Numerical Transform Inversion to Analyze Teletraffic Models

Gagan L. Choudhury,[a] David M. Lucantoni[a] and Ward Whitt[b]

[a]AT&T Bell Laboratories, Holmdel, NJ 07733-3030, USA

[b]AT&T Bell Laboratories, Murray Hill, NJ 07974-0636, USA

We describe recently developed algorithms for numerically inverting transforms to calculate cumulative distribution functions and moments of random quantities of interest in teletraffic models. This new work goes considerably beyond direct one-dimensional numerical inversion. We give a broad overview and describe two developments in some detail. The first combines numerical transform inversion with numerical integration of Pollaczek's classical contour integrals to treat the general GI/G/1 queue. The second combines numerical transform inversion (ordinary and multidimensional) and matrix-analytic methods to describe the transient behavior of the BMAP/G/1 queue (with a batch Markovian arrival process).

1. INTRODUCTION

It is well known that many mathematical results for queueing and teletraffic models can be expressed in the form of transforms. Thirty years ago, this caused Kendall [1] to make his celebrated remark about ''the Laplacian curtain which has obscured much of the detail of the queue-theoretic scene.'' However, it is becoming more and more clear that these transforms can be numerically inverted, so that they can provide the numbers needed for practical engineering studies.

The first numerical inversion algorithm used to study teletraffic models at AT&T Bell Laboratories was developed by M. Eisenberg in about 1970; it is closely related to Weeks' [2] method, which he helped develop. Other algorithms were subsequently developed by Jagerman [3]–[5]. Teletraffic applications of these algorithms appear in [6]–[9]. Recently we have also begun working with numerical transform inversion. We started, with J. Abate, by applying the Dubner-Abate [10] version of the Fourier-series method to study the transient behavior of the M/M/1 queue [11], [12]. (However, many quantities associated with this particular model can actually be effectively computed by direct numerical integration [13].) We then developed new versions of the Fourier-series method and did a fairly extensive survey [14]–[16].

Our purpose here is to briefly describe some of our recent contributions to numerical inversion. This new work goes significantly beyond the one-dimensional numerical inversion in the papers above. In this introduction we give a broad overview. In the following two sections we describe two developments in more detail. Even more detail is provided in the cited references.

First, we have combined numerical transform inversion [14] with numerical integration of Pollaczek's classical contour integrals [17], [18] to calculate the steady-state waiting-time distribution in the general GI/G/1 queue, where neither the interarrival-time distribution nor the service-time distribution need have rational Laplace transform [19]. (We discuss this further in Section 2.) Since the transforms need not be rational, this algorithm is convenient for studying the impact of long-tail distributions [20]. This is important because long-tail distributions have been observed in teletraffic systems [21], [22]. Our numerical results show that the quality of approximations for steady-state waiting-time tail probabilities in typical regions of interest (e.g., 10^{-2}) based on small-tail asymptotics can be remarkably bad when there are long-tail service-time distributions.

We have also combined numerical inversion [14] with matrix-analytic methods [23]–[25] to calculate the steady-state distributions in the BMAP/G/1 queue, which has a batch Markovian arrival process (BMAP) [26]. The BMAP/G/1 queue can be regarded as a matrix generalization of the M/G/1 queue, so it is remarkably tractable. The BMAP is a convenient representation of the versatile Markovian point process or Neuts process which is a powerful model with many applications [23]–[30]. For example, since superpositions of independent BMAPs are BMAPs, they are useful for studying statistical multiplexing. We have applied the algorithms for BMAP/G/1 queues to study the performance of approximations for steady-state distributions based on small-tail asymptotics [29]–[33]. We have found that the quality of these asymptotic approximations is often remarkably good when the service-time distribution has a finite moment generating function [29]–[31], but that the quality can deteriorate dramatically when the arrival process is the superposition of many independent sources, which is a case of great interest for ATM networks [32]–[33]. We provide numerical examples for fairly complicated models, including arrival processes with 64 component two-state Markov modulated Poisson process (MMPP) sources. (We discuss this work further in another paper in these proceedings [33].)

