The Annals of Applied Probability
1993, Vol. 3, No. 2, 277–295

CAPACITY OF ATM SWITCHES

BY I. ISCOE,[1] D. MCDONALD[2] AND K. QIAN

McMaster University, University of Ottawa and University of Ottawa

When traffic sources are statistically multiplexed over a common link, the sum of the peak rates of the sources exceeds the throughput of the link. The excess may be stored in a buffer, but when this overflows, information is lost. When a source is bursty, the peak rate is attained only for very short periods of time, whereas between bursts the source is idle. Because the sources are independent, the chance that many bursts arrive simultaneously is small, but these rare events do occur and the mean time until overload is a key design parameter.

Here we model the multiplexor as a multidimensional Markov process with a set of forbidden states that represent the exceedance of the link capacity. We use the theory of induced Dirichlet forms to estimate the Laplace transform of the hitting time of this forbidden set. We obtain an upper bound on the probability that the link capacity is exceeded during a fixed time interval along with a lower bound for the mean time until the link capacity is exceeded. This provides the network designer with a degree of assurance about the probability and frequency of overloads.

1. Introduction. The asynchronous transfer mode (ATM) is currently being considered as the preferred transport method for the broad-band integrated services digital network [see Woodruff and Kositpaiboon (1990) for a general overview]. ATM is suitable for multimedia traffic because it offers greater flexibility in bandwidth allocation by transmitting information in fixed length packets, called cells, through virtual network connections.

To achieve maximum bandwidth efficiency, bursty traffic is statistically multiplexed. When traffic sources are statistically multiplexed over a common link, the sum of the peak rates of the sources, in cells per second, exceeds the throughput of the link. The excess cells may be stored in a buffer, but when this overflows, cells are lost. The results in Li (1989) suggest, moreover, that when transmission rates are high, no practical buffering will prevent the loss of cells when the link rate is exceeded. When a source is bursty, cells are generated at the peak rate only for very short periods of time. Immediately afterward the source becomes idle and generates no cells. Because the sources are independent, the chance that many sources transmit simultaneously at the peak rate is small.

Received October 1990; revised March 1992.

[1]Research supported in part by Natural Sciences and Engineering Research Council of Canada Grant A 9351.

[2]Research supported in part by Natural Sciences and Engineering Research Council of Canada Grant A4551, as well as a senior industrial research fellowship held at Bell Northern Research.

AMS 1991 *subject classifications*. Primary 60K30; secondary 60J45, 60K25.

Key words and phrases. Capacities, ATM network overload.

We assume that the link rate is $l - 1$ cells per second. Further we assume that traffic sources belong to n distinct, independent service categories (voice, text, video, etc.) and that traffic sources in category i may be described as an alternating series of idle and bursty periods. A burst from a source in category i produces cells at a rate of d_i cells per second. This means that during a burst from source i, approximately every $(l-1)/d_i$th cell leaving the link comes from source i. We must say approximately because cells that arrive simultaneously at the link from different sources must be slotted one after the other. This is accomplished by buffering and it results in jitter or a slight delay in the arrival of one cell relative to others in the burst. We assume that bursts of category i arrive according to a Poisson process having a rate of a_i bursts per second. We also assume that the burst periods are independent (and independent of the arrival process) and are exponentially distributed with a mean burst length of $1/b_i$.

The aggregate of the n different source categories represents the total load at the link. In particular, if we let $N_i(t)$ represent the number of bursts from category i sources being multiplexed at the link at time t, then the total load at time t may be represented by

$$N(t) := \sum_{i=1}^{n} d_i N_i(t).$$

When the load exceeds the link rate we say the multiplexor is congested. Define

$$\tau = \inf\{t \geq 0 : N(t) \geq l\},$$

so τ is the time until congestion occurs.

$N_i(t)$ is statistically equivalent, up to time τ, to an $M/M/\infty$ queue [which we still denote by $N_i(t)$] with arrival rate a_i and service rate b_i, so assuming each category is in equilibrium, the mean load is $\sum_{i=1}^{n} d_i a_i / b_i$. To characterize τ, the time until congestion, we describe the traffic at the multiplexor by the Markov process $\mathbf{N}(t) := (N_1(t), \dots, N_n(t))$ defined on the state space $S := \{0, 1, 2, \dots\}^n$. Let $\mathscr{D}_0$ denote those real-valued functions that are constant outside a finite subset of S. $\mathbf{N}$ has infinitesimal generator $-\mathscr{L}$, having $\mathscr{D}_0$ as a core, given at $u \in \mathscr{D}_0$ by

$$-\mathscr{L}u(\mathbf{x}) = \sum_{i=1}^{n} \big[(u(\mathbf{x} + \delta_i) - u(\mathbf{x})) a_i + (u(\mathbf{x} - \delta_i) - u(\mathbf{x})) x_i b_i \big], \qquad \mathbf{x} = (x_1, x_2, \dots, x_n) \in S,$$

where δ_i is the ith basis vector in S having all its components equal to 0 except the ith, which is 1; τ is the first time the process $\mathbf{N}(t)$ reaches the forbidden region $F = \{\mathbf{x} \in S : \sum_i^n d_i x_i \geq l\}$.

For bounded $g: S \to \mathbb{R}$ we can define $\alpha(\mathbf{x}) = E_{\mathbf{x}} \int_0^{\tau} g(\mathbf{N}(t))\, dt$, which if $g = \chi_{F^c}$ (χ is the indicator function) represents the mean time to reach the forbidden region starting at $\mathbf{x} \in S$, but in general gives some measure of the occupation time in S before hitting the forbidden set. α satisfies

$$\begin{aligned} \mathscr{L}\alpha(\mathbf{x}) &= g(\mathbf{x}) \quad \text{for } \mathbf{x} \notin F, \\ \alpha(\mathbf{x}) &= 0 \qquad \text{for } \mathbf{x} \in F. \end{aligned} \tag{1}$$

This linear system can be solved, but the number of variables is of the order l^n, so large systems are intractable. Similarly define $\kappa_\theta(\mathbf{x}) = E_\mathbf{x} \exp(-\theta\tau)$ to be the Laplace transform of τ. It is easy to check that

$$\begin{aligned} -\mathscr{L}\kappa_\theta(\mathbf{x}) &= \theta\kappa_\theta(\mathbf{x}) \quad \text{for } \mathbf{x} \notin F, \\ \kappa_\theta(\mathbf{x}) &= 1 \qquad \text{for } \mathbf{x} \in F. \end{aligned} \tag{2}$$

Again this linear system is only tractable for small n.

