THE PHYSICS OF THE $M_t/G/\infty$ QUEUE

STEPHEN G. EICK, WILLIAM A. MASSEY and WARD WHITT

AT&T Bell Laboratories, Murray Hill, New Jersey

(Received February 1991; revision received November 1991; accepted December 1991)

We establish some general structural results and derive some simple formulas describing the time-dependent performance of the $M_t/G/\infty$ queue (with a nonhomogeneous Poisson arrival process). We know that, for appropriate initial conditions, the number of busy servers at time t has a Poisson distribution for each t. Our results show how the time-dependent mean function m depends on the time-dependent arrival-rate function λ and the service-time distribution. For example, when λ is quadratic, the mean $m(t)$ coincides with the pointwise stationary approximation $\lambda(t)E[S]$, where S is a service time, except for a time lag and a space shift. It is significant that the well known insensitivity property of the stationary $M/G/\infty$ model does not hold for the nonstationary $M_t/G/\infty$ model; the time-dependent mean function m depends on the service-time distribution beyond its mean. The service-time stationary-excess distribution plays an important role. When λ is decreasing before time t, $m(t)$ is increasing in the service-time variability, but when λ is increasing before time t, $m(t)$ is decreasing in service-time variability. We suggest using these infinite-server results to approximately describe the time-dependent behavior of multiserver systems in which some arrivals are lost or delayed.

We want to gain a better understanding of the behavior of queues with time-dependent arrival rates. For example, we want to understand how the peak congestion lags behind the peak arrival rate. We begin here by considering a model for which very nice analytical results are available, namely, the $M_t/G/\infty$ model, which has a Poisson arrival process with time-dependent deterministic arrival-rate function $\lambda \equiv \{\lambda(t)\}$, independent and identically distributed (i.i.d.) service times that are independent of the arrival process, and infinitely many servers.

Palm (1943) and Khintchine (1955) showed that (for appropriate initial conditions) the number $Q(t)$ of busy servers at time t has a Poisson distribution with a mean $m(t)$ that is easily expressed in terms of the arrival rate function λ and the service-time distribution. Moreover, it is known that nice properties also hold for $M_t/G/\infty$ networks, i.e., open networks of $\cdot/G/\infty$ queues with M_t arrival processes. First, the departure process from an $M_t/G/\infty$ queue is again M_t (nonhomogeneous Poisson); see p. 405 of Doob (1953), Mirasol (1963), Newell (1966), Brown (1969), Section 1 of Daley (1976), and Foley (1982, 1986). Second, for appropriate initial conditions, the vector $\mathbf{Q}(t) \equiv (Q_1(t), \ldots, Q_n(t))$ representing the number of busy servers at each queue in an $M_t/G/\infty$ network at time t has independent Poisson marginals with a time-dependent mean vector $\mathbf{m}(t) \equiv (m_1(t), \ldots, m_n(t))$ that can be expressed in terms of the vector arrival rate function $\boldsymbol{\lambda} \equiv \{\boldsymbol{\lambda}(t)\} \equiv \{(\lambda_1(t), \ldots, \lambda_n(t))\}$ and the service-time distributions; see Section 4 of Kingman (1969), Harrison and Lemoine (1981), Foley (1982), and Keilson and Servi (1989). There is a long history for these results, which is somewhat hard to trace because essentially the same model appears under many names, e.g., see subsection 9.6 of Cox and Miller (1965).

The nice results for $M_t/G/\infty$ networks occur, in part, because different customers do not interact. Having infinitely many servers at each queue means that the customers do not get in each other's way; having Poisson arrivals means that the arrival time of one customer carries no information about the arrival times of other customers. Expressed more abstractly, the nice results occur because the input (arrival times and service times) can be represented as a Poisson process or Poisson random measure on Euclidean space; see Prékopa (1958), Foley (1982, 1986), Daley and Vere-Jones (1988), pp. 26–31 of Serfozo (1990), and Section 1 below. The various random quantities of interest, such as $Q_i(t)$, correspond to the Poisson random number of points in a particular subset. Independence occurs for different random variables when the subsets are disjoint.

Our purpose in this paper and its sequels, Eick, Massey and Whitt (1993a, b), is to show that a much

Subject classifications Queues, approximations: infinite-server queues with time-dependent arrival rates. Queues, nonstationary: simple formulas for infinite-server models. Queues, transient results. the magnitude and timing of peak congestion.
Area of review Stochastic Processes and Their Applications.

0030-0364X/93/4104-0731$01.25

better understanding of the time-dependent behavior of $M_t/G/\infty$ models and $M_t/G/s/r$ models (with s servers and r extra waiting spaces) can be obtained by looking more carefully at the $M_t/G/\infty$ results. We obtain remarkably informative simple formulas for the $M_t/G/\infty$ system with special arrival rate functions and special service-time distributions; e.g., see (14) and (20).

The rest of this paper is organized as follows. In Section 1 we review the basic $M_t/G/\infty$ theory. In Section 2 we establish some stochastic comparisons and other general structural results for $M_t/G/\infty$ models. In Section 3 we consider the special case of polynomial arrival rate functions, focusing especially on the quadratic case. In Section 4 we consider step functions, and in Section 5 we consider spikes (a constant times the indicator function of an interval). The case of a step function embodies the transient analysis of the stationary model. The spike represents a sudden transient traffic surge. Finally, in Section 6 we state our conclusions.

In Sections 1–3, we also introduce and investigate various approximations for the mean number of busy servers in the $M_t/G/\infty$ model. For example, in Section 3 we study Taylor series approximations, which are equivalent to the uniform acceleration approximation in Massey (1981, 1985) (see Remark 15). We are interested in these approximations, not because they are needed to calculate the time-dependent mean for the $M_t/G/\infty$ model, but because they help us understand the behavior of the $M_t/G/\infty$ model and to understand corresponding approximations for the more difficult $M_t/G/s/r$ model.

In Eick, Massey and Whitt (1993a) we consider the special case of $M_t/G/\infty$ queues with sinusoidal arrival rate functions. We also apply those results to treat $M_t/G/\infty$ models with general periodic arrival rate functions using Fourier series. In Eick, Massey and Whitt (1993b) we consider applications of the infinite-server results to approximately describe the performance of $M_t/G/s/r$ models. In Massey and Whitt (1993) we establish additional results for $M_t/G/\infty$ networks. For other work on queues with time-dependent arrival rates, see Brown and Ross (1969), Newell (1973), Jagerman (1975), Duda (1986), Ong and Taaffe (1989), Rolski (1989), Green and Kolesar (1991), and Green, Kolesar and Svoronos (1991). For work related to the $M_t/G/\infty$ queue in inventory theory, see Hillestad and Carrillo (1980) and Carrillo (1991).

1. REVIEW OF THE $M_t/G/\infty$ THEORY

In this section, we review established results for $M_t/G/\infty$ queues. To obtain simple formulas, we assume that our $M_t/G/\infty$ system started empty in the distant past, i.e., at $t = -\infty$. (Thorisson 1985 has developed theoretical support for initializing nonstationary models at $t = -\infty$.) We have a nonhomogeneous Poisson arrival process on $(-\infty, \infty)$ with deterministic arrival rate function $\lambda \equiv \{\lambda(t): -\infty < t < \infty\}$. (We assume that λ is nonnegative, measurable, and integrable over any bounded arrival.) Consequently, the number of arrivals in any interval $[s, t]$ has a Poisson distribution with mean $\int_s^t \lambda(u)\,du$.

The service times are i.i.d. and independent of the arrival process. Let S be a generic service-time random variable and let G be its cumulative distribution function (cdf). An important role is played by a random variable S_e with the associated stationary-excess or equilibrium-residual-lifetime cdf

$$G_e(t) \equiv P(S_e \leq t) \equiv \frac{1}{E[S]} \int_0^t G^c(u)\,du, \quad t \geq 0, \tag{1}$$

where $G^c(t) = 1 - G(t)$. See Whitt (1985) and (16) and (37) of Serfozo (1990). The moments of S_e are related to the moments of S by

$$E[S_e^k] = \frac{E[S^{k+1}]}{(k+1)E[S]}, \quad k \geq 1. \tag{2}$$

Let $Q(t)$ represent the number of busy servers at time t and let $m(t) = E[Q(t)]$. Here is the basic $M_t/G/\infty$ result.