Next, we have developed an algorithm for numerically computing an arbitrary number of moments of a probability distribution from its moment generating function (which is assumed to be well defined) [34]. The algorithm requires computation of the moment generating function at several complex values of its argument, but it does not require knowledge about the location and type of its singularities. A straightforward inversion of the moment generating function can be difficult because the n^{th} moment can increase or decrease rapidly with n. A key ingredient of the algorithm is dynamic scaling based on the last two moments to control errors. The algorithm remains accurate even when the moment order gets large. (e.g., several hundreds), or the moment itself gets very large (e.g., 2^{1023}) or very small (e.g., 2^{-1023}). The algorithm is useful for computing the first several moments that are routinely needed in telecommunication applications but are difficult to compute directly for complex stochastic models. The algorithm is also useful to do asymptotic analysis, as in [29]-[33]. The algorithm can be used to predict the asymptotic form of a probability distribution and to calculate the asymptotic parameters. We also use the moments to approximate a probability distribution accurately over its entire range by matching both low-order and high-order moments [35].

We have also developed numerical inversion algorithms for multidimensional transforms [36]. In addition to computing multivariate probability distributions, such as the joint distribution of the length of a busy period and the number of customers served in that busy period in the M/G/1 queue, we have applied the multivariate transform inversion to compute the transient distributions in queueing systems that are not in steady state. We have combined multivariate transform inversion with matrix-analytic methods to calculate transient distributions for the BMAP/G/1 queue [37]. (We discuss this work further in Section 3.) This work shows that transient distributions in complex models are well within our computational reach. These transient results are important for studying the real-time control of communication networks and other teletraffic systems. We provide numerical examples for fairly complicated models, including ten-state BMAPs.

We have extended the transient analysis of the M/G/1 queue to produce an algorithm for computing the time-dependent workload distribution in a piecewise-stationary $M_t/G_t/1$ queue, where the arrival rates and service-time distributions change only at finitely many time points [38]. (Corresponding algorithms for the $BMAP_t/G_t/1$ queue are being developed.) The problem reduces to a nested family of transient solutions of stationary models, where the initial workload distribution in any interval where the model is stationary is the final workload distribution in the previous interval where again the model is stationary. We develop an effective recursive algorithm, exploiting both one-dimensional and two-dimensional inversions, by storing and reusing intermediate values, and by increasing the precision of the inversions. A straightforward application of numerical inversion would cause the computation to grow exponentially in the number n of intervals and precision to be quickly lost. In our improved algorithm the computation grows only quadratically with n and high precision is maintained even after 20 intervals.

We have shown how to iteratively calculate transform values for transforms that are implicitly defined via functional equations [39], [40]. This approach is important for calculating the busy-period distribution in the M/G/1 queue and its BMAP/G/1 generalization. These busy-period results have important applications to priority queues. The transient analysis of the M/G/1 and BMAP/G/1 queues also requires the busy-period transform [36]–[38].

Finally, we mention recent work with K. K. Leung [41] in which we calculate normalization constants (or partition functions) in closed queueing networks and related product-form models by numerically inverting their generating functions. We have found that this approach can be an attractive alternative to available recursive algorithms. Numerical inversion enables us to compute the normalization constant at a single population vector without recursively computing all smaller values. The computation grows exponentially in the dimension of the generating function, but great computational savings can sometimes be achieved by dimension reduction and Euler summation. A key ingredient here is scaling for error control.

In the remainder of this paper we describe two of these developments in somewhat greater detail.

2. EXPLOITING POLLACZEK'S GI/G/1 CONTOUR INTEGRAL

Consider the standard GI/G/1 queue with one server, unlimited waiting room, the first-in first-out service discipline and i.i.d. service times that are independent of i.i.d. interarrival times. Let U and V be generic interarrival times and service times and assume that $0 < EV < EU < \infty$, so that $\rho \equiv EV/EU < 1$ and the steady-state waiting time, say W, has a proper probability distribution. Let $\phi(z)$ be the moment generating function of $V - U$, i.e.,

$$\phi(z) = Ee^{z(V-U)} = Ee^{zV}Ee^{-zU} , \tag{2.1}$$

which we assume is analytic for complex z in the strip $|\mathrm{Re}\ z| < \delta$ for some positive δ. Under this condition, Pollaczek [17], [18] showed that the Laplace transform of W is

$$Ee^{-sW} = \exp\left\{ - \frac{1}{2\pi i} \int_C \frac{s}{z(s-z)} \log[1-\phi(-z)]\, dz \right\} , \tag{2.2}$$

where s is a complex number with $\mathrm{Re}(s) \geq 0$ and C is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of $\log[1-\phi(-z)]$ in the left half plane [19].

