Queueing Systems, 6 (1990) 335–352

QUEUES WITH SERVICE TIMES AND INTERARRIVAL TIMES DEPENDING LINEARLY AND RANDOMLY UPON WAITING TIMES

Ward WHITT

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

Received 9 December 1988; revised 8 January 1990

Abstract

We consider a modification of the standard $G/G/1$ queue with unlimited waiting space and the first-in first-out discipline in which the service times and interarrival times depend linearly and randomly on the waiting times. In this model the waiting times satisfy a modified version of the classical Lindley recursion. We determine when the waiting-time distributions converge to a proper limit and we develop approximations for this steady-state limit, primarily by applying previous results of Vervaat [21] and Brandt [4] for the unrestricted recursion $Y_{n+1} = C_n Y_n + X_n$. Particularly appealing for applications is a normal approximation for the stationary waiting time distribution in the case when the queue only rarely becomes empty. We also consider the problem of scheduling successive interarrival times at arrival epochs, with the objective of achieving nearly maximal throughput with nearly bounded waiting times, while making the interarrival time sequence relatively smooth. We identify policies depending linearly and deterministically upon the work in the system which meet these objectives reasonably well; with these policies the waiting times are approximately contained in a specified interval a specified fraction of time.

Keywords: State-dependent service and interarrival times, Lindley equation, recursive stochastic equations, stability, stochastic comparisons, normal approximations, scheduling arrivals.

1. Introduction

In this paper we consider a modification of the standard $G/G/1$ queue (with unlimited waiting space and the first-in first-out discipline) in which the service times and the interarrival times depend *linearly* and *randomly* on the waiting times. Our model is specified by a stationary and ergodic sequence of four-tuples of nonnegative random random variables $\{(U_n, V_n, A_n, B_n): n \geqslant 0\}$. (We do not assume independence among different vectors or within each vector, although we will at various points below.) We study the sequence $\{W_n: n \geqslant 0\}$, which is defined recursively by

$$W_{n+1} = \left[W_n + \overline{V}_n - \overline{U}_n\right]^+, \quad n \geqslant 0, \tag{1.1}$$

where $[x]^+ = \max\{x, 0\}$,

$$\bar{U}_n = U_n + A_n W_n, \tag{1.2}$$

$$\bar{V}_n = V_n + B_n W_n, \tag{1.3}$$

and W_0 is a nonnegative random variable.

We interpret W_n as the waiting time and $\bar{V}_n$ as the service time of customer n; we interpret $\bar{U}_n$ as the interarrival time between customers n and $n+1$. We call V_n the *nominal service time* of customer n and U_n the *nominal interarrival time* between customers n and $n+1$, because these would be the actual times if the state-dependent behavior were omitted, i.e., if $A_n = B_n = 0$ w.p.1. We assume that $0 < E[U_0] < \infty$ and $E[V_0] < \infty$, and define the *nominal traffic intensity* in the usual way as $\rho = E[V_0]/E[U_0]$.

We analyze this model by recognizing that the waiting times satisfy the generalized Lindley recursion

$$W_{n+1} = [C_n W_n + X_n]^+, \quad n \geq 0, \tag{1.4}$$

where

$$C_n = 1 + B_n - A_n \quad \text{and} \quad X_n = V_n - U_n, \quad n \geq 0. \tag{1.5}$$

Equation (1.4) reduces to the classical Lindley recursion when $P(C_0 = 1) = 1$. As in the classical case, our analysis depends on (1.4) and the sequence $\{(C_n, X_n)\}$, and not on the specific way (C_n, X_n) is defined in terms of (U_n, V_n, A_n, B_n) in (1.5). Recursion (1.4) is a special case of more general recursions that have been analyzed in the literature (e.g., Borovkov [3], Lisek [14] and references cited there), but it seems that we obtain stronger results for (1.4) by exploiting the special structure. There also has been considerable previous work on queues with state-dependent service and arrival processes; e.g., see Brill [5], Callahan [6], Harris [10,11], Laslett, Pollard and Tweedie [13], Mudrov [16], Posner [17], Rosenshine [18], and Sugawara and Takahashi [20].