Theorem 1. *For each t, $Q(t)$ has a Poisson distribution with mean*

$$m(t) = E\left[\int_{t-S}^{t} \lambda(u)\,du\right] = E[\lambda(t - S_e)]E[S]. \tag{3}$$

The departure process is a Poisson process with time-dependent rate function δ, where

$$\delta(t) = E[\lambda(t - S)]. \tag{4}$$

For each t, $Q(t)$ is independent of the departure process in the interval $(-\infty, t]$.

Proof. An appealing probabilistic proof used by Prékopa, who credits the idea to a 1953 lecture by C. Ryll-Nardzewski, exploits the fact that the arrival and service times generate a Poisson random measure on $(-\infty, \infty) \times [0, \infty)$; i.e., put a point at (u, v) if there is an arrival at time u with service time v. The number of points in $(a, b] \times (c, d]$ is then Poisson with mean $\int_a^b \lambda(u)\,du[G(d) - G(c)]$. The numbers of points in two disjoint rectangles $(a, b] \times (c_1, d_1]$ and $(a, b] \times (c_2, d_2]$, where $c_1 < d_1 < c_2 < d_2$, are independent Poisson random variables because independent splitting of a Poisson process produces independent

Poisson processes. Consequently, the number of points in any finite collection of disjoint rectangles are independent Poisson variables; this determines the distribution of the Poisson random measure (see Daley and Vere-Jones and pp. 27–31 of Serfozo). This argument is also used by Foley (1982). Since $Q(t)$ is just the number of pairs (u, v) with $u \leqslant t$ and $u + v > t$, see Figure 1, $Q(t)$ is Poisson with mean

$$\begin{aligned} m(t) &= \int_{-\infty}^{t} G^c(t-u)\lambda(u)\,du \\ &= \int_0^\infty \int_{t-s}^{t} \lambda(u)du\; dG(s) = E\left[\int_{t-S}^{t} \lambda(u)\,du\right] \\ &= \int_0^\infty G^c(u)\lambda(t-u)\,du = E[\lambda(t-S_e)]E[S], \end{aligned}$$

which establishes (3). Similarly, we see that the number $D(s, t)$ of departures in $[s, t]$ is just the number of pairs (u, v) with $s \leqslant u + v \leqslant t$, which is Poisson with mean $\int_s^t \delta(u)\,du$, where

$$\delta(t) = \int_0^\infty \lambda(t-u)\,dG(u) = E[\lambda(t-S)],$$

which establishes (4). The final independence property in Theorem 1 holds because the number of points in disjoint sets are independent Poisson variables with a Poisson random measure (see Figure 2).

Remark 1. Our assumptions in Theorem 1 do not guarantee that $m(t)$ in (3) is finite or that δ in (4) is integrable over bounded intervals. Sufficient conditions are for either λ to be integrable over $(-\infty, t]$ or the service-time distribution to have bounded support.

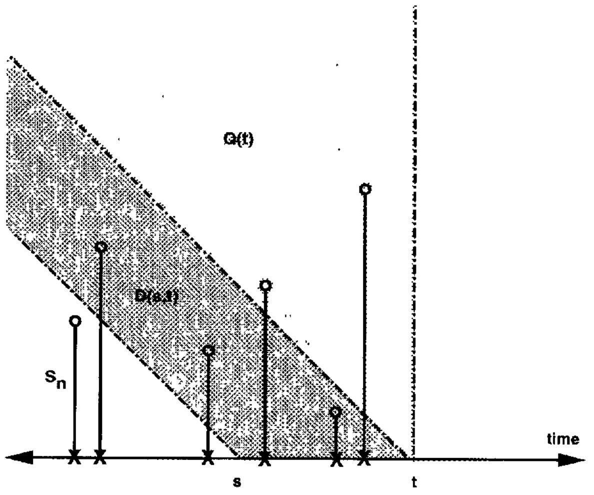

Figure 1. A possible realization of the Poisson random measure for Theorem 1; the random variables $Q(t)$ and $D(s, t)$ count the number of points in the designated subset.

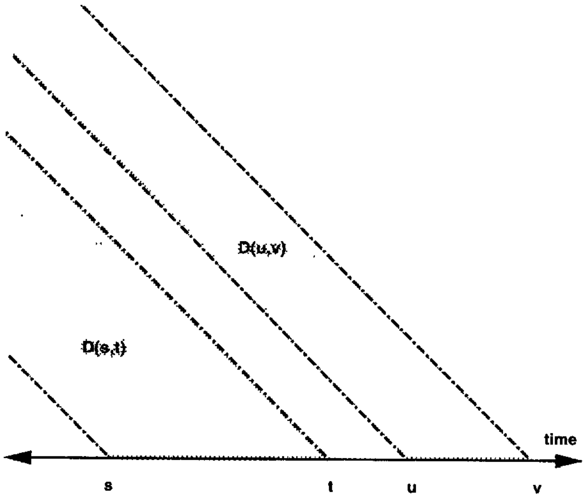

Figure 2. Two disjoint sets in the plane for calculating the joint distribution of the number of departures in two disjoint time intervals for Theorem 1 using the Poisson random measure; the random variables count the number of points in the designated subset.

Remark 2. Extensions to other initial conditions are easy, but sometimes messy. For example, suppose that the $M_t/G/\infty$ system starts out at $t = 0$ with k customers. In general, we need to know the remaining service times of these customers. However, assuming that the k customers all start service at $t = 0$, $Q(t)$ is the independent sum of a Poisson random variable with mean $m(t)$ in (3), where $\lambda(t) = 0$ for $t < 0$, and a binomial random variable with parameters k and $p = G^c(t)$. If the initial number is random with a Poisson distribution with mean θ, then the binomial random variable is replaced by a Poisson random variable with mean $\theta G^c(t)$, in which case the overall distribution is again Poisson.

Remark 3. For interpretation, it is useful to relate $m(t)$ in (3) to the *instantaneous offered load* $\lambda(t)E[S]$, because $m(t)$ would equal $\lambda(t)E[S]$ if the arrival process were homogeneous with constant arrival rate $\lambda(t)$. Following Green and Kolesar, we call $\lambda(t)E[S]$ the *pointwise stationary approximation* (**PSA**) for $m(t)$; also see Palm (1943), Massey (1981, 1985), and Whitt (1991). The last expression in (3) says that **PSA** is correct except for a random time lag in $\lambda(t)$. Of course, if λ changes very little before t, then $E[\lambda(t - S_e)] \approx \lambda(t)$ and $m(t) \approx \lambda(t)E[S]$. The formulas in (3) indicate that, unlike **PSA**, the true formula for $m(t)$ involves averaging λ before t (as in subsection 5.3 of Whitt 1991).

Of course, given the relatively simple formula for $m(t)$ in (3), approximations such as **PSA** are *not important for calculating* $m(t)$ in the $M_t/G/\infty$ model. We study these approximations, first, to develop a better understanding of $m(t)$ in the $M_t/G/\infty$ model and, second, to gain insight into the way corresponding approximations will perform for the more complicated $M_t/G/s/r$ model.

Remark 4. Throughout this paper, we assume that the service-time distributions are not time dependent, which is realistic for many applications, but the theory extends to time-dependent, service-time distributions. If the service-time distribution of an arrival at time t has cdf G_t, then the time-dependent mean is

$$m(t) = \int_{-\infty}^{t} G_u^c(t - u)\lambda(u)\, du,$$

assuming, of course, that $G_u^c(t - u)$ is integrable (see Brown and Ross 1969, Carrillo 1991, and Massey and Whitt 1993).