To calculate the probabilities $P(W < x)$, we use numerical transform inversion [14], numerically integrating (2.2) in order to obtain the required transform values. A key step in carrying out this program is identifying an appropriate contour to convert the integral in (2.2) into a familiar integral of a real-valued function. For this purpose, we note that the singularity of $\log[1-\phi(-z)]$ in the left halfplane closest to the origin corresponds to the singularity of Ee^{-sW} in the left halfplane closest to the origin. We find this singularity by finding the unique solution of the equation $\phi(\eta) = 1$ in the range $(0,\eta_s)$, where $\eta_s = \sup\{s \geq 0\colon\ Ee^{sV} < \infty\}$, then we let the contour be a vertical line at $-\eta/2$.

Equation (2.2) becomes $Ee^{-sW} = \exp(-I)$, where

$$I = \frac{1}{2\pi} \int_0^\infty \left[\frac{s}{\bar{z}(s-\bar{z})} \log[1-\phi(-\bar{z})] + \frac{s}{s(s-z)} \log[1-\phi(-z)] \right] dy \tag{2.3}$$

with $z = -\eta/2 + iy$ and $\bar{z} = -\eta/2 - iy$. Since I in (2.3) is in general complex, we compute its real and imaginary parts by integrating the real and imaginary parts of the integrand, respectively. We can also compute $P(W > 0)$ and all the cumulants of W by evaluating integrals similar to (2.3); no transform inversion is required for these.

The specific numerical integration procedure we used is fifth-order Romberg integration. For greater efficiency, the integration interval $(0,\infty)$ in (2.3) is divided into $m + 1$ subintervals: $(0,b_1)$, (b_1,b_2), . . . , (b_{m-1},b_m), (b_m,∞). The last subinterval is transformed into the finite interval $(0,b_m^{-1})$. Within each subinterval, fifth-order Romberg integration is performed. The program generates successive partitions (going from n to $2n$ points) until the estimated improvement is no more than an error tolerance (chosen to be 10^{-12}). The subintervals are chosen to prevent the ratio of the maximum to the minimum value of the integrand within each subinterval from being too great; see [19] for more details.

The effectiveness of our algorithm is illustrated in Table 1 where we present the solution of the $E_k/E_k/1$ queue for $k = 10^j$ for $j = 1,2,3$ and 4. Since these Erlang distributions are nearly deterministic, the traffic intensity is set at $\rho = 1-k^{-1}$. The exponential distribution with mean 1 is approached as $k \rightarrow \infty$.

In [19] we also show how to treat service-time distributions that do not have moment generating functions that are finite in a neighborhood of the origin, such as the Pareto distribution and other long-tail service-time distributions. We approximate these distributions by other distributions that do have this analyticity property. (We do successive refinements to check that the approximation is good enough.) The approximation is done by exponential damping (i.e., we replace a density $f(x)$ by $e^{-\alpha x} f(x)$ and renormalize to keep the same mean), so that we can compute the Laplace transform whenever the original transform is known. By choosing α suitably small, we make the damped distribution close to the original distribution.

3. THE TRANSIENT BMAP/G/1 QUEUE

The *batch Markovian arrival process* (BMAP) is a natural generalization of the Poisson process. It can be constructed by considering a two-dimensional Markov process $\{[N(t), J(t)] : t \geq 0\}$ on the state space $\{(i,j) : i \geq 0, 1 \leq j \leq m\}$ with an *infinitesimal generator* Q having the structure

$$Q = \begin{bmatrix} D_0 & D_1 & D_2 & D_3 & \ldots \\ & D_0 & D_1 & D_2 & \ldots \\ & & D_0 & D_1 & \ldots \\ & & & D_0 & \ldots \end{bmatrix} , \tag{3.1}$$

where D_k, $k \geq 0$, are $m \times m$ matrices; D_0 has negative diagonal elements and nonnegative off-diagonal elements; D_k, $k \geq 1$, are nonnegative; and $D \equiv \sum_{k=0}^{\infty} D_k$ is an irreducible infinitesimal generator. We assume that $D \neq D_0$, so that arrivals do occur.