Our analysis of (1.4) is primarily based on relating it to the unrestricted recursion

$$Y_{n+1} = C_n Y_n + X_n, \quad n \geq 0, \tag{1.6}$$

(without the positive-part operator corresponding to the barrier at the origin) which has been studied by Vervaat [21] (who reviews the extensive earlier literature) and Brandt [4]. In the more general framework of (1.4) and (1.6), W_n or Y_n may represent an inventory in time period n (e.g., cash), C_n may represent a multiplicative, possibly random, decay or growth factor between times n and $n+1$ (e.g., interest rate) and X_n may represent a quantity that is added or subtracted between times n and $n+1$ (e.g., deposit minus withdrawal). Obviously the positive-part operator in (1.4) is appropriate for many applications.

A major conclusion of this paper (developed in section 2) is that the system studied here has dramatically different stability conditions than the nominal system in which $C_n = 1$ w.p.1 in (1.4). In the nominal system, the familiar stability condition is $\rho < 1$. In contrast, when $P(C_n = 1) < 1$, stability depends on the multiplicative factor C_n instead of ρ (see theorem 1). Moreover, for this model, the concept of stability is only a limited partial characterization. It is possible to have instability, even though the time required to reach a high level, from which the process can diverge to $+\infty$, can be extraordinarily long with high probability. On the other hand, it is possible to have stability even though the limiting distribution can concentrate on very high values. In section 3 we establish stochastic comparison results that enable us to compare different systems.

We also focus on stable systems with $\rho > 1$. Having $\rho > 1$ can tend to keep the process $\{W_n\}$ in (1.4) away from the origin, so that $\{W_n\}$ behaves much like $\{Y_n\}$ in (1.6). In section 4 we show that a normal approximation for $\{Y_n\}$ developed by Vervaat [21] also applies to $\{W_n\}$ when $\rho > 1$ under appropriate conditions. In section 5 we apply this normal approximation to determine specific policies of the form (1.2) for scheduling interarrival times under which the waiting times are approximately contained in a specified interval a specified fraction of time. These policies have the property that the interarrival times change smoothly, which is desirable in some contexts, e.g., in production smoothing (see pp. 400–413 of Heyman and Sobel [12]). Our analysis in section 5 is somewhat in the spirit of recent work by Denardo and Tang [8,9], Although this paper was done independently, Denardo and Tang consider controlled service processes in queueing networks using something like a vector generalization of (1.6) as a direct approximation. A multidimensional version of this paper is an intended sequel.

Something like the linear control in (1.4) is automatically achieved in queueing systems with c exponential servers when c is large. Then the arrival rate is fixed, say at λ, but the service rate is $k\mu$ when there are k busy servers; i.e., the service rate is state-dependent. If c is large, then typically $\lambda \gg k\mu$ for small k, but $\lambda < c\mu$, so that the number of busy servers tends to concentrate around $k^* < c$ such that $\lambda = k^*\mu$. Assuming that k^* is fairly high but not too close to c, the steady-state number of customers tends to be approximately normally distributed with mean k^* and standard deviation of order $\sqrt{k^*}$ (see Whitt [24]).

2. Stability

We start with two preliminary lemmas. The first relates to recursion (1.4) to an associated unrestricted recursion of the form (1.6). We say that a sequence $\{W_n: n \geq 0\}$ is *stochastically bounded* if for all $\epsilon > 0$ there exists a constant K such that $P(|W_n| > K) < \epsilon$ for all n, i.e., if the sequence $\{W_n\}$ is tight (see p. 90 of Chung [7]). A sequence is stochastically bounded if and only if every subsequence has a subsequence converging to a proper limit. Let $\Rightarrow$ denote convergence in distribution.