Remark 5. Since the mean function m is a linear function of the arrival rate function λ (also see Theorem 8), we can regard the $M_t/G/\infty$ model as a linear system and apply linear system theory (e.g., see Chapter 2 of Ziemer and Tranter 1976). The linear system view is discussed further in Eick, Massey and Whitt (1993a).

It is important to note that, even though $Q(t)$ has a Poisson distribution for each t, $Q \equiv \{Q(t): -\infty < t < \infty\}$ is *not* a Poisson process. However, it is not difficult to determine the finite-dimensional distributions of Q. We illustrate with the case of the two-dimensional distributions. For previous work on the stationary model, see Riordan (1951) and Beneš (1957).

Let $\overset{d}{=}$ denote equality in distribution. Let $(x)^+ = \max\{x, 0\}$. For random variables X and Y, let $\text{Cov}[X, Y]$ be the covariance, i.e., $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$. Recall that two random variables X and Y are *associated* if $\text{Cov}[f(X), g(Y)] \leq 0$ for all nondecreasing real-valued functions f and g (see p. 29 of Barlow and Proschan 1975).

Theorem 2. *For $u > 0$ and for each t,*

$$\begin{aligned}\text{Cov}&[Q(t), Q(t + u)]\\ &= E\left[\int_{t-(S-u)^+}^{t} \lambda(s)\, ds\right]\\ &= E[(\lambda(t - (S - u)_e^+)]E[(S - u)^+]. \qquad (5)\end{aligned}$$

Moreover, $Q(t)$ and $Q(t + u)$ are associated random variables. Indeed,

$$\begin{aligned}&[Q(t), Q(t + u)]\\ &\overset{d}{=} [A(t, u) + C(t, u), B(t, u) + C(t, u)], \qquad (6)\end{aligned}$$

where $A(t, u)$, $B(t, u)$, and $C(t, u)$ are independent Poisson random variables with $E[C(t, u)] = Cov(Q(t), Q(t + u))$. Consequently, $Q(t)$ and $Q(t + u)$ are independent if $P(S > u) = 0$.

Proof. Let $A(t, u)$ be the number of arrivals up to time t that depart in the interval $[t, t + u)$; let $B(t, u)$ be the number of arrivals in $(t, t + u]$ that are still in the system at time $t + u$; and let $C(t, u)$ be the number of arrivals up to time t that are still in the system at time $t + u$. These random variables correspond to the Poisson random numbers of points in the respective subsets of $(-\infty, t] \times [0, \infty)$ in Figure 3. Obviously, $Q(t) = A(t, u) + C(t, u)$ and $Q(t + u) = B(t, u) + C(t, u)$. Since the subsets corresponding to $A(t, u)$, $B(t, u)$, and $C(t, u)$ are disjoint, the random variables $A(t, u)$, $B(t, u)$, and $C(t, u)$ are independent.

From (6), it is easy to see that $Q(t)$ and $Q(t + u)$ are associated; first condition on $A(t, u)$ and $B(t, u)$, and then uncondition (see pp. 30–31 of Barlow and Proschan). The covariance formula in (5) follows from (6) and the fact that the variance of a Poisson random variable equals its mean. In particular,

$$\begin{aligned}E[C(t, u)] &= \int_{-\infty}^{t} \lambda(s)P(S > t + u - s)\, ds\\ &= \int_{-\infty}^{t} \lambda(s)P((S - u)^+ > t - s)\, ds\\ &= E[\lambda(t - (S - u)_e^+)]E[(S - u)^+].\end{aligned}$$

Moreover, when $P(S > u) = 0$, $E[C(t, u)] = 0$, so that $P(C(t, u) = 0) = 1$, which implies that $Q(t)$ and $Q(t + u)$ are independent.

Remark 6. In the homogeneous case with $\lambda(t) = \lambda(0)$ for all t, (5) becomes

$$\begin{aligned}\text{Cov}[(Q(t), Q(t + u)] &= \lambda(0)E[(S - u)_+]\\ &= \lambda(0)P(S_e > u)E[S].\end{aligned}$$

2. STOCHASTIC COMPARISONS AND OTHER STRUCTURAL RESULTS

In this section, we establish some general results about the way the mean function m in (3) depends on the arrival rate function λ and the service-time distribution G. Let $\leq_{st}$ and $\leq_c$ denote the usual stochastic ordering and convex stochastic ordering, respectively;

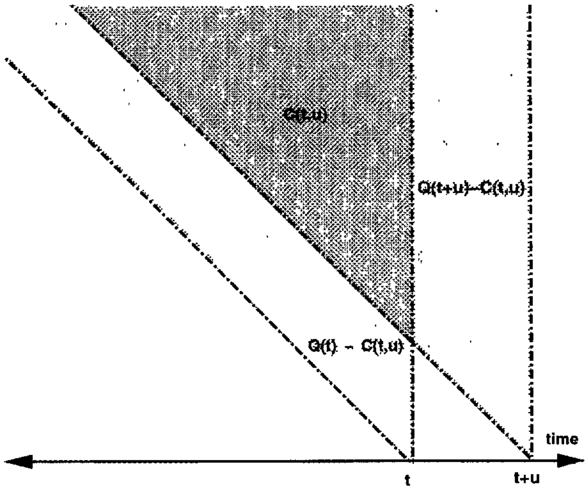

Figure 3. Three disjoint sets in the plane for calculating the joint distribution of $Q(t)$ and $Q(t + u)$ for Theorem 2 using the Poisson random measure; the random variables count the number of points in the designated subset.

$S_1 \leqslant_{st} S_2$ ($S_1 \leqslant_c S_2$) if $E[f(S_1)] \leqslant E[f(S_2)]$ for all increasing (all convex) real-valued functions f such that the expectations are well defined (see Ross 1983 and Stoyan 1983). (Throughout this paper we use increasing and decreasing in the weak sense; we say strictly increasing for the stronger property.) The ordering $\leqslant_{st}$ ($\leqslant_c$) is a strong expression for the distribution of one random variable being less (less variable) than another.

The comparison results for m below immediately imply stochastic comparisons for the random variable $Q(t)$ because two Poisson distributions with different means are stochastically ordered. Indeed, if X_i is Poisson with mean α_i where $\alpha_1 < \alpha_2$, then $P(X_2 = k)/P(X_1 = k)$ is easily seen to be strictly increasing in k, so that X_1 is less than X_2 in the likelihood ratio ordering, which implies the usual stochastic ordering ($\leqslant_{st}$) (see Ross 1983).

In the following, the *support* of (the distribution of) S plays an important role. We assume that this support of S is in the interval $[0, \sigma]$ where $\sigma \leqslant \infty$, which means that $G(\sigma) = 1$. From (1) it follows that the support of S_e is in $[0, \sigma]$ too. From (3) it is apparent that $m(t)$ depends on λ only over the interval $(t - \sigma, t]$ when S_e has support in $[0, \sigma]$.

Theorem 3. *Consider two $M_t/G/\infty$ models with a common arrival process but different generic service times S_1 and S_2, both with support in* $[0, \sigma]$. *Let λ be increasing (decreasing) over the interval $(t - \sigma, t]$.*

a. *If $S_1 \leqslant_{st} S_2$, then $\delta_1(t) \geqslant (\leqslant)\, \delta_2(t)$.*
b. *If $S_1 \leqslant_c S_2$, then $S_{1e} \leqslant_{st} (\geqslant_{st})\, S_{2e}$, so that $m_1(t) \geqslant (\leqslant)\, m_2(t)$.*

Proof. Part a is an immediate consequence of (4) and the definition of $\leqslant_{st}$. For part b, we use the fact that $S_1 \leqslant_c S_2$ implies $E[S_1] = E[S_2]$ and $S_{1e} \leqslant_{st} S_{2e}$, because $S_1 \leqslant_c S_2$ is equivalent to $E[S_1] = E[S_2]$ and $\int_t^\infty G_1^c(u)\, du \leqslant \int_t^\infty G_2^c(u)\, du$ for all t (e.g., see Theorem 3.3 and Section 5 of Whitt 1985). Finally, apply (3) and the definition of $\leqslant_{st}$.