The variable $N(t)$ counts the number of arrivals in the interval $(0,t]$, and the variable $J(t)$ represents an auxiliary state. Transitions from (i,j) to $(i+k,l)$, $k \geq 0$, $1 \leq j, l \leq m$, correspond to batch arrivals of size k along with a change of state from j to l, and these occur with intensity $(D_k)_{jl}$. Note that it is possible to have any of: (i) arrivals without change of auxiliary state, (ii) arrivals with change of auxiliary state, and (iii) change of auxiliary state without arrivals. A familiar special case is the *Markov modulated Poisson process* (MMPP) having an m-dimensional Markovian environment with infinitesimal generator M and an m-dimensional (diagonal) rate matrix Λ. (The environment is governed by the Markov chain with generator M. When the chain is in state j, arrivals occur according to a Poisson process with rate λ_j.) An MMPP is a BMAP with $D_0 = M - \Lambda$, $D_1 = \Lambda$ and $D_k = 0$ for $k \geq 2$.

A key quantity for the BMAP is the *matrix generating function*

$$D(z) = \sum_{k=0}^{\infty} D_k z^k \quad \text{for} \quad |z| \leq 1 \ . \tag{3.2}$$

Let $P_{ij}(n,t)$ be the *transition function* of the Markov process (N,J), i.e.,

$$P_{ij}(n,t) = P(N(t) = n, J(t) = j \,|\, N(0) = 0, J(0) = i) \tag{3.3}$$

and let $P(n,t)$ be the $m \times m$ matrix with elements $P_{ij}(n,t)$. Then the matrix generating function of $P(n,t)$, defined by $P^*(z,t) = \sum_{n=0}^{\infty} P(n,t) z^n$ for $|z| \leq 1$, can be shown to be given explicitly by $P^*(z,t) = e^{D(z)t}$, where $e^{D(z)t}$ is an exponential matrix [24].

As shown in [37], just as for the M/G/1 queue, it is possible to derive transforms of the transient distributions of the standard stochastic processes of interest in the BMAP/G/1 queue. We illustrate by discussing the workload distribution. The workload at time t is the remaining service time of all customers in the system at time t.

For this purpose, let F be the cdf of the initial work at time 0 and let f be its Laplace-Stieltjes transform; let H be the cdf of a service time and let h be its Laplace-Stieltjes transform; and let $W(t,x)$ be the matrix whose $(i,j)^{\text{th}}$ element is the joint probability that the work in the system is less than or equal to x and the phase is j at time t, given that at time 0 the phase was i and the initial workload (including the customer in service, if any) was distributed according to F. Our object is to compute $W(t,x)$.

Let $w(t,s)$ be the (matrix) Laplace-Stieltjes transform of $W(t,x)$ and let $\tilde{w}(\xi,s)$ be the Laplace transform of $w(t,s)$, i.e.,

$$w(t,s) = \int_0^{\infty} e^{-sx} d_x W(t,x) \quad \text{and} \quad \tilde{w}(\xi,s) = \int_0^{\infty} e^{-\xi t} w(t,s)\, dt \ . \tag{3.4}$$

In [37] we show that the double transform $\tilde{w}(\xi,s)$ can be expressed as

$$\tilde{w}(\xi,s) = (f(s)I - sp_0(\xi))[\xi I - sI - D(h(s))]^{-1} \ , \tag{3.5}$$

where f and h are the Laplace-Stieltjes transforms introduced above, I is the $m \times m$ identity matrix, $D(z)$ is the matrix generating function in (3.2), and $p_0(\xi)$ is the Laplace transform of the emptiness matrix $P_0(t)$. The elements $P_0^{jk}(t)$ of the emptiness matrix $P_0(t)$ give the probability that at time t the system is empty and in phase k given that at time 0 it was in phase j starting with the initial workload distributed according to F. In [37] we show that

$$p_0(\xi) \equiv \int_0^{\infty} e^{-\xi t} P_0(t)\, dt = f(\xi I - D[G(\xi)])(\xi I - D[G(\xi)])^{-1} \ . \tag{3.6}$$