We first remark that the N_i are independent and each is reversible with respect to the stationary Poisson measure having mean $\lambda_i := a_i/b_i$ assumed less than one. Hence $\mathbf{N}(t)$ is also reversible with respect to the stationary product measure π given by

$$\pi(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i}.$$

The reversibility of $\mathbf{N}(t)$ with respect to π means that for all $1 \le i \le n$,

$$b_i x_i \pi(\mathbf{x}) = a_i \pi(\mathbf{x} - \delta_i) \quad \text{if } x_i > 0.$$

Define the Dirichlet (zero) form

$$\begin{aligned} \mathscr{E}(u,u) &:= \sum_{\mathbf{x}\in S} u(\mathbf{x})\mathscr{L}u(\mathbf{x})\pi(\mathbf{x}) \quad (\text{for } u \in \mathscr{D}_0) \\ &= \sum_{\mathbf{x}\in S} \sum_{i=1}^{n} \frac{1}{2}\Big[(u(\mathbf{x}+\delta_i) - u(\mathbf{x}))^2 a_i \\ &\qquad + (u(\mathbf{x}-\delta_i) - u(\mathbf{x}))^2 x_i b_i\Big]\pi(\mathbf{x}) \\ &= \sum_{\mathbf{x}\in S} \sum_{i=1}^{n} [u(\mathbf{x}+\delta_i) - u(\mathbf{x})]^2 a_i \pi(\mathbf{x}) \end{aligned} \tag{3}$$

and define

$$\mathscr{E}_\theta(u,u) := \mathscr{E}(u,u) + \theta \sum_{\mathbf{x}\in S} u(\mathbf{x})^2 \pi(\mathbf{x}),$$

$$\mathscr{A}(u,u) := \frac{1}{2}\mathscr{E}(u,u) - \sum_{\mathbf{x}\in S} u(\mathbf{x}) g(\mathbf{x}) \pi(\mathbf{x}).$$

Let $\mathscr{H}$ be the set of functions defined on S that equal 1 on F. Because $\mathbf{N}(t)$ is reversible, the function κ_θ satisfies a variational principle.

THEOREM 1.1. *Among $u \in \mathscr{H}$, κ_θ minimizes $\mathscr{E}_\theta(u,u)$. Moreover,*

$$\begin{aligned}\operatorname{Cap}_\theta(F) &:= \inf\{\mathscr{E}_\theta(u,u);\, u \in \mathscr{H}\}\\ &= \theta E_\pi \exp(-\theta\tau)\\ &\equiv \theta \sum_{\mathbf{x}\in S} \kappa_\theta(\mathbf{x})\pi(\mathbf{x}).\end{aligned}$$

PROOF. For a proof that is valid for a general reversible process, see Fukushima [(1980), Lemmas 3.1.1 and 4.3.1]. For an elementary proof, differentiate the form $\mathscr{E}_\theta(u,u)$ at any $u(\cdot)$ and follow Liggett (1985). □

Let $\mathscr{K}$ be the convex set of functions defined on S that equal 0 on F. Again because $\mathbf{N}(t)$ is reversible, the function α satisfies a variational principle.

THEOREM 1.2. *Among $u \in \mathscr{K}$, α minimizes $\mathscr{A}(u,u)$. Moreover*

$$\inf\{\mathscr{A}(u,u);\, u \in \mathscr{K}\} = -\frac{1}{2}\sum_{\mathbf{x}\in S} \alpha(\mathbf{x})g(\mathbf{x})\pi(\mathbf{x}).$$

PROOF. Note that there exists an $m > 0$ such that $\mathscr{E}(u,u) \geq m$ if $u \in \mathscr{K}$ and $\sum_{\mathbf{x}\in S} u(\mathbf{x})^2\pi(\mathbf{x}) = 1$. For otherwise, if $\mathscr{E}(u,u) = 0$, then, by (3), necessarily u is constant and hence 0 by the boundary condition. This gives the coercivity of $\mathscr{A}$. To find the minimum, differentiate $\mathscr{A}(u,u)$ at any $u(\cdot)$. □

In particular, if $g = \chi_{F^c}$, then $E_\pi\tau = -2\mathscr{A}(\alpha,\alpha)$.

We now map these complicated minimization problems onto simpler ones. Specifically, define the map f from S into $\mathscr{R}_+$ by $f(x_1, x_2, \dots, x_n) = \sum_{j=1}^n d_j x_j$. First f induces a measure π^*, having (countable) support $S^* \subset \mathscr{R}_+$, defined by

$$\pi^*(r) \equiv \pi(\{\mathbf{x}: f(\mathbf{x}) = r\}) = \sum_{\mathbf{x}:\, \sum_{j=1}^n d_j x_j = r} \pi(\mathbf{x}).$$

By Corollary 1.12 in Iscoe and McDonald (1990) this map also induces a regular Dirichlet form $\mathscr{E}^*$ on S^*. For any function $h \in \mathscr{D}_0^*$, where $\mathscr{D}_0^*$ is the set of real-valued functions defined on S^* that are constant outside a finite subset of S^*,

$$
\begin{aligned}
\mathscr{E}^*(h,h) &:= \mathscr{E}(h \circ f, h \circ f) \\
&= \frac{1}{2} \sum_{r \in S^*} \sum_{\mathbf{x} \in S:\, \sum_{j=1}^n d_j x_j = r} \sum_{i=1}^n \Big[(h(r+d_i) - h(r))^2 a_i \\
&\qquad\qquad + (h(r-d_i) - h(r))^2 x_i b_i \Big] \pi(\mathbf{x}) \\
&= \frac{1}{2} \sum_{r \in S^*} \sum_{i=1}^n (h(r+d_i) - h(r))^2 a_i \pi^*(r) \\
&\quad + \frac{1}{2} \sum_{r \in S^*} \sum_{i=1}^n (h(r-d_i) - h(r))^2 \\
&\quad \times \left[\sum_{\mathbf{x} \in S:\, \sum_{j=1}^n d_j x_j = r} x_i b_i \frac{\pi(\mathbf{x})}{\pi^*(r)} \right] \pi^*(r) \qquad (4) \\
&= \frac{1}{2} \sum_{r \in S^*} \sum_{i=1}^n (h(r+d_i) - h(r))^2 a_i \pi^*(r) \\
&\quad + \frac{1}{2} \sum_{r \in S^*} \sum_{i=1}^n (h(r-d_i) - h(r))^2 \\
&\quad \times \left[\frac{a_i \pi^*(r-d_i)}{\pi^*(r)} \right] \pi^*(r) \\
&= \sum_{r \in S^*} \sum_{i=1}^n (h(r+d_i) - h(r))^2 a_i \pi^*(r).
\end{aligned}
$$

The form $\mathscr{E}^*$ is associated with a Markov jump process $N^*(t)$ on S^*, having stationary measure π^* and generator $-\mathscr{L}^*$, which jumps from $r \in S^*$ to the right to $r + d_i \in S^*$ with intensity a_i and to the left to $r - d_i$ with intensity $a_i \pi^*(r - d_i)/\pi^*(r)$. It is not, in general, the same process as $N(t)$ because the latter is not typically Markovian.