Remark 7. The conditions in Theorem 3 are necessary and sufficient to have the conclusions for all λ increasing (decreasing) over the interval $(t - \sigma, t]$.

Remark 8. There are many results in the queueing literature stating that the congestion increases in a stationary model when the service-time distribution becomes more variable, with the convex stochastic ordering being a typical characterization of greater variability; see Whitt (1984), especially Section 6, for references and a discussion of exceptions. Results of this kind for the stationary $M/G/\infty$ queue appear in Huffer (1987). Theorem 3b shows that the time-dependent mean $m(t)$ in the $M_t/G/\infty$ model has this property provided that λ is decreasing before t, but $m(t)$ is *decreasing* in the service-time variability when λ is increasing before t. Theorem 3b shows that the response of $m(t)$ to service-time variability depends on the shape of λ.

We now consider the effect of changing λ. The following is an elementary consequence of (3) and (4).

Theorem 4. *Consider two $M_t/G/\infty$ systems with a common service-time distribution having support in* $[0, \sigma]$ *and two possible arrival rate functions λ_1 and λ_2.*

a. *If $\lambda_1(s) \leqslant \lambda_2(s)$ for $t - \sigma \leqslant s \leqslant t$, then $\delta_1(t) \leqslant \delta_2(t)$.*
b. *If $\int_s^t \lambda_1(u)\, du \leqslant \int_s^t \lambda_2(u)\, du$ for $t - \sigma \leqslant s \leqslant t$, then $m_1(t) \leqslant m_2(t)$. If the distribution of S has a decreasing right-continuous density, then $\delta_1(t) \leqslant \delta_2(t)$ also.*

Proof. We only discuss the last statement. If S has a decreasing right-continuous density, then the distribution of S is the stationary-excess of another nonnegative random variable S^* which has support in $[0, \sigma]$. By (3) and (4), then

$$\delta(t) = \frac{1}{E[S^*]}\, E\left[\int_{t-S^*}^{t} \lambda(u)\, du\right],$$

so that the proof is the same as for $m(t)$.

Corollary 1. *In an $M_t/G/\infty$ system with S having* *support in* $[0, \sigma]$,

$$\inf_{t-\sigma<s\leq t}\{\lambda(s)E[S]\} \leq m(t) \leq \sup_{t-\sigma<s\leq t}\{\lambda(s)E[S]\}$$

and similarly for $\delta(t)$.

Remark 9. When λ is monotone over $[t-\sigma, t]$, Corollary 1 gives direct comparisons between $m(t)$ and **PSA,** i.e., $\lambda(t)E[S]$ (see Remark 3). In particular, **PSA** is too high when λ is increasing and too low when λ is decreasing.

We now relate the shape of λ to the shape of m and δ.

Theorem 5. *Consider an* $M_t/G/\infty$ *system with* S *having support in* $[0, \sigma]$. *Then* m *and* δ *have the following properties*:

a. *If, for any* $s \leq t$, λ *is increasing* (*decreasing*) *on* $[s-\sigma, t]$, *then both* m *and* δ *are increasing* (*decreasing*) *on* $[s, t]$, *and for all* r *belonging to* $[s, t]$,

$$m(r) \leq (\geq)\ \lambda(r)E[S] \quad \text{and} \quad \delta(r) \leq (\geq)\lambda(r).$$

b. *If, for any* $s \leq t$, λ *is convex* (*concave*) *on* $[s-\sigma, t]$, *then both* m *and* δ *are convex* (*concave*) *on* $[s, t]$, *and for all* r *belonging to* $[s, t]$,

$$m(r) \geq (\leq)\ \lambda(r-E[S_e])E[S] \text{ and } \delta(r) \geq (\leq) \lambda(r-E[S]).$$

Proof. These results are elementary consequences of (3) and (4). For part b, apply Jensen's inequality.

Corollary 2. *If* λ *is unimodal with a maximum at* t_λ, *then the maximum of* m *occurs after* t_λ.

Remark 10. Since a linear function is both concave and convex, part b of Theorem 5 implies that

$$m(t) = \lambda(t - E[S_e])E[S] \tag{7}$$

when λ is linear on $(t-\sigma, t]$. An expression equivalent to (7) is

$$\begin{aligned} m(t) &= (\lambda(t) - \lambda'(t)E[S_e])E[S] \\ &= \lambda(t)E[S] - \frac{\lambda'(t)E[S^2]}{2}; \end{aligned} \tag{8}$$

see (2). More generally, (7) and (8) represent *linear approximations* for $m(t)$ when λ is approximately linear before t. (The linear approximation for $m(t)$ in (7) was previously discussed by Newell 1973, p. 31.) We call (7) **LIN-S** for the linear approximation with time shift, and we call (8) **LIN-D** for the linear approximation with derivative. An obvious refinement of (8) is to take the maximum of (8) and zero, and we do this.

Theorem 5b shows how **LIN-S** should perform when the arrival rate function is primarily concave or convex before time t. Note that **LIN-S** agrees with **PSA** in Remark 2 except for a *time lag* of $E[S_e] = E[S](c_s^2+1)/2$, where c_s^2 is the squared coefficient of variation (variance divided by the square of the mean) of S. Also note that this time lag $E[S_e]$ is independent of λ when λ is linear. Similarly, **LIN-D** agrees with **PSA** except for a *space shift* of

$$\lambda'(t)E[S^2]/2 = \lambda'(t)(c_s^2+1)(E[S])^2/2.$$

When λ is approximately linear before t, the error in **PSA** for the mean $m(t)$ is

$$\begin{aligned} &|m(t) - \lambda(t)E[S]| \\ &\quad \approx |\lambda(t - E[S_e])E[S] - \lambda(t)E[S]| \\ &\quad \approx \lambda'(t)E[S_e]E[S] \end{aligned} \tag{9}$$

and the relative error is

$$\begin{aligned} &\frac{|m(t) - \lambda(t)E[S]|}{m(t)} \\ &\quad \approx \frac{|\lambda(t - E[S_e])E[S] - \lambda(t)E[S]|}{\lambda(t)E[S]|} \\ &\quad \approx \left[\frac{|\lambda'(t)|}{\lambda(t)}\right]E[S_e]. \end{aligned} \tag{10}$$

(An exact expression for the error in (9) is given by Theorem 10.) Also note that if we consider **PSA** as an approximation for the average $\bar{m} \equiv (b-a)^{-1}\int_a^b m(t)\,dt$ over some time interval $[a, b]$, then the error is still given by (9) when λ is approximately linear before b; the relative error is then approximately $\lambda'(t)E[S_e]/\bar{\lambda}$, where $\bar{\lambda} = (b-a)^{-1}\int_a^b \lambda(t)\,dt$.

Example 1. Suppose that $\lambda(t) = t$ for $t \geq 0$, but $\lambda(t) = 0$ for $t < 0$, and that the service-time distribution is exponential with mean 1. Then (7) and (8) do not hold exactly, but we anticipate that **LIN-S** and **LIN-D** will perform well for suitably large t. (For this example, **LIN-S** and **LIN-D** agree.) Moreover, to illustrate an infinite-server approximation for an $M_t/M/s/0$ loss model, we consider the cases $s = 20$ and 50. (The "exact" values of $m(t)$ for these $M_t/M/s/0$ models were computed using the algorithm described in Section 1 of Eick, Massey and Whitt 1993b.) For $t \leq 20$, the case $s = 50$ corresponds approximately to an infinite-server model, as can be seen by considering the stationary model with $\lambda = 20$ and $s = 50$.

First, we can see the effect of the initial nonlinearity by comparing **LIN-S** with the exact results for $s = 50$

in Table I. Indeed, the error due to the initial nonlinearity decreases for t above 1, reaching 12% at $t = 2$ and 2.5% at $t = 3$.

For the case $s = 20$, we begin to see the effect of the blocking for $t \geq 12$. **LIN-S** has an error of only 8% at $t = 18$, but 14% at $t = 20$. For this example, **PSA** is consistently an inferior approximation, but it is useful together with the linear approximation, because they bracket the exact value; i.e., for the infinite-server model, $\lambda(t) \geq m(t) \geq \lambda(t - 1)$ for all t, because λ is increasing and convex (see Theorem 5).

