MANAGEMENT SCIENCE
Vol. 25, No. 6, June 1979
Printed in U.S.A.

A CLOSURE APPROXIMATION FOR THE NONSTATIONARY $M/M/s$ QUEUE*

MICHAEL H. ROTHKOPF† AND SHMUEL S. OREN†

Typical queues do not have constant arrival rates. This paper discusses effective computational methods for dealing with queues having nonstationary arrival processes. It presents a computationally undemanding approximate method for finding the time dependent mean and standard deviation of the number of customers in an s server queueing system with time-varying arrival and service rates. Results are exhibited for various 1 and 3 server queues with constant service rates and sinusoidal components in the arrival rate.
(QUEUES–NONSTATIONARY; QUEUES–TRANSIENT RESULTS; QUEUES–NUMERICAL METHODS)

1. Introduction

Queueing models were among the first models of operations research. The leading introductory texts of operations research have chapters that develop the elementary theory and give examples of actual situations involving queues. These examples involve situations in which the arrival process typically has significant and predictable time-dependent characteristics. Wagner [26], for example, lists grocery store checkout stands, bank tellers' windows, service station pumps and attendants, telephone trunk lines, and photocopy equipment maintenance men. Most service facilities of these kinds experience predictable demand fluctuations. Although elementary queueing theory dealing with stationary behavior is useful in many areas, it is inadequate for dealing with situations in which the time dependent behavior is important. This is often true for queues that go through periods in which the arrival rate exceeds the service rate. There are a variety of ways of attempting to bridge the gap between elementary theory and practical, nonstationary problems. This paper reviews several of them and presents a new "closure" approximation method for computing the time-dependent mean and variance of the number in the system of a nonstationary $M/M/s$ queue. This method is computationally tractable and usable with any dynamic pattern of arrival (and/or service) rates and with any number of servers. Experiments show that the method gives close approximations when used for calculating transient behavior and limiting periodic behavior. We illustrate the use of the method by computing graphs showing the average number in one and three server systems when the arrival rate has both constant and sinusoidal components.

This work has been used to model the capacity acquisition problems faced by the managers of a variety of central document reproduction departments. These departments typically have predictable fluctuations—time of day, day of the week, and sometimes day of the month—in the rate of arrival of their work. In some reproduction departments these fluctuations are mild, but often they are large enough that the peak period arrival rate significantly exceeds an economically reasonable capacity level. Detailed data on particular central reproduction departments is expensive to obtain, but estimates are usually readily available for the average arrival rate of jobs, the average time to do a job, and on the period and relative magnitude of systematic fluctuations in workload. For capacity acquisition decisions it is usually adequate to assume that job times are exponentially distributed and that arrivals are generated by

* Accepted by Marcel F. Neuts; received December 1, 1977. This paper has been with the authors 6 months for 3 revisions.

† Xerox Palo Alto Research Center.

0025-1909/79/2506/0522$01.25

a nonhomogeneous Poisson process, i.e., a Poisson process with time dependent arrival rate. Making these decisions requires only a few simple measures of effectiveness—primarily the average wait in the system (which can be calculated from the average number in the system).

Problems of this type can be approached in various ways. The simplest way is to ignore the fluctuations and use standard steady state analysis. This is appropriate if the fluctuation are mild or if the period of fluctuation is sufficiently short that only a few arrivals are likely to occur during a peak period. If the fluctuations do not let the arrival rate get as large as the service rate and if the period of the fluctuation is sufficiently long, then the queue can be analyzed as one that goes through a sequence of steady states. A different approach described by Neuts [19], [20] is to model the fluctuations in the arrival rate as being generated by an underlying Markov process.

Monte Carlo simulation is a common approach to such problems. It can handle virtually any complication and can compute virtually any measure of effectiveness if enough computer time is used. The required computer time, however, can be substantial, and it grows as the square of the required precision.

The general transient solution for the stationary single server queue is known [3]. Unfortunately, it involves an infinite sum of Bessel functions and would not normally be used for numerical computation. Morse's book [17] contains closed form transient solutions for the finite state single server queue with constant arrival rate. By discretizing the dynamic change of the arrival rate, this could conceivably be adapted to the dynamic case, but we are not aware of any attempt to implement this approach. Commonly, numerical approaches deal directly with the differential equation of the queue. An important paper by Leese and Boyd [13] examined seven such numerical methods for the nonstationary single server queue. These methods, including Clarke's [2], Luchak's [15], and Wragg's [28], are compared to Monte Carlo simulation. Of the seven, only Leese and Boyd's adaptation of Wragg's method seemed to them computationally superior to simulation. Koopman [11], and more recently Kolesar et al. [10], have solved nonstationary queueing problems by numerically integrating a truncated set of the state probability equations. This appears to be a useful method for problems in which one can assume a reasonably low upper limit on queue length. Rather than numerically integrating the state probability differential equations, it is possible to approximate a queue by a discrete time Markov process and solve numerically the resulting state probability difference equations. Neuts [18] discusses the computational aspects of such discrete time Markov queues.

When a queue experiences an arrival rate substantially above its service rate for a period sufficiently long relative to the average interarrival time to allow the law of large numbers to apply, a deterministic model of queue build up and decay will provide a good approximation. Such a method is described by Oliver and Samuel [23], by Newell [21] and in the booklet by Howard [7] commissioned and published by the Xerox Corporation for distribution to reproduction center managers. When the bulk of the delay in a system is caused by predictable surges of demand beyond service capacity, such deterministic models provide good approximations to the overall delay. Newell [21], [22] presents a diffusion equation that approximates the behavior of a heavily used queue as the arrival rate passes through the service rate. Newell's method was intended to be useful for the design of transportation facilities in which there are many servers and hundreds or thousands of arrivals during a rush hour surge. For any such heavy traffic situations his methods are valuable. Because he assumes heavy traffic, his approximations do not depend upon the service time distribution or the number of servers.

There are also transform inversion approximation approaches for analyzing nonstationary queues. Such methods have been suggested by Gaver [4] and recently by Kotiah [17].

The approach taken in this paper belongs to a general class of methods known as *closure techniques* for approximating solutions of infinite systems of equations. These are commonly used in physics. The basic strategy of a closure technique is to reduce an infinite system of equations to a finite system by making a "closure assumption" in the form of a functional relationship between the variables of the system. For instance, the well known Maxwell-Boltzman equations are obtained by closing an infinite hierarchy of differential equations for the probability density functions over the position and momentum of particles in a fluid [8]. The key to the success of a closure method is its closure assumption, while the test of any such assumption is the empirical usefulness of the resulting equations. The experience in physics is that systematic approaches to derive such assumptions have not been as effective as intuition [9].

The finite state approximation technique employed by Koopman [11] and Kolesar et al. [10] can be viewed as an elementary closure method in which the infinite system of equations is closed by assuming zero probability beyond some finite number in the system. Another closure method for approximating the time dependent mean of a nonstationary $M/M/1$ queue has been recently published by Rider [24]. He combines the state equations to obtain the differential equation for the mean in terms of the idle probability. He then makes a closure assumption relating the idle probability to the mean number in the system, the arrival rate and the service rate. Our approach is somewhat similar to Rider's approach but it is more general and appears to produce better approximations. We use the differential equations for the mean and variance for the $M/M/1$ queue due to Clark [2] and generalize them to the $M/M/s$ case. These equations are expressed in terms of the state probabilities for states with fewer customers than servers. Our closure assumption employs the negative binomial distribution to approximate the state probability distribution in terms of the mean and variance of the number in the system. This leaves us with a pair of differential equations for the mean and variance of the number in the system. We integrate these equations by standard numerical methods.

At the time of the final revision of this paper, we learned of recent unpublished work by Chang [1] and his student Wang [27]. Chang analyzes networks of $M/M/1$ queues using an approach similar to ours but with a different closure assumption relating the idle probability to the mean and variance of the number in the system. Wang's work attempts to improve that closure assumption and generalize the results to other queueing systems.

The next section of this paper describes in detail the method we propose and its characteristics. The following section displays and evaluates some results we have obtained using this method. Finally, the appendix contains theorems stating its theoretical properties.

2. The Method

For ease of exposition, we consider first the single server queue with Poisson arrivals at a nonnegative time varying rate $\lambda(t)$, and exponential service at a time varying nonnegative rate, $\mu(t)$. Let $P_n(t)$ be the probability that there are n customers in the system at time t, and let $P_n'(t)$ be the derivative of $P_n(t)$ with respect to time. For convenience, we will henceforth omit the argument t except where it is required to avoid ambiguity. The standard differential-difference equations for the $M/M/1$ queue are

$$\begin{aligned} P_n' &= -(\lambda+\mu)P_n + \lambda P_{n-1} + \mu P_{n+1}, \qquad n = 1,2,\dots, \\ P_0' &= -\lambda P_0 + \mu P_1. \end{aligned} \tag{1}$$

Denoting the average number in the system as $m(t)$ and its derivative by m', we can obtain from (1) the intuitively obvious equation

$$m' = \sum_{n=0}^{\infty} nP_n' = \lambda - \mu(1 - P_0). \tag{2}$$

Similarly, the derivative of mean squared number in the system is given by

$$\overline{n^2}' = \sum_{n=0}^{\infty} n^2 P_n' = \lambda + \mu(1 - P_0) + 2m(\lambda - \mu). \tag{3}$$

Hence the derivative of the variance, v, of the number in the system is

$$v' = \overline{n^2}' - 2mm' = \lambda + \mu - \mu P_0(2m + 1). \tag{4}$$

Equations (2) and (4) were obtained by Clarke [2]. He used them to derive limiting formulas for the mean and variance of the number in the system for the $M/M/1$ queue with λ/μ constant and $\lambda \geqslant \mu$. In order to analyze more general nonstationary behavior, we introduce our closure assumption expressing P_0 in terms of m and v, using the negative binomial distribution.

The negative binomial distribution given by

$$P_n = \binom{r + n - 1}{n} p^r (1 - p)^n, \qquad n = 0, 1, 2, \ldots, \tag{5}$$

is fully specifiable by its mean, $r(1 - p)/p$, and variance, $r(1 - p)/p^2$. In particular, P_0 can be expressed as

$$P_0(m, v) = p^r, \quad \text{where} \quad p = m/v \text{ and } r = m^2/(v - m). \tag{6}$$

The negative binomial distribution has the desirable property that when its variance equals its mean plus the square of its mean it specializes to the geometric distribution on nonnegative integers. This is the steady state distribution of the stationary $M/M/1$ queue. Thus, if this method is applied to a stationary $M/M/1$ queue, then the assumed negative binomial distribution will asymptotically approach the correct steady state distribution of the number in the system.

It can be shown that for all nonnegative m and v the function $P_0(m, v)$ defined by (6) is monotonically increasing in v and monotonically decreasing in m and that

$$0 \leqslant \exp(-m^2/v) \leqslant P_0 \leqslant \exp(-m) \leqslant 1.$$

These properties of $P_0(m, v)$ are compatible with the way one would intuitively expect the idle probability of an $M/M/1$ queue to behave. Furthermore, the fact that these properties hold for all nonnegative values of m and v allows the use of $P_0(m, v)$ as an approximation to the idle probability even for $m > v$ when the negative binomial distribution is not well defined.

For $M/M/1$ queues, our calculation scheme amounts to integrating the pair of differential equations (2) and (4) with P_0 given by (6) when $v \neq m$ and by the limit of (6), $\exp(-m)$, when $v = m$. This scheme proceeds as follows:

1. Use $m(t)$ and $v(t)$ to obtain $P_0(t)$ using (6) if $v \neq m$ and $\exp(-m)$ if $v = m$.
2. Use this P_0 in (2) and (4) to obtain $m'(t^+)$ and $v'(t^+)$.
3. For some small $\Delta > 0$, set $\Delta_t = \text{Min}[\Delta, \delta_t]$, where δ_t is the time increment to the next discontinuity in $\mu(t)$ or $\lambda(t)$.
4. Approximate $m(t + \Delta_t)$ by $\text{Max}[0, m(t) + \Delta_t m'(t^+)]$ and $v(t + \Delta_t)$ by $\text{Max}[0, v(t) + \Delta_t v'(t^+)]$.
5. Increment t by Δ_t and repeat.

This computational scheme is just forward Euler integration applied piecewise over continuous segments of $\mu(t)$ and $\lambda(t)$ [6].

When the queue has s servers, each working at rate μ, equations (2) and (4) generalize to

$$m' = \lambda - \mu s + \mu \sum_{n=0}^{s-1} (s-n)P_n \tag{7}$$

and

$$v' = \lambda + \mu s - \mu \sum_{n=0}^{s-1} (2m+1-2n)(s-n)P_n. \tag{8}$$

To "close" equations (7) and (8), we need the probabilities $P_0, P_1, \ldots, P_{s-1}$ which we approximate as before using (5) with $p = m/v$ and $r = m^2/(v-m)$. Unfortunately for $s > 1$, such an approximation no longer produces the correct steady state distribution for the stationary case. Our experiments indicate, however, that the error in the mean and variance resulting from this discrepency is small. To remove this error, we add a correction term to the value of m obtained from (5), (7) and (8). This correction is determined so as to offset the steady state error for the stationary case and is calculated by comparing standard analytical results for stationary queues with the results of uncorrected calculations with the algorithm for the same queues. Table 1 gives the correction for various two and three server queues. As an example of the use of the table, when we calculate below the dynamic steady state mean number in the system for a particular three server periodic queue with 60% time average utilization to be 4.308, we increase this by 0.070 to 4.378. A similar correction is made in the variance. Aside from these corrections, no other changes have been made in going from the single to the multiple server case.

TABLE 1

Amount by which True Mean and Standard Deviation of Number in the System Exceeds Calculated Value in Stationary Steady State

Number of Servers	Utilization	Excess m	Excess σ_n	Percent of True m	Percent of True σ_n
1	any	0	0	0	0
2	0.1	0.00055	0.0023	0.27	0.44
	0.2	0.0036	0.0101	0.86	1.5
	0.3	0.0100	0.026	1.5	2.9
	0.4	0.020	0.051	2.1	4.5
	0.5	0.032	0.091	2.4	6.0
	0.6	0.045	0.14	2.4	7.1
	0.7	0.059	0.23	2.1	8.0
	0.8	0.065	0.37	1.5	8.2
	0.9	0.0102	0.69	0.11	7.2
3	0.1	0.00011	0.00046	0.038	0.079
	0.2	0.0016	0.0044	0.27	0.59
	0.3	0.0072	0.018	0.77	1.8
	0.4	0.0193	0.045	1.5	3.7
	0.5	0.040	0.098	2.3	6.1
	0.6	0.070	0.18	3.0	8.8
	0.7	0.108	0.32	3.3	11.2
	0.8	0.155	0.59	3.1	13.0
	0.9	0.154	1.21	1.5	12.7

It is interesting to note that the summation in (7) is just the mean number of idle servers and that the summation in (8) can be replaced by an expression involving the mean and variance of the number of idle servers.

