and service time distribution $B(\cdot)$ with first moment β and finite second moment $\beta^{(2)}$. If $C^{(n)} \leq n\lambda\beta$, then

$$\Pr\{\mathbf{Y}^{(n)} \leq C^{(n)} + m - \epsilon n^{\alpha+1}\} \leq \frac{\lambda\beta}{\lambda\beta + m^2 n^{-1} - 2m\epsilon n^{\alpha} + \epsilon^2 n^{2\alpha+1}},$$

for any $\epsilon > 0$, $n > 0$, and $m \leq \epsilon n^{\alpha+1}$.

Proof

See Appendix B.

□

Theorem 3.2

For any point $x^* \in S \cap T$, $\epsilon > 0$, *and* $n > 0$ *with* $r_k := |\mathcal{R}_k| \le \epsilon n^{\alpha+1}$,

$$\max_{\pi \in \Pi^{(n)}} \Pr\{\mathbf{Z}^\pi \in B^{(n)}_{\alpha,\epsilon}(x^*)\} \ge 1 - \sum_{k=1}^{K} \frac{\rho_k}{\rho_k + r_k^2 n^{-1} - 2r_k \epsilon n^\alpha/\sqrt{K} + \epsilon^2 n^{2\alpha+1}/K}.$$

Proof

Define policy $\pi^{(n)}_{x^*} \in \Pi^{(n)}$ as follows. Suppose the network occupancy is $x \in P^{(n)}$ when a class-k call arrives. Policy $\pi^{(n)}_{x^*}$ accepts it onto route $p \in \mathcal{R}_k$ if $x_{(k,m)} = \lfloor n x^*_{(k,m)} \rfloor$ for all $m = 1, \ldots, p-1$ and $x_{(k,p)} < \lfloor n x^*_{(k,p)} \rfloor$, and rejects it if $x_{(k,m)} = \lfloor n x^*_{(k,m)} \rfloor$ for all $m \in \mathcal{R}_k$. Define $U^{(n)}_{\alpha,\epsilon} := \prod_{k=1}^{K} [\gamma^{(n)}_k - \epsilon n^{\alpha+1}/\sqrt{K}, \infty)$, with $\gamma^{(n)}_k := n \sum_{m \in \mathcal{R}_k} x^*_{(k,m)}$. It is easily verified that if $y \in U^{(n)}_{\alpha,\epsilon}$ with $y_k = \sum_{m \in \mathcal{R}_k} x_{(k,m)}$, $0 \le x_{(k,m)} \le \lfloor n x^*_{(k,m)} \rfloor$ for all $(k,m) \in \mathcal{I}$, then $\frac{x}{n} \in B^{(n)}_{\alpha,\epsilon}(x^*)$, so

$$\Pr\{\mathbf{Z}^{\pi^{(n)}_{x^*}} \in B^{(n)}_{\alpha,\epsilon}(x^*)\} \ge \Pr\{\mathbf{Y}^{\pi^{(n)}_{x^*}} \in U^{(n)}_{\alpha,\epsilon}\}. \tag{8}$$

Note that

$$\Pr\{\mathbf{Y}^{\pi^{(n)}_{x^*}} \in U^{(n)}_{\alpha,\epsilon}\} \ge 1 - \sum_{k=1}^{K} \Pr\{\mathbf{Y}_k^{\pi^{(n)}_{x^*}} \le \gamma^{(n)}_k - \epsilon n^{\alpha+1}/\sqrt{K}\}. \tag{9}$$

Define $C^{(n)}_k := \sum_{m \in \mathcal{R}_k} \lfloor n x^*_{(k,m)} \rfloor$. Observe that $\mathbf{Y}_k^{\pi^{(n)}_{x^*}}$ has the distribution of the number of busy servers in an $C^{(n)}_k$-server system with a Poisson arrival process of rate $n\lambda_k$ and exponentially distributed service times with mean $1/\mu_k$. Note that $C^{(n)}_k \le n\rho_k$ and $\gamma^{(n)}_k \le C^{(n)}_k + r_k$. Hence, it follows from Lemma 3.2 that

$$\Pr\{\mathbf{Y}_k^{\pi^{(n)}_{x^*}} \le \gamma^{(n)}_k - \epsilon n^{\alpha+1}/\sqrt{K}\} \le \frac{\rho_k}{\rho_k + r_k^2 n^{-1} - 2r_k \epsilon n^\alpha/\sqrt{K} + \epsilon^2 n^{2\alpha+1}/K}, \tag{10}$$

for all $k = 1, \ldots, K$. Combining (8)-(10) completes the proof.

□

Remark 3.1

We could also define policy $\pi^{(n)}_{x^*}$ as follows. Subject to admissibility of the network occupancy, policy $\pi^{(n)}_{x^*}$ accepts an arriving class-k call onto route $m \in \mathcal{R}_k$ with probability $x^*_{(k,m)}/\rho_k$, and rejects it with probability $1 - \sum_{m \in \mathcal{R}_k} x^*_{(k,m)}/\rho_k$.

□

The next corollary summarizes the crucial information from Theorem 3.2.

Corollary 3.2

For any point $x^* \in S \cap T$, $\alpha > -1/2$, and $\epsilon > 0$,

$$\lim_{n\to\infty} \max_{\pi \in \Pi^{(n)}} \Pr\{\mathbf{Z}^\pi \in B_{\alpha,\epsilon}^{(n)}(x^*)\} = 1.$$

Note that for any $\alpha < 0$, the set $B_{\alpha,\epsilon}^{(n)}(x^*)$ 'shrinks' to the point x^* as $n \to \infty$. So, asymptotically, any normalized network occupancy vector in $S \cap T$ can all the time be approximated arbitrarily close by some strategy.