where $G(\xi)$ is the Laplace-Stieltjes transform matrix of the busy-period matrix $\hat{G}(x)$;

i.e., $\hat{G}_{jk}(x)$ is the probability that the first passage from state $(i+1,j)$ at time 0 to state (i,k) occurs no later than time x and k is the first state visited at level i. By [24], the matrix $G(s)$ is the minimal solution to the functional equation

$$G(s) \; = \; \int_0^\infty e^{-sx} e^{D[G(s)]x} dH(x) \; = \; h(sI - D[G(s)]) \; . \tag{3.7}$$

In [40] we show that an iteration for $G(s)$ using (3.7) converges (even when server utilization is bigger than 1) provided we start the iteration with $G(s) = 0$ where 0 represents the null matrix. In general, computation of the right side of (3.7) is difficult. However, in [37] we show that it may be computed efficiently (involving one matrix inversion and a few matrix multiplications) if the Laplace transform h is rational.

Given that we can obtain $G(s)$, we solve (3.6) to get the emptiness transform values and, finally, (3.5) to get the workload double transform values $\tilde{w}(\xi,s)$. We actually calculate the *complementary cdf* $W^c(t,x) \equiv 1 - W(t,x)$ by inverting its two-dimensional Laplace transform

$$\tilde{w}^c(\xi,s) \; \equiv \; \int_0^\infty \int_0^\infty e^{-(\xi t + sx)} \, W^c(t,x) \, dt \; dx \; = \; \frac{1}{s\xi} - \frac{\tilde{w}(\xi,s)}{s} \; . \tag{3.8}$$

We calculate $W^c(t,x)$ by numerically inverting the two-dimensional transform $\tilde{w}^c(\xi,s)$ in (3.8), which uses only $\tilde{w}(\xi,s)$ in (3.5). See [36] for the details of the inversion algorithm. It requires $\tilde{w}(\xi,s)$ for several complex values of its argument pairs. A straightforward evaluation of $\tilde{w}^c(\xi,s)$ using (3.5), (3.6) and (3.7) would be too slow mainly because $G(\xi)$ needs to be computed iteratively. However, usually only a small fraction of the computations of $G(\xi)$ are at distinct ξ values. Therefore, we pre-compute and store $G(\xi)$ for all needed distinct values of ξ and use those stored values during the computation of $\tilde{w}^c(\xi,s)$. This technique greatly speeds up the computations.

In Figure 1 we show the transient workload distribution in an $\sum MMPP_i / E_{16} / 1$ queue with four MMPPs. Each MMPP alternates between a high-rate and a low-rate state where the ratio of the arrival rates in the two states is 4:1. The durations of each state are such that there is an average of five arrivals during the sojourns in each state. We assume that at $t = 0$, there is a departure, 2 sources are in the high-rate state, and there are i_0 (0 or 32) customers in the system. The numerical examples not only show that the algorithm is effective; they also show that the transient behavior can be dramatically different from the steady-state behavior. Thus, we feel that the numerical inversion can help us obtain a much better understanding of the performance of complex systems.

ACKNOWLEDGMENT. We thank Joseph Abate for his important contribution to our numerical transform inversion work.