The schemes described above for the $M/M/1$ and the $M/M/s$ queues can be used for two kinds of problems. Given specific intitial conditions for m and v, this method enables us to approximate the transient behavior of a queue by tracking the time dependent mean and variance. For queues with periodic arrival and/or service rates (with some common period) the method can be used to approximate the limiting periodic mean and variance of the number in the system. (Recent work by Harrison and Lemoine [5] proves the existence of a periodic limiting distribution for the periodic $M/G/1$ queue.) To obtain such limiting behavior, we start with some arbitrary nonnegative initial values satisfying $v(0) \geqslant m(0) \geqslant 0$ and track $m(t)$ and $v(t)$ in time until we detect periodicity. Specifically, using the smallest common period of the arrival and service rates, we track the system through several periods until the beginning values of m and v for two succesive common periods are the same (within a small tolerance). We then compute m and v over the following period.

The appendix to this paper contains three theorems summarizing our theoretical analysis of the proposed method for the $M/M/1$ queue. The proofs to these theorems along with related results are given in [25]. In our analysis, we focus on two aspects of the method. First we examine issues concerning our closure assumption, and then we analyze the error propagation resulting from our numerical integration. In particular, we show that for any nonnegative initial values $m(0)$, $v(0)$, the solution $m(t)$, $v(t)$ of (2), (4) and (6) will be nonnegative for all t and consequently $0 \leqslant P_0(t) \leqslant 1$ for all t. We also show that under reasonable assumptions on $\mu(t)$ and $\lambda(t)$ any trajectory of $m(t)$ and $v(t)$ generated by (2), (4) and (6) with $v(0) \geqslant m(0) \geqslant 0$ will asymptotically approach any other trajectory generated by these equations with nonnegative initial values. This result justifies our assumption $v(0) \geqslant m(0) \geqslant 0$ when using the procedure to obtain limiting periodic solutions. Finally we show that with the above restriction on the initial values, one can always select a sufficiently small Δ so that the numerical integration will approximate the exact solution to (2), (4) and (6) on any finite time interval within any given tolerance. While we have not established the above results for the general $M/M/s$ case, our computational experience leads us to believe that similar results hold for it.

3. Results and Discussion

Since this method is approximate and we have no analytic bounds on the approximation, it is necessary to test the method's accuracy by means of computational comparisons. We have performed several kinds of comparisons. First, we obtained from Eric Leese and ran the test problem used in [13]. That problem involved a widely fluctuating arrival rate and one server. Unfortunately, only the plots of the results reported in [13] (shown there as Figures 5 and 7) are now available. The results we obtained for the mean number in the system agree everywhere with Leese and Boyd's exact result to within the accuracy of their plot. This is not true for the standard deviation of the number in the system. However, the nature of the deviation and a comparison with the figures in [14] and Figure 8 in [13] leads us to believe that Figure 7 in [13] is an incorrectly labled plot of the expected number in the *queue*. Figure 1 presents a comparison of our results with the plot of the standard deviation of the number in the system for this test problem as reported in Figure 6 of [14].

The only other published test problem we are aware of is given by Rider [24]. It is a single server periodic queue with widely varying arrival and service rates. The solution Rider reports is a periodic limiting result. Our method solved that problem more accurately than Rider's method. Table 2 gives the comparison. In that table, the "exact" solution is the one reported by Rider and verified by us by numerical integration of a truncated version of (1).

We have also run a number of tests of our method with the arrival rate given by

$$\lambda(t) = \lambda_0(1 + A \sin 2\pi t). \tag{9}$$

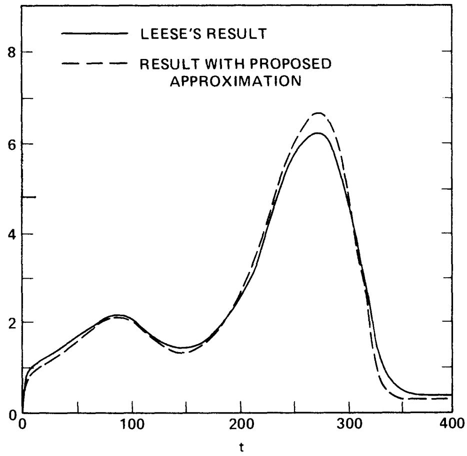

FIGURE 1. Comparison with Leese's Result for Standard Deviation of Number in the System.

TABLE 2
Comparison with Rider's Results

t	Average Number in the System Rider's Solutions Exact	Approximate with $T=0$	Approximate with $T=0.4/\mu(t)$	Our Approximation
1	4.52	4.95	4.65	4.47
2	3.63	3.72	3.86	3.67
3	2.66	2.35	2.80	2.76
4	1.29	0.81	1.15	1.39
5	0.69	0.55	0.58	0.66
6	1.39	1.43	1.28	1.33
7	1.74	1.84	1.67	1.69
8	1.28	1.24	1.28	1.28
9	1.49	1.52	1.46	1.49
10	1.17	1.14	1.16	1.18
11	1.17	1.16	1.16	1.17
12	0.90	0.88	0.89	0.90
13	0.88	0.88	0.88	0.88
14	0.88	0.88	0.88	0.88
15	0.88	0.88	0.88	0.88
16	2.40	2.69	2.27	2.34
17	3.14	3.61	3.05	3.06
18	2.81	3.07	2.84	2.81
19	3.41	3.77	3.44	3.40
20	3.77	4.17	3.83	3.74
21	2.70	2.68	2.83	2.77
22	3.18	3.20	3.12	3.22
23	4.03	4.32	3.98	3.99
24	4.53	4.97	4.54	4.47
Absolute Deviation				
Mean		0.18	0.07	0.03
Maximum		0.48	0.23	0.10

We have tested both dynamic tracking from given initial conditions and periodic limiting results. The results appear to be accurate enough for most application purposes. Figure 2 gives a typical comparison of dynamic tracking from given initial conditions. Note that the method works well starting with initial conditions in which m exceeds v. Figures 3 and 4 display typical comparisons for periodic limiting results with one and three servers. Occasionally, errors build up to a larger extent during the

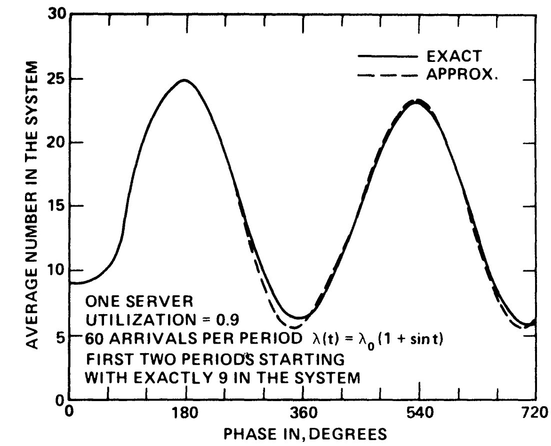

FIGURE 2. Dynamic Tracking from Initial Conditions.

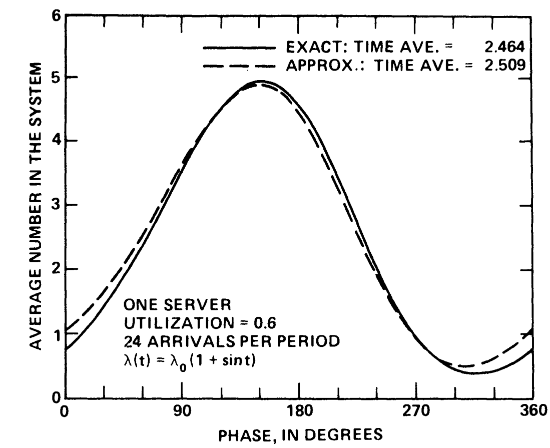

FIGURE 3. Typical Dynamic Steady State Comparison.

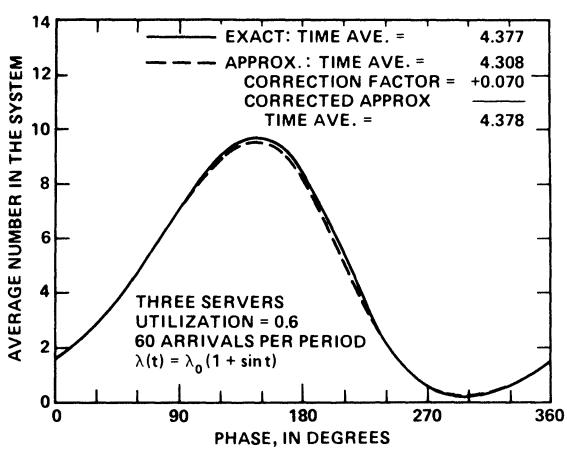

FIGURE 4. Multiserver Dynamic Steady State Comparison.

convergence to periodic limiting results. Figure 5 gives the worst example of this we have encountered. Such a buildup cannot occur if somewhere during the period the system approaches a steady state (e.g. an empty queue). Figure 6 gives a typical comparison for the standard deviation of the number in the system. In general, standard deviation fits are worse than fits for the mean. In each of these comparisons, the "exact" results were obtained by numerical integration of a truncated set of state equations. The criterion for truncation was that the long run fraction of arrivals "lost" be less than 10^{-6}.

Next, we describe the time average behavior of the number in the queuing system when the service rate is constant and the arrival rate has a constant and a sinusoidal component as in (9) with relative weight $A \leqslant 1$. These results were calculated using our approximation method. We found them useful in the motivating application of this work described above. Figure 7 shows how the relative magnitude of the sinusoidal fluctuation, A, and the arrival rate (holding utilization constant) affect the time average mean number in the system. When $A = 0$, this average is independent of the number of arrivals per "period," λ_0, provided the ratio of λ_0 to μ is held constant. When $A > 0$ but $(1 + A)\lambda_0/s\mu < 1$, the average increases with A but approaches a maximum. When $(1 + A)\lambda_0/s\mu \geqslant 1$, the average increases without bound as λ_0 increases since part of each period is spent with the arrival rate at or above service capacity. It is helpful to note that λ_0 can be interpreted both as arrivals per period and period length expressed in units of average interarrival time.

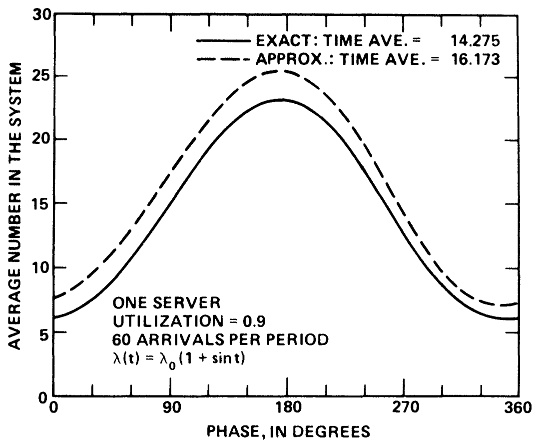

FIGURE 5. "Worst Case" Dynamic Steady State Comparison.

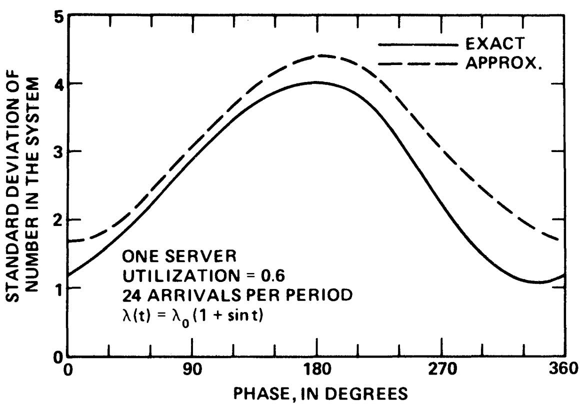

FIGURE 6. Standard Deviation Comparison for Dynamic Steady State.

Figure 8 shows how the time average number in the system varies with the utilization. Figure 9 shows how it varies with the number of servers. Notice that when the utilization multiplied by $1 + A$ exceeds unity, heavy traffic conditions apply for large λ_0 and the number of servers becomes unimportant.

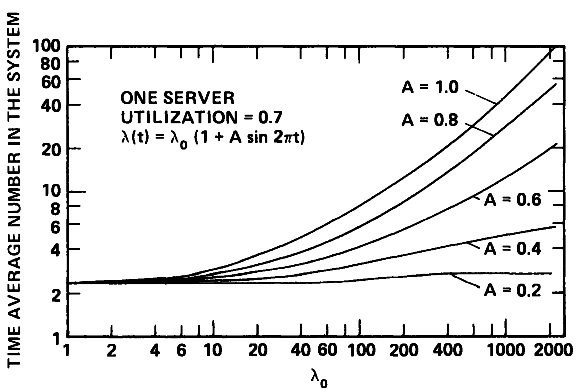

FIGURE 7. Time Average Number in the System: Various A.

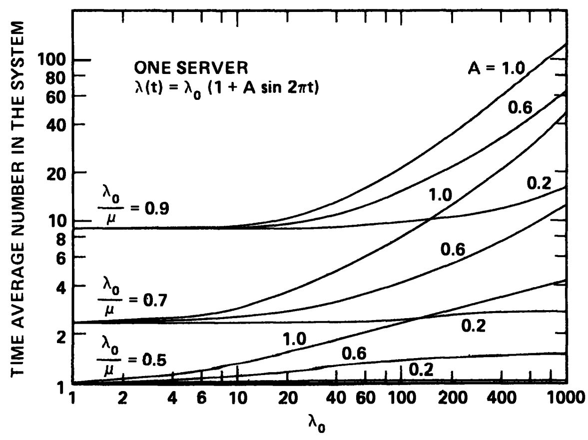

FIGURE 8. Time Average Number in the System: Various Utilizations.

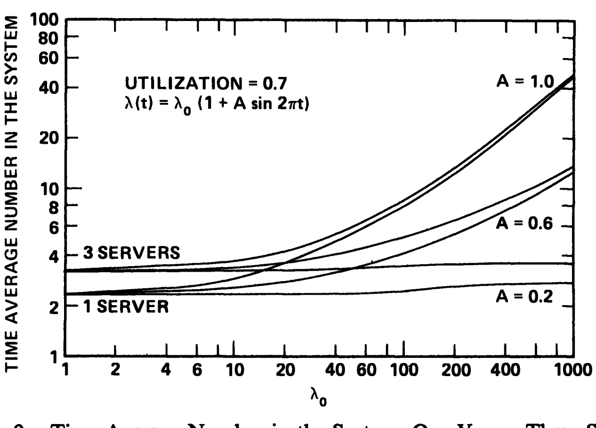

FIGURE 9. Time Average Number in the System: One Versus Three Servers.

Note that measures of system effectiveness other than the mean and standard deviation of the number in the system can easily be calculated or approximated from them. For example, the average time spent in the system by customers is just the time averaged number in the system divided by the arrival rate. For another example, the average number of customers in the queue at any time can be approximated by subtracting from the average number in the system the average number of busy servers as we have approximated it using the negative binomial distribution.

This work suggests a number of new areas for further theoretical work. First, we would like to see a more extensive analysis of the errors introduced by our approximation with the hope that simple changes in our method could improve its accuracy. These changes could consist of overall correction terms for systematically observed errors or modifications of the negative binomial approximation we have used in step 1 of our procedure. The numerical integration procedure could obviously be replaced by more advanced methods that would enable us to increase the step size without affecting convergence or losing accuracy. Our procedure for finding a limiting periodic result for periodic queues by tracking the system until it settles down could be improved in those cases in which many periods are required for the settling down. A generalization of the theorems in the appendix to multiple servers would be welcome. Finally, extension of our method to other queueing systems could make models of those systems useful in situation involving systematic arrival rate fluctuation. Priority queues, in particular, could be treated.