We now consider a continuous mapping $f(\cdot)$ from S to $\mathbb{R}^M$, which may be thought of as describing the performance as a function of the normalized network occupancy. A performance vector $v \in \mathbb{R}^M$ is said to be asymptotically achievable, if there are strategies $\pi^{(1)}, \pi^{(2)}, \ldots$ with $\pi^{(n)} \in \Pi^{(n)}$, such that $v = \lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{\pi^{(n)}})$. Denote by $R(f)$ the set of asymptotically achievable performance vectors. For any set $A \subseteq S$, denote by $f(A) := \bigcup_{x\in A} \{f(x)\}$ the range of $f(\cdot)$ on A.

Theorem 3.3

$$R(f) \subseteq \mathrm{conv}(f(S \cap T)).$$

Proof

Assume $v \in R(f)$. By definition, there are strategies $\pi^{(1)}, \pi^{(2)}, \ldots$ with $\pi^{(n)} \in \Pi^{(n)}$, such that $\lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{\pi^{(n)}}) = v$. Corollary 3.1 says that $\lim_{n\to\infty} \Pr\{\mathbf{Z}^{\pi^{(n)}} \in B_{\alpha,\epsilon}^{(n)}(S \cap T)\} = 1$ for any $\alpha > -1/2$ and $\epsilon > 0$. The next lemma then implies that $v \in \mathrm{conv}(f(S \cap T))$.

□

Lemma 3.3

Let $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots$ be a sequence of stochastic variables on S, such that $\lim_{n\to\infty} \Pr\{\mathbf{Z}^{(n)} \in B_{\alpha,\epsilon}^{(n)}(A)\} = 1$ for some closed set $A \subseteq S$, with $\alpha < 0$, $\epsilon > 0$, and such that $v = \lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{(n)})$ exists. Then $v \in \mathrm{conv}(f(A))$.

Proof

See Appendix C.

□

Theorem 3.4

$$f(S \cap T) \subseteq R(f).$$

Proof

Assume $v \in f(S \cap T)$. By definition, there is an $x^* \in S \cap T$ such that $f(x^*) = v$. Corollary 3.2 says that there are strategies $\pi^{(1)}, \pi^{(2)}, \ldots$ with $\pi^{(n)} \in \Pi^{(n)}$, such that $\lim_{n\to\infty} \Pr\{\mathbf{Z}^{\pi^{(n)}} \in B^{(n)}_{\alpha,\epsilon}(x^*)\} = 1$ for any $\alpha > -1/2$ and $\epsilon > 0$. It then follows from the next lemma that $\lim_{n\to\infty} \mathrm{E}f(\mathbf{Z}^{\pi^{(n)}}) = f(x^*)$, i.e., $v \in R(f)$.

Lemma 3.4

Let $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots$ be a sequence of stochastic variables on S, such that $\lim_{n\to\infty} \Pr\{\mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(x^*)\} = 1$ for some point $x^* \in S$, with $\alpha < 0$ and $\epsilon > 0$. Then $\lim_{n\to\infty} \mathrm{E}f(\mathbf{Z}^{(n)}) = f(x^*)$.

Proof

See Appendix D.

□

We now particularize the above results to the asymptotically achievable normalized mean reward rate. A normalized mean reward rate r is said to be asymptotically achievable if there are strategies $\pi^{(1)}, \pi^{(2)}, \ldots$ with $\pi^{(n)} \in \Pi^{(n)}$, such that $\lim_{n\to\infty} \mathrm{E}u(\mathbf{Z}^{\pi^{(n)}}) = r$. Define $R(u)$ as the set of asymptotically achievable normalized mean reward rates. Combining Theorems 3.3 and 3.4 with $f(\cdot) = u(\cdot)$, and noting that $\mathrm{conv}(u(S \cap T)) = u(S \cap T)$ since $S \cap T$ is connected, gives the following corollary.

Corollary 3.3

$$R(u) = u(S \cap T).$$

Denote by $r_{\max}^{(n)}$ the maximum achievable normalized mean reward rate in the n-th model. Taking the maximum element of both sets in the above corollary, we obtain the following characterization of $\lim_{n\to\infty} r_{\max}^{(n)}$ as the solution to a mathematical programming problem.

Corollary 3.4

$$\lim_{n\to\infty} r_{\max}^{(n)} = \max_{x\in S\cap T} u(x).$$

4 Linearization of the boundary

We now focus on the case that the admissible region may be characterized as the *intersection of several regions, each with a convex complement*, the boundaries possibly being highly non-linear. We will describe how the boundaries may be linearized – thus reducing the admissible region – without any significant reduction in the potential revenue.

So suppose there are regions $S_1, \ldots, S_L$ in $\mathbb{R}_+^{\mathcal{I}}$ such that $S = \bigcap_{l=1}^{L} S_l$, with the complement $\mathbb{R}_+^{\mathcal{I}} \setminus S_l$ convex for all $l = 1, \ldots, L$. Denote by ∂S_l the boundary of S_l. Let the point $x^* \in S \cap T$ be such that $u(x^*) = \max_{x\in S\cap T} u(x)$. Corollary 3.4 says that the maximum asymptotically achievable normalized mean reward rate is given by $\lim_{n\to\infty} r_{\max}^{(n)} = u(x^*)$, *which characterizes x^* as the optimal normalized operating point.*

Since the complement $\mathbb{R}_+^{\mathcal{I}} \setminus S_l$ is convex, there is a hyperplane H_l tangent to ∂S_l, separating x^* from $\mathbb{R}_+^{\mathcal{I}} \setminus S_l$ (passing through x^* in case x^* lies on ∂S_l). In other words, there is a linear inequality such that $\sum_{\sigma\in\mathcal{I}} b_{l\sigma} x_\sigma \geq C_l$ for all $x \in \mathbb{R}_+^{\mathcal{I}} \setminus S_l$, with $C_l \geq \sum_{\sigma\in\mathcal{I}} b_{l\sigma} x_\sigma^*$, and $b_{l\sigma} \geq 0$ for all $\sigma \in \mathcal{I}$.