Now let h be any function in $\mathscr{H}^*$, those functions on S^* taking the value 1 on the image of the forbidden set $F^* = f(F) = [l, \infty)$. Clearly,

$$
\begin{aligned}
\mathrm{Cap}_\theta(F) &= \inf_{u \in \mathscr{H}} \mathscr{E}_\theta(u,u) \\
&\le \inf_{h \in \mathscr{H}^*} \mathscr{E}_\theta(h \circ f, h \circ f) \\
&= \inf_{h \in \mathscr{H}^*} \left[\mathscr{E}^*(h,h) + \theta \sum_{r \in S^*} h(r)^2 \pi^*(r) \right].
\end{aligned}
$$

Hence, defining

$$
\mathscr{E}^*_\theta(h,h) := \mathscr{E}^*(h,h) + \theta \sum_{r \in S^*} h(r)^2 \pi^*(r)
$$

and

$$\mathrm{Cap}_\theta^*(F^*) := \inf_{h \in \mathscr{H}^*} \mathscr{E}_\theta^*(h,h),$$

we have

$$\mathrm{Cap}_\theta(F) \le \mathrm{Cap}_\theta^*(F^*).$$

If we define

$$\tau^* := \inf\{t \ge 0 : N^*(t) \ge l\}$$

and we denote by E_{π^*} the expectation associated with N^* started with its stationary measure, π^*, then the analogue of Theorem 1.1 is valid for the Markov process $N^*(t)$; and we have the following proposition.

PROPOSITION 1.3. *For all* $\theta \ge 0$,

$$E_\pi \exp(-\theta\tau) \le E_{\pi^*} \exp(-\theta\tau^*)$$

and

$$E_{\pi^*}\tau^* \le E_\pi \tau.$$

PROOF. The second inequality follows from the first by subtracting 1 from both sides, dividing by θ and letting θ tend to 0. □

Using Chebyshev's inequality, we immediately have an upper bound on the probability of congestion occurring in a fixed time interval $[0, T]$.

COROLLARY 1.4. *For any* $\theta > 0$,

$$P_\pi(\tau \le T) \le T\theta^{-1} e^{\theta}\, \mathrm{Cap}_{\theta/T}^*(F^*).$$

We may numerically evaluate the overestimate in Corollary 1.4 of the probability of severe congestion in a given time interval and we may do the same for the underestimate of the mean time until severe congestion sets in. Because the induced process is a one-dimensional jump process, the computation of the Laplace transform and the mean level-crossing time is feasible and is essentially independent of the number of sources. In the next section, comparison is made for four cases (see Tables 1–5) with the real time until the forbidden region is reached, obtained by solving (1) and (2). In the four cases $n = 3$, so the number of queues in the system is three; $l = 11$ is the maximum link capacity; d_i is the burst rate of queue i ($d_1 = 1$, $d_2 = 3$, $d_3 = 5$ are fixed); a_i is the arrival rate of queue i, b_i is the service rate of queue i; $T = 10$ is the time. Rates for the four different models are shown in Table 1.

In all cases the bounds given by Proposition 1.3 are very close. For practical purposes we may take the burst rates d_i to be integers. Indeed, rounding up the d_is will decrease τ and hence increase our overestimate of $P_\pi(\tau \le T)$ and decrease our underestimate of $E_\pi\tau$. Also one should, through scaling, ensure the d_is and l have g.c.d. = 1, thereby making l as small as possible to minimize computation.

TABLE 1
Rates for four different models

Model	Queue 1		Queue 2		Queue 3	
	a_1	b_1	a_2	b_2	a_3	b_3
1	0.1	12	0.2	20	0.3	30
2	0.3	6	0.2	9	0.1	10
3	0.3	12	0.2	18	0.1	20
4	0.3	16	0.2	22	0.1	30

Now, to apply Corollary 1.4 we may simply set $\theta = 1$ and, by defining $\theta' = 1/T$, the bound above may be written $(e/\theta')\mathrm{Cap}^*_{\theta'}(F^*)$. Alternatively, we may optimize in θ. Let $G_{\pi^*}(\theta) = E_{\pi^*}\exp(-\theta\tau^*)$ and define $\mu^*_{\pi}(\theta) := \log(G_{\pi^*}(\theta))$. The bound above is now equivalent to $P_{\pi}(\tau \le T) \le \exp(\theta T + \mu^*_{\pi^*}(\theta))$. For $T \le -(\mu^*_{\pi^*})'(0) = E_{\pi^*}\tau^*$, the value θ that produces the tightest upper bound is found by solving $(\mu^*_{\pi^*})'(\theta_0) = -T$. This equation has a unique solution because $\mu^*_{\pi^*}(\theta)$ is strictly convex, strictly decreasing and analytic on the interior of its domain of convergence, $(\bar{\theta}, \infty)$ for some $\bar{\theta} \le 0$. In Table 2 under the columns labelled Upper Bound 2, we find this minimum numerically. This involves fitting a quadratic to the function $\theta T + \mu^*_{\pi^*}(\theta)$ and finding the minimum for this quadratic. Each evaluation of this function at a given θ involves the solution of the induced linear system (7). We see this is moderately successful, but although the Laplace transform $E_{\pi}\exp(-\theta\tau)$ is well approximated, the Chebyshev inequality is rather rough.

Better upper bounds on $P_{\pi}(\tau \le T)$ and lower bounds on $E_{\pi}\tau$ can be found by applying Theorem 1.5 below. The upper bounds are given in Table 2 under the column "Upper bound 1." The corresponding lower bounds on $E_{\pi}\tau$ are given in Table 3 under the column "Lower bound."

When l is large, even solving the induced problem may become troublesome. In Section 3 we give a closed-form upper bound on the probability of severe congestion in a given time interval and a closed-form lower bound on the mean time until severe congestion. In particular we prove the following theorem.

TABLE 2
Upper bounds and lower bounds for the probabilities

Model	Upper bound 1	Upper bound 2		Lower bound	
		Original	Induced	$1 - \exp(-\Lambda T)$	$1 - \exp(-\Lambda^* T)$
1	0.00118269	0.00314619	0.00314619	0.00116212	0.00116212
2	0.00166685	0.00438615	0.00438622	0.00161618	0.00161620
3	0.00042244	0.00112713	0.00112714	0.00041598	0.00041598
4	0.00022586	0.00060497	0.00060497	0.00022338	0.00022338

TABLE 3
Mean exit times and their lower bounds, principal eigenvalues

Model	Mean $E_\pi \tau$	Mean $E_{\pi^*} \tau^*$	Lower bound	$\bar{\kappa}$	Λ	Λ^*	$\pi(F) = \pi^*(F^*)$
1	8599.98	8599.96	8520.63	1.17361E-4	1.16279E-4	1.16279E-4	1.55686E-6
2	6182.41	6182.31	6094.06	1.64090E-4	1.61748E-4	1.61751E-4	6.03955E-6
3	24034.6	24034.4	23864.4	4.19033E-5	4.16067E-5	4.16070E-5	7.67553E-7
4	44761.5	44761.4	44546.1	2.24486E-5	2.23406E-5	2.23407E-5	2.94977E-7