This work also facilitates some interesting research questions related to applications. In those queueing situations in which dynamic aspects have been ignored or grossly approximated, how good have the approximations been? Most important, however, is the contribution of this work to new application opportunties.

Appendix

This appendix presents three theorems, proved in [25], describing the properties of the proposed approximation scheme for calculating $m(t)$ and $v(t)$ in a nonstationary $M/M/1$ queue. Theorems 1 and Theorem 2 with its corollary, state some of the properties of the exact solution to the differential equations (2) and (4) with the closure assumption (6). The third theorem concerns the error propagation resulting from the numerical integration scheme applied to equations (2), (4) and (6).

THEOREM 1. *Let $m(t)$, $v(t)$ be defined by* (2) *and* (4) *with $P_0(m,v)$ given by* (6) *and with initial values $m(0) \geqslant 0$, $v(0) \geqslant 0$. Assume $\lambda(t)$ and $\mu(t)$ to be bounded nonnegative piecewise continuously differentiable functions of t such that $m(t)$ is uniformly bounded above and so that for every $t \in [0,\infty)$ and any $0 \leqslant M < \infty$, there exist a finite τ (depending on t and M) for which $\int_t^\tau \lambda(\theta)\,d\theta \geqslant M$. Then*

1. *$m(t) \geqslant 0$ for all t, and $m(t) > 0$ for almost every $t \in \{t \mid \lambda(t) > 0,\ t \geqslant 0\}$.*
2. *$v(t) \geqslant 0$ for all t, and $v(t) > 0$ for almost every $t \in \{t \mid \lambda(t) + \mu(t) > 0,\ m(t) > 0,\ t \geqslant 0\}$.*
3. *$0 \leqslant \exp(-m(t)^2/v(t)) \leqslant P_0(m(t),v(t)) \leqslant \exp(-m(t)) \leqslant 1$ for all t.*
4. *There exists a $T \in [0,\infty)$ such that $v(T) \geqslant m(T)$. Furthermore, given such a T, $v(t) \geqslant m(t)$ for $t \geqslant T$ and $v(t) > m(t)$ for almost every $t \in \{t \mid \mu(t) > 0,\ m(t) > 0,\ t \geqslant T\}$.*

THEOREM 2. *Let $m(t)$, $v(t)$ and $m(t) + \Delta m(t)$, $v(t) + \Delta v(t)$ be the respective trajectories obtained by integating* (2) *and* (4) *with P_0 given by* (6), *with nonnegative initial values $m(0)$, $v(0)$ and $m(0) + \Delta m(0)$, $v(0) + \Delta v(0)$ and with the same forcing functions $\lambda(t)$, $\mu(t)$. Assume $\lambda(t)$ and $\mu(t)$ to be bounded nonnegative piecewise continuously differentiable functions of t such that $m(t)$ is uniformly bounded above and so that for* *every* $t \in [0, \infty)$ *and any* $0 \leqslant M < \infty$, *there exist a finite* τ (*depending on* t *and* M) *for which* $\int_t^\tau \lambda(\theta)d\theta \geqslant M$. *Then*,

$$\lim_{t\to\infty} \Delta m(t) = \lim_{t\to\infty} \Delta v(t) = 0.$$

COROLLARY. *Corresponding to any pair of forcing functions* $\lambda(t)$, $\mu(t)$ *having the properties stated in Theorem* 2; *equations* (2), (4) *and* (6) *have a unique asymptotic solution* $m^*(t)$, $v^*(t)$ *satisfying* $v^*(t) \geqslant m^*(t) \geqslant 0$, *for any nonnegative initial values* $m(0)$, $v(0)$.

THEOREM 3. *Let* $m(t)$, $v(t)$ *be defined by* (2), (4) *and* (6) *with initial values* $v(0) > m(0) > 0$ *and assume* $\lambda(t)$ *and* $\mu(t)$ *are uniformly bounded and piecewise continuously differentiable with uniformly bounded derivatives on any continuous segment. Furthermore, assume* $\lambda(t)$ *and* $\mu(t)$ *are such that* $m(t)$ *is uniformly bounded by* η. *Let* m_k *and* v_k *denote the approximations to* $m(t_k)$ *and* $v(t_k)$ *obtained by Euler's numerical integration method restricted to positive values of* m_k *and* v_k, *and applied piecewise on continuously differentiable segments of* $\lambda(t)$ *and* $\mu(t)$ *with stepsizes* $\Delta_k \leqslant \Delta$ *for all* k. *Then, the error vector* z_k *between the exact and approximate trajectories satisfies*:

$$\|z_k\|_2 = \|(m(t_k) - m_k), (v(t_k) - v_k)\|_2 \leqslant \Delta L\{\exp[2Mk\Delta(2+\eta)] - 1\}/2M(2+\eta),$$

where M *is a uniform upper bound on* $\mu(t)$ *and* L *is a constant such that*

$$\|(m''(\xi), v''(\zeta))\|_2 \leqslant 2L$$

for any ξ *and* ζ.[1]

[1] We wish to acknowledge helpful communications with Charles Feinstein, J. Michael Harrison, Joseph Keller, Ernest Koenigsberg, Eric Leese, Marcel Neuts, Gordon Newell, Richard Smallwood and Stephen Smith. Ward Whitt who served as a referee provided many valuable suggestions and comments and pointed out a minor error in an earlier version of Theorems 1 and 2. We are grateful to the Xerox Corporation for support and research freedom.

References

1. CHANG, S. S. L., "Simulation of Transcient and Time Varying Conditions in Queueing Networks," Proceedings of the Seventh Annual Pittsburg Conference on Modeling and Simulation, University of Pittsburg, 1977, 1075–1078.
2. CLARKE, A. B., "A Waiting Line Process of Markov Type," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 452–459.
3. COX, D. R. AND SMITH, W. L., *Queues*, Metheuen, London, 1961, Section 3.1.
4. GAVER, D. P., "Observing Stochastic Processes and Approximate Transform Inversion," *Operations Res.*, Vol. 14 (1966), pp. 444–459.
5. HARRISON, J. M. AND LEMOINE, A. J., "Limit Theorems for Periodic Queues," *J. Appl. Probability*, Vol. 14 (1977), pp. 566–576.
6. HENRICI, P., *Elements of Numerical Analysis*, Wiley, New York, 1964.
7. HOWARD, W. J., *Turn Around Time: A Key Factor in Printing Productivity*, Xerox Corporation, Rochester, N.Y., 1976.
8. ISIHARA, A., *Statistical Physics*, Academic Press, New York, 1971, Section 6.1.
9. KELLER, J., private communication.
10. KOLSSAR, P. J., RIDER, K. L., CRABILL, T. B. AND WALKER, W. E., "A Queueing-Linear Programming Approach to Scheduling Police Patrol Cars," *Operations Res.*, Vol. 23 (1975), pp. 1045–1062.
11. KOOPMAN, B. O., "Air-Terminal Queues Under Time-Dependent Conditions," *Operations Res.*, Vol. 20 (1972), pp. 1089–1114.
12. KOTIAH, T. C., "Approximate Transient Analysis of Some Queueing Systems," *Operations Res.*, Vol. 26 (1978), pp. 333–346.
13. LEESE, E. L. AND BOYD, D. W., "Numerical Methods for Determining the Transient Behavior of Queues with Variable Arrival Rates," *J. Canad. Operational Res. Soc.*, Vol. 4 (1966), pp. 1–13.

14. LEESE, E. L., MACKAY, J. G. I. AND DEWAR, D. W., "Comparison of Analytic and Monte Carlo Methods of Numerical Computation of the Transient Behavior of Queues with Variable Arrival Rate," Canadian Department of National Defence, Operational Research Division, Directorate of Mathematics and Statistics, ORD Report No. 65/R11, October, 1965.

1 LUCHAK, G., "The Solution of the Single-Channel Queuing Equations Characterized by a Time-Dependent Arrival Rate and a General Class of Holding Times," *Operations Res.*, Vol. 4 (1956), pp. 711–732.

1 Moore, S. C., "Approximating the Behavior of Nonstationary Single-Server Queues," *Operations Res.* Vol. 23 (1975), pp. 1011–1032.

17. MORSE, P. M., *Queues, Inventories and Maintenance*, Wiley, New York, 1958.

18. NEUTS, M. F., "The Single Server Queue in Discrete Time Analysis I," *Naval. Res. Logist. Quart.*, Vol. 20 (1973), pp. 297–304.

19. ———, "The $M/M/1$ Queue with Randomly Varying Arrival and Service Rates," *Opsearch*, Vol. 15 (1978), pp. 139–157.

20. ———, "Further Results on the $M/M/1$ Queue and Randomly Varying Rates," *Opsearch*, Vol. 15 (1978), pp. 158–168.

2 NEWELL, G. F., "Queues with Time-Dependent Arrival Rates: I, II, and III," *J. Appl. Probability*, Vol. 5 (1968), pp. 436–451; 579–590; 591–606.

22. ———, *Applications of Queueing Theory*, Chapman and Hall, London, 1971.

2 OLIVER, R. M. AND SAMUEL, A. H., "Reducing Letter Delays in Post Offices," *Operations Res.*, Vol. 10 (1962), pp. 839–892.

24. RIDER, K. L., "A Simple Approximation to the Average Queue Size in the Time-Dependent $M/M/1$ Queue," *J. Assoc. Comput. Mach.*, Vol. 23 (1976), pp. 361–367.

25. ROTHKOPF, M. H. AND OREN, S. S., "The Nonstationary $M/M/s$ Queue," Report ARG 77-10, Xerox Palo Alto Research Center, Palo Alto, Calif., October 1977; revised February 1979.

26. WAGNER, H. M., *Principles of Operations Research*, 2nd edition, Prentice Hall, Englewood Cliffs, N.J., 1975.

27. WANG, K.-K. G., "Continuous Simulation of Queueing Networks with Transient and Time Varying Conditions," Ph.D. Dissertation, State University of New York at Stoneybrook, forthcoming.

28. WRAGG, A., "The Solution of an Infinite Set of Differential-Difference Equations Occurring in Polymerization and Queueing Problems," *Proc. Cambridge Philos. Soc.*, Vol. 59 (1963), pp. 117–124.

XEROX ANALYSIS RESEARCH GROUP
PALO ALTO RESEARCH CENTER
3333 COYOTE HILL ROAD, PALO ALTO, CA 94304

THE NONSTATIONARY M/M/S QUEUE

by

Michael H. Rothkopf and Shmuel S. Oren

October 1977
(revised February 1979)

ARG Technical Report No. 77-10

THE NONSTATIONARY M/M/s QUEUE

by

Michael H. Rothkopf and Shmuel S. Oren
Analysis Research Group

October 1977
Revised February 1979

Abstract

Typical queues do not have constant arrival rates. This paper discusses effective computational methods for dealing with queues having nonstationary arrival processes. It presents a computationally undemanding approximate method for finding the time dependent mean and standard deviation of the number of customers in an s server queueing system with time-varying arrival and service rates. Results are exhibited for various 1 and 3 server queues with constant service rates and sinusoidal components in the arrival rate.

Introduction

Queueing models were among the first models of operations research. The leading introductory texts of operations research have chapters that develop the elementary theory and give examples of actual situations involving queues. These examples involve situations in which the arrival process typically has significant and predictable time-dependent characteristics. Wagner [26], for example, lists grocery store checkout stands, bank tellers' windows, service station pumps and attendants, telephone trunk lines, and photocopy equipment maintenance men. Most service facilities of these kinds experience predictable demand fluctuations. Although elementary queueing theory dealing with stationary behavior is useful in many areas, it is inadequate for dealing with situations in which the time dependent behavior is important. This is often true for queues that go through periods in which the arrival rate exceeds the service rate. There are a variety of ways of attempting to bridge the gap between elementary theory and practical, nonstationary problems. This paper reviews several of them and presents a new "closure" approximation method for computing the time-dependent mean and variance of the number in the system of a nonstationary M/M/s queue. This method is computationally tractable and usable with any dynamic pattern of arrival (and/or service) rates and with any number of servers. Experiments show that the method gives close approximations when used for calculating transient behavior and limiting periodic behavior. We illustrate the use of the method by computing graphs showing the average number in one and three server systems when the arrival rate has both constant and sinusoidal components.

This work has been used to model the capacity acquisition problems faced by the managers of a variety of central document reproduction departments. These departments typically have predictable fluctuations - time of day, day of the week, and sometimes day of the month - in the rate of arrival of their work. In some reproduction departments these fluctuations are mild, but often they are large enough that the peak period arrival rate significantly exceeds an economically reasonable capacity level. Detailed data on particular central reproduction departments is expensive to obtain, but estimates are usually readily available for the average arrival rate of jobs, the average time to do a job, and on the period and relative magnitude of systematic fluctuations in workload. For capacity acquisition decisions it is usually adequate to assume that job times are exponentially distributed and that arrivals are generated by a nonhomogeneous Poisson process, i.e, a Poisson process with time dependent arrival rate. Making these decisions requires only a few simple measures of effectiveness - primarily the average wait in the system (which can be calculated from the average number in the system).

Problems of this type can be approached in various ways. The simplest way is to ignore the fluctuations and use standard steady state analysis. This is appropriate if the fluctuation are mild or if the period of fluctuation is sufficiently short that only a few arrivals are likely to occur during a peak period. If the fluctuations do not let the arrival rate get as large as the service rate and if the period of the fluctuation is sufficiently long, then the queue can be analyzed as one that goes through a sequence of steady states. A different approach described by Neuts [20,21] is to model the fluctuations in the arrival rate as being generated by an underlying Markov process.

Monte Carlo simulation is a common approach to such problems. It can handle virtually any complication and can compute virtually any measure of effectiveness if enough computer time is used. The required computer time, however, can be substantial, and it grows as the square of the required precision.

The general transient solution for the stationary single server queue is known [4]. Unfortunately, it involves an infinite sum of Bessel functions and would not normally be used for numerical computation. Morse's book [18] contains closed form transient solutions for the finite state single server queue with constant arrival rate. By discretizing the dynamic change of the arrival rate, this could conceivably be adapted to the dynamic case, but we are not aware of any attempt to implement this approach. Commonly, numerical approaches deal directly with the differential equation of the queue. An important paper by Leese and Boyd [14] examined seven such numerical methods for the nonstationary single server queue. These methods, including Clarke's [3], Luchak's [16], and Wragg's [28], are compared to Monte Carlo simulation. Of the seven, only Leese and Boyd's adaptation of Wragg's method seemed to them computationally superior to simulation. Koopman [12], and more recently Kolesar *et al.* [11], have solved nonstationary queueing problems by numerically integrating a truncated set of the state probability equations. This is appears to be a useful method for problems in which one can assume a reasonably low upper limit on queue length. Rather than numerically integratating the state probability differential equations, it is possible to approximate a queue by a discrete time Markov process and solve numerically the resulting state probability difference equations. Neuts [19] discusses the computational aspects of such discrete time Markov queues.