Our next results describe the derivative of $m(t)$.

Theorem 6. *If the departure rate function δ in* (4) *is integrable in a neighborhood of t, then the mean function m in* (3) *is absolutely continuous with respect to a Lebesgue measure in a neighborhood of t, with density*

$$m'(t) = \lambda(t) - \delta(t). \tag{11}$$

If, in addition, λ is continuous and bounded on $[a, b]$, *where* $(t - \sigma, t] \subseteq (a, b)$, *then δ is continuous and m is continuously differentiable at t.*

Table I
A Comparison Between Infinite-Server Approximations and Exact Values for the Mean Number of Busy Servers at Time t, $m(t)$, in an $M_t/M/s/0$ Loss Model With Arrival Rate Function $\lambda(t) = t$ for $t \geq 0$ and $\lambda(t) = 0$ for $t < 0$, for Example 1. (For these times, the case $s = 50$ corresponds to an infinite-server (**IS**) model, while $s = 20$ shows the effect of blocking. The service-time distribution is exponential with mean 1.)

Time	$m(t)$ Exact		
$t = \lambda(t)$			**LIN-S = LIN-D**
IS-PSA	$s = 20$	$s = 50$	$\lambda(t - 1)$
0.1	0.005	0.005	0.0
0.5	0.107	0.107	0.0
1.0	0.368	0.368	0.0
1.5	0.724	0.724	0.5
2.0	1.14	1.14	1.0
2.5	1.58	1.58	1.5
3.0	2.05	2.05	2.0
4.0	3.02	3.02	3.0
5.0	4.01	4.01	4.0
6.0	5.00	5.01	5.0
8.0	7.00	7.00	7.0
10.0	9.00	9.00	9.0
12.0	10.98	11.00	11.0
14.0	12.84	13.00	13.0
15.0	13.69	14.00	14.0
16.0	14.46	15.00	15.0
17.0	15.13	16.00	16.0
18.0	15.72	17.00	17.0
19.0	16.22	18.00	18.0
20.0	16.65	19.00	19.0

Proof. Note that

$$Q(t + h) - Q(t)$$
$$= [A(t + h) - A(t)] - [D(t + h) - D(t)],$$

where $\{A(t){:}t \geq 0\}$ and $\{D(t){:}t \geq 0\}$ are the arrival and departure counting processes. Since these counting processes are Poisson (the first by assumption and the second by Theorem 1), we can take exceptations and write

$$m(t + h) - m(t) = \int_t^{t+h} [\lambda(u) - \delta(u)]\, du.$$

By assumption, $\int_t^{t+h} \lambda(u)\, du < \infty$, so that the expectation is well defined, but possibly $-\infty$. The assumption on δ rules out $-\infty$. The integral representation expresses the first conclusion. By the bounded convergence theorem, $E[\lambda(t - S)]$ is continuous in t if λ is continuous and bounded in the specified way.

Here are simple applications of Theorem 6.

Corollary 3. *If the distribution of S is deterministic* $(G = D)$, *then* $m'(t) = \lambda(t) - \lambda(t - S)$ a.s. *If λ is unimodal with maximum* (*minimum*) *at t_λ, then m has a maximum* (*minimum*) *at $t_m = t_\lambda + x$, where $0 \leq x \leq S$ and if, in addition, λ is continuous, then $\lambda(t_m) = \lambda(t_m - S)$. Furthermore, if λ is continuous and λ attains an extreme value over $[t_\lambda - S, t_\lambda + S]$ at t_λ, then $m'(t) = 0$ for some t in $[t_\lambda, t_\lambda + S]$.*

Corollary 4. *If S has an exponential distribution* $(G = M)$, *then*

$$m'(t) = \lambda(t) - \frac{m(t)}{E[S]} \quad \text{a.s.}$$

Remark 11. Corollary 4 implies that for the $M_t/M/\infty$ model **PSA** will be too high (low) precisely when $m(t)$ is increasing (decreasing). Moreover, **PSA** and $m(t)$ will coincide precisely when $m'(t) = 0$. In linear system theory, the differential equation in Corollary 4 corresponds to the low-pass RC filter (see pages 50 and 55 of Ziemer and Tranter).

Theorem 6 suggests comparing $\delta(t)$ and $m(t)$. When λ is monotone, such comparisons follow from (3), (4), and established comparisons between the distributions of S and S_e. For this purpose, recall that the distribution of S is new worse than used in expectation (NWUE) if

$$E[S] \leq E[S - t \mid S > t] \quad \text{for all } t. \tag{12}$$

The distribution is new better than used in expectation (NBUE) if the inequality in (12) is reversed.

Theorem 7. *Suppose that* λ *is increasing (decreasing) in* $(t - s, t]$ *for* $s > \sigma$ *and that* (11) *is valid at* t. *If the distribution of* S *is NWUE, then*

$$m(t) \leqslant (\geqslant) \delta(t) E[S] \quad \text{and} \quad m'(t) \leqslant (\geqslant) \lambda(t) - \frac{m(t)}{E[S]}.$$

If the distribution of S *is NBUE, then the inequalities are reversed.*

Proof. The NWUE property in (12) implies that $S \leqslant_{st} S_e$. The NBUE property implies $S \geqslant_{st} S_e$ (see Theorem 3.1 of Whitt 1985). Finally, apply (3) and Theorem 6.

In examples, the arrival rate function λ may be represented as the sum of two or more pieces, e.g., a linear piece plus a sinusoidal piece. It is significant that the decomposition is inherited by m. Moreover, this remains true for subtraction, so that we can regard the map from λ to m as a linear operator on a function space, with the appropriate function space depending on the structure assumed for λ. For example, if the function space is $L_\infty \equiv L_\infty(-\infty, \infty)$, the space of bounded measurable real-valued functions on $(-\infty, \infty)$ with the supremum norm $\|\cdot\|_\infty$, then the operator is bounded. We can thus say two mean functions are close when the corresponding arrival rate functions are close.

Theorem 8. *If an arrival rate function* λ *for an* $M_t/G/\infty$ *model can be represented as* $\lambda = a_1\lambda_1 + a_2\lambda_2$ *for arrival rate functions* λ_i *and constants* a_i, *then* $m = a_1 m_1 + a_2 m_2$, *where* m_i *is the time-dependent mean function associated with* λ_i *via* (3). *Consequently, if* λ *is bounded on* $(-\infty, \infty)$, *then the map in* (3) *taking* λ *into* m *is a bounded linear operator on the space* L_∞ *with the supremum norm* $\|\cdot\|_\infty$ *and* $\|m_1 - m_2\|_\infty \leqslant \|\lambda_1 - \lambda_2\|_\infty E[S]$.

Proof. Without loss of generality, suppose that a_1 and a_2 are positive. (Otherwise, move the negative term to the other side.) The Poisson process with arrival rate function λ may be split into two independent Poisson processes with arrival rate functions $a_1\lambda_1$ and $a_2\lambda_2$ by thinning with probability $a_1\lambda_1(t)/\lambda(t)$ at time t (see Theorem 2.8 of Serfozo). The sum of two independent Poisson random variables is again Poisson. We apply Theorem 1 to treat m_{i1}. The operator properties then follow easily from (3).

A result closely related to Theorem 8 gives the mean of an $M_t/G/\infty$ model when the service-time distribution is a mixture; then the system is equivalent to the sum of independent $M_t/G/\infty$ systems with reduced arrival rate functions and the component service-time cdf's.

3. POLYNOMIAL ARRIVAL RATE FUNCTIONS

We discussed linear arrival rate functions and linear approximations in Remark 10. Now we consider more general polynomial arrival rate functions and polynomial approximations. Recall that we examine approximations in this relatively simple $M_t/G/\infty$ setting primarily to gain insight into corresponding approximations for the more difficult $M_t/G/s/r$ model.