THEOREM 1.5. *If the integers d_j are aperiodic (the greatest common denominator of the d_j's is 1), then as $r \to \infty$,*

$$\pi^*(r) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sum_{j=1}^n d_j^2 \lambda_j s^{d_j}}} s^{-r} \exp\left(\sum_{j=1}^n \lambda_j (s^{d_j} - 1) \right)$$

where $s \equiv s(r)$ is the positive solution of $r = \sum_{j=1}^n d_j \lambda_j s^{d_j}$ and hence s is asymptotic to $(r/\lambda d)^{1/d}$, where $d := \max\{d_j\colon j = 1, \dots, n\}$ and $\lambda := \sum_{j:\, d_j = d} \lambda_j$. Moreover,

$$P_\pi(\tau \le T) \le 1 - \left(1 - \frac{\bar{\kappa}}{\mathrm{Gap}(\mathscr{L})}\right) e^{-\bar{\kappa} T}, \qquad E_\pi \tau \ge \frac{1}{\bar{\kappa}} - \frac{1}{\mathrm{Gap}(\mathscr{L})},$$

where $\mathrm{Gap}(\mathscr{L}) = \min\{b_i;\ i = 1, \dots, n\}$ *and*

$$\bar{\kappa} = \sum_{k=1}^{d} \sum_{j:\, d_j \ge k} a_j \pi^*(l-k) \Big/ \sum_{r=0}^{l-1} \pi^*(r).$$

These bounds are, in fact, asymptotically accurate estimates as $l \to \infty$.

The induced process is in fact the aggregated process that arises in the aggregation–disaggregation method for finding the steady state of a large Markov system [see Schweitzer (1990)]. The engineering approach, then, is to replace any quantity associated with the process $\mathbf{N}(t)$ by the corresponding quantity for the induced process $N^*(t)$. One might, for instance, use the small induced process rather than the huge original process to drive a buffer in the multiplexor [see Qian (1992)]. Hong and Perros (1992) similarly use the induced or aggregated processes associated with a number of interrupted Bernoulli processes to drive a multiplexor buffer. If one takes this approach, we should estimate $P_\pi(\tau \le T)$ by $P_{\pi^*}(\tau^* \le T)$. For small problems we may calculate the latter directly, as in Table 4 under Exact Probability. The results are excellent. For larger problems, a saddle point approximation could be tried [see Daniels (1954)]. In particular,

$$P_{\pi^*}(\tau^* \le T) \approx \frac{1}{\sqrt{2\pi\theta_0^2 (\mu_{\pi^*}^*)''(\theta_0)}} \exp(\theta_0 T + \mu_{\pi^*}^*(\theta_0))$$

TABLE 4
Exact probabilities and approximations

Model	Exact probability $P_\pi(\tau \le T)$	$P_{\pi^*}(\tau^* \le T)$	Saddle point approximation Original	Induced	Approximation 2 $T/E_\pi\tau$	$T/E_{\pi^*}\tau^*$
1	0.00116369	0.00116370	0.00108643	0.00108643	0.00116279	0.00116280
2	0.00162234	0.00162236	0.00151901	0.00151903	0.00161749	0.00161752
3	0.00041676	0.00041676	0.00038878	0.00038878	0.00041607	0.00041607
4	0.00022368	0.00022368	0.00020850	0.00020850	0.00022341	0.00022341

for $(\pi^*_{\pi^*})'(\theta_0) = -T$. In Table 4, under "Saddle point approximation," we see the approximation works well for T and L fixed. We may, moreover, estimate $E_\pi \tau$ by $E_{\pi^*}\tau^*$. In Table 3 we see that results are excellent. Similarly, if we compare the original and induced capacities as in Table 5 we again see excellent results.

The variational method allows one to get bounds on other quantities aside from $P_\pi(\tau \le T)$ and $E_\pi\tau$. In particular, consider $g := g^* \circ f$, where $g^* \in L^2(\pi^*)$; $g^* \equiv \chi_{[F^*]^c}$ is such a function. Let h be any function in $\mathscr{K}^*$, those functions on S^* taking the value 0 on F^*. Clearly,

$$\begin{aligned}\inf_{u \in \mathscr{K}} \mathscr{A}(u,u) &\le \inf_{h \in \mathscr{K}^*} \mathscr{A}(h \circ f, h \circ f)\\ &= \inf_{h \in \mathscr{K}^*}\left[\tfrac{1}{2}\mathscr{E}^*(h,h) - \sum_{r \in S^*} h(r) g^*(r)\pi^*(r)\right]\\ &\equiv \inf_{h \in \mathscr{K}^*} \mathscr{A}^*(h,h),\end{aligned}$$

where

$$\mathscr{A}^*(h,h) := \tfrac{1}{2}\mathscr{E}^*(h,h) - \sum_{r \in S^*} h(r) g^*(r)\pi^*(r).$$

Because $\mathscr{A}^*$ is coercive on $\mathscr{K}^*$, we have a result analogous to that of Theorem

TABLE 5
Capacities at different θ's

Model	$\theta_1 = 0.05$ Original	Induced	$\theta_2 = 0.1$ Original	Induced	$\theta_3 = 0.15$ Original	Induced
1	0.00011609	0.00011609	0.00011630	0.00011630	0.00011643	0.00011643
2	0.00016153	0.00016154	0.00016210	0.00016211	0.00016250	0.00016250
3	0.00004161	0.00004161	0.00004167	0.00004167	0.00004171	0.00004171
4	0.00002235	0.00002235	0.00002237	0.00002237	0.00002238	0.00002238

1.2; that is,

$$\sum_{r \in S^*} \alpha^*(r) g^*(r) \pi^*(r) \le \sum_{\mathbf{x} \in S} \alpha(\mathbf{x}) g^* \circ f(\mathbf{x}) \pi(\mathbf{x}), \tag{5}$$

where $\alpha^*(r) := E_r \int_0^{\tau^*} g^*(N^*(t))\, dt$.

Taking $g = \chi_{F^c}$ in the foregoing inequality, we can give another proof of the second part of Proposition 1.3. If we take g to be the indicator of those $\mathbf{x}$ such that $\sum_{i=1}^n d_i x_i = l - 1$, then we can give a lower bound on the mean time spent on this hyperplane before hitting the forbidden region. If $\alpha(\mathbf{x})$ is this mean hitting time, then

$$\alpha^*(l-1)\pi^*(l-1) \le \sum_{\mathbf{x}:\ \sum_{i=1}^n d_i x_i = l-1} \alpha(\mathbf{x}) \pi(\mathbf{x}),$$

where $\alpha^*(l-1)$ is the mean time spent by the induced process at $l-1$ before hitting l.

Reaching the forbidden region F is a rare event. The probability of a large deviation into the forbidden set during a given time interval is so small that traditional simulation techniques are impractical. If we consider a very long period of time T, the ergodic theorem implies that the expected amount of time spent in the forbidden set is $T\pi(F)$, where $\pi(F)$ is the stationary measure of the forbidden set. On the other hand, the expected number of visits to the forbidden set during this interval is approximately $T/E_\pi \tau$. If we denote by σ the sojourn into the forbidden states, then we have that $T\pi(F)$ is approximately $E_\pi \sigma (T/E_\pi \tau)$, which in turn implies $E_\pi \sigma \sim \pi(F) \cdot E_\pi \tau$. Typically, however, the probability assigned to the forbidden set by the stationary measure is orders of magnitude less than the inverse of the mean time of a large deviation into it. This is borne out in the examples in the next sections (see Table 3, 4). One may conclude that there are many very brief sojourns into the forbidden set as opposed to a few longer sojourns. Because any sojourn into the forbidden states may result in lost cells, we conclude it is dangerous to measure performance using only a steady state analysis.