When a queue experiences an arrival rate substantially above its service rate for a period sufficiently long relative to the average interarrival time to allow the law of large numbers to apply, a deterministic model of queue build up and decay will provide a good approximation. Such a method is described by Oliver and Samuel [24], by Newell [22] and in the booklet by Howard [8] commissioned and published by the Xerox Corporation for distribution to reproduction center managers. When the bulk of the delay in a system is caused by predictable surges of demand beyond service capacity, such deterministic models provide good approximations to the overall delay.

Newell [22,23] presents a diffusion equation that approximates the behavior of a heavily used queue as the arrival rate passes through the service rate. Newell's method was intended to be useful for the design of transportation facilities in which there are many servers and hundreds or thousands of arrivals during a rush hour surge. For any such heavy traffic situations his methods are valuable. Because he assumes heavy traffic, his approximations do not depend upon the service time distribution or the number of servers.

There are also transform inversion approximation approaches for analyzing nonstationary queues. Such methods have been suggested by Gaver [5] and recently by Kotiah [13].

The approach taken in this paper belongs to a general class of methods known as *closure techniques* for approximating solutions of infinite systems of equations. These are commonly used in physics. The basic strategy of a closure technique is to reduce an infinite system of equations to a finite system by making a "closure assumption" in the form of a functional relationship between the variables of the system. For instance, the well known Maxwell-Boltzman equations are obtained by closing an infinite hierarchy of differential equations for the probability density functions over the position and momentum of particles in a fluid [9]. The key to the success of a closure method is its closure assumption, while the test of any such assumption is the empirical usefulness of the resulting equations. The experience in physics is that systematic approaches to derive such assumptions have not been as effective as intuition [10].

The finite state approximation technique employed by Koopman [12] and Kolesar [11] can be viewed as an elementary closure method in which the infinite system of equations is closed by assuming zero probability beyond some finite number in the system. Another closure method for approximating the time dependent mean of a nonstationary M/M/1 queue has been recently published by Rider [25]. He combines the state equations to obtain the differential equation for the mean in terms of the idle probability. He then makes a closure assumption relating the idle probability to the mean number in the system, the arrival rate and the service rate. Our approach is somewhat similar to Rider's approach but it is more general and appears to produce better approximations. We use the differential equations for the mean and variance for the M/M/1 queue due to Clarke [3] and generalize them to the M/M/s case. These equations are expressed in terms of the state probabilities for states with fewer customer than servers. Our closure assumption employs the negative binomial distribution to approximate the state probability distribution in terms of the mean and variance of the number in the system. This leaves us with a pair of differential equations for the mean and variance of the number in the system. We integrate these equations by standard numerical methods.

At the time of the final revision of this paper, we learned of recent unpublished work by Chang [2] and his student Wang [27]. Chang analyzes networks of M/M/1 queues using an approach similar to ours but with a different closure assumption relating the idle probability to the mean and variance of the number in the system. Wang's work attempts to improve that closure assumption and generalize the results to other queueing systems.

The next section of this paper describes in detail the method we propose and its characteristics. The following section displays and evaluates some results we have obtained using this method. Finally, the appendix contains theorems stating its theoretical properties and their proofs.

The Method

For ease of exposition, we consider first the single server queue with Poisson arrivals at a nonnegative time varying rate $\lambda(t)$, and exponential service at a time varying nonnegative rate, $\mu(t)$. Let $P_n(t)$ be the probability that there are n customers in the system at time t, and let $P_n'(t)$ be the derivative of $P_n(t)$ with respect to time. For convenience, we will henceforth omit the argument t except where it is required to avoid ambiguity. The standard differential-difference equations for the M/M/1 queue are

$$P_n' = -(\lambda+\mu)P_n + \lambda P_{n-1} + \mu P_{n-1}, \qquad n = 1,2,\ldots$$

(1)

$$P_0' = -\lambda P_0 + \mu P_1 \quad .$$

Denoting the average number in the system as $m(t)$ and its derivative by m', we can obtain from (1) the intuitively obvious equation

(2) $$m' = \sum_{n=0}^{\infty} nP_n' = \lambda - \mu(1-P_0).$$

Similarly, the derivative of mean squared number in the system is given by

(3) $$\overline{n^2}' = \sum_{n=0}^{\infty} n^2 P'n = \lambda + \mu(1-P_0) + 2m(\lambda-\mu).$$

Hence the derivative of the variance, v, of the number in the system is

(4) $$v' = \overline{n^2}' - 2mm' = \lambda + \mu - \mu P_0(2m + 1).$$

Equations (2) and (4) were obtained by Clarke [3]. He used them to derive limiting formulas for the mean and variance of the number in the system for the M/M/1 queue with λ/μ constant and $\lambda \geq \mu$. In order to analyze more general nonstationary behavior, we introduce our closure assumption expressing P_0 in terms of m and v, using the negative binomial distribution.

The negative binomial distribution given by

(5) $$P_n = \binom{r + n - 1}{n} p^r(1 - p)^n, \qquad n = 0,1,2, \ldots$$

is fully specifiable by its mean, $r(1-p)/p$, and variance, $r(1-p)/p^2$. In particular, P_0 can be expressed as

(6) $P_0(m,v) = p^r$, where $p = m/v$ and $r = m^2/(v-m)$.

The negative binomial distribution has the desirable property that when its variance equals its mean plus the square of its mean it specializes to the geometric distribution on nonnegative integers. This is the steady state distribution of the stationary M/M/1 queue. Thus, if this method is applied to a stationary M/M/1 queue, then the assumed negative binomial distribution will asymptotically approach the correct steady state distribution of the number in the system.

It can be shown that for all nonnegative m and v the function $P_0(m,v)$ defined by (6) is monotonically increasing in v and monotonically decreasing in m and that $0 \leq \exp(-m^2/v) \leq P_0 \leq \exp(-m) \leq 1$. These properties of $P_0(m,v)$ are compatible with the way one would intuitively expect the idle probability of an M/M/1 queue to behave. Furthermore, the fact that these properties hold for all nonnegative values of m and v allows the use of $P_0(m,v)$ as an approximation to the idle probability even for m>v when the negative binomial distribution is not well defined.

For M/M/1 queues, our calculation scheme amounts to integrating the pair of differential equations (2) and (4) with P_0 given by (6) when $v \neq m$ and by the limit of (6), exp(-m), when $v = m$. The scheme proceeds as follows:

1. Use m(t) and v(t) to obtain $P_0(t)$ using (6) if $v \neq m$ and exp(-m) if $v = m$.
2. Use this P_0 in (2) and (4) to obtain $m'(t^+)$ and $v'(t^+)$.
3. For some small $\Delta > 0$, set $\Delta_t = \text{Min}[\Delta, \delta_t]$, where δ_t is the time increment to the next discontinuity in . $\mu(t)$ or $\lambda(t)$.
4. Approximate $m(t+\Delta_t)$ by $\text{Max}[0, m(t)+\Delta_t m'(t^+)]$ and $v(t+\Delta_t)$ by $\text{Max}[0, v(t)+\Delta_t v'(t^+)]$.
5. Increment t by Δ_t and repeat.

This computational scheme is just forward Euler integration applied piecewise over continuous segments of $\mu(t)$ and $\lambda(t)$ [7].

When the queue has s servers, each working at rate μ, equations (2) and (4) generalize to

$$(7)\ m' = \lambda = \mu s + \mu \sum_{n=0}^{s-1} (s-n)P_n$$

and

$$(8)\quad v' = \lambda + \mu s - \mu \sum_{n=0}^{s-1} (2m + 1-2n)(s-n)P_n.$$

To "close" equations (7) and (8), we need the probabilities P_0, P_1, ... P_{s-1} which we approximate as before using (5) with $p=m/v$ and $r=m^2/(v-m)$. Unfortunately for s>1, such an approximation no longer produces the correct steady state distribution for the stationary case. Our experiments indicate, however, that the error in the mean and variance resulting from this discrepency is small. To remove this error, we add a correction term to the value of m obtained from (5), (7) and (8). This correction is determined so as to offset the steady state error for the stationary case and is calculated by comparing standard analytical results for stationary queues with the results of uncorrected calculations with the algorithm for the same queues. Table 1 gives the correction for various two and three server queues. As an example of the use of the table, when we below calculate the dynamic steady state mean number in the system for a particular three server periodic queue with 60% time average utilization to be 4.308, we increase this by .070 to 4.378. A similar correction is made in the variance. Aside from these corrections, no other changes have been made in going from the single to the multiple server case.

It is interesting to note that the summation in (7) is just the mean number of idle servers and that the summation in (8) can be replaced by an expression involving the mean and variance of the number of idle servers.

The schemes described above for the M/M/1 and the M/M/s queues can be used for two kinds of problems. Given specific initial conditions for m and v, this method enables us to approximate the transient behavior of a queue by tracking the time dependent mean and variance. For queues with periodic arrival and/or service rates (with some common period) the method can be used to approximate the limiting periodic mean and variance of the number in the system. (Recent work by Harrison and Lemoine [6] proves the existence of a periodic limiting distribution for the periodic M/G/1 queue.). To obtain such limiting behavior, we start with some arbitrary nonnegative initial values satisfying $v(0) \geq m(0) \geq 0$ and track m(t) and v(t) in time until we detect periodicity. Specifically, using the smallest common period of the arrival and service rates, we track the system

Table 1

Amount by Which True Mean and Standard Deviation of Number in System Exceeds Calculated Value in Stationary Steady State

Number of Servers	Utilization	Excess m	Excess σ_n	Percent of True m	Percent of True σ_n
1	any	0	0	0	0
2	.1	.00055	.0023	.27	.44
	.2	.0036	.0101	.86	1.5
	.3	.0100	.026	1.5	2.9
	.4	.020	.051	2.1	4.5
	.5	.032	.091	2.4	6.0
	.6	.045	.14	2.4	7.1
	.7	.059	.23	2.1	8.0
	.8	.065	.37	1.5	8.2
	.9	.0102	.69	.11	7.2
3	.1	.00011	.00046	.038	.079
	.2	.0016	.0044	.27	.59
	.3	.0072	.018	.77	1.8
	.4	.0193	.045	1.5	3.7
	.5	.040	.098	2.3	6.1
	.6	.070	.18	3.0	8.8
	.7	.108	.32	3.3	11.2
	.8	.155	.59	3.1	13.0
	.9	.154	1.21	1.5	12.7

through several periods until the beginning values of m and v for two succesive common periods are the same (within a small tolerance). We then compute m and v over the following period.

The appendix to this paper proves three theorems that give a theoretical analysis of the proposed method for the M/M/1 queue. In this analysis, we focus on two aspects of the method. First we examine issues concerning our closure assumption, and then we analyze the error propagation resulting from our numerical integration. In particular, we show that for any nonnegative initial values m(0), v(0), the solution m(t), v(t) of (2), (4) and (6) will be nonnegative for all t and consequently $0 \leq P_0(t) \leq 1$ for all t. We also show that under reasonable assumptions on $\mu(t)$ and $\lambda(t)$ any trajectory of m(t) and v(t) generated by (2), (4) and (6) with $v(0) \geq m(0) \geq 0$ will asymptotically approach any other trajectory generated by these equations with nonnegative initial values. This result justifies our assumption $v(0) \geq m(0) \geq 0$ when using the procedure to obtain limiting periodic solutions. Finally we show that with the above restriction on the initial values, one can always select a sufficiently small Δ so that the numerical integration will approximate the exact solution to (2), (4) and (6) on any finite time interval within any given tolerance. While we have not established the above results for the general M/M/s case, our computational experience leads us to believe that similar results hold for it.

Results and Discussion

Since this method is approximate and we have no analytic bounds on the approximation, it is necessary to test the method's accuracy by means of computational comparisons. We have performed several kinds of comparisons. First, we obtained from Eric Leese and ran the test problem used in [14]. That problem involved a widely fluctuating arrival rate and one server. Unfortunately, only the plots of the results of reported in [14] (shown there as Figures 5 and 7) are now available. The results we obtained for the mean number in the system agree everywhere with Leese and Boyd's exact result to within the accuracy of their plot. This is not true for the standard deviation of the number in the system. However, the nature of the deviation and a comparison with the figures in [15] and Figure 8 in [14] leads us to believe that Figure 7 in [14] is an incorrectly labled plot of the expected number in the queue. Figure 1 presents a comparison of our results with the plot of the standard deviation of the number in the system for this test problem as reported in

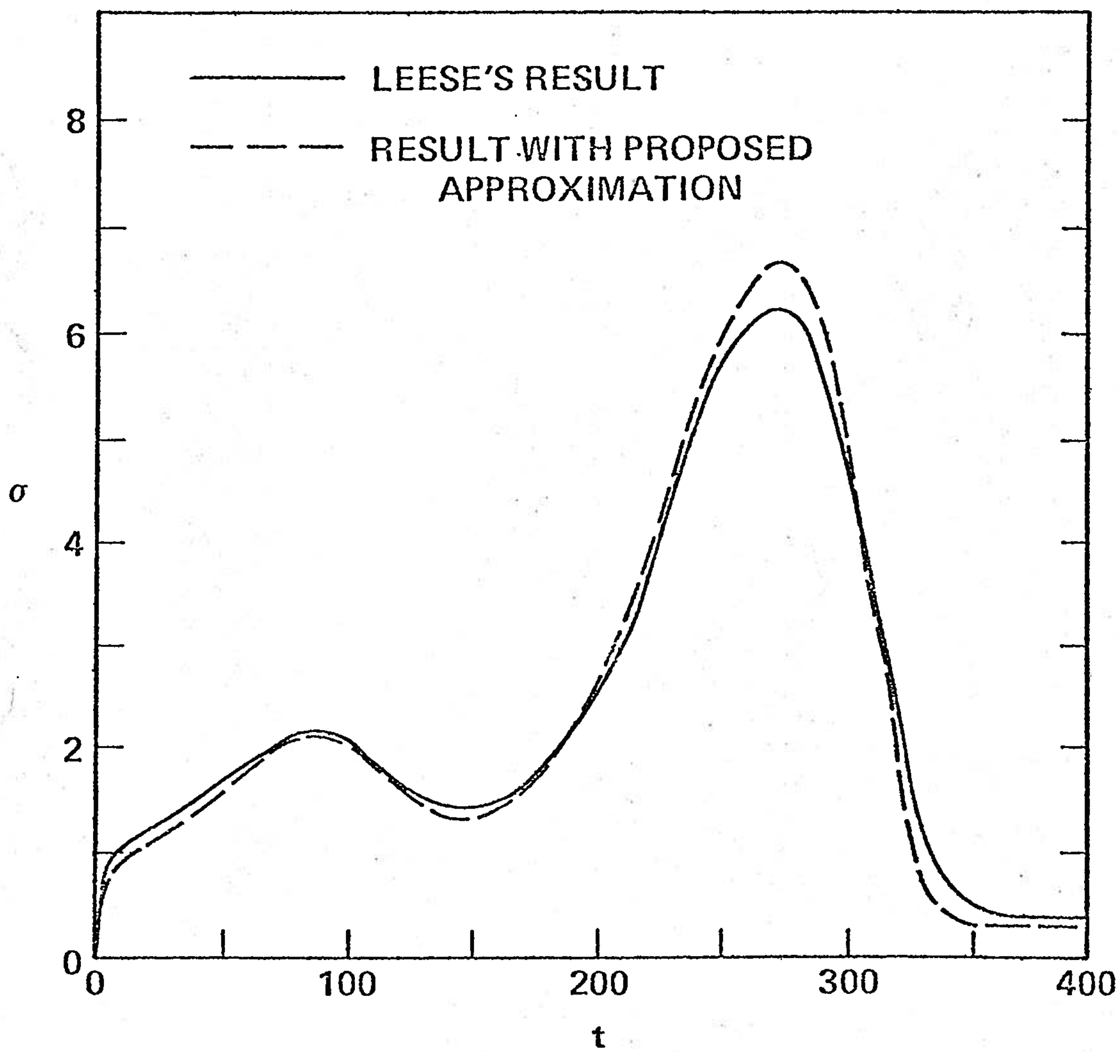

Figure 1: Comparison with Lees's Result for Standard Deviation of Number in the System

Figure 6 of [15].