Define $\tilde{S}_l := \{x \in \mathbb{R}_+^{\mathcal{I}} : b_{l\sigma} x_\sigma \leq C_l \text{ for all } l = 1, \ldots, L\}$, $\tilde{S} := \bigcap_{l=1}^{L} \tilde{S}_l$, $\tilde{P}^{(n)} := n\tilde{S} \cap \mathbb{N}^{\mathcal{I}}$. By construction, $\tilde{P}^{(n)} \subseteq P^{(n)}$, i.e., the admissible region has been reduced, but still $x^* \in \tilde{S}$, i.e., the optimal operating point has been retained. Without any essential loss of generality, we may assume the coefficients C_l and $b_{l\sigma}$ for all $\sigma \in \mathcal{I}$ to be integers. Thus, we may interpret $\tilde{P}^{(n)}$ as the feasible region of a circuit-switched loss network, in which each call from stream σ requires $b_{l\sigma}$ circuits on link l, comprising nC_l circuits.

Now suppose we were able to construct a close-to-optimal policy $\tilde{\pi}^{(n)}$ for the n-th model in the sequence defined in Section 3, with the admissible region $P^{(n)}$ replaced by $\tilde{P}^{(n)}$.

Since $\tilde{P}^{(n)} \subseteq P^{(n)}$, policy $\tilde{\pi}^{(n)}$ is guaranteed to be feasible for the original model with admissible region $P^{(n)}$ as well. The inherent disadvantage is that there may be far better policies for the model with feasible region $P^{(n)}$, i.e., that $\mathrm{E}u(\mathbf{Z}^{\tilde{\pi}^{(n)}})$ could be far from $r_{\max}^{(n)}$. By assumption, however, $\tilde{\pi}^{(n)}$ is a close-to-optimal policy, i.e., $\mathrm{E}u(\mathbf{Z}^{\tilde{\pi}^{(n)}})$ is close to $\tilde{r}_{\max}^{(n)}$, the maximum achievable normalized mean reward rate for the model with admissible region $\tilde{P}^{(n)}$. So, to disprove that $\mathrm{E}u(\mathbf{Z}^{\tilde{\pi}^{(n)}})$ could be far from $r_{\max}^{(n)}$, it suffices to show that the difference $\Delta^{(n)} := r_{\max}^{(n)} - \tilde{r}_{\max}^{(n)}$ is not too big. Since $x^* \in \tilde{S}$, Corollary 3.4 says that $\lim_{n\to\infty} \tilde{r}_{\max}^{(n)} = u(x^*)$, which implies that $\Delta^{(n)}$ must be immaterial for n large enough, i.e., in high-capacity networks.

We have performed several numerical experiments, reported in the next section, to investigate how fast $\Delta^{(n)}$ tends to 0. The results provide strong evidence that $\Delta^{(n)}$ is negligible, even for relatively small n, i.e., in networks of moderate capacity.

Remark 4.1

In fact, $\lim_{n\to\infty} \tilde{r}_{\max}^{(n)} = u(x^*)$ for any region $\tilde{S}$ with $x^* \in \tilde{S}$. Instead, we could take for example $\tilde{S} = \{x \in \mathbb{R}_+^{\mathcal{I}} : x_\sigma \leq x_\sigma^* \text{ for all } \sigma \in \mathcal{I}\}$. However, the speed of convergence of $\tilde{r}_{\max}^{(n)}$ to $u(x^*)$, and hence of $\Delta^{(n)}$ to 0, is likely to increase with the size of $\tilde{S}$. The size of the polytope $\tilde{S}$ defined in the above paragraph is maximal, in the sense that it is not contained in any other polytope contained in S.

□

5 Numerical experiments

We now describe the numerical experiments we have conducted to examine how fast $\Delta^{(n)}$ converges to 0. We consider four different base cases, indexed I through IV, each with two streams, i.e., $\mathcal{I} = \{1, 2\}$. Each base case is defined in terms of the admissible region S, the reward function $u(\cdot)$, and the reciprocal-mean-holding time vector $\mu = (\mu_1, \mu_2)$. In each of the base cases, we consider four different scenarios for the arrival rate vector $\lambda := (\lambda_1, \lambda_2)$, indicated by A through D.

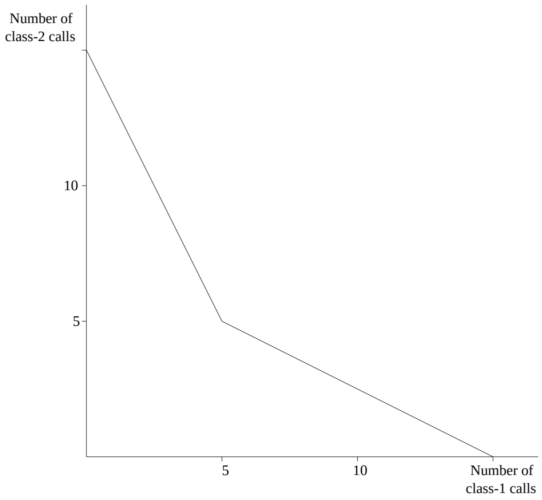

Figure 3: Admissible region for case I.