From (2) and (3), we see that we can write down an explicit expression for $m(t)$ in terms of λ and the moments of S when λ is a polynomial. However, in the queueing context λ must be positive in the interval $(t - \sigma, t]$ when S has support in $[0, \sigma]$. To avoid this qualification, in this section we regard (3) as the definition of $m(t)$ without insisting that λ be positive in $(t - \sigma, t]$. When λ is indeed not positive everywhere in $(t - \sigma, t]$, the results may still be useful as approximations.

First consider the quadratic arrival rate function

$$\lambda(t) = a + bt + ct^2. \tag{13}$$

The principal case in (13) is when the coefficients a and b are positive, but c is negative. Then λ first increases and then decreases, attaining a maximum of $a + (b^2/2c)$ at time $t_\lambda = -b/2c$. From (3) and (13), we obtain the following nice formula.

Theorem 9. *Suppose that* λ *is quadratic as in* (13). *If* $E[S^3] < \infty$, *then*

$$m(t) = \lambda(t - E[S_e])E[S] + c\,\mathrm{Var}(S_e)E[S]. \tag{14}$$

From (14), we obtain a general principle: *If* λ *is approximately quadratic before time* t, *then* $m(t)$ *is approximately equal to the* ***PSA*** *value* $\lambda(t)E[S]$ *except for a time lag of* $E[S_e] = E[S](c_s^2 + 1)/2$ *and a space shift by*

$$c\,\mathrm{Var}(S_e)E[S] = c[(E[S^3]/3) - (E[S^2]^2/4E[S])].$$

As in the linear case, the time lag $E[S_e]$ is independent of the coefficients a, b, and c in (13). (This does not hold for higher polynomials.) For given first two moments of S, the magnitude of the space shift is increasing in $E[S^3]$.

Paralleling the linear approximations **LIN-S** and **LIN-D** in (7) and (8) we suggest the *quadratic* *approximations* **QUAD-S**

$$m(t) \approx \lambda(t - E[S_e])E[S] + \frac{\lambda''(t)}{2}\operatorname{Var}(S_e)E[S] \tag{15}$$

and **QUAD-D**

$$m(t) \approx \left[\lambda(t) - \lambda'(t)E[S_e] + \frac{\lambda''(t)E[S_e^2]}{2}\right]E[S]$$
$$\approx \lambda(t)E[S] - \frac{\lambda'(t)E[S^2]}{2} + \frac{\lambda''(t)E[S^3]}{6} \tag{16}$$

for more general arrival rate functions.

Formula 16 is based on taking the first three terms of the Taylor series expansion of $\lambda(t - S_e)$ about t. The following theorem treats the general case and gives an explicit probabilistic expression for the remainder term. In our expressions a key role is played by the k-fold iterate of the stationary-excess operator in (1), i.e., $S_e^{(k)} = (S_e^{(k-1)})_e$ (see Whitt 1985). Let $\lambda^{(k)}$ denote the kth derivative of the arrival-rate function λ.

Theorem 10. *For any $n \geqslant 0$, suppose that λ is $(n + 1)$-times differentiable and the $(n + 1)^{st}$ derivative is Riemann integrable on $[t - x, t]$ for all x, $0 < x < \sigma$. If $E[S^{n+2}] < \infty$ and $E[|\lambda^{(k)}(t - S_e^{(k+1)})|] < \infty$, $0 \leqslant k \leqslant n + 1$, then $m(t) = m_n(t) + R_n(t)$, $n \geqslant 0$, where*

$$m_n(t) = \sum_{j=0}^{n} (-1)^j \frac{\lambda^{(j)}(t)E[S^{j+1}]}{(j + 1)!}$$

and

$$R_n(t) = (-1)^{n+1}E[\lambda^{(n+1)}(t - S_e^{(n+2)})]\,\frac{E[S^{n+2}]}{(n + 2)!}.$$

Remark 12. Note that Theorem 10 can be regarded as a probabilistic generalization of Taylor's theorem and (3) can be regarded as a probabilistic generalization of the fundamental theorem of calculus.

Remark 13. Note that approximation **PSA, LIN-D,** and **QUAD-D** introduced in Remarks 2 and 10 and (16) are the Taylor series for $n = 0$, 1, and 2, respectively.

Remark 14. From Theorem 10, we see that the Taylor series for $m(t)$ is absolutely convergent (assuming that λ has bounded derivatives of all orders) if and only if

$$\sum_{j=0}^{\infty} \frac{|\lambda^{(j)}(t)|E|S^{j+1}]}{(t + 1)!} < \infty.$$

For example, if S is exponential with mean 1, then $E[S^j] = (E[S])^j j! = j!$, so that the Taylor series is absolutely convergent then if and only if $\sum_{j=0}^{\infty} |\lambda^{(j)}(t)| < \infty$.

Remark 15. It is significant that the Taylor series approximation for $m(t)$ in Theorem 10 can be interpreted as a uniform acceleration asymptotic expansion for $m(t)$, as in Massey (1981, 1985). In particular, we construct a family of systems indexed by ε and let $\varepsilon \to 0$. In system ε, the service time is $S_\varepsilon \equiv \varepsilon S$ (which has stationary-excess $(S_\varepsilon)_e = \varepsilon S_e$) and the arrival rate of time t is $\lambda_\varepsilon(t) \equiv \lambda(t)/\varepsilon$. Then

$$m_\varepsilon(t) = E[\lambda_\varepsilon(t - (S_\varepsilon)_e]E[S_\varepsilon]$$
$$= E[\lambda(t - \varepsilon S_e)]E[S]. \tag{17}$$

From (17), we see that if λ is bounded (or, more generally, if the limit of the expectations is the expectation of the limit), then $m_\varepsilon(t) \to \lambda(t)E[S]$ as $\varepsilon \to 0$, so that **PSA** is asymptotically correct as $\varepsilon \to 0$. More generally, under the conditions of Theorem 10, we obtain an asymptotic expansion in powers of ε. For example,

$$\begin{aligned} m_\varepsilon(t) &= (\lambda_\varepsilon(t) - \lambda_\varepsilon'(t)E[(S_\varepsilon)_e] \\ &\quad + \frac{\lambda_\varepsilon''(t)}{2}E[(S_\varepsilon)_e^2])E[S_\varepsilon] + O(\varepsilon^3) \\ &= (\lambda(t) - \varepsilon\lambda'(t)E[S_e] \\ &\quad + \varepsilon^2\frac{\lambda''(t)}{2}E[(S_e)^2])E[S] + O(\varepsilon^3). \end{aligned} \tag{18}$$

Before proving Theorem 10, we state a consequence.

Corollary 5. *Assume the conditions of Theorem* 10.

a. *If $(-1)^{n+1}\lambda^{(n+1)}$ is positive (negative), then $m(t) \geqslant (\leqslant)\ m_n(t)$;*

b. *If $(-1)^{n+1}\lambda^{(n+1)}$ is increasing (decreasing), then*

$$m(t) \leqslant (\geqslant)\ m_n(t) + (-1)^{n+1}\lambda^{(n+1)}(t)\,\frac{E[S^{n+2}]}{(n + 2)!};$$

c. *If $(-1)^{n+1}\lambda^{(n+1)}$ is convex (concave), then*

$$m(t) \geqslant (\leqslant)\ m_n(t) + (-1)^{n+1}\lambda^{(n+1)}(t - E[S_e^{(n+2)}])\,\frac{E[S^{n+2}]}{(n + 2)!}.$$

In order to prove Theorem 10, we need two lemmas.