The only other published test problem we are aware of is given by Rider [25]. It is a single server periodic queue with widely varying arrival and service rates. The solution Rider reports is a periodic limiting result. Our method solved that problem more accurately than Rider's method. Table 2 gives the comparison. In that table, the "exact" solution is the one reported by Rider and verified by us by numerical integration of a truncated version of (1).

We have also run a number of tests of our method with the arrival rate given by

(9) $\lambda(t) = \lambda_0(1+A \sin 2\pi t)$.

We have tested both dynamic tracking from given initial conditions and periodic limiting results. The results appear to be accurate enough for most application purposes. Figure 2 gives a typical comparison of dynamic tracking from given initial conditions. Note that the method works well starting with initial conditions in which m exceeds v. Figures 3 and 4 display typical comparisons for periodic limiting results with one and three servers. Occasionally, errors build up to a larger extent during the convergence to periodic limiting results. Figure 5 gives the worst example of this we have encountered. Such a buildup cannot occur if somewhere during the period the system approaches a steady state (e.g. an empty queue). Figure 6 gives a typical comparison for the standard deviation of the number in the system. In general, standard deviation fits are worse than fits for the mean. In each of these comparisons, the "exact" results were obtained by numerical integration of a truncated set of state equations. The criterion for truncation was that the long run fraction of arrivals "lost" be less than 10^{-6}.

Next, we describe the time average behavior of the number in the queuing system when the service rate is constant and the arrival rate has a constant and a sinusoidal component as in (9) with relative weight $A \leq 1$. These results were calculated using our approximation method. We found them useful in the motivating application of this work described above. Figure 7 shows how the relative magnitude of the sinusoidal fluctuation, A, and the arrival rate (holding utilization constant) affect the time average mean number in the system. When $A = 0$, this average is independent of the number of arrivals per "period", λ_0, provided the ratio of λ_0 to μ is held constant. When $A > 0$ but $(1+A)\lambda_0/s\mu < 1$, the average increases with A but approaches a maximum. When $(1+A)\lambda_0/s\mu \geq 1$, the average increases without bound as λ_0 increases since part

Table 2

Comparison with Rider's Results

Average Number in the System

t	Rider's Solutions			Our Approximation
	Exact	Approximate with T=0	Approximate with $T=.4/\mu(t)$	
1	4.52	4.95	4.65	4.47
2	3.63	3.72	3.86	3.67
3	2.66	2.35	2.80	2.76
4	1.29	.81	1.15	1.39
5	.69	.55	.58	.66
6	1.39	1.43	1.28	1.33
7	1.74	1.84	1.67	1.69
8	1.28	1.24	1.28	1.28
9	1.49	1.52	1.46	1.49
10	1.17	1.14	1.16	1.18
11	1.17	1.16	1.16	1.17
12	.90	.88	.89	.90
13	.88	.88	.88	.88
14	.88	.88	.88	.88
15	.88	.88	.88	.88
16	2.40	2.69	2.27	2.34
17	3.14	3.61	3.05	3.06
18	2.81	3.07	2.84	2.81
19	3.41	3.77	3.44	3.40
20	3.77	4.17	3.83	3.74
21	2.70	2.68	2.83	2.77
22	3.18	3.20	3.12	3.22
23	4.03	4.32	3.98	3.99
24	4.53	4.97	4.54	4.47
Absolute Deviation				
Mean		.18	.07	.03
Maximum		.48	.23	.10