From (1.4), it is obvious that the waiting times will be at least as large if we replace C_n by $[C_n]^+$ and X_n by $[X_n]^+$. Moreover, when we make this change, the positive-part operator in (1.4) becomes unnecessary.

LEMMA 1

If W_n satisfies (1.4), then $W_n \leqslant Y_n$ for all n w.p.1, where

$$Y_{n+1} = [C_n]^+ Y_n + [X_n]^+, \quad n \geqslant 0, \tag{2.1}$$

and $Y_0 = W_0 \geqslant 0$.

COROLLARY

If W_n satisfies (1.4), Y_n satisfies (2.1) and $Y_n \Rightarrow Y$ as $n \to \infty$ where Y is proper, then $\{W_n\}$ is stochastically bounded and $P(W > t) \leqslant P(Y > t)$ for all t, where W is the limit (in $\Rightarrow$) of any convergent subsequence of $\{W_n\}$.

We now obtain an explicit expression for W_n in the case $P(C_0 \geqslant 0) = 1$. Without loss of generality, we assume that the stationary sequence $\{(C_n, X_n): n \geqslant 0\}$ has been extended to $-\infty < n < \infty$. Let $\stackrel{d}{=}$ denote equality in distribution. We say that a sequence $\{W_n: n \geqslant 0\}$ is *stochastically increasing* if $W_n \leqslant_d W_{n+1}$ for all n, where $\leqslant_d$ denotes stochastic order, as on p. 4 of Stoyan [19].

LEMMA 2

If $P(C_0 \geqslant 0) = 1$, then

$$\begin{aligned} W_{n+1} = \max\{0,\ &X_n,\ X_n + C_n X_{n-1},\ X_n + C_n X_{n-1} + C_n C_{n-1} X_{n-2}, \\ &\ldots,\ X_n + C_n X_{n-1} + \cdots + C_n \ldots C_2 X_1, \\ &X_n + C_n X_{n-1} + \cdots + C_n \ldots C_2 X_1 + C_n \ldots C_1 X_0 + C_n \ldots C_0 W_0\} \end{aligned} \tag{2.2}$$

$$\begin{aligned} \stackrel{d}{=} M_{n+1} \equiv \max\{0,\ &X_0,\ X_0 + C_0 X_{-1},\ X_0 + C_0 X_{-1} + C_0 C_{-1} X_{-2}, \\ &\ldots,\ X_0 + C_0 X_{-1} + \cdots + C_0 \ldots C_{-(n-2)} X_{-(n-1)}, \\ &X_0 + C_0 X_{-1} + \cdots + C_0 \ldots C_{-(n-2)} X_{-(n-1)} \\ &+ C_0 \ldots C_{-(n-1)} X_{-n} + C_0 \ldots C_{-n} W_0\}. \end{aligned} \tag{2.3}$$

If $W_0 = 0$, then W_n is stochastically nondecreasing in n, so that $W_n \Rightarrow W \stackrel{d}{=} M$ as $n \to \infty$ where $M_n \Rightarrow M$ as $n \to \infty$, with M possibly being improper.

Proof

Since $P(C_0 \geqslant 0) = 1$, we can apply mathematical induction in the usual way, following Loynes [15] and chap. 1 of Borovkov [2]. Since $W_n \stackrel{d}{=} M_n$ and M_n is nondecreasing in n w.p.1, W_n is stochastically increasing.

Remarks

(2.1) It is easy to see that the nonnegativity condition on C_0 is needed in lemma 2; note that (2.2) fails for $n = 2$ in example 2.4 below.