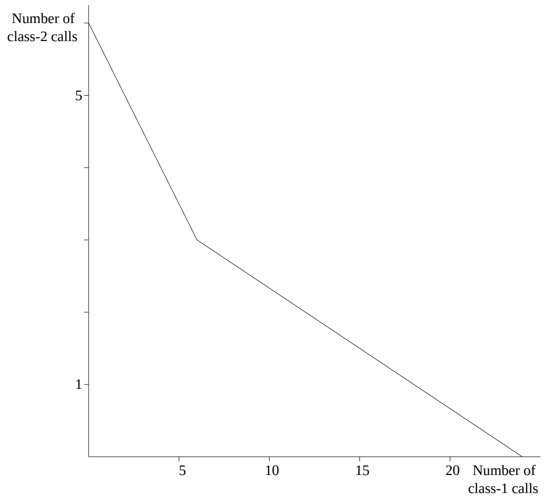

Figure 4: Admissible region for case II.

In case I, the admissible region is symmetric, with a piece-wise linear boundary, see Figure 3; the reward function is symmetric as well:

I. $\mu = (1.0, 1.0)$, $u(x) = x_1 + x_2$, $S = \{x \in \mathbb{R}_+^2 : x_1 + 2x_2 \leq 15 \vee 2x_1 + x_2 \leq 15\}$

A. $\lambda = (5.0, 5.0)$, $x^* = (5.0, 5.0)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : x_1 + 2x_2 \leq 15\}$

B. $\lambda = (7.5, 7.5)$, $x^* = (7.5, 3.75)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : x_1 + 2x_2 \leq 15\}$

C. $\lambda = (10.0, 2.5)$, $x^* = (10.0, 2.5)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : x_1 + 2x_2 \leq 15\}$

D. $\lambda = (15.0, 3.75)$, $x^* = (15.0, 0.0)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : x_1 + 2x_2 \leq 15\}$

For each of the four traffic scenarios, we computed $r_{\max}^{(n)}$ and $\tilde{r}_{\max}^{(n)}$ for n ranging from 1 to 10, using value iteration with $10,000n$ iterations. The results are tabulated below. The bottom row marked with $n = \infty$ shows $\lim_{n\to\infty} r_{\max}^{(n)} = \lim_{n\to\infty} \tilde{r}_{\max}^{(n)} = u(x^*)$.

I	A		B		C		D	
n	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$
1	8.2514	7.8950	9.6858	9.4648	10.2730	10.2730	12.3323	12.3323
2	8.7220	8.4994	10.1626	10.0892	10.9197	10.9197	13.0522	13.0522
3	8.9465	8.7706	10.3930	10.3541	11.2109	11.2109	13.3868	13.3868
4	9.0840	8.9332	10.5341	10.5173	11.3908	11.3908	13.5906	13.5906
5	9.1788	9.0446	10.6350	10.6257	11.5129	11.5129	13.7322	13.7322
6	9.2491	9.1270	10.7063	10.7010	11.6023	11.6023	13.8376	13.8376
7	9.3040	9.1912	10.7597	10.7570	11.6716	11.6716	13.9202	13.9202
8	9.3483	9.2430	10.8058	10.8044	11.7282	11.7282	13.9872	13.9872
9	9.3852	9.2869	10.8441	10.8433	11.7754	11.7754	14.0929	14.0429
10	9.4164	9.3246	10.8754	10.8749	11.8152	11.8152	14.0903	14.0903
∞	10.0000		11.2500		12.000		14.2500	

Table 1: Numerical results for case I.

In case II, the admissible region is no longer symmetric, but still has a piece-wise linear boundary, see Figure 4; the reward function is now asymmetric also:

II. $\mu = (1.0, 1.0)$, $u(x) = x_1 + 4x_2$, $S = \{x \in \mathbb{R}_+^2 : x_1 + 6x_2 \leq 24 \vee 2x_1 + 4x_2 \leq 24\}$

A. $\lambda = (6.0, 3.0)$, $x^* = (6.0, 3.0)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : 2x_1 + 4x_2 \leq 24\}$

B. $\lambda = (9.0, 4.5)$, $x^* = (3.0, 4.5)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : 2x_1 + 4x_2 \leq 24\}$

C. $\lambda = (12.0, 2.0)$, $x^* = (12.0, 2.0)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : x_1 + 6x_2 \leq 24\}$

D. $\lambda = (18.0, 3.0)$, $x^* = (18.0, 1.0)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : x_1 + 6x_2 \leq 24\}$

Note that the revenues are assumed to be proportional to the capacity requirements for both streams when carried in isolation.

II	A		B		C		D	
n	$r^{(n)}_{\max}$	$\tilde{r}^{(n)}_{\max}$	$r^{(n)}_{\max}$	$\tilde{r}^{(n)}_{\max}$	$r^{(n)}_{\max}$	$\tilde{r}^{(n)}_{\max}$	$r^{(n)}_{\max}$	$\tilde{r}^{(n)}_{\max}$
1	13.4986	13.4280	15.8065	15.7676	15.3558	15.3504	18.4889	18.4889
2	14.8290	14.7672	17.3392	17.3361	16.6018	16.6015	19.6427	19.6427
3	15.4193	15.3687	18.0960	18.0952	17.1833	17.1833	20.1594	20.1594
4	15.7808	15.7508	18.5368	18.5365	17.5388	17.5388	20.4655	20.4655
5	16.0368	16.0143	18.8543	18.8542	17.7859	17.7859	20.6713	20.6713
6	16.2237	16.2054	19.0815	19.0815	17.9715	17.9715	20.8217	20.8217
7	16.3685	16.3518	19.2564	19.2563	18.1177	18.1177	20.9365	20.9365
8	16.4839	16.4684	19.4004	19.4003	18.2362	18.2362	21.0282	21.0282
9	16.5786	16.5641	19.5189	19.5189	18.3347	18.3347	21.1032	21.1032
10	16.6580	16.6443	19.6157	19.6157	18.4183	18.4183	21.1656	21.1656
∞	18.0000		21.0000		20.0000		22.0000	

Table 2: Numerical results for case II.