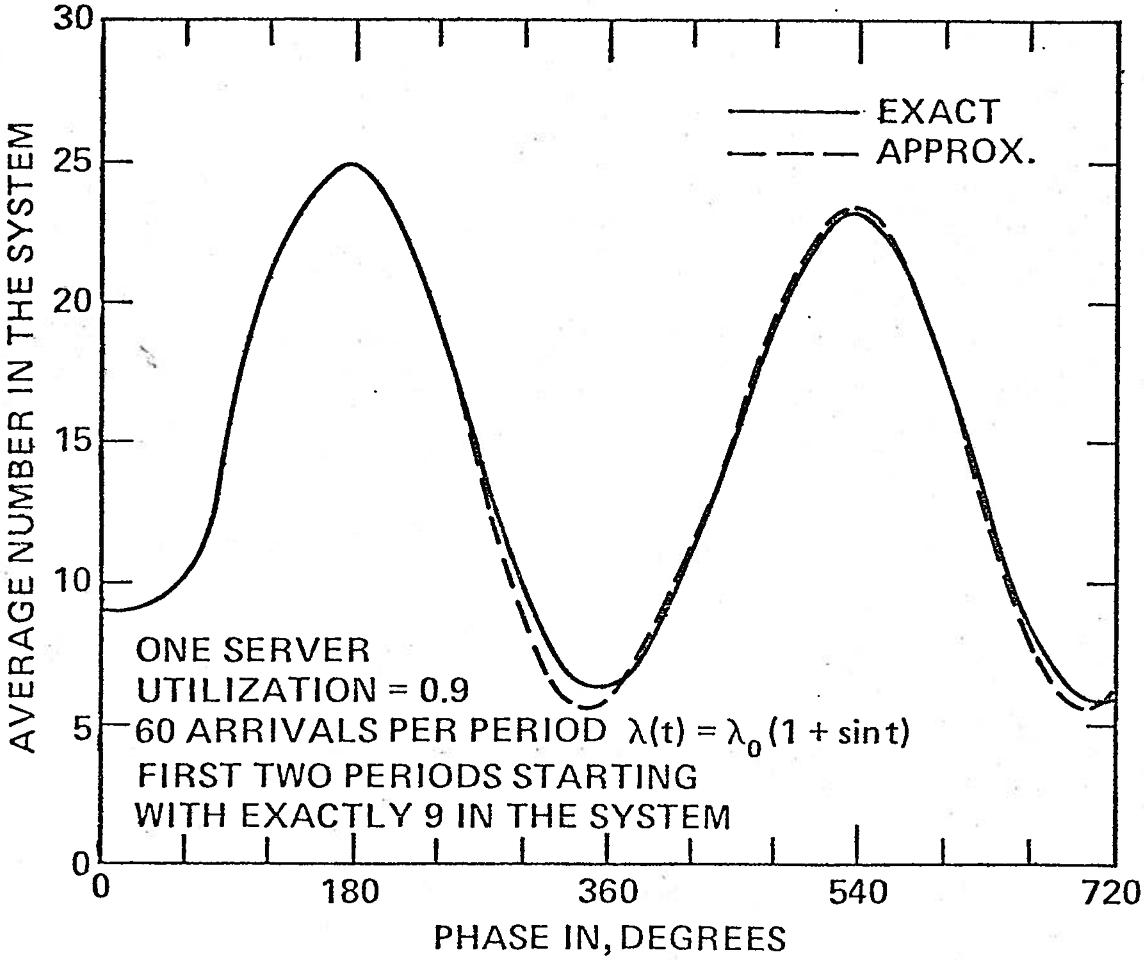

Figure 2

Dynamic Tracking from Initial Conditions

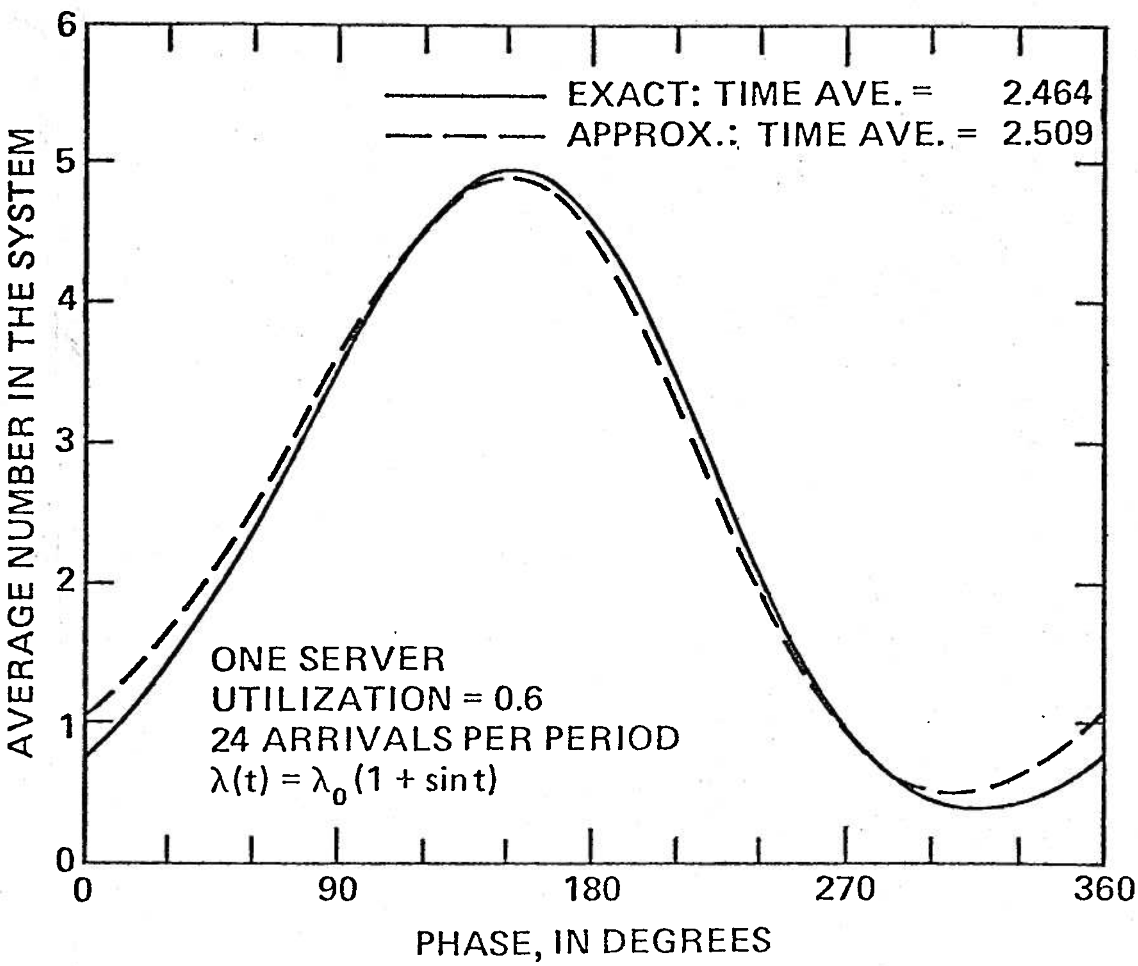

Figure 3

Typical Dynamic Steady State Comparison

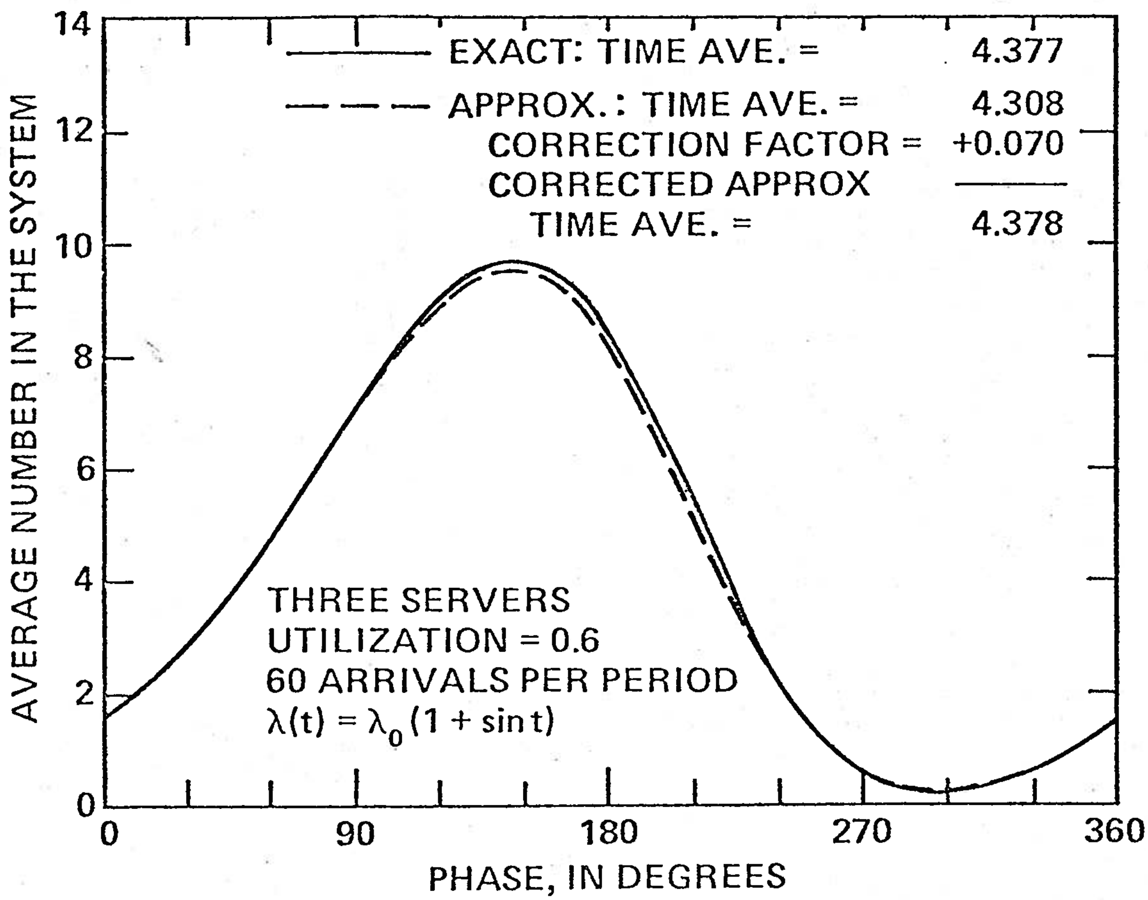

Figure 4

Multiserver Dynamic Steady State Comparison

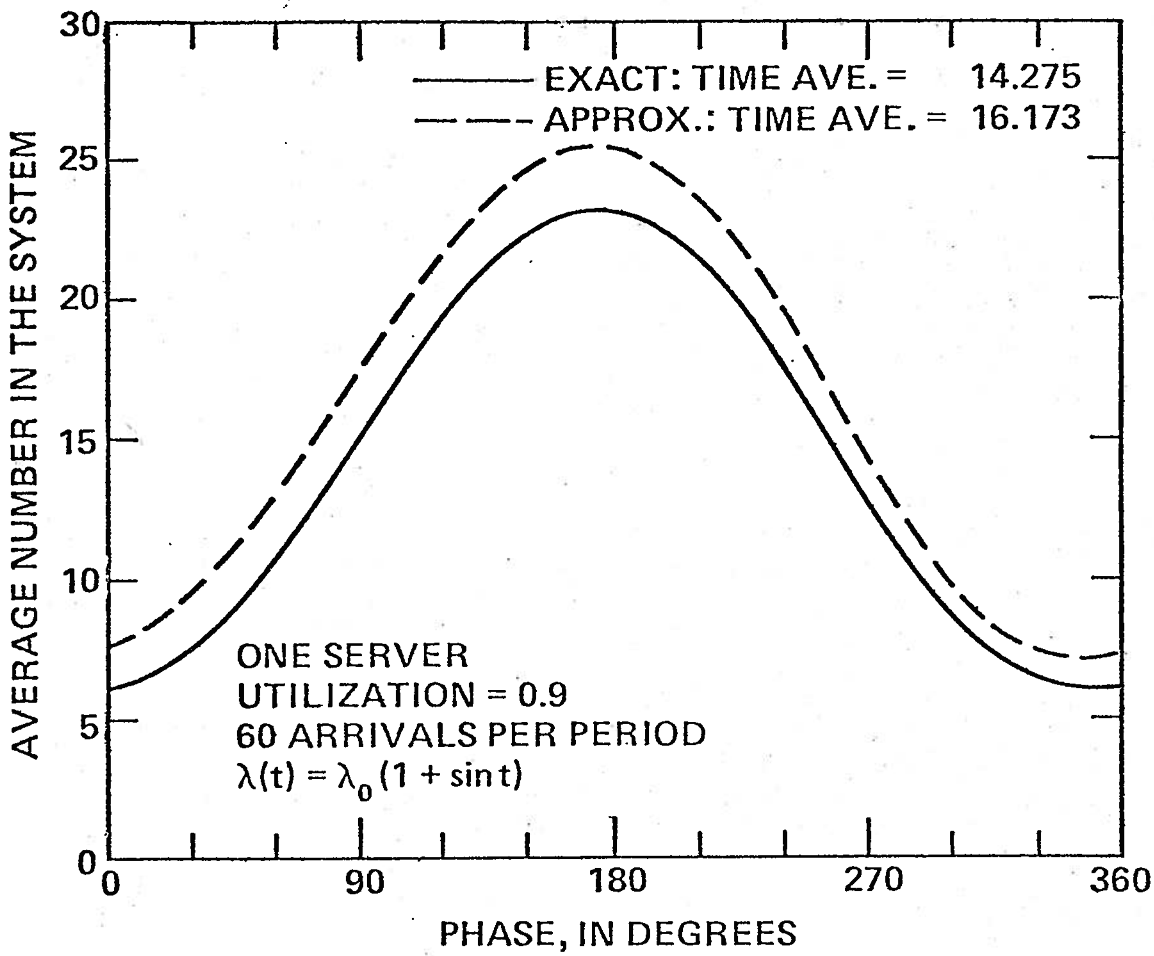

Figure 5

"Worst Case" Dynamic Steady State Comparison

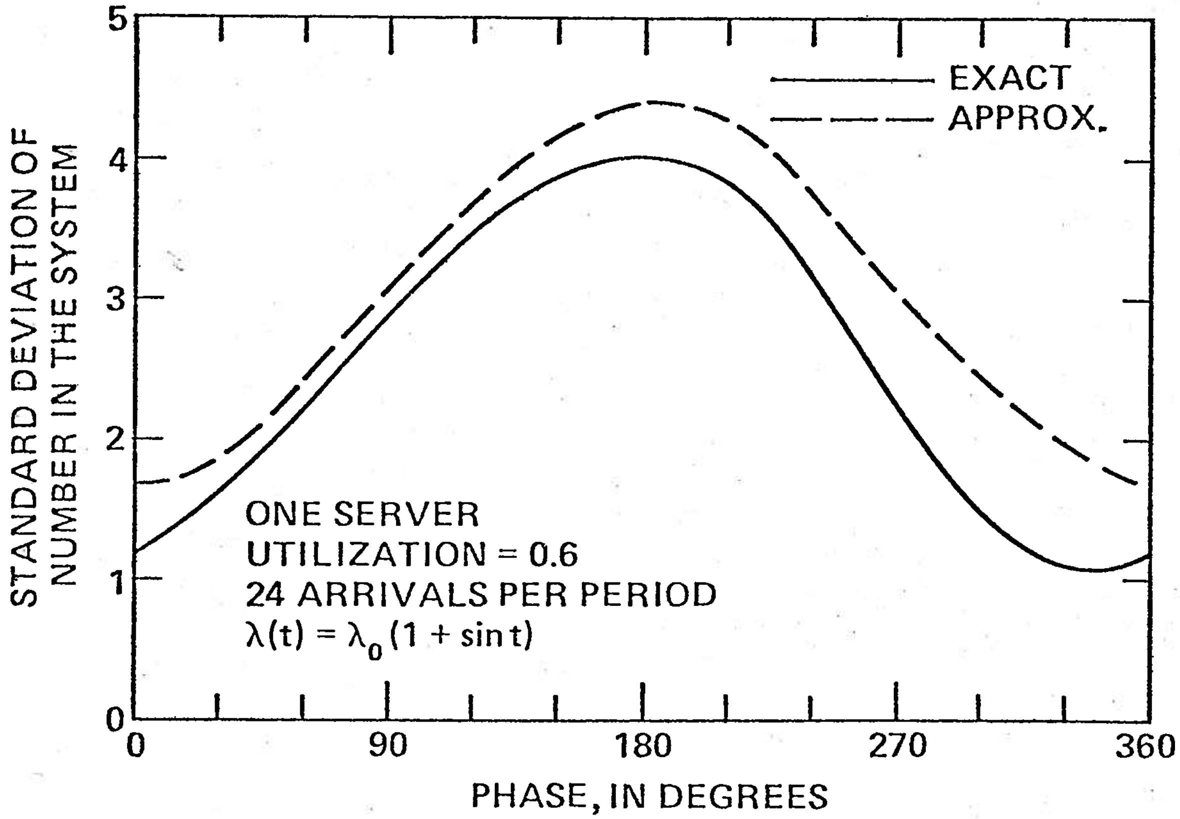

Figure 6

Standard Deviation Comparison for Dynamic Steady State

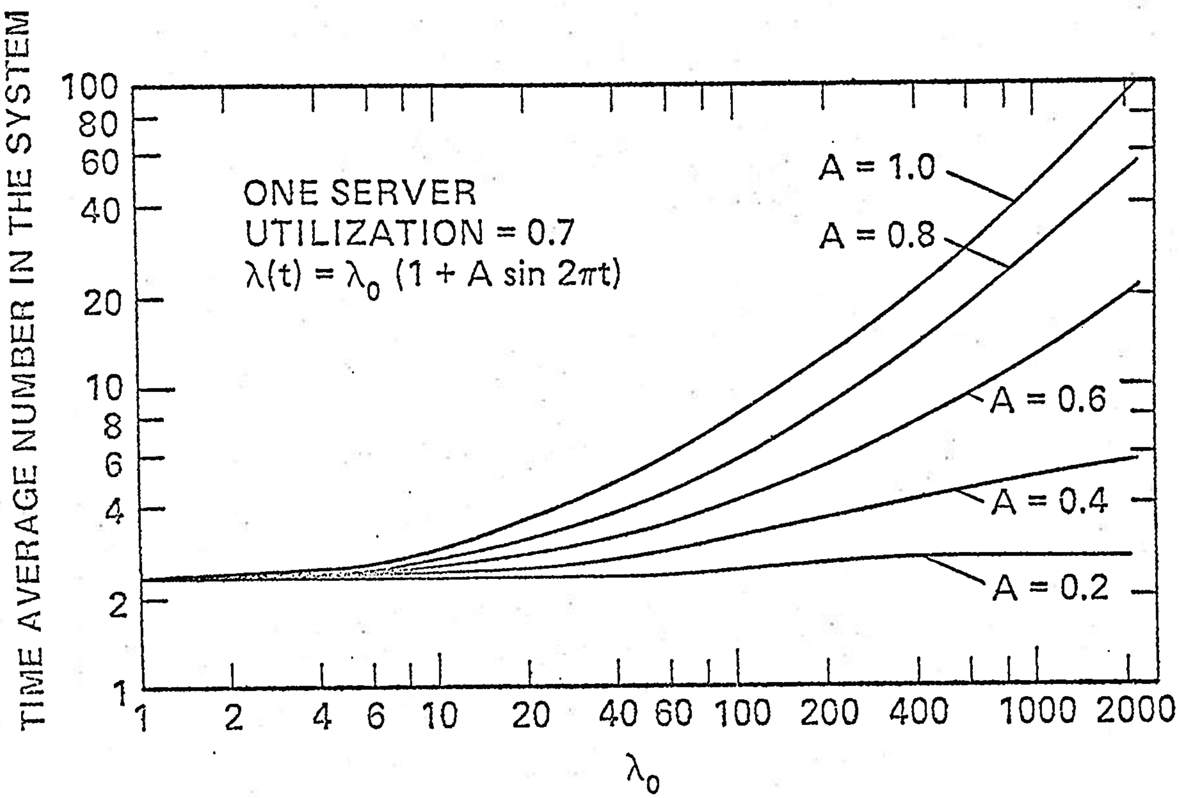

Figure 7: Time Average Number in the System: Various A

of each period is spent with the arrival rate at or above service capacity. It is helpful to note that λ_0 can be interpreted both as arrivals per period and period length expressed in units of average interarrival time.

Figure 8 shows how the time average number in the system varies with the utilization. Figure 9 shows how it varies with the number of servers. Notice that when the utilization multiplied by $1+A$ exceeds unity, heavy traffic conditions apply for large λ_0 and the number of servers becomes unimportant.

Note that measures of system effectiveness other than the mean and standard deviation of the number in the system can easily be calculated or approximated from them. For example, the average time spent in the system by customers is just the time averaged number in the system divided by the arrival rate. For another example, the average number of customers in the queue at any time can be approximated by subtracting from the average number in the system the average number of busy servers as we have approximated it using the negative binomial distribution.

This work suggests a number of new areas for further theoretical work. First, we would like to see a more extensive analysis of the errors introduced by our approximation with the hope that simple changes in our method could improve its accuracy. These changes could consist of overall correction terms for systematically observed errors or modifications of the negative binomial approximation we have used in step 1 of our procedure. The numerical integration procedure could obviously be replaced by more advanced methods that would enable us to increase the step size without affecting convergence or loosing accuracy. Our procedure for finding a limiting periodic result for periodic queues by tracking the system until it settles down could be improved in those cases in which many periods are required for the settling down. A generalization of the theorems in the appendix to multiple servers would be welcome. Finally, extension of our method to other queueing systems could make models of those systems useful in situation involving systematic arrival rate fluctuation. Priority queues, in particular, could be treated.

This work also facilitates some interesting research questions related to applications. In those queueing situations in which dynamic aspects have been ignored or grossly approximated, how good have the approximations been? Most important, however, is the contribution of this work to new application opportunities.