(2.2) As on pp. 9–13 of Borovkov [2], under the condition of lemma 2, we can show that the entire sequence $\mathbf{W}_k \equiv \{W_{k+n}\colon n \geqslant 0\}$ converges in distribution to the stationary sequence $\mathbf{W}_s \equiv \{W_{sn}\colon n \geqslant 0\}$ as $k \to \infty$ when $W_0 = 0$, where

$$W_{sn} = \sup\left\{0,\ \sum_{j=0}^{k}\left(\prod_{j=n+1-j}^{n} C_i\right) X_{n-j}\colon k \geqslant 0\right\}, \quad n \geqslant 0, \tag{2.4}$$

where $\prod_{i=n+1}^{n} C_i = 1$. □

Here is our main stability result.

THEOREM 1

There are five possible cases:

(a) If $P(C_0 < 0) > 0$, then W_n is stochastically bounded for all ρ and W_0. If, in addition, $\{(C_n, X_n)\}$ is a sequence of independent vectors with $P(C_0 \leqslant 0,\ X_0 \leqslant 0) > 0$, then the events $\{W_{n+1} = 0\}$ are regeneration points with finite mean time and $W_n \Rightarrow W$ as $n \to \infty$, where W is proper for all ρ and W_0.

(b) If $P(C_0 \geqslant 0) = 1$ and $P(C_0 = 0) > 0$, then $W_n \Rightarrow W$ as $n \to \infty$, where W is proper for all ρ and W_0.

(c) If $P(C_0 > 0) = 1$ and $E[\log C_0] < 0$, then $W_n \Rightarrow W$ as $n \to \infty$, where W is proper for all ρ and W_0.

(d) If $P(C_0 > 0) = 1$ and $E[\log C_0] > 0$, then $W_n/(C_0 \dots C_{n-1}) \to \hat{W}$ w.p.1 as $n \to \infty$, where $(C_0 \dots C_{n-1})^{1/n} \to \mathrm{e}^{E[\log C_0]} > 1$, w.p.1 as $n \to \infty$ and $\hat{W}$ is proper for all ρ and W_0. Moreover, there is a proper initial workload W_0 such that W_n diverges to $+\infty$ w.p.1. If, in addition, $\{(C_n, X_n)\}$ is a sequence of independent random vectors with $P(X_0 > 0) > 0$, then W_n diverges to $+\infty$ w.p.1 for all ρ and W_0.

(e) If $P(C_0 > 0) = 1$ and $E[\log C_0] = 0$, then $W_n \Rightarrow W$ when $W_0 = 0$, where W may be proper or improper. If $P(C_0 = 1) = 1$, then $W_n \Rightarrow W$ for all W_0, where W is proper (improper) for $\rho < 1$ ($\rho > 1$).

Proof

(a) Theorem 1 of Brandt [4] implies that Y_n satisfying (2.1) converges to a proper limit if $P([C_0]^+ = 0) = P(C_0 \leqslant 0) > 0$, which with lemma 1 implies that W_n is stochastically bounded. To establish the second statement, note that $W_{n+1} = 0$ whenever $C_n \leqslant 0$ and $X_n \leqslant 0$. Hence, if $P(C_0 \leqslant 0,\ X_0 \leqslant 0) = p > 0$, then $W_{n+1} = 0$ infinitely often w.p.1. If, in addition, $\{(C_n, X_n)\}$ is a sequence of independent vectors, then the events $\{W_{n+1} = 0\}$ are regeneration points for the process. Let T be a generic time between regenerations, i.e., the first passage time

$$T = \min\{n \geqslant 1\colon W_n = 0 \mid W_0 = 0\}. \tag{2.5}$$

Since T is stochastically dominated by a geometrically distributed random variable with parameter p, $E[T] < \infty$. Moreover, $P(T = 1) = P(W_1 = 0 \mid W_0 = 0)$