Lemma 1. *Suppose that f is differentiable with a Riemann integrable derivative on $[t - x, t]$ for all $x > 0$. If $E[S] < \infty$, $E[|f(t - S)|] < \infty$ and $E[|f'(t - S_e)|] < \infty$, then*

$$E[f'(t - S_e)] = \frac{E[f(t) - f(t - S)]}{E[S]}.$$

Proof. We apply the fundamental theorem of the calculus and Fubini's theorem (with the moment conditions) to get

$$\begin{aligned} E[f(t) - f(t - S)] &= E\left[\int_0^S f'(t - s)\, ds\right] \\ &= E\left[\int_0^\infty 1_{\{S \geqslant s\}} f'(t - s)\, ds\right] \\ &= \int_0^\infty P(S \geqslant s) f'(t - s)\, ds \\ &= E[f'(t - S_e)]E[S]. \end{aligned}$$

We also need expressions for the moments of the iterated stationary-excess variables. Let $(n)_k = n(n - 1) \dots (n - k + 1)$.

Lemma 2. *For $n \geqslant 1$ and $k \geqslant 1$, if $E[S^{n+k}] < \infty$, then $S_t^{(n)}$ has a proper distribution and*

$$E[(S_e^{(n)})^k] = \frac{n!E[S^{n+k}]}{(n + k)_n E[S^n]}.$$

Proof. Use (2) plus induction on n and k.

Proof of Theorem 10. Apply Taylor's theorem to obtain

$$\lambda(t - S_e) = \sum_{j=0}^{n} \frac{\lambda^{(j)}(t)(-S_e)^j}{j!} + r_n(t),$$

where

$$\begin{aligned} r_n(t) &= \int_t^{t-S_e} \frac{(t - S_e - u)^{n-1}}{(n - 1)!} [\lambda^{(n)}(u) - \lambda^{(n)}(t)]\, du \\ &= \int_t^{t-S_e} \frac{(t - S_e - u)^n}{n!} \lambda^{(n+1)}(u)\, du \\ &= \int_{t-S_e}^{t} \frac{(t - v)^n}{n!} \lambda^{(n+1)}(v - S_e)\, dv, \end{aligned}$$

using integration by parts and the change of variables $v = S_e + u$. The result then follows from Lemmas 1 and 2 and induction on n. Indeed, for $n = 0$,

$$\begin{aligned} R_0 &\equiv E[r_0(t)] \\ &= E\left[\int_{t-S_e}^{t} \lambda^{(1)}(v - S_e)\, dv\right] = E[\lambda^{(1)}(t - S_e^{(2)})]E[S_e], \end{aligned}$$

by virtue of Lemma 1 with $f(t) = \int_0^t \lambda^{(1)}(v - S_e)\, dv$. (The moment conditions in Theorem 10 imply the moment conditions in Lemma 1.) Given that the result has been established for all $k \leqslant n - 1$, it suffices to show that

$$\frac{(-1)^n \lambda^{(n)}(t) E[S^{n+1}]}{(n + 1)!} + R_n(t) = R_{n-1}(t)$$

or

$$\begin{aligned} &E[\lambda^{(n+1)}(t - S_e^{(n+2)})] \\ &\quad = (n + 2) \frac{E[S^{n+1}]}{E[S^{n+2}]} (\lambda^{(n)}(t) - E[\lambda^{(n)}(t - S_e^{(n+1)})]), \end{aligned}$$

but this follows from Lemma 1 using $f(t) = \lambda^{(n)}(t)$. We use Lemma 2 to obtain $E[S_e^{(n+1)}] = E[S^{n+2}]/(n + 2)E[S^{n+1}]$.

4. STEP FUNCTIONS

We now consider arrival rate functions of the form

$$\lambda(t) = \begin{cases} b, & t < 0 \\ a + b, & t \geqslant 0 \end{cases} \tag{19}$$

for $b = 0$, which corresponds to the transient behavior of a stationary model starting out empty. It also covers the more general case of an instantaneous shift in rates of a homogeneous arrival process ($b \geqslant 0$) by virtue of Theorem 8. Hence, we restrict attention to (19) with $b = 0$.

From (3) and (19), we see that

$$m(t) = \lambda(0)P(S_e \leqslant t)E[S]. \tag{20}$$

Formula 20 obviously provides a remarkably clear description of the approach to steady state. As in Abate and Whitt (1987) and Mitra and Weiss (1989), $m(t)/m(\infty)$ is a cdf, indeed, the cdf of S_e. Its moments are given in (2). The first moment $E[S_e] = (c_s^2 + 1)/2$ is one useful representation of the notion of relaxation time. From it, we see that the approach to steady state is slower when the service-time distribution is more variable.

For the case of deterministic service times, S_e is uniform on [0, 1], so that

$$m(t) = \begin{cases} 0, & t < 0 \\ \lambda(0)t, & 0 \leqslant t \leqslant 1 \\ \lambda(0), & t \geqslant 1. \end{cases} \tag{21}$$

Moreover, since λ in (19) is increasing and the deterministic distribution is a lower bound in the convex stochastic order, (21) is an upper bound for m for all service-time distributions with mean 1 (see Theorem 3). Moreover, by Theorem 3, as the service-time distribution increases in the convex stochastic order, m decreases, so that steady state is approached more slowly.

In addition to the transient start-up of a stationary model represented by (19) with $b = 0$, we can consider the transient shut-down behavior of a stationary model characterized by an arrival rate function of the form

$$\lambda(t) = \begin{cases} \lambda(0-), & t < 0 \\ 0, & t \geqslant 0. \end{cases} \tag{22}$$

Since the homogeneous Poisson process with rate $\lambda(0)$ is the sum of the arrival rate functions in (19) with $b = 0$ and (22), we can apply Theorem 8 to immediately obtain

$$m(t) = \lambda(0-)P(S_e > t)E[S]. \qquad (23)$$

More generally, we can consider the effect of turning off an arbitrary arrival process at time t. For this purpose, let

$$\tilde{\lambda}(u) = \begin{cases} \lambda(u), & u < t \\ 0, & u \geq t. \end{cases} \qquad (24)$$

The following is an elementary consequence of (3) and Theorem 2.

Theorem 11. *The mean function $\tilde{m}$ associated with arrival rate function $\tilde{\lambda}$ in (24) satisfies $\tilde{m}(u) = \mathrm{Cov}[Q(t), Q(t+u)]$, where $Q(t)$ is the number of busy servers at time t in the $M_t/G/\infty$ model with arrival rate function λ.*

Note that (23) is covered by Theorem 11 and Remark 6.

5. SPIKES

In this section, we consider arrival rate functions of the form

$$\lambda(t) = \begin{cases} \lambda(0), & 0 \leq t < s \\ 0, & \text{otherwise.} \end{cases} \qquad (25)$$

These arrival functions are of interest to represent transient traffic surges, which can be added to other arrival rate functions by virtue of Theorem 8. They are also of interest for piecewise constant approximations of general arrival rate functions.

From (3) and (25), we see that

$$m(t) = \lambda(0)P(t - s \leq S_e \leq t)E[S]. \qquad (26)$$

As $s \to \infty$ in (25), (26) approaches (20).

6. CONCLUSIONS

Unlike almost any other queueing model with time-dependent arrival rates, the $M_t/G/\infty$ model is remarkably amenable to analysis. As illustrated here, the $M_t/G/\infty$ theory can be applied to obtain simple formulas that provide considerable insight into the time-dependent congestion, e.g., (7) and (8) in the linear case, (14) in the quadratic case, and (20) and (23) in the step function case. Simple formulas for sinusoidal arrival rates are given in Eick, Massey and Whitt (1993a).

In contrast to the steady-state distribution of the number $Q(t)$ of busy servers in a stationary $M/G/\infty$ model, the time-dependent distribution in an $M_t/G/\infty$ does not have the insensitivity property, i.e., the distribution of $Q(t)$ depends on the service-time distribution beyond its mean. From (3) and the other formulas here, it is clear that the service-time stationary-excess distribution in (1) plays a vital role. Indeed, if the arrival rate function is sufficiently smooth so that we can use the quadratic approximation (14), then the lag in peak congestion behind the peak arrival rate is approximately equal to the mean of the service-time stationary-excess distribution, i.e., $E[S_e] = E[S](c_s^2 + 1)/2$; see (2), (3), and (14).