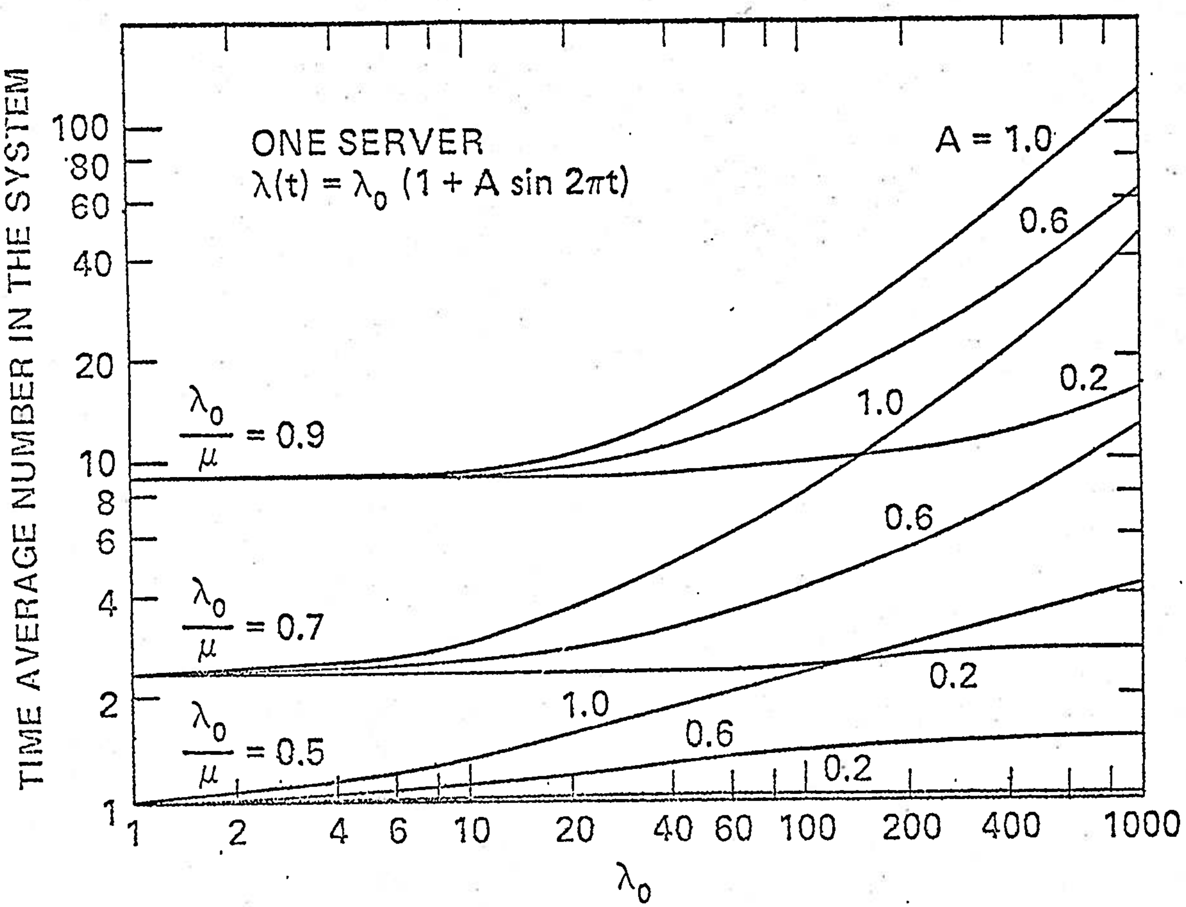

Figure 8: Time Average Number in the System: Various Utilizations

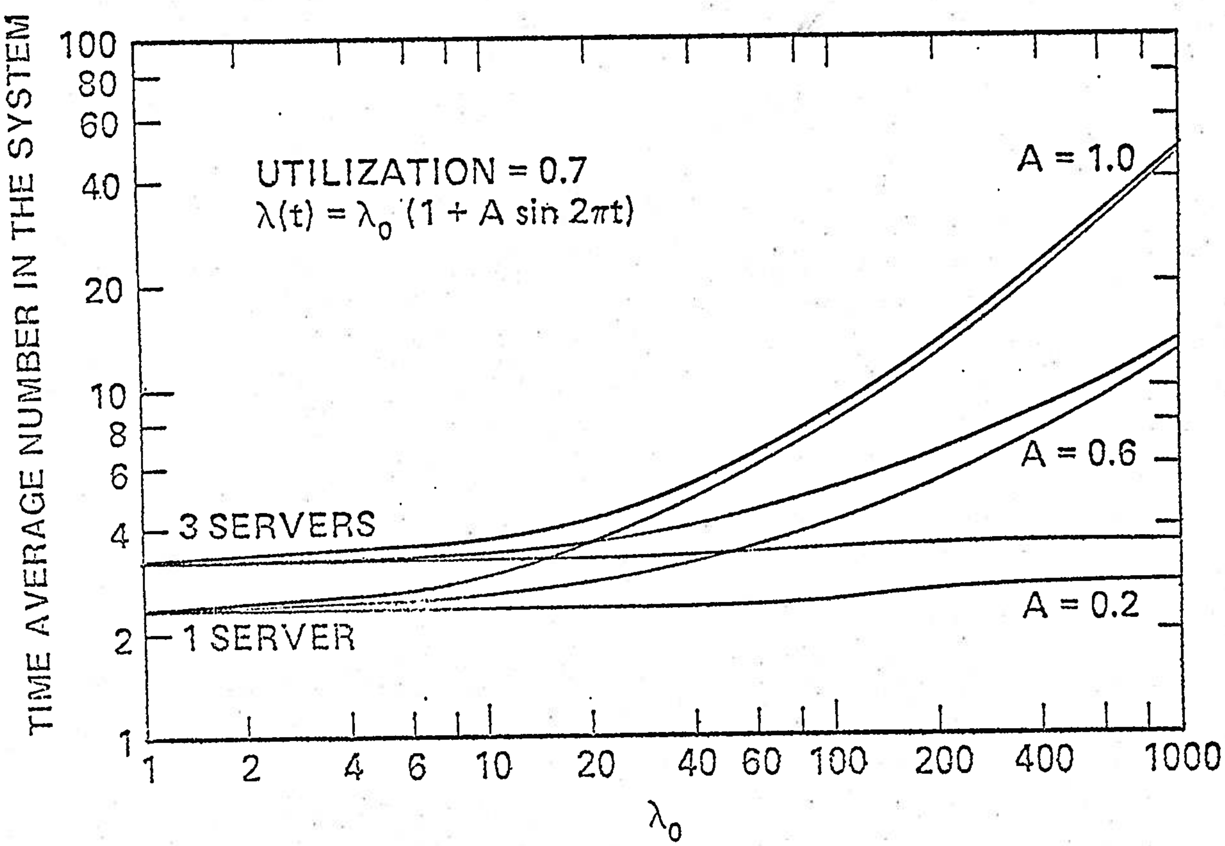

Figure 9: Time Average Number in the System: One Versus Three Servers

Appendix

In this Appendix we analyze some theoretical properties of the proposed approximation scheme for calculating m(t) and v(t) in a nonstationary M/M/1 queue. Three theorems are proved: First we prove in Theorem 1 that any trajectory (m(t), v(t)) generated by equations (2) and (4) and initialized in the positive orthant will stay positive. Furthermore, under reasonable conditions on $\mu(t)$ and $\lambda(t)$, the variance v(t) will eventually exceed m(t) and continue to do so from then on. The above properties guarantee that P_0 is always well defined along the trajectory and that the physical interpretations of m and v make sense.

Theorem 2 proves that the system of differential equations given by (2) and (4) is asymptotically stable in the sense that any discrepency caused by a perturbation in the initial conditions decays after sufficient time. This implies that trajectories corresponding to different initial conditions will merge asymptotically.

The third theorem concerns the effect of using Euler's numerical integration scheme in integrating equation (2) and (4). To avoid accummulation of error due to discontinuities in $\lambda(t)$ and $\mu(t)$ we assume that the numerical integration is performed piecewise on continuous segments of $\mu(t)$ and $\lambda(t)$, i.e. at any discontinuity point t the integration is performed up to t^- and continued from t^+ with $m(t^+)=m(t^-)$ and $v(t^+)=v(t^-)$. Under these assumptions and the additional provision that $v(0)\geq m(0)\geq 0$, we obtain in Theorem 3 an upper bound of order Δ on the error between the continuous solution and its discrete approximation. The above condition on the initial values of m and v is not an important restriction since Theorem 1 assures us that any trajectory initiated in the positive orthant can be approached asymptotically starting with $v(0)\geq m(0)$.

The following Lemma is used in the proof of Theorem 1.

Lemma 1

Let x(t) be a continuous function of t satisfying the following conditions:

1. $x'(t)$ is piecewise continuous on $\tau \subset [0,\infty)$,
2. $x'(t) \geq 0$ for all t such that $x(t) = 0$,
3. $x'(t) > 0$ for all $t \in \tau$ such that $x(t) = 0$,

4. $x(T) \geq 0$ for some $T \in [0,\infty)$.

Then, $x(t) \geq 0$ for all $t \geq T$, and $x(t) > 0$ for "almost every" $t \in \tau$ (i.e. excluding isolated points).

Proof

Since $x(t)$ is continuous and $x'(t) > 0$ on the boundary, $x(t)$ cannot cross the boundary from the positive to the negative region. Thus, since $x(t)$ is nonnegative at $t=T$ it will remain so for all $t \geq T$. The second assertion is proved by contradiction. Suppose there exist an interval $I \subset \tau$ such that $x(t) = 0$ for $t \subset I$. We can assume without loss of generality that $x'(t)$ is continuous on I. Thus, by the Mean Value Theorem, for any t_1, $t_2 \in I$, such that, $t_2 > t_1$ there exist a $\theta \in [t_1,t_2] \subset I$ such that

$$x(t_2) - x(t_1) = x'(\theta)(t_2-t_1) > 0 . \qquad \text{(A.1)}$$

This, however, contradicts the hyphothesis $x(t_1)=x(t_2)=0$ implied by assuming t_1, $t_2 \in I$.

Theorem 1

Let $m(t)$, $v(t)$ be defined by (2) and (4) with $P_0(m,v)$ given by (6) and with initial values $m(0) \geq 0$, $v(0) \geq 0$. Assume $\lambda(t)$ and $\mu(t)$ to be bounded nonegative piecewise continuously differentiable functions of t such $m(t)$ is uniformly bounded above and so that for every $t \in [0,\infty)$ and any $0 \leq M < \infty$, there exist finite τ (depending on t and M) for which $\int_t^T \lambda(\theta)d\theta \geq M$. Then,

1. $m(t) \geq 0$ for all t, and $m(t) > 0$ for almost every $t \in \{t|\lambda(t) > 0,\ t \geq 0\}$.
2. $v(t) \geq 0$ for all t, and $v(t) > 0$ for almost every $t \in \{t|\lambda(t)+\mu(t) > 0,\ m(t) > 0,\ t \geq 0\}$.
3. $0 \leq \exp(-m(t)^2/v(t)) \leq P_0(m(t),v(t)) \leq \exp(-m(t)) \leq 1$ for all t.
4. There exist a $T \in [0,\infty)$ such that $v(T) \geq m(T)$. Furthermore, given such a T, $v(t) \geq m(t)$ for $t \geq T$ and $v(t) > m(t)$ for almost every $t \in \{t|\mu(t) > 0,\ m(t) > 0,\ t \geq T\}$.

Proof

For the sake of compactness we shall ommit the argument t whenever such ommision does not compromise clarity.

1. By (6)

$$P_0(m,v) = (m/v)^{(m^2/(v-m))} \quad . \tag{A.2}$$

For $m=0$, $v>0$, (A.2) yields $P_0=1$

For $m=v$, we obtain by L'Hospital's rule

$$\lim_{m \to v} P_0(m,v) = \exp\{\lim_{m \to v} [m^2(\log(m/v))/(v-m)]\}$$
$$= \exp\{-m^2/v\}|_{m=v} = \exp(-v) = \exp(-m) \tag{A.3}$$

Hence, for $m=v=0$, $P_0=1$ and consequently by (2), $m'(t)=\lambda(t)\geq 0$ for all t such that $m(t)=0$. Thus by Lemma 1, $m(t)\geq 0$ for $t\geq 0$ and $m(t)>0$ for almost every $t\in\{t|\lambda(t)>0,\ t\geq 0\}$.

2. For $v=0$, $m=0$ we have by (A.3) $P_0=1$ and thus (4) reduces to $v'(t)=\lambda(t)\geq 0$. On the other hand for $v=0$, $m>0$ so that (A.2) implies $P_0=0$ and consequently (4) reduces to $v'(t) = \lambda(t)+\mu(t)\geq 0$. By Lemma 1 we thus have $v(t)\geq 0$ for $t\geq 0$ and $v(t)>0$ for $t\in\{t|\lambda(t)+\mu(t)>0,\ m(t)>0\}$.

3. From (A.2) we have

$$P_0 = \exp\{-mp(\log p)/(p-1)\} \tag{A.4}$$

where $p=m/v$. Substituting $x=1/p$ and $x=p$ into the well known inequality $\log x\leq x-1$, yields

$$p \log p \geq p-1 \geq p-1$$

Consequently, (A.4) implies that for any nonnegative m and v,

$$0 \leq \exp(-m^2/v) \leq P_0 \leq \exp(-m) \leq 1 \quad . \tag{A.5}$$

4. Subtracting (2) from (4) yields

$$v'(t) - m'(t)=2\mu(t)(1 - P_0 - mP_0). \tag{A.6}$$

Thus, by (A.5), (A.3) and the property $\exp(m)\leq 1+m+m^2/2$, we obtain

$$v'(t) - m'(t) \geq 2\mu \exp(-m) (\exp(m) - 1 - m) \geq \mu m^2 \exp(-m). \qquad \text{(A.7)}$$

Suppose now that $v(t) \leq m(t)$ for $t \geq 0$. Then

$$\int_0^T \frac{d}{dt}(v(t)/m(t))dt \leq v(T)/m(T) \leq 1 \quad \text{for all } T \geq 0 \qquad \text{(A.8)}$$

But by (A.7), for $v \leq m$;

$$\frac{d}{dt}(v(t)/m(t)) = (mv' - vm')/m^2 \geq (v' - m')/m \geq \mu m \exp(-m). \qquad \text{(A.9)}$$

On the other hand by (4) and (A.5),

$$v(t)' = \lambda + \mu(1 + P_0) - 2m\mu P_0 \geq \lambda - 2\mu m \exp(-m). \qquad \text{(A.10)}$$

By (A.8), (A.9) and (A.13) we have

$$\int_0^T \lambda(t)dt - (v(T) - v(0)) \leq 2\int_0^T \mu m \exp(-m)dt \leq 2\int_0^T \frac{d}{dt}(v/m)dt \leq 2. \qquad \text{(A.11)}$$

But since $v(T) \leq m(T) \leq \eta$, where η is a uniform upper bound on $m(t)$, (A.11) implies that

$$\int_0^T \lambda(t)dt \leq 2 + v(T) - v(0) \leq 2 + \eta \quad \text{for all } T \geq 0. \qquad \text{(A.12)}$$

However, (A.12) contradicts our assumption on λ, implying that for some finite $T \in [0,\infty)$, $v(T) \geq m(T)$. By (A.7) however, $v'(t)-m'(t) \geq 0$ with strict inequality when $\eta \geq 0$ and $m \geq 0$. hence, the result follows by Lemma 1.

Theorem 2

Let $m(t)$, $v(t)$ and $m(t)+\Delta m(t)$, $v(t)+\Delta v(t)$ be the respective trajectories obtained by integrating (2) and (4) with P_0 given by (6), with nonnegative initial values $m(0)$, $v(0)$ and $m(0)+\Delta m(0)$, $v(0) + \Delta v(0)$ and with the same forcing functions $\lambda(t)$, $\mu(t)$. Assume $\lambda(t)$ and $\mu(t)$ to be bounded nonegative piecewise continuously differentiable functions of t such $m(t)$ is uniformly bounded above and so that for every $t \in [0,\infty)$ and any $0 \leq M < \infty$, there exist finite τ (depending on t and M) for which $\int_t^\tau \lambda(\theta)d\theta \geq M$. Then,

$$\lim_{t \to \infty} \Delta m(t) = \lim_{t \to \infty} \Delta v(t) = 0. \tag{A.13}$$

Proof

Again for the sake of compactness we shall ommit the argument t whereever possible. From (2) and (4) we obtain

$$\left.\begin{aligned} \Delta m' &= \mu[P_0(m+\Delta m,\ v+\Delta v) - P_0(m,v)], \\ \Delta v' &= \mu\{[2m+2\Delta m+1]P_0(m+\Delta m,\ v+\Delta v) - [2m+1]P_0(m,v)\}. \end{aligned}\right\} \tag{A.14}$$

Let $\tau = \{t|\mu(t) > 0,\ t \geq 0\}$. Since for $t \notin \tau$, $\Delta m' = \Delta v' = 0$, it is sufficient to prove (A.13) for $t \in \tau$. By (2) we have $m' = \lambda - \mu(1 - P_0) \geq \lambda - \mu$. Thus, for any $T \in [0,\infty)$, we have,

$$\int_0^T \lambda(t)dt \leq m(T) - m(0) + \int_0^T \mu(t)dt \leq \eta + \int_0^T \mu(t)dt$$

where η is a uniform upper bound on m. From the asumption on $\lambda(t)$ it thus follow that for any $0 \leq M < \infty$, there exist finite $T \in [0,\infty)$ for which $\int_0^T \mu(\theta)d\theta \geq M$. Since μ is bounded above, this implies that the set τ can be maped onto the real line. Thus, without loss of generality, we shall assume $\mu(t) > 0$ for all t. Furthermore, since m(0), v(0) are arbitrary nonnegative values, it is sufficient to consider small perturbations $\Delta m(0)$, $\Delta v(0)$.