It is interesting to note that the θ-capacity $\mathrm{Cap}_\theta(F)$ of the forbidden set F is subadditive. So even if the forbidden set were very complex (say there are many multiplexors in a reversible Markovian network and the forbidden set represents overload at any of the multiplexors), then the capacity of the forbidden set is less than or equal to the sum of the capacities of its parts. This means that the network designer can assure a desired grade of service for cell losses by decoupling the components of the network. In particular, each multiplexor may be considered separately and the sum of the capacities of the forbidden sets of each individual multiplexor estimates the desired total capacity.

2. Numerical evaluation of the induced chain. For simplicity we assume the d_i are integer-valued; for otherwise we can round up to the next integer and this has the effect of reducing τ. This is acceptable because we are looking for underestimates of $E_\pi \tau$ and overestimates of $P_\pi(\tau \le T)$.

Unfortunately π^* is quite complicated in general. The asymptotic behaviour is calculated in the Appendix and the following recursion relation is shown:

$$\pi^*(r) = \frac{1}{r}\sum_{j=1}^{n} d_j\lambda_j\pi^*(r-d_j), \qquad r \in S^* \setminus \{0\}, \tag{6}$$

with $\pi^*(0) = \exp(-A)$, where $A := \sum_{j=1}^{n}\lambda_j$. Note that $\pi^*(r-d_j)$ is 0 in the recursion if $r - d_j$ is not in S^*.

This recursion provides a practical means of calculating π^*. With it we may evaluate the jump rates for the induced Markov process having Dirichlet form (4). The following linear systems, which are analogous to (1) and (2), can be solved by computer:

$$\begin{aligned} \mathscr{L}^*\alpha^*(r) &= g^*(r) && \text{for } r \notin F^*, \\ \alpha^*(r) &= 0 && \text{for } r \in F^*, \end{aligned}$$

for $\alpha^*(r) = E_r\tau^*$ and

$$\begin{aligned} -\mathscr{L}^*\kappa_\theta^*(r) &= \theta\kappa_\theta^*(r) && \text{for } r \notin F^*, \\ \kappa_\theta^*(r) &= 1 && \text{for } r \in F^*, \end{aligned} \tag{7}$$

for $\kappa_\theta^*(r) = E_r\exp(-\theta\tau^*)$. From this exact solution, $\mathrm{Cap}_\theta^*(F^*)$ and $E_{\pi^*}\tau^*$ follow. This is feasible even if n is arbitrarily large, because the preceding systems are at most of dimension l. As mentioned earlier we may solve the systems (1) and (2) if l is small. Tables 1 and 2 show that the induced quantities yield exceedingly close bounds for the cases considered!

For these small systems we can even calculate $P_\pi(\tau \le T)$ and $P_{\pi^*}(\tau^* \le T)$. We can calculate the absorption probabilities directly by calculating $\exp(-T\hat{\mathscr{L}})$ and $\exp(-T\hat{\mathscr{L}}^*)$, where $-\hat{\mathscr{L}}$ and $-\hat{\mathscr{L}}^*$ are, respectively, the infinitesimal of $\mathbf{N}$ and N^* killed when they reach their forbidden sets. The expressions $\exp(-T\hat{\mathscr{L}})$ and $\exp(-T\hat{\mathscr{L}}^*)$ are calculated by an eigenvector expansion. These exact results are given in Table 4.

3. Special cases and analytic results.

3.1. *Identical burst rates.* In the case where the burst rates d_j from all sources are identical (and without loss of generality equal to 1), everything simplifies. From (4), $\mathscr{E}^*$ becomes ($S^* = \mathscr{N}$):

$$\mathscr{E}^*(h,h) = \sum_{r\in S^*}\left[(h(r+1)-h(r))^2 B\pi^*(r)\right],$$

where $B = \sum_{i=1}^{n} a_i$. Because $\sum_{i=1}^{n} N_i(t)$ is a Poisson random variable with mean $A = \sum_{j=1}^{n}\lambda_j$, it follows that $\pi^*(r) = \exp(-A)A^r/r!$. By reversibility, it follows that the jump rate from r to $r-1$ is

$$\frac{B\pi^*(r-1)}{\pi^*(r)} = \frac{Br}{A}.$$

We conclude that the induced form is that of a $M/M/\infty$ queue having constant birth rate B, linear death rate Br/A and equilibrium measure $\pi^*(r) = \exp(-A)A^r/r!$.

Consider any positive recurrent, irreducible, birth and death process $(X(t); t \geq 0)$ with generator $-\mathscr{L}$ with birth rates $B(x) > 0$, $x \in \mathscr{N}$, and death rates $D(x) > 0$, $x \in \mathscr{N} \setminus \{0\}$, $D(0) = 0$; and stationary measure π^* given by

$$\pi^*(x) = \pi_0^*(x) \Big/ \sum_{y=0}^{\infty} \pi_0^*(y),$$

$$\pi_0^*(x) := \begin{cases} 1, & \text{if } x = 0, \\ \dfrac{B(0) \cdots B(x-1)}{D(1) \cdots D(x)}, & \text{if } x \geq 1. \end{cases}$$

The reversibility property is

$$B(x)\pi^*(x) = D(x+1)\pi^*(x+1). \tag{8}$$

As usual we set

$$\tau_l = \min\{t \geq 0: X(t) \geq l\}. \tag{9}$$

Let $\alpha(x) = E_x[\tau_l]$. α satisfies

$$\begin{aligned} \mathscr{L}\alpha(x) &= 1 \quad \text{for } 0 \leq x \leq l-1, \\ \alpha(x) &= 0 \quad \text{for } x \geq l. \end{aligned} \tag{10}$$

Also define

$$M^*(r) = \sum_{s=0}^{r} \pi^*(s) \quad \text{and} \quad \nu_1(x) = \sum_{r=0}^{x-1} \frac{M^*(r)}{B(r)\pi(r)}.$$

It is straightforward to verify directly that $\nu_1(l) - \nu_1(x)$ solves the problem (10). Consequently we have that

$$E_x\tau_l = \nu_1(l) - \nu_1(x) \tag{11}$$

and, in particular, $\nu_1(l) = E_0\tau_l$. Also,

$$\begin{aligned} E_{\pi^*}\tau_l &= \sum_{x=0}^{l-1} E_x\tau_l\pi^*(x) = \nu_1(l)M^*(l-1) \\ &\quad - \sum_{x=0}^{l-1} \nu_1(x)\pi^*(x) \\ &= \sum_{x=0}^{l-1} \frac{M^*(x)^2}{B(x)\pi^*(x)}. \end{aligned} \tag{12}$$

The last equality follows from a summation by parts. Alternative formulations of these results are given in Karlin and Taylor (1975).