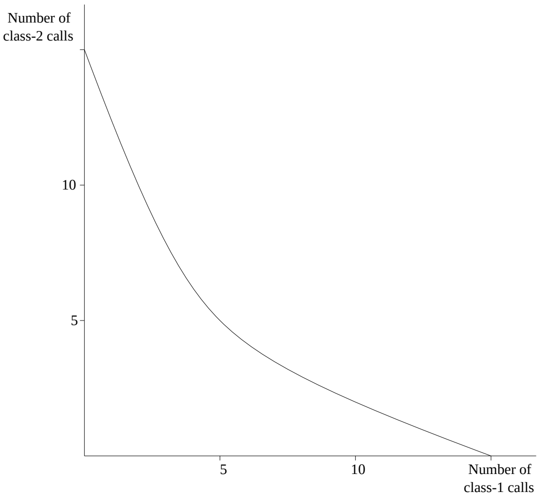

Figure 5: Admissible region for case III.

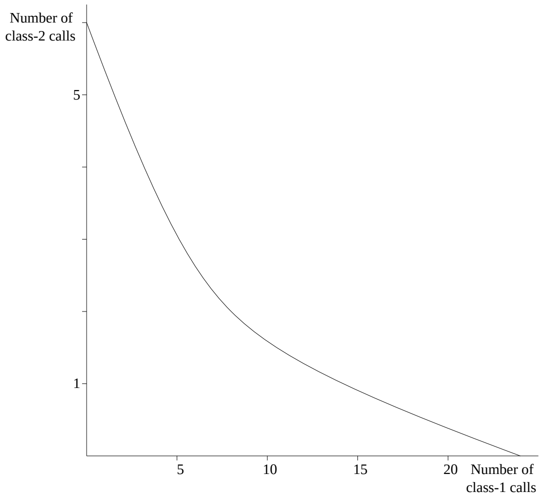

Figure 6: Admissible region for case IV.

In case III, the admissible region is symmetric, with a spheric boundary, see Figure 5; the reward function is symmetric as well:

III. $\mu = (1.0, 1.0)$, $u(x) = x_1 + x_2$, $S = \{x \in \mathbb{R}^2_+ : x_1, x_2 \leq 15, (x_1 - 17.5)^2 + (x_1 - 17.5)^2 \geq 312.5\}$

A. $\lambda = (5.0, 5.0)$, $x^* = (5.0, 5.0)$, $\tilde{S} = \{x \in \mathbb{R}^2_+ : x_1 + x_2 \leq 10\}$

B. $\lambda = (7.5, 7.5)$, $x^* = (7.5, 17.5 - 2.5\sqrt{34})$, $\tilde{S} = \{x \in \mathbb{R}^2_+ : 4x_1 + \sqrt{34}x_2 \leq 17.5\sqrt{34} - 55\}$

C. $\lambda = (10.0, 1.5)$, $x^* = (10.0, 17.5 - 2.5\sqrt{41})$, $\tilde{S} = \{x \in \mathbb{R}^2_+ : 3x_1 + \sqrt{41}x_2 \leq 17.5\sqrt{41} - 72.5\}$

D. $\lambda = (15.0, 2.25)$, $x^* = (15.0, 0.0)$, $\tilde{S} = \{x \in \mathbb{R}^2_+ : x_1 + 7x_2 \leq 15\}$

III	A		B		C		D	
n	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$
1	7.8982	7.8541	8.9672	8.3297	9.7010	9.3254	12.2951	12.2951
2	8.4210	8.4111	9.3428	9.0877	10.1279	9.9370	13.0131	13.0131
3	8.6822	8.6754	9.5422	9.4112	10.3561	10.2424	13.3497	13.3497
4	8.8434	8.8384	9.6700	9.5893	10.5031	10.4119	13.5560	13.5560
5	8.9554	8.9521	9.7442	9.6880	10.6119	10.5296	13.6993	13.6993
6	9.0397	9.0373	9.8176	9.7765	10.6828	10.6238	13.8064	13.8064
7	9.1074	9.1043	9.8820	9.8415	10.7497	10.6919	13.8905	13.8905
8	9.1610	9.1588	9.9314	9.8967	10.7990	10.7505	13.9587	13.9587
9	9.2060	9.2043	9.9657	9.9339	10.8424	10.7957	14.0156	14.0156
10	9.2449	9.2430	9.9988	9.9681	10.8840	10.8352	14.0640	14.0640
∞	10.0000		10.4226		11.2829		15.0000	

Table 3: Numerical results for case III.

In case IV, the admissible region is asymmetric, with an elliptic boundary, see Figure 6; the reward function is asymmetric also:

IV. $\mu = (1.0, 1.0)$, $u(x) = x_1 + 4x_2$, $S = \{x \in \mathbb{R}_+^2 : x_1 \leq 24, x_2 \leq 6, (x_1 - 28)^2 + 16(x_2 - 7)^2 \geq 800\}$

A. $\lambda = (8.0, 2.0)$, $x^* = (8.0, 2.0)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : x_1 + 4x_2 \leq 16\}$

B. $\lambda = (12.0, 3.0)$, $x^* = (12.0, 7 - \sqrt{34})$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : x_1 + \sqrt{34}x_2 \leq 7\sqrt{34} - 22\}$