When the arrival rate function λ is nearly constant before t, the mean number of busy servers at time t, $m(t)$, is approximately equal to $\lambda(t)E[S]$, which is the pointwise stationary approximation (**PSA**, see Remark 2), but as λ changes more before t, $m(t)$ begins to differ more from $\lambda(t)E[S]$. Then there tends to be a pronounced lag and the extremes of $m(t)$ tend to become less extreme than the extremes of $\lambda(t)E[S]$ (e.g., see (14)).

We believe that the simple formulas here, such as (14) for the quadratic case and (20) in the step function case, can quickly provide a rough idea about the time-dependent congestion in light-to-moderately loaded $M_t/G/s/r$ models. More generally, the infinite-server results can serve as a basis for more sophisticated approximations for these systems. Infinite-server approximations for loss models are investigated in Eick, Massey and Whitt (1993b).

ACKNOWLEDGMENT

We thank Erhan Çinlar and Andras Prékopa for bringing the Prékopa (1958) paper to our attention.

REFERENCES

ABATE, J., AND W. WHITT. 1987. Transient Behavior of the M/M/1 Queue: Starting at the Origin. *Queue. Syst* **2**, 41–65.

BARLOW, R. E., AND F. PROSCHAN. 1975. *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart and Winston, New York.

BENEŠ, V. 1957. Fluctuations of Telephone Traffic. *Bell Sys Tech. J.* **36**, 965–973.

BROWN, M. 1969. An Invariance Property of a Poisson Process. *J. Appl Prob.* **6**, 453–458.

BROWN, M., AND S. M. ROSS. 1969. Some Results for Infinite Server Poisson Queues. *J. Appl Prob.* **6**, 604–611.

CARRILLO, M. J. 1991. Extensions of Palm's Theorem: A Review. *Mgmt. Sci* **37**, 739–744.

COX, D. R., AND H. D. MILLER. 1965. *The Theory of Stochastic Processes.* John Wiley, New York.

DALEY, D. J. 1976. Queueing Output Processes. *Adv Appl. Prob.* **8,** 395–415.

DALEY, D. J., AND D. VERE-JONES. 1988. *An Introduction to the Theory of Point Processes.* Springer-Verlag, New York.

DOOB, J. L. 1953. *Stochastic Processes.* John Wiley, New York.

DUDA, A. 1986. Diffusion Approximations for Time-Dependent Queueing Systems. *IEEE J. Selected Areas Commun.* **4,** 905–918.

EICK, S. G., W. A. MASSEY AND W. WHITT. 1993a. $M_t/G/\infty$ Queues With Sinusoidal Arrival Rates. *Mgmt. Sci* **39,** 241–252.

EICK, S. G., W. A. MASSEY AND W. WHITT. 1993b. Infinite-Server Approximations for Multi-Server Loss Models With Time-Dependent Arrival Rates. AT&T Bell Laboratories, Murray Hill, N.J. (in preparation).

FOLEY, R. D. 1982. The Nonhomogeneous $M/G/\infty$ Queue. *Opsearch* **19,** 40–48.

FOLEY, R. D. 1986. Stationary Poisson Departure Processes From Non-Stationary Queues. *J. Appl. Prob* **23,** 256–260.

GREEN, L., AND P. KOLESAR. 1991. The Pointwise Stationary Approximation for Queues With Nonstationary Arrivals. *Mgmt Sci* **37,** 84–97.

GREEN, L., P. KOLESAR AND A. SVORONOS. 1991. Some Effects of Nonstationarity on Multiserver Markovian Queueing Systems. *Opns. Res.* **39,** 502–511.

HARRISON, J. M., AND A. J. LEMOINE. 1981. A Note on Networks of Infinite-Server Queues. *J. Appl. Prob.* **18,** 561–567.

HILLESTAD, R. J., AND M. J. CARRILLO. 1980. Models and Techniques for Recoverable Item Stockage When Demand and the Repair Process Are Nonstationary—Part I: Performance Measurement. The Rand Corporation, Santa Monica, Calif.

HUFFER, F. W. 1987. Inequalities for the $M/G/\infty$ Queue and Related Shot Noise Processes. *J. Appl. Prob.* **24,** 978–989.

JAGERMAN, D. L. 1975. Nonstationary Blocking in Telephone Traffic. *Bell Syst. Tech. J.* **54,** 625–661.

KEILSON, J., AND L. D. SERVI. 1989. Networks of Nonhomogeneous $M/G/\infty$ Systems. GTE Laboratories, Waltham, Mass.

KHINTCHINE, A. Y. 1955. *Mathematical Methods in the Theory of Queueing* (in Russian). *Trudy Mat Inst. Steklov* 49 (English translation by Charles Griffin and Co., London, 1960).

KINGMAN, J. F. C. 1969. Markov Population Processes. *J. Appl. Prob.* **6,** 1–18.

MASSEY, W. A. 1981. Nonstationary Queues. Ph.D. Dissertation, Department of Mathematics, Stanford University, Stanford, Calif.

MASSEY, W. A. 1985. Asymptotic Analysis of the Time Dependent M/M/1 Queue. *Math. Opns. Res.* **10,** 305–327.

MASSEY, W. A., AND W. WHITT. 1993. Networks of Infinite-Server Queues With Nonstationary Poisson Input. *Queue Syst.* (to appear).

MIRASOL, N. M. 1963. The Output of an $M/G/\infty$ Queueing System is Poisson. *Opns. Res.* **11,** 282–284.

MITRA, D., AND A. WEISS. 1989. The Transient Behavior in Erlang's Model for Large Trunk Groups and Various Traffic Conditions. In *Teletraffic Science for the New Cost-Effective Systems, Networks and Services,* ITC 12, M. Bonatti (ed.). North-Holland, Amsterdam, 1367–1374.

NEWELL, G. F. 1966. The $M/G/\infty$ Queue. *SIAM J. Appl. Math.* **14,** 86–88.

NEWELL, G. F. 1973. *Approximate Stochastic Behavior of n-Server Service Systems With Large n.* Lecture Notes in Econ. and Math. Systems 87, Springer-Verlag, Berlin.

ONG, K. L., AND M. R. TAAFFE. 1989. Nonstationary Queues With Interrupted Poisson Arrivals and Unreliable/Repairable Servers. *Queue. Syst* **4,** 27–46.

PALM, C. 1943. Intensity Variations in Telephone Traffic. *Ericsson Technics* **44,** 1–189 (in German). (English translation by North-Holland, Amsterdam, 1988).

PRÉKOPA, A. 1958. On Secondary Processes Generated by a Random Point Distribution of Poisson Type. *Annales Univ. Sci Budapest de Eotvos Nom. Sectio Math.* **1,** 153–170.

RIORDAN, J. 1951. Telephone Traffic Time Averages. *Bell Syst. Tech. J.* **30,** 1129–1144.

ROLSKI, T. 1989. Queues With Nonstationary Inputs. *Queue. Syst* **5,** 113–130.

ROSS, S. M. 1983. *Stochastic Processes.* John Wiley, New York.

SERFOZO, R. F. 1990. Point Processes. In *Handbooks in OR and MS,* Vol. 2, D. P. Heyman and M. J. Sobel (eds.). Elsevier Science Publishers, Amsterdam, 1–93.

STOYAN, D. 1983. *Comparison Methods for Queues and Other Stochastic Models.* John Wiley, New York.

THORISSON, H. 1985. On Regenerative and Ergodic Properties of the k-Server Queue With Nonstationary Poisson Arrivals. *J. Appl. Prob.* **22,** 893–902.

WHITT, W. 1984. Minimizing Delays. *Opns. Res.* **32,** 41–51.

WHITT, W. 1985. The Renewal-Process Stationary-Excess Operator. *J. Appl. Prob.* **22,** 156–167.

WHITT, W. 1991. The Pointwise Stationary Approximation for $M_t/M_t/s$ Queues is Asymptotically Correct as the Rates Increase. *Mgmt. Sci.* **37,** 307–314.

ZIEMER, R. E., AND W. H. TRANTER. 1976. *Principles of Communications Systems, Modulation and Noise.* Houghton Mifflin, Boston.