REFERENCES

1. D. G. Kendall, ‘‘Some Recent Work and Further Problems in the Theory of Queues,’’ *Theor. Probability Appl.*, vol. 9, 1964, pp. 1-13.
2. W. T. Weeks, ‘‘Numerical Inversion of Laplace Transforms Using Laguerre Functions,’’ *J. ACM*, vol. 13, 1966, pp. 419-426.
3. D. L. Jagerman, ‘‘An Inversion Technique for the Laplace Transform with Applications,’’ *Bell System Tech. J.*, vol. 57, 1978, pp. 669-710.
4. D. L. Jagerman, ‘‘An Inversion Technique for the Laplace Transform,’’ *Bell System Tech. J.*, vol. 61, 1982, pp. 1995-2002.
5. D. L. Jagerman, *MATHCALC*, AT&T Bell Laboratories, Holmdel, NJ, 1987.
6. B. T. Doshi, ‘‘Analysis of Clocked Schedules–High Priority Tasks,’’ *AT&T Tech. J.*, vol. 64, 1985, pp. 633-659.
7. B. T. Doshi and J. Kaufman, ‘‘Sojourn Times in an M/G/1 Queue with Bernoulli Feedback,’’ pp. 207-233 in *Queueing Theory and its Applications, Liber Amicorum for J. W. Cohen*, eds. O. J. Boxma and R. Syski, North Holland, Amsterdam, 1988.
8. H. Heffes and D. M. Lucantoni, ‘‘A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance,’’ *IEEE J. Sel. Areas Commun.*, vol. SAC-4, 1986, pp. 856-868.
9. D. P. Heyman, ‘‘Mathematical Models of Database Degradation,’’ *ACM Trans. Database Sys.*, vol. 7, 1982, pp. 615-631.
10. H. Dubner and J. Abate, ‘‘Numerical Inversion of Laplace Transforms by Relating Them to the Finite Fourier Cosine Transform,’’ *J. ACM*, vol. 15, 1968, pp. 115-123.
11. J. Abate and W. Whitt, ‘‘Transient Behavior of the M/M/1 Queue: Starting at the Origin,’’ *Queueing Systems*, vol. 2, 1987, pp. 41-65.
12. J. Abate and W. Whitt, ‘‘Approximations for the M/M/1 Busy-Period Distribution,’’ pp. 149-191 in *Queueing Theory and its Applications, Liber Amicorum for J. W. Cohen*, eds. O. J. Boxma and R. Syski, North Holland, Amsterdam, 1988.
13. J. Abate and W. Whitt, ‘‘Calculating Time-Dependent Performance Measures for the M/M/1 Queue,’’ *IEEE Trans. Commun.*, vol. 37, 1989, pp. 1102-1104.
14. J. Abate and W. Whitt, ‘‘The Fourier-Series Method for Inverting Transforms of Probability Distributions,’’ *Queueing Systems*, vol. 10, 1992, pp. 5-88.
15. J. Abate and W. Whitt, ‘‘Numerical Inversion of Probability Generating Functions,’’ *Oper. Res. Letters*, vol. 12, 1992, pp. 245-251.
16. J. Abate and W. Whitt, ‘‘Numerical Inversion of Laplace Transforms of Probability Distributions,’’ *ORSA J. Comput.*, to appear.
17. F. Pollaczek, ‘‘Fonctions caractéristiques de certaines répartitions définies au moyen de la notion d’ordre. Application à la théorie des attentes. *C. R. Acad. Sci. Paris*, vol. 234, 1952, pp. 2334-2336.
18. F. Pollaczek, *Problèmes Stochastiques Posés par le Phénomène de Formation d’une Queue d’Attente à un Guichet et par des Phènomènes Apparentés*, Mémorial des Sciences Mathématiques, Gauthier-Villars, Paris, 1961.