For the preceding $M/M/\infty$ queue this gives

$$\alpha(r) = \sum_{k=r}^{l-1} \frac{k!}{B} A^k \exp(A) M^*(k),$$

where $M^*(k) := \sum_{i=0}^{k} \pi^*(i) = \exp(-A)\sum_{i=0}^{k}(1/i!)A^i$. By Proposition 1.3, it follows that

$$E_\pi \tau \geq E_{\pi^*}\tau^* = \frac{1}{B}\exp(A) \sum_{k=0}^{l-1} A^{-k} k! M^*(k)^2 \tag{13}$$

$$\geq \frac{M^*(l-1)^2}{Dl\pi^*(l)}. \tag{14}$$

The lower bound (13) provides a slight improvement over that given in Theorem 1.5 when the integers d_i are all 1. The bound (14) is asymptotically equivalent to the bound given in Theorem 1.5.

3.2. *Nonidentical burst rates*: *proof of Theorem* 1.5. If the d_j are integer-valued but not identical, then $-\mathscr{L}^*$ is the generator of a jump process that is not a birth and death process. The lower bounds given in Theorem 1.5 provide a practical means of assuring the grade of service of the multiplexor but we can also show they are asymptotically accurate as $l \to \infty$.

Let Λ be the smallest eigenvalue of the Dirichlet problem:

$$\begin{aligned} \mathscr{L}\rho(\mathbf{x}) &= \Lambda\rho(\mathbf{x}) \quad \text{for } \mathbf{x} \notin F, \\ \rho(\mathbf{x}) &= 0 \qquad\quad \text{for } \mathbf{x} \in F. \end{aligned} \tag{15}$$

By Theorem 3.2 in Iscoe and McDonald (1993) we have

$$P_\pi(\tau \leq T) \geq 1 - e^{-\Lambda T}$$

and

$$P_\pi(\tau \leq T) \leq 1 - \left(1 - \frac{\Lambda}{\mathrm{Gap}(\mathscr{L})}\right)\exp(-\Lambda T),$$

where $\mathrm{Gap}(\mathscr{L})$ is the second largest eigenvalue of the operator $\mathscr{L}$. Integrating with respect to T gives

$$E_\pi \tau \geq \frac{1}{\Lambda} - \frac{1}{\mathrm{Gap}(\mathscr{L})}.$$

Alternately we may use Theorem 3 in Aldous and Brown (1992) after reduction to a finite state space.

We can easily give an upper bound on Λ using the Rayleigh–Ritz principle; that is,

$$\Lambda \leq \sum_{\mathbf{x} \in S} h(\mathbf{x}) \mathscr{L} h(\mathbf{x}) \pi(\mathbf{x}),$$

where h is a function that is 0 on F and such that $\sum_{\mathbf{x}} h^2(\mathbf{x})\pi(\mathbf{x}) = 1$. Let $h = \chi_{F^c}/\pi(F^c)$. It follows that

$$\Lambda \le \sum_{i=1}^{n} \sum_{\mathbf{x}:\, \sum_{j=1}^{n} d_j x_j + d_i \ge l} \frac{a_i \pi(\mathbf{x})}{\pi(F^c)}$$

$$= \sum_{k=1}^{d} \sum_{j:\, d_j \ge k} a_j \pi^*(l-k) \Big/ \sum_{r=0}^{l-1} \pi^*(r) \equiv \bar{\kappa}.$$

Moreover by Proposition 3.1 below, $\mathrm{Gap}(\mathscr{L}) = \min_{j=1,\dots,n} b_j$. Replacing Λ by $\bar{\kappa}$ and $\mathrm{Gap}(\mathscr{L})$ by $\min_{j=1,\dots,n} b_j$ in the bounds in Theorem 3.2 in Iscoe and McDonald (1993) we get the bounds in Theorem 1.5.

PROPOSITION 3.1. $\mathrm{Gap}(\mathscr{L}) = \min_{j=1,\dots,n} b_j$.

PROOF. By Theorem 2.6 in Liggett (1989), $\mathrm{Gap}(\mathscr{L}) = \min_{j=1,\dots,n} \mathrm{Gap}(\mathscr{L}_j)$, where $\mathrm{Gap}(\mathscr{L}_j)$ is the second smallest eigenvalue of the generator $-\mathscr{L}_j$ of an $M/M/\infty$ queue with arrival rate a_j and service rate b_j. We shall show that $\mathrm{Gap}(\mathscr{L}_j) = b_j$ [see Kosten (1974) for a similar argument]. For simplicity, let a, b and λ be the arrival rate, service rate and load a/b, respectively, of the $M/M/\infty$ queue. Consequently, the generator $-L$ of this queue is a self-adjoint operator on $L^2(\pi)$, where $\pi(k) = \exp(-\lambda)\lambda^k/k!$. Let α be an eigenvalue of the generator $-L$ and $\phi \in L^2(\pi)$ be the corresponding right eigenvector. Then,

$$(\alpha + a + bk)\phi(k) = a\phi(k+1) + bk\phi(k+1). \tag{16}$$

Define the weighted generating function of ϕ as

$$\Phi(z) := \sum_{k=0}^{\infty} \phi(k) z^k \pi(k).$$

Using the Cauchy–Schwarz inequality,

$$|\Phi(z)|^2 \le \sum_{k=0}^{\infty} |\phi(k)|^2 \pi(k) \sum_{k=0}^{\infty} |z|^{2k} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Because $\phi \in L^2(\pi)$, it is clear that Φ is entire. Multiplying $z^k\pi(k)$ on both sides of (16) and summing over k from 0 to ∞, we get

$$(\alpha + a)\Phi(z) + bz\Phi'(z) = \frac{a}{\lambda}\Phi'(z) + b\lambda z\Phi(z).$$

This is equivalent to

$$\frac{\Phi'(z)}{\Phi(z)} = \lambda + \frac{\alpha}{b}\frac{1}{1-z}.$$

The only solutions to this equation, normalized so that $\Phi(0) = 1$, are

$$\Phi(z) = e^{\lambda z}|1 - z|^{\alpha/b},$$

where α/b must be a nonnegative integer. Hence, reversing the preceding argument, the spectrum of L is $\sigma(L) = \{0, b, 2b, 3b, \dots\}$. In particular, $\mathrm{Gap}(L) = b$. □

We conclude by giving some asymptotic results obtained by connecting the Laplace transform, $\theta^{-1}\,\mathrm{Cap}_\theta(F)$, of τ_l and $P_\pi(\tau_l \le T)$. We need the following propositions of a general nature.