C. $\lambda = (4.0, 3.5)$, $x^* = (28 - 2\sqrt{151}, 3.5)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : \sqrt{151}x_1 + 28x_2 \leq 28\sqrt{151} - 204\}$

D. $\lambda = (6.0, 5.25)$, $x^* = (28 - \sqrt{751}, 5.25)$, $\tilde{S} = \{x \in \mathbb{R}_+^2 : \sqrt{751}x_1 + 28x_2 \leq 28\sqrt{751} - 604\}$

IV	A		B		C		D	
n	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$	$r_{\max}^{(n)}$	$\tilde{r}_{\max}^{(n)}$
1	11.6564	11.5622	13.7293	13.0807	12.8553	12.2404	16.5777	14.5988
2	12.7798	12.7462	14.6167	14.3809	14.0389	13.6420	18.0780	17.1216
3	13.3211	13.2992	15.0757	14.9220	14.6857	14.3622	18.7900	18.1812
4	13.6479	13.6377	15.3263	15.2332	15.0690	14.8209	19.2246	18.7920
5	13.8841	13.8726	15.5097	15.4383	15.3345	15.1204	19.5233	19.1981
6	14.0601	14.0482	15.6403	15.5871	15.5512	15.3403	19.7437	19.3650
7	14.1944	14.1858	15.7508	15.6995	15.6942	15.5170	19.9135	19.4530
8	14.3026	14.2976	15.8205	15.7877	15.8406	15.6619	20.0494	19.6684
9	14.3965	14.3907	15.8906	15.8603	15.9262	15.7934	20.1611	19.8424
10	14.4757	14.4699	15.9496	15.9202	16.0096	15.8897	20.2528	19.9857
∞	16.0000		16.6762		17.4236		21.5956	

Table 4: Numerical results for case IV.

Note that the convergence of both $r_{\max}^{(n)}$ and $\tilde{r}_{\max}^{(n)}$ to $u(x^*)$ is quite slow. In all cases, however, the *difference* $\Delta^{(n)}$ vanishes fairly quickly, so that the reduction in revenue is insignificant, even if the admissible region is substantially diminished.

6 Conclusion

In the present paper we have characterized the achievable performance of ATM networks in terms of the offered traffic, the admissible region, and the revenue measure. We have described how the boundaries of admissible regions with convex complements may be linearized – thus reducing the admissible region – so as to obtain a convenient loss network representation.

As mentioned earlier, knowledge of the achievable performance is of critical importance in constructing admission control policies. The knowledge should be used with some caution though. Suppose we have some method of constructing admission control policies, which was to produce some policy $\pi^{(n)}$ for the n-th model in the sequence defined in Section 3. Corollary 3.4 says that the maximum asymptotically achievable mean normalized reward rate is given by $u(x^*) = \max_{x \in S \cap T} u(x)$. Clearly, if $\mathrm{E}u(\mathbf{Z}^{\pi^{(n)}})$ does not converge to $u(x^*)$, the method should be regarded with some suspicion. However, even if $\mathrm{E}u(\mathbf{Z}^{\pi^{(n)}})$ *does* tend to $u(x^*)$, the method is not necessarily adequate, since the convergence may be quite slow.

To illustrate this, in circuit-switched loss networks, trunk reservation policies are known to perform very well, see for instance Key [10]. A trunk reservation policy is specified through parameters $t_1, \ldots, t_K$, with the interpretation that a class-k call, requiring b_k circuits, is accepted only if the number of free circuits is at least $b_k + t_k$. Although asymptotic optimality may be harder to establish due to the lack of a product-form solution, trunk reservation policies typically outperform the complete-partitioning and probabilistic-filtering policies introduced in the proof of Theorem 3.2 and Remark 3.1, respectively. Under specific assumptions, the optimal policy may be shown to belong to the class of trunk reservation policies. Even if these assumptions are relaxed, trunk reservation policies are typically found to perform very well. In addition, trunk reservation policies are known to be extremely robust with regard to perturbations in the offered traffic, which makes them less vulnerable to inaccurate traffic forecasts.

It is not immediately clear how trunk reservation policies need to be modified when leaving the territory of circuit-switched networks. As we have demonstrated, however, in those cases we may simply transfer admission control problems to that domain without any significant reduction in the potential revenue.

Acknowledgment

The first author would like to thank Bill Massey, Marty Reiman, and Alan Weiss for several helpful discussions.

References

[1] Bean, N.G. (1993). Effective bandwidths with different quality of service requirements. In: *Integrated Broadband Communication Networks and Services*, ed. V.B. Iversen, IFIP, 241-252.

[2] Boorstyn, R.R., Suh, J.-K., Kershenbaum, A., Chennikara, J. (1995). An algorithm for call acceptance based on characterization of sources. Preprint Polytechnic University, Brooklyn, New York.

[3] Elwalid, A., Mitra, D. (1993). Effective bandwidth of general Markovian traffic sources and admission control of high speed networks. IEEE/ACM *Trans. Netw.* **1**, 329-343.

[4] Elwalid, A., Mitra, D. (1995). Analysis, approximations and admission control of a multi-service multiplexing system with priorities. In: *Proc. INFOCOM '95, Boston MA*, 463-472.

[5] Elwalid, A., Mitra, D., Wentworth, R.H. (1995). A new approach for allocating buffers and bandwidth to heterogeneous, regulated traffic in an ATM node. IEEE *J. Sel. Ar. Commun.* **13**, *Special Issue on the Fundamentals of Networking*, 1115-1127.

[6] Faragó, A., Blaabjerg, S., Ast, L., Gordos, G., Henk, T. (1995). A new degree of freedom in ATM network dimensioning: optimizing the local configuration. IEEE *J. Sel. Ar. Commun.* **13**, *Special Issue on the Fundamentals of Networking*, 1199-1206.