$= p > 0$, so that the regenerations are aperiodic. Hence, $W_n \Rightarrow W$ where W is proper. Indeed, for any bounded measurable real-valued function f,

$$E[f(W)] = E\left[\sum_{k=1}^{T} f(W_k)\right] / E[T]. \tag{2.6}$$

(b) Since $P(C_0 = 0) > 0$, $\{W_n\}$ is stochastically bounded just as in the proof of part (a). Since $P(C_0 \geqslant 0) = 1$, $\{W_n\}$ is stochastically increasing in n when $W_0 = 0$ by lemma 2. Hence, W_n converges to a proper limit when $W_0 = 0$. Let $W_n(W_0)$ represent W_n as a function of the initial workload W_0. Since $P(C_0 \geqslant 0) = 1$, $W_n(W_0) \geqslant W_n(0)$ for all n, but $W_n(W_0) = W_n(0)$ for all $n > n_0$ where $C_{n_0} = 0$. Since $P(C_0 = 0) > 0$, $C_n = 0$ infinitely often w.p.1. Hence, W_n converges to the same proper limit independent of the initial workload W_0.

(c) Since $P(C_0 \geqslant 0) = 1$, W_n is stochastically increasing when $W_0 = 0$ by lemma 2. By theorem 1 of Brandt [4] and lemma 1 above, W_n is stochastically bounded if $E[\log C_0] < 0$ and $E(\log[X_0]^+) < \infty$. Since

$$\log[X_0]^+ \leqslant [X_0]^+ \leqslant V_0 \tag{2.7}$$

and we have assumed $E[V_0] < \infty$, we have $E[\log[X_0]^+) < \infty$. Hence, W_n converges to a proper limit when $W_0 = 0$. From (1.4), by mathematical induction,

$$|W_n(W_0^1) - W_n(W_0^2)| \leqslant (C_{n-1} \cdots C_0)|W_0^2 - W_0^2|, \quad n \geqslant 1. \tag{2.8}$$

By the strong law of large numbers,

$$(C_{n-1} \cdots C_0)^{1/n} = e^{\log[(C_{n-1} \cdots C_0)^{1/n}]} = e^{n^{-1}\sum_{i=1}^{n-1} \log C_i} \to e^{E[\log C_0]} < 1 \text{ w.p.1}$$
$$\text{as } n \to \infty, \tag{2.9}$$

so that $(C_{n-1} \cdots C_0) \to 0$ as $n \to \infty$ w.p.1 and W_n converges to a proper limit for any initial workload.

(d) Note that

$$\frac{W_{n+1}}{C_0 \cdots C_n} = \max\{0, Z_n, Z_n + Z_{n-1}, Z_n + Z_{n-1} + Z_{n-2}, \ldots, Z_n + Z_{n-1} + \ldots + Z_0 + W_0\}, \tag{2.10}$$

where

$$Z_n = \frac{X_n}{C_0 \cdots C_n}, \quad n \geqslant 0. \tag{2.11}$$

The claimed convergence of (2.10) follows because $\sum_{n=0}^{\infty} Z_n$ converges absolutely w.p.1. To see this, recall that, by (2.9), $(C_0 \ldots C_n) > \alpha^{n+1}$ for some $\alpha > 1$ and all $n \geqslant n_0$ w.p.1 (where n_0 depends on the sample path). Then $\sum_{n=0}^{\infty} \alpha^{-(n+1)} X_n$ converges absolutely w.p.1 because $E[|X_n|] \leqslant E[V_n] + E[U_n] < \infty$ (see theorem 5.3.4 and exercise 7, pp. 120–122, of Chung [7]). If

$$W_0 > \sum_{i=0}^{\infty} |Z_i|, \tag{2.12}$$

then the limit of (2.10) is always strictly positive, implying that $W_n \to \infty$ w.p.1 because

$$\frac{W_{n+1}}{C_0 \ldots C_n} \geqslant \sum_{i=0}^{n} Z_i + W_0, \quad n \geqslant 0. \tag{2.13}$$

To establish the final conclusion under the independence condition, note that from the limit for (2.10) we have $W_n \to W^*$ w.p.1 where $P(W^* = 0) = 1 - P(W^* = \infty)$. However