PROPOSITION 3.2. *Let $\{X_n;\ n \ge 1\}$ be a sequence of nonnegative random variables and $\{c_n;\ n \ge 1\}$ a sequence of positive constants such that for each $\theta > 0$, $\lim_{n\to\infty} c_n E\exp[-\theta X_n] = \theta^{-1}$. Then for all $x > 0$, $\lim_{n\to\infty} c_n P(X_n \le x) = x$. The converse is also valid.*

PROOF. Set

$$U_n = c_n P(X_n \le x) \quad \text{and} \quad \hat{U}_n(\theta) = \int_0^\infty e^{-\theta x}\, dU_n(x), \qquad \theta \ge 0.$$

By the extended continuity theorem for Laplace transforms [see Feller (1971)], $\lim_{n\to\infty} U_n(x) = x$ iff $\lim_{n\to\infty} \hat{U}_n(\theta) = \int_0^\infty e^{-\theta x}\,dx \equiv \theta^{-1}$, for all $\theta > 0$. But

$$\hat{U}_n(x) \equiv c_n E[e^{-\theta X_n}] \to \theta^{-1}, \quad \text{as } n \to \infty,$$

by hypothesis. □

COROLLARY 3.3. *Let $\{x(t);\ t \ge 0\}$ be a reversible, positive recurrent Markov process with stationary distribution (measure) π. Consider forbidden sets F_l and define $\tau_l = \inf\{t \ge 0\colon x(t) \in F_l\}$. Suppose that for some sequence of constants $\{c_l;\ l \ge 1\}$ [or functions $c(l)$, $l \in \mathscr{R}_+$], $\lim_{l\to\infty} c_l\,\mathrm{Cap}_\theta(F_l) = 1$ for each $\theta > 0$. Then*

$$\lim_{l\to\infty} c_l P_\pi(\tau_l \le T) = T.$$

The converse is also valid.

PROOF. It suffices to make the identifications $P = P_\pi$, $\tau_l \leftrightarrow X_l$, $T \leftrightarrow x$ and $\mathrm{Cap}_\theta(F_l) = \theta E_\pi e^{-\theta\tau_l}$ in order to apply Proposition 3.2. □

We now give estimates on $\mathrm{Cap}_\theta(F)$ for the ATM model.

PROPOSITION 3.4. *The capacity $\mathrm{Cap}_\theta(F)$ of $F = \{\mathbf{x} \in S\colon \Sigma_i^n d_i x_i \ge l\}$ and the capacity $\mathrm{Cap}_\theta^*(F^*)$ of $F^* = \{r \ge l\}$ satisfy*

$$\frac{\theta\Lambda}{\theta + \Lambda} + \theta\pi(F) \le \mathrm{Cap}_\theta(F) \le \mathrm{Cap}_\theta^*(F^*) \le \bar{\kappa}\pi(F^c) + \theta\pi(F).$$

PROOF. The lower bound is obvious because $P_\pi(\tau \le T) \ge 1 - e^{-\Lambda T}$ by Theorem 3.2 in Iscoe and McDonald (1993). The upper bound follows by the variational characterization of $\mathrm{Cap}_\theta(F)$ and $\mathrm{Cap}^*_\theta(F^*)$:

$$\mathrm{Cap}_\theta(F) \le \mathrm{Cap}^*_\theta(F^*) := \inf\{\mathscr{E}^*_\theta(u,u);\, u \in \mathscr{H}^*\},$$

where $\mathscr{H}^*$ are functions that are 1 on F^*. It suffices to let $u := \chi_{F^*}$, in which case

$$\begin{aligned}\mathrm{Cap}^*_\theta(F^*) &\le \sum_{i=1}^{n} \sum_{j:\, d_j \ge d_i} a_j \pi^*(l - d_i) + \theta\pi(F)\\ &= \bar{\kappa}\pi(F^c) + \theta\pi(F).\end{aligned}$$

□

In Iscoe, McDonald and Qian (1992) we show that, in fact, $\Lambda \sim \Lambda^* \sim \bar{\kappa}$ as $l \to \infty$. By Corollary 2.3 in Iscoe and McDonald (1991) it follows that $E_\pi\tau \sim \Lambda^{-1}$, so $E_\pi\tau$ is asymptotically equal to $(\Lambda^*)^{-1} \sim E_{\pi^*}\tau^*$. Moreover, taking $c_l \equiv \Lambda^{-1}$ in Corollary 3.3 we have, from Proposition 3.4, that $\Lambda^{-1}P_\pi(\tau \le T) \to T$ as $l \to \infty$. Similarly $(\Lambda^*)^{-1}P_{\pi^*}(\tau^* \le T) \to T$ as $l \to \infty$. Equivalently,

$$E_\pi\tau P_\pi(\tau \le T) \to T \quad \text{and} \quad E_{\pi^*}\tau^* P_{\pi^*}(\tau^* \le T) \to T$$

as $l \to \infty$. In Table 4, Approximation 2 shows that even for $l = 11$ this approximation is excellent. Note that these asymptotic results are not implied by the bounds in Iscoe and McDonald (1993).

APPENDIX

On the induced stationary measure. In this Appendix we derive the recursion relation (6) for the induced probability π^* and study its asymptotic behaviour. For each $t \ge 0$, the weighted sum, $\sum_{i=1}^n d_i N_i(t)$, of independent Poisson random variables, where $N_i(t)$ has mean λ_i, has a compound Poisson distribution with characteristic function

$$\phi(t) := \exp\left(\sum_{j=1}^{n} \lambda_j [e^{itd_j} - 1]\right).$$

For any $r \in S^*$,

$$\begin{aligned}\pi^*(r) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-irt}\phi(t)\,dt\\ &= \frac{1}{2\pi i}\oint z^{-(r+1)} \exp\left\{\sum_{j=1}^{n} \lambda_j [z^{d_j} - 1]\right\} dz\end{aligned}$$

after substituting $z = \exp(it)$, where $\oint$ denotes complex integration around the unit circle. Next, integrating by parts, we get, for $r > 0$,

$$\begin{aligned}\pi^*(r) &= \frac{1}{2\pi i r}\oint z^{-(r+1)} \exp\left\{\sum_{j=1}^{n} \lambda_j [z^{d_j} - 1]\right\} \sum_{j=1}^{n} \lambda_j d_j z^{d_j}\,dz\\ &= \frac{1}{r}\sum_{j=1}^{n} \lambda_j d_j \pi^*(r - d_j).\end{aligned}$$

This recursion, concentrated on $S^* \subset \mathscr{R}_+$ with $\pi^*(0) = \exp(-A)$, where $A := \sum_{i=1}^n \lambda_i$, provides a practical means of calculating π^*.

We may, moreover, derive the asymptotics of π^*.

THEOREM A.1. *If the integers d_j are aperiodic, then as $r \to \infty$,*

$$\pi^*(r) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sum_{j=1}^n d_j^2 \lambda_j s^{d_j}}} s^{-r} \exp\left(\sum_{j=1}^n \lambda_j (s^{d_j} - 1) \right),$$

where $s = s(r)$ is the positive solution of $r = \sum_{j=1}^n d_j \lambda_j s^{d_j}$; so $s \sim (r/\lambda d)^{1/d}$.