19. J. Abate, G. L. Choudhury and W. Whitt, ''Calculation of the GI/G/1 Waiting-Time Distribution and Its Cumulants from Pollaczek's Formulas,'' *Archiv für Elektronik und Übertragungstechnik*, vol. 47, 1993, pp. 311-321.
20. J. Abate, G. L. Choudhury and W. Whitt, ''Waiting-Time Tail Probabilities in Queues with Long-Tail Service-Time Distributions,'' *Queueing Systems* to appear.
21. K. S. Meier-Hellstern, P. E. Wirth, Y. L. Lan and D. A. Hoeflin, ''Traffic Models for ISDN Data Users: Office Automation Application,'' *Teletraffic and Datatraffic in a Period of Change, ITC 13*, A. Jensen and V. B. Iversen (eds.), North-Holland, 1991, pp. 167-172.
22. D. E. Duffy, A. A. McIntosh, M. Rosenstein and W. Willinger, ''Analyzing Telecommunications Traffic Data from Working Common Channel Signaling Subnetworks,'' *Proc. of INTERFACE '93,* to appear.
23. M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications,* Marcel Dekker, New York, 1989.
24. D. M. Lucantoni, ''New Results for the Single Server Queue with a Batch Markovian Arrival Process,'' *Stochastic Models,* vol. 7, 1991, pp. 1-46.
25. D. M. Lucantoni, ''The BMAP/G/1 Queue: A Tutorial,'' in *Models and Techniques for Performance Evaluation of Computer and Communications Systems,* L. Donatiello and R. Nelson eds., Springer Verlag, 1993, 330-358.
26. G. L. Choudhury, D. M. Lucantoni and W. Whitt, ''An Algorithm for a Large Class of G/G/1 Queues,'' in preparation.
27. M. F. Neuts, ''A Versatile Markovian Point Process,'' *J. Appl. Prob.,* vol. 16, 1979, pp. 764-779.
28. V. Ramaswami, ''The N/G/1 Queue and its Detailed Analysis,'' *Adv. Appl. Prob.,* vol. 12, 1980, pp. 222-261.
29. J. Abate, G. L. Choudhury and W. Whitt, ''Asymptotics for Steady-State Tail Probabilities in Structured Markov queueing Models,'' *Stochastic Models,* vol. 10, 1994.
30. G. L. Choudhury and W. Whitt, ''Heavy-Traffic Asymptotic Expansions for the Asymptotic Decay Rates in the BMAP/G/1 Queue,'' *Stochastic Models,* to appear.
31. J. Abate, G. L. Choudhury and W. Whitt, ''Exponential Approximations for Tail Probabilities in Queues, I: Waiting Times,'' *Oper. Res.,* to appear.
32. G. L. Choudhury, D. M. Lucantoni and W. Whitt, ''Tail Probabilities in a Queue with Many Independent Sources,'' in preparation.
33. G. L. Choudhury, D. M. Lucantoni and W. Whitt, ''On the Effectiveness of Effective Bandwidths for Admission Control in ATM Networks,'' ITC 14.
34. G. L. Choudhury and D. M. Lucantoni, ''Numerical Computation of Moments With Application to Asymptotic Analysis,'' *Oper. Res.,* to appear.
35. G. L. Choudhury, D. M. Lucantoni and W. Whitt, ''Refined Approximations for G/G/1 Queues,'' in preparation.
36. G. L. Choudhury, D. M. Lucantoni and W. Whitt, ''Multidimensional Transform Inversion with Applications to the Transient M/G/1 Queue,'' *Ann. Appl. Prob.*, to appear.
37. D. M. Lucantoni, G. L. Choudhury and W. Whitt, ''The Transient BMAP/G/1 Queue,'' *Stochastic Models,* vol. 10, 1994.

38. G. L. Choudhury, D. M. Lucantoni and W. Whitt, ‘‘Numerical Solution of $M_t/G_t/1$ Queues,’’ 1993, submitted for publication.
39. J. Abate and W. Whitt, ‘‘Solving Probability Transform Functional Equations for Numerical Inversion,’’ *Oper. Res. Letters*, vol. 12, 1992, pp. 275-281.
40. G. L. Choudhury, D. M. Lucantoni and W. Whitt, ‘‘The Distribution of the Duration and Number Served During a Busy Period in the BMAP/G/1 Queue,’’ in preparation.
41. G. L. Choudhury, K. K. Leung and W. Whitt, ‘‘Calculating Normalization Constants of Closed Queueing Networks by Numerically Inverting Their Generating Functions,’’ 1993, submitted for publication.

congestion measure	k				
	10	100	1,000	10,000	∞
$P(W > 0)$	0.7102575	0.9035808	0.9687712	0.9900406	1.0000000
$P(W > 1)$	0.2780070	0.3385844	0.3584117	0.3648607	0.3678794
$P(W > 3)$	0.0376169	0.0458224	0.0485057	0.0493785	0.0497871
$P(W > 5)$	0.0050909	0.0062014	0.0065645	0.0066825	0.0067379
$P(W > 7)$	0.0006890	0.0008393	0.0008884	0.0009044	0.0009119

Table 1. Tail probabilities of the steady-state waiting time in the $E_k/E_k/1$ model with traffic intensity $\rho = 1 - k^{-1}$, as a function of k. The case $k = \infty$ is an exponential with mean 1.

Figure 1. Numerical results for the workload tail probabilities as a function of the time t and the initial queue length i_0 in the $\sum MMPP_i/E_{16}/1$ queue with traffic intensity $\rho = 0.7$ and 2 of the four MMPPs starting in the high-rate state.