[7] Gibbens, R.J., Hunt, P.J. (1991). Effective bandwidths for the multi-type UAS channel. *Queueing Systems* **9**, 17-28.

[8] Hui, J.Y. (1990). *Switching and Traffic Theory for Integrated Broadband Networks*, Kluwer.

[9] Kelly, F.P. (1991). Effective bandwidths at multi-type queues. *Queueing Systems* **9**, 5-15.

[10] Key, P.B. (1990). Optimal control and trunk reservation in loss networks. *Prob. Eng. Inf. Sc.* **4**, 203-242.

[11] Lippman, S.A. (1975). Applying a new device in the optimization of exponential queuing systems. *Oper. Res.* **23**, 687-710.

[12] Miller, B. (1969). A queuing reward system with several customer classes. *Mgmt. Sc.* **16**, 234-245.

[13] Mitra, D., Morrison, J.A., Ramakrishnan, K.G. (1995). ATM network design and optimization: a multirate loss network framework. *Proc. INFOCOM '96, San Francisco CA*, 994-1003.

[14] Reiman, M.I., Wang, J., Mitra, D. (1995). Dynamic call admission control of an ATM multiplexer with on/off sources. *Proc. 34th CDC, New Orleans LA*, 1382-1388.

[15] Ross, K.W. (1995). *Multirate Loss Models for Broadband Telecommunication Networks*, Springer.

[16] Ross, K.W., Tsang, D.H.K. (1989). Optimal circuit access policies in an ISDN environment: a Markov decision approach. IEEE *Trans. Commun.* **37**, 934-939.

[17] Weiss, A. (1994). Equivalent bandwidth for a priority system. Unpublished notes.

A Proof of Lemma 3.1

Lemma 3.1

Let $\mathbf{Y}^{(n)}$ be the number of busy servers in an infinite-server system with a Poisson arrival process of rate $n\lambda$ and service time distribution $B(\cdot)$ with first moment β and finite second moment $\beta^{(2)}$. Then

$$\Pr\{\mathbf{Y}^{(n)} \geq n\lambda\beta + \epsilon n^{\alpha+1}\} \leq \frac{\lambda\beta + \epsilon n^{\alpha}}{\lambda\beta + \epsilon n^{\alpha} + \epsilon^2 n^{2\alpha+1}},$$

for any $\epsilon > 0$ and $n > 0$.

Proof

Observe that $\mathbf{Y}^{(n)}$ has a Poisson distribution with parameter $n\lambda\beta$, so

$$\Pr\{\mathbf{Y}^{(n)} \geq m\} = \frac{\sum_{i=m}^{\infty} \frac{(n\lambda\beta)^i}{i!}}{\sum_{j=0}^{\infty} \frac{(n\lambda\beta)^j}{j!}} = \frac{\sum_{i=m}^{\infty} \frac{(n\lambda\beta)^i}{i!}}{\sum_{j=0}^{m-1} \frac{(n\lambda\beta)^j}{j!} + \sum_{j=m}^{\infty} \frac{(n\lambda\beta)^j}{j!}}, \tag{11}$$

for any integer $m \geq 0$.

Now

$$\sum_{i=m}^{\infty} \frac{(n\lambda\beta)^i}{i!} \leq \frac{(n\lambda\beta)^m}{m!} \sum_{i=0}^{\infty} \left(\frac{n\lambda\beta}{m}\right)^i = \frac{m}{m - n\lambda\beta} \frac{(n\lambda\beta)^m}{m!}, \tag{12}$$

while

$$\sum_{j=0}^{m-1} \frac{(n\lambda\beta)^j}{j!} \geq \sum_{j=\lfloor n\lambda\beta \rfloor}^{m-1} \frac{(n\lambda\beta)^j}{j!} \geq (m - \lfloor n\lambda\beta \rfloor) \frac{(n\lambda\beta)^m}{m!}, \tag{13}$$

for any integer $m \geq n\lambda\beta$.

Using (11)-(13),

$$\Pr\{\mathbf{Y}^{(n)} \geq y\} = \Pr\{\mathbf{Y}^{(n)} \geq \lceil y \rceil\} \leq \frac{1}{1 + \frac{(\lceil y \rceil - n\lambda\beta)(\lceil y \rceil - \lfloor n\lambda\beta \rfloor)}{\lceil y \rceil}} = \tag{14}$$

$$\frac{\lceil y \rceil}{\lceil y \rceil + (\lceil y \rceil - n\lambda\beta)(\lceil y \rceil - \lfloor n\lambda\beta \rfloor)} \leq \frac{y}{y + (y - n\lambda\beta)^2},$$

for any $y \geq n\lambda\beta$. Taking $y = n\lambda\beta + \epsilon n^{\alpha+1}$ in (14) completes the proof.

$\square$

B Proof of Lemma 3.2

Lemma 3.2

Let $\mathbf{Y}^{(n)}$ be the number of busy servers in an $C^{(n)}$-server loss system with a Poisson arrival process of rate $n\lambda$ and service time distribution $B(\cdot)$ with first moment β and finite second moment $\beta^{(2)}$. If $C^{(n)} \leq n\lambda\beta$, then

$$\Pr\{\mathbf{Y}^{(n)} \leq C^{(n)} + m - \epsilon n^{\alpha+1}\} \leq \frac{\lambda\beta}{\lambda\beta + m^2 n^{-1} - 2m\epsilon n^{\alpha} + \epsilon^2 n^{2\alpha+1}},$$

for any $\epsilon > 0$, $n > 0$, and $m \leq \epsilon n^{\alpha+1}$.