PROOF. See Moser and Wyman (1956), expansion (3.49). □

LEMMA A.2. *If the d_j's are aperiodic, $d := \max\{d_i;\ i = 1, \dots, n\}$ and $\lambda := \sum_{j:\, d_j = d} \lambda_j$, then*

$$\lim_{r \to \infty} \frac{\pi^*(r) r^{1/d}}{\pi^*(r-1)} = (d\lambda)^{1/d}.$$

PROOF. Using the previous theorem we have

$$\frac{\pi^*(r) s(r)}{\pi^*(r-1)} \sim \left(\frac{s(r-1)}{s(r)} \right)^{r-1} \times \frac{\exp\left(\sum_{j=1}^n \lambda_j s(r)^{d_j}\right)}{\exp\left(\sum_{j=1}^n \lambda_j s(r-1)^{d_j}\right)} \sqrt{\frac{\sum_{j=1}^n d_j^2 \lambda_j s(r)^{d_j}}{\sum_{j=1}^n d_j^2 \lambda_j s(r-1)^{d_j}}}. \tag{17}$$

Let $r = f(s) := \sum_{j=1}^n d_j \lambda_j s^{d_j}$. Clearly $f(s)/s^d \to d\lambda$, so $s(r) \sim (r/d\lambda)^{1/d}$. The result will follow if we show the right-hand side of (17) tends to 1.

Expanding s around r, we get

$$s(r-1) = s(r) - s'(r) + \tfrac{1}{2} s''(\tilde{r}),$$

where $r - 1 < \tilde{r} < r$. Now

$$s'(r) = \frac{1}{f'(s)} = \frac{s}{\sum_{j=1}^n d_j^2 \lambda_j s^{d_j}}$$

and

$$s''(r) = \frac{d}{ds} \left(\frac{s}{\sum_{j=1}^n d_j^2 \lambda_j s^{d_j}} s'(r) \right),$$

so $|s''(r)| = \mathscr{O}(s(r)/r^2)$. Hence,

$$\begin{aligned} \frac{s(r-1)}{s(r)} &= 1 - \frac{1}{d^2 \lambda s^d (1 + \mathscr{O}(1/s))} + \mathscr{O}\left(\frac{1}{r^2}\right) \\ &= 1 - \frac{1}{d^2 \lambda s^d}\left(1 + \mathscr{O}\left(\frac{1}{s}\right)\right). \end{aligned}$$

We conclude

$$\begin{aligned}\left(\frac{s(r-1)}{s(r)}\right)^{r} &= \prod_{j=1}^{n}\left(\frac{s(r-1)}{s(r)}\right)^{d_j\lambda_j s^{d_j}}\\ &\sim \left(1-\frac{1}{d^2\lambda s^d}\left(1+\mathscr{O}\left(\frac{1}{s}\right)\right)\right)^{d\lambda s^d}\\ &\to \exp(-1/d)\end{aligned}$$

as $s \to \infty$. Also,

$$\begin{aligned}\frac{\exp\left(\sum_{j=1}^{n}\lambda_j s(r)^{d_j}\right)}{\exp\left(\sum_{j=1}^{n}\lambda_j s(r-1)^{d_j}\right)} &= \exp\left(\sum_{j=1}^{n}\lambda_j s(r)^{d_j}\left[1-\left(\frac{s(r-1)}{s(r)}\right)^{d_j}\right]\right)\\ &= \exp\left(\sum_{j=1}^{n}\lambda_j s(r)^{d_j}\frac{d_j}{d^2\lambda s^d}\left(1+\mathscr{O}\left(\frac{1}{s}\right)\right)\right)\\ &\to \exp\left(\frac{1}{d}\right).\end{aligned}$$

Finally, the last term on the right-hand side of (17) tends to 1 so the proof is complete. □

COROLLARY A.3. *If the d_j's are aperiodic, then there is an R such that $r\pi^*(r)$ is decreasing for $r \geq R - d$. R may be identified as the smallest value r such that $\pi^*(r)$ is decreasing on $[r-d, r]$.*

PROOF. By the previous lemma we have that, for r sufficiently large, $\pi^*(r)/\pi^*(r-1) < 1$; so there is an R such that $\pi^*(r)$ is decreasing for $r \geq R - d$. By the recursion formula for π^*, if $r \geq R$, then

$$\begin{aligned}(r+1)\pi^*(r+1) &= \sum_{j=1}^{n}\lambda_j d_j \pi^*(r-d_j)\\ &\leq \sum_{j=1}^{n}\lambda_j d_j \pi^*(r-d_j-1)\\ &= r\pi^*(r).\end{aligned}$$

We can also see from the foregoing that if $\pi^*(r)$ is decreasing on $[r-d, r]$, then it also is decreasing on $[r+1, \infty)$. □

Acknowledgments. We thank Lucien Delbrouck at Bell Northern Research for suggesting this problem. We also thank the referee for suggesting improvements in the presentation of the numerical results and for his valuable comments on our asymptotic results.

REFERENCES

ALDOUS, D. and BROWN, M. (1992). Inequalities for rare events in time-reversible markov processes I. In *Inequalities in Statistics and Probability* (M. Shaked, Y. L. Tong, eds.). IMS, Hayward, CA.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–650.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

FUKUSHIMA, M. (1980). *Dirichlet Forms and Markov Processes*. North-Holland, Amsterdam.

HONG, S. and PERROS, H. G. (1992). An approximate analysis of an ATM multiplexor with multiple heterogeneous bursty arrivals. Preprint.

ISCOE, I. and MCDONALD, D. (1989). Large deviations for l^2-valued Ornstein–Uhlenbeck processes. *Ann. Probab.* **17** 58–73.

ISCOE, I. and MCDONALD, D. (1990). Induced Dirichlet forms and capacitary inequalities. *Ann. Probab.* **18** 1195–1221.

ISCOE, I. and MCDONALD, D. (1993). Asymptotics of exit times for Markov jump processes. *Ann. Probab.* To appear.

ISCOE, I., MCDONALD, D. and QIAN, K. (1992). Asymptotics of the exit distribution for Markov jump processes; application to ATM. Unpublished manuscript.

KARLIN, S. and TAYLOR, H. (1975). *A First Course in Stochastic Processes*, 2nd ed. Academic, New York.

KOSTEN, L. (1974). Stochastic theory of a multi-entry buffer (I). *Delft Progr. Rep. Ser. F* **1** 10–18.

LI, S. (1989). Study of information loss in packet voice systems. *IEEE Trans. Comm.* **37** 1192–1202.

LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, New York.

MOSER, L. and WYMAN, M. (1956). Asymptotic expansions. *Canad. J. Math.* **8** 225–233.

QIAN, K. (1992). On the asymptotics of exit times for Markov jump processes with applications to ATM. Thesis, Univ. Ottawa.

SCHWEITZER, P. (1990). A survey of aggregation–disaggregation in large Markov chains. First International Workshop on the Numerical Solution of Markov Chains, Raleigh, NC.

WOODRUFF, G. M. and KOSITPAIBOON, R. (1990). Multimedia traffic management principles for guaranteed ATM network performance. *IEEE J. Select. Areas Commun.* **8** 437–446.

I. ISCOE
DEPARTMENT OF MATHEMATICS
AND STATISTICS
MCMASTER UNIVERSITY
HAMILTON, ONTARIO
CANADA L8S 4K1

D. MCDONALD
K. QIAN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OTTAWA
585 KING EDWARD
OTTAWA, ONTARIO
CANADA K1N 6N5