The proof of (A.13) thus amounts to proving that $\Delta m(t) = \Delta v(t) = 0$ is an asymptotically stable solution of the system (A.14). Following Lyapunov's first method (see e.g.[1]) this can be proved by showing that the Jacobian of the right hand side in (A.14) evaluated at $\Delta m(t) = \Delta v(t) = 0$ has strictly negative eigenvalues for almost every t (i.e. excluding isolated points). Evaluating the above Jacobian yields

$$J(m,v) = -\mu \begin{bmatrix} -\partial P_0/\partial m & -\partial P_0/\partial v \\ 2P_0 + [2m+1](\partial P_0/\partial m) & [2m+1](\partial P_0/\partial v) \end{bmatrix} . \tag{A.15}$$

The eigenvalues of J(m,v) are strictly negative if and only if the determinant and the trace of -J(m,v) are strictly positive. Since $\mu(t) > 0$, it is sufficient that

$$\left.\begin{aligned} &\partial P_0/\partial m<0 \\ \text{and} \\ &P_0(\partial P_0/\partial v)>0 \quad . \end{aligned}\right\} \tag{A.16}$$

From (6) we have

$$\log P_0 = (mp/(1-p))\log p = (vp^2/(1-p))\log p, \tag{A.17}$$

where $p = m/v$. Thus,

$$P_0(\partial P_0/\partial v) = P_0{}^2(\partial \log P_0/\partial p)(-m/v^2) = P_0{}^2(p^2/(1-p)^2)(p-1-\log p) \tag{A.18}$$

$$(\partial P_0/\partial m) = P_0(\partial \log P_0/\partial p)/v = P_0(p/(1-p)^2)((1-p)\log p-(p-1-\log p)). \tag{A.19}$$

From the series expansion of log p around $p=1$ we have

$$\left.\begin{aligned} &\log p \leq p-1 \qquad \text{for all } p \\ \text{and} \\ &\log p \leq -(1-p + (1-p)^2/2 + (1-p)^3/3) \quad \text{for} \quad 0 \leq p \leq 1 \end{aligned}\right. \tag{A.20}$$

with strict inequalities for $p \neq 1$. Hence by (A.19)(A.20) and (A.3) we obtain

$$P_0\,(\partial P_0/\partial v) \quad \left.\begin{aligned} &= 0 && \text{for } P_0 = 0 \\ &= (1/2)\exp(-2v) > 0 && \text{for } p = 1 \\ &> 0 && \text{for } p \neq 1 \\ &> (P_0 p)^2/2 && \text{for } 0 \leq p < 1 \quad . \end{aligned}\right\} \tag{A.21}$$

Similarly from (A.19) and (A.20) and (A.3)

$$\partial P_0/\partial m \quad \left.\begin{aligned} &= 0 && \text{for } P_0 = 0 \\ &= -\exp(-v) < 0 && \text{for } p = 1 \\ &< -P_0 p/2 && \text{for } p \neq 1 \quad . \end{aligned}\right\} \tag{A.22}$$

It is evident from (6) that $P_0 = 0$ only if $m > 0$ and $v=0$. Theorem 1, however, excludes this possibility for almost every $t \in \{t \mid \lambda(t) + \mu(t) > 0,\ t \geq 0\} \subset \tau$. Consequently, condition (A.16) is satisfied for almost every $t \in \tau$.

Corollary 1

Corresponding to any pair of forcing functions $\lambda(t)$, $\mu(t)$ having the properties stated in Theorem 2; equations (2), (4) and (6) have a unique asymptotic solution $m^*(t)$, $v^*(t)$ satisfying $v^*(t) \geq m^*(t) \geq 0$, for any nonnegative initial values $m(0)$, $v(0)$.

Proof

Follows from the results of Theorems 1 and 2.

The result of the following Lemma is needed for Theorem 3.

Lemma 2

Let $m(t)$, $v(t)$ be defined by (2) (4) and (6) with initial values $v(0) \geq m(0) \geq 0$ and assume $\lambda(t)$, $\mu(t)$ are uniformly bounded continuously differentiable with uniformly bounded derivative over the open interval (t_1, t_2). Furthermore, assume $\lambda(t)$, $\mu(t)$ are such that $m(t)$, $v(t)$ are uniformly bounded. Then d^2m/dt^2 and d^2v/dt^2 are uniformly bounded over (t_1, t_2).

Proof:

From (A.18) and (A.19) we have

$$\left.\begin{aligned} \partial \log P_0/\partial v &= (p/(1-p))^2(p-1-\log p) \qquad \geq 0 \\ \partial \log P_0/\partial m &= (p/(1-p)^2)((1-p)\log p - (p-1-\log p)) \leq 0 \end{aligned}\right\} \tag{A.23}$$

where $p = m/v$. For $0 \leq p \leq 1$ we then obtain by (A.23) and (A.20):

$$\begin{aligned} \partial^2 \log P_0/\partial v^2 &= p^2((\partial^2 \log P_0/\partial m^2) \\ &= -(p^2/(1-p)^3)(2(p-1-\log p)-(1-p)^2)/v \leq -2p^2/3v \leq 0. \end{aligned} \tag{A.24}$$

Using the mean value theorem with (A.24) and the fact that $P_0 \leq 1$, we get for $v \geq m$

$$0 \leq \partial P_0/\partial v = P_0 \partial \log P_0/\partial v \leq P_0 \partial \log P_0/\partial v|_{v=m} = P_0/2 \leq 1/2 . \tag{A.25}$$

Similarly,

$$0 \geq \partial P_0/\partial m = P_0 \partial \log P_0/\partial m \geq P_0 \,\partial \log P_0/\partial m|_{m=v} = -P_0/2 \geq -1/2 \,. \qquad \text{(A.26)}$$

The assumptions on the initial values of m and v imply, however, by Theorem 1 that $v(t) \geq m(t) \geq 0$ for $t \geq 0$. Thus, the inequalities (A.25) and (A.26) hold for all $t \geq 0$.

Differentiating (2) and (4) with respect to t yields

$$d^2m/dt^2 = \lambda'(t) - \mu'(t)(1-P_0) + \mu(t)(dP_0/dt) \qquad \text{(A.27)}$$

$$d^2v/dt^2 = \lambda'(t) + \mu'(t)(1-P_0-(2m(t)+1))-2m'(t)P_0-(2m(t)+1)(dP_0/dt) \quad , \qquad \text{(A.28)}$$

where

$$dP_0/dt = (\partial P_0/\partial m)\, m'(t) + (\partial P_0/\partial v)\; v'(t). \qquad \text{(A.29)}$$

It is evident from (2) and (4) that uniform boundness of $\lambda(t)$, $\mu(t)$ and $m(t)$ over (t_1,t_2) implies uniform boundness of $m'(t)$ and $v'(t)$ over that interval. Hence, by (A.25) (A.26) (A.29) and the assumptions of this lemma it follows that all the terms on the right hand side of (A.27) and (A.28) are uniformly bounded on (t_1, t_2). This proves the boundness of of the second derivatives of $m(t)$ and $v(t)$.

Theorem 3.

Let $m(t)$, $v(t)$ be defined by (2), (4) and (6) with initial values $v(0) > m(0) > 0$ and assume $\lambda(t)$ and $\mu(t)$ are uniformly bounded and piecewise continuously differentiable with uniformly bounded derivatives on any continuous segment. Furthermore, assume $\lambda(t)$ and $\mu(t)$ are such that $m(t)$ is uniformly bounded by η. Let m_k and v_k denote the approximations to $m(t_k)$ and $v(t_k)$ obtained by Euler's numerical integration method restricted to positive values of m_k and v_k, and applied piecewise on continuously differentiable segments of $\lambda(t)$ and $\mu(t)$ with stepsizes $\Delta_k \leq \Delta$ for all k. Then, the error vector z_k between the exact and approximate trajectories satisfies:

$$\|z_k\|_2 = \|(m(t_k)-m_k),\ (v(t_k)-v_k)\|_2 \;\leq\; \Delta L\{\exp[2Mk\Delta(2+\eta)]-1\}/2M(2+\eta),$$

where M is a uniform upper bound on $\mu(t)$ and L is a constant such that $\|(m''(\xi),v''(\zeta))\|_2 \leq 2L$ for any ξ and ζ.

Proof:

Let $f_1(m,v,t)$ and $f_2(m,v,t)$ denote the right hand side of (2) and (4) respectively. Then by Euler's method restricted to positive values

$$\left.\begin{aligned} m_{k+1} &= \max[0, (m_k + f_1(m_k, v_k, t_k)\Delta_k)], \\ v_{k+1} &= \max[0, (v_k + f_2(m_k, v_k, t_k)\Delta_k)], \end{aligned}\right\} \tag{A.31}$$

with $m_0 = n(0)$ and $v_0 = v(0)$. By the Mean Value Theorem there exist $\xi, \zeta \in [t_k, t_k+\Delta_k]$ such that

$$m(t_{k+1}) = m(t_k + \Delta_k) = m(t_k) + f_1(m(t_k), v(t_k), t_k)\Delta_k + (1/2)\, m''(\xi)\Delta_k^2$$

and (A.32)

$$v(t_{k+1}) = v(t_k+\Delta_k) = v(t_k) + f_2(m(t_k), v(t_k), t_k)\Delta_k + (1/2)\, v''(\zeta)\Delta_k^2 \quad .$$

Let $z_k = ((m(t_k) - m_k), (v(t_k)-v_k))$ and $(N(\rho),V(\rho)) = (m(t_k), v(t_k)) + \rho z_k$; then from (A.31) and (A.32) we can write

$$|z_{k+1}| \leq |z_k + \Delta_k[\int_0^1 J(N(\rho), V(\rho))d\rho]z_k + (1/2)\,\Delta_k^2(m''(\xi), v''(\zeta))|, \tag{A.33}$$

with $z_0 = 0$. Thus,

$$\begin{aligned} \|z_{k+1}\|_2 &\leq \| I+\Delta_k \int_0^1 J(N(\rho), V(\rho))d\rho\|_2 \, \|z_k\|_2 + \Delta_k^2 L \\ &\leq (1+\Delta \int_0^1 \|J(N(\rho), V(\rho))\|_2 d\rho) \, \|z_k\|_2 + \Delta^2 L \ . \end{aligned} \tag{A.34}$$

Using the fact that the spectral norm of a matrix is bounded above by the sum of the absolute values of its elements, we obtain from (A.15), (A.25) and (A.26):

$$\begin{aligned} \|J(N(\rho), V(\rho))\|_2 &\leq 2\mu\{P_0+[N(\rho)+1][|\partial P_0/\partial m|+|\partial P_0/\partial v|]\} \\ &\leq 2\mu\{P_0+[N(\rho)+1]\} \leq 2M(2+\eta) \quad . \end{aligned} \tag{A.35}$$

From (A.34) and (A.35) we then get:

$$\|z_{k+1}\|_2 \leq [1+2\Delta M(2+\eta)] \, \|z_k\|_2 + \Delta^2 L \tag{A.36}$$

with $\|z_0\|_2 = 0$. Equation (A.30) follows from the solution to the above difference equation (See, for example Lemma 14.3 in [7]).

Acknowledgements

We wish to acknowledge helpful communications with Charles Feinstein, J. Michael Harrison, Joseph Keller, Ernest Koenigsberg, Eric Leese, Marcel Neuts, Gordon Newell, Richard Smallwood and Stephen Smith. An anonymous referee provided many valuable suggestions and comments and pointed out a minor error in an earlier version of theorems 1 and 2. We are grateful to the Xerox Corporation for support and research freedom.

Bibliography

1. F. Brauer and J. A. Nahal, *Qualitative Theory of Ordinary Differential Equations*, W. A. Benjamin, Inc. N.Y, 1969.

2. S. S. L. Chang, "Simulation of Transcient and Time Varying Conditions in Queueing Networks," Proceedings of the Seventh Annual Pittsburg Conference on Modeling and Simulation, University of Pittsburg, 1977, 1075-1078.

3. A. B. Clarke, "A Waiting Line Process of Markov Type," *Annals of Math. Stat.* 27, 452-459 (1956).

4. D. R. Cox and W. L. Smith, *Queues*, London: Metheuen, 1961, Section 3.1.

5. D. P. Gaver, "Observing Stochastic Processes and Approximate Transform Inversion," *Operations Research* 14, 444-459 (1966).

6. J. M. Harrison and A. J. Lemoine, "Limit Theorems for Periodic Queues," Technical Report No. 42, Department of Operations Research, Stanford University, September, 1976.

7. P. Henrici, *Elements of Numerical Analisis* , John Wiley & Sons, Inc., NY, 1964.

8. W. J. Howard, *Turn Around Time: A Key Factor in Printing Productivity*, Rochester, NY, Xerox Corporation, 1976.

9. A. Isihara, *Statistical Physics*, New York: Academic Press, 1971, Section 6.1.

10. J. Keller, private communication..

11. P. J. Kolesar, K. L. Rider, T. B. Crabill and W. E. Walker, "A Queuing-Linear Programming Approach to Scheduling Police Patrol Cars," *Operations Research* 23, 1045-1062 (1975).

12. B. O. Koopman, "Air-Terminal Queues Under Time-Dependent Conditions," *Operations Research* 20, 1089-1114 (1972).

13. T. C. Kotiah, "Approximate Transient Analysis of Some Queueing Systems," *Operations Research* 26, 333-346 (1978).

14. E. L. Leese and D. W. Boyd, "Numerical Methods for Determining the Transient Behavior of Queues with Variable Arrival Rates," *J. of the Canadian Operational Research Society*, 4, 1-13 (1966).

15. E. L. Leese, J. G. I. MacKay and D. W. Dewar, "Comparison of Analytic and Monte Carlo Methods of Numerical Computation of the Transient Behaviour of Queues with Variable Arrival Rate," Canadian Department of National Defence, Operational Research Division, Directorate of Mathematics and Statistics, ORD Report No. 65/R11, October, 1965.

16. G. Luchak, "The Solution of the Single-Channel Queuing Equations Characterized by a Time-Dependent Arrival Rate and a General Class of Holding Times," *Operations Research* **4**, 711-732 (1956).

17. S. C. Moore, "Approximating the Behavior of Nonstationary Single-Server Queues," *Operations Research* 23, 1011-1032 (1975).

18. P. M. Morse, *Queues, Inventories and Maintenance*, New York: John Wiley and Sons, 1958.

19. M. F. Neuts, "The Single Server Queue in Discrete Time Analysis I," *Naval Res. Logistics Quart.* 20, 297-304 (1973).

20. M. F. Neuts, "The M/M/1 Queue with Randomly Varying Arrival and Service Rates," Dept. of Statistics and Computer Science, Univ. of Delaware, Tech. Report 77/11, July 1977.

21. M. F. Neuts, "Further Results on the M/M/1 Queue with Randomly Varying Rates," Dept. of Statistics and Computer Science, Univ. of Delaware, Tech. Report 78/4, 1978.

22. G. F. Newell, "Queues with Time-Dependent Arrival Rates: I, II, and III," *Journal of Applied Probability* 5, 436-451, 579-590, and 591-606 (1968).

23. G. F. Newell, *Applications of Queueing Theory*, London: Chapman and Hall, 1971.

24. R. M. Oliver and A. H. Samuel, "Reducing Letter Delays in Post Offices," *Operations Research* **10**, 839-892 (1962).

25. K. L. Rider, "A Simple Approximation to the Average Queue Size in the Time-Dependent M/M/1 Queue," *Journal of the Association for Computing Machinery* 23, 361-367 (1976).

26. H. M. Wagner, *Principles of Operations Research* 2nd edition, Englewood Cliffs, NJ: Prentice Hall, 1975.

27. K.-K. G. Wang, "Continuous Simulation of Queueing Networks with Transient and Time Varying Conditions," Ph.D. Dissertation, State University of New York at Stoneybrook, forthcoming.

28. A. Wragg, "The Solution of an Infinite Set of Differential-Difference Equations Occurring in Polymerization and Queueing Problems," *Proc. Comb. Phil. Soc.* **59**, 117-124 (1963).