Proof

Observe that $\mathbf{Y}^{(n)}$ has a Poisson distribution with parameter $n\lambda\beta$, truncated at the level $C^{(n)}$, so

$$\Pr\{\mathbf{Y}^{(n)} \leq m\} = \frac{\sum_{i=0}^{m} \frac{(n\lambda\beta)^i}{i!}}{\sum_{j=0}^{C^{(n)}} \frac{(n\lambda\beta)^j}{j!}} = \frac{\sum_{i=0}^{m} \frac{(n\lambda\beta)^i}{i!}}{\sum_{j=0}^{m} \frac{(n\lambda\beta)^j}{j!} + \sum_{j=m+1}^{C^{(n)}} \frac{(n\lambda\beta)^j}{j!}}, \tag{15}$$

for any integer $m = 0, \ldots, C^{(n)}$.

Now

$$\sum_{i=0}^{m} \frac{(n\lambda\beta)^i}{i!} \leq \frac{(n\lambda\beta)^m}{m!} \sum_{i=0}^{m} \left(\frac{m}{n\lambda\beta}\right)^i \leq \frac{(n\lambda\beta)^m}{m!} \sum_{i=0}^{\infty} \left(\frac{m}{n\lambda\beta}\right)^i = \frac{n\lambda\beta}{n\lambda\beta - m} \frac{(n\lambda\beta)^m}{m!}, \tag{16}$$

while

$$\sum_{j=m+1}^{C^{(n)}} \frac{(n\lambda\beta)^j}{j!} \geq (C^{(n)} - m)\frac{(n\lambda\beta)^m}{m!}, \tag{17}$$

for any integer $m = 0, \ldots, C^{(n)}$.

Using (15)-(17),

$$\Pr\{\mathbf{Y}^{(n)} \leq y\} = \Pr\{\mathbf{Y}^{(n)} \leq \lfloor y \rfloor\} \leq \frac{1}{1 + \frac{(n\lambda\beta - \lfloor y \rfloor)(C^{(n)} - \lfloor y \rfloor)}{n\lambda\beta}} = \tag{18}$$

$$\frac{n\lambda\beta}{n\lambda\beta + (n\lambda\beta - \lfloor y \rfloor)(C^{(n)} - \lfloor y \rfloor)} \leq \frac{n\lambda\beta}{n\lambda\beta + (C^{(n)} - y)^2},$$

for any $y \leq C^{(n)}$. Taking $y = C^{(n)} + m - \epsilon n^{\alpha+1}$ in (18) completes the proof.

$\square$

C Proof of Lemma 3.3

Lemma 3.3

Let $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots$ *be a sequence of stochastic variables on* S*, such that* $\lim_{n\to\infty} \Pr\{\mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(A)\} = 1$ *for some closed set* $A \subseteq S$*, with* $\alpha < 0$*,* $\epsilon > 0$*, and such that* $v = \lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{(n)})$ *exists. Then* $v \in \mathrm{conv}(f(A))$*.*

Proof

By definition,

$$\begin{aligned}
\mathrm{E} f(\mathbf{Z}^{(n)}) &= \mathrm{E} f(\mathbf{Z}^{(n)} \mid \mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(A)) \Pr\{\mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(A)\} \\
&+ \mathrm{E} f(\mathbf{Z}^{(n)} \mid \mathbf{Z}^{(n)} \notin B^{(n)}_{\alpha,\epsilon}(A)) \Pr\{\mathbf{Z}^{(n)} \notin B^{(n)}_{\alpha,\epsilon}(A)\}.
\end{aligned}$$

By continuity of $f(\cdot)$, $\lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{(n)} \mid \mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(A)) \in \mathrm{conv}(f(A))$, since the set A and hence the set $f(A)$ is closed, while $\|\, \mathrm{E} f(\mathbf{Z}^{(n)} \mid \mathbf{Z}^{(n)} \notin B^{(n)}_{\alpha,\epsilon}(A)) \,\| \le \max_{x\in S} \|\, f(x) \,\| < \infty$, with $\|\cdot\|$ denoting a norm in $\mathbb{R}^M$. Hence, $\lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{(n)}) \in \mathrm{conv}(f(A))$.

$\square$

D Proof of Lemma 3.4

Lemma 3.4

Let $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots$ *be a sequence of stochastic variables on* S*, such that* $\lim_{n\to\infty} \Pr\{\mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(x^*)\} = 1$ *for some point* $x^* \in S$*, with* $\alpha < 0$ *and* $\epsilon > 0$*. Then* $\lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{(n)}) = f(x^*)$*.*

Proof

By definition,

$$\begin{aligned}
\mathrm{E} f(\mathbf{Z}^{(n)}) &= \mathrm{E} f(\mathbf{Z}^{(n)} \mid \mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(x^*)) \Pr\{\mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(x^*)\} \\
&+ \mathrm{E} f(\mathbf{Z}^{(n)} \mid \mathbf{Z}^{(n)} \notin B^{(n)}_{\alpha,\epsilon}(x^*)) \Pr\{\mathbf{Z}^{(n)} \notin B^{(n)}_{\alpha,\epsilon}(x^*)\}.
\end{aligned}$$

By continuity of $f(\cdot)$, $\lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{(n)} \mid \mathbf{Z}^{(n)} \in B^{(n)}_{\alpha,\epsilon}(x^*)) = f(x^*)$, while $\|\, \mathrm{E} f(\mathbf{Z}^{(n)} \mid \mathbf{Z}^{(n)} \notin B^{(n)}_{\alpha,\epsilon}(x^*)) \,\| \le \max_{x\in S} \|\, f(x) \,\| < \infty$. Hence, $\lim_{n\to\infty} \mathrm{E} f(\mathbf{Z}^{(n)}) = f(x^*)$.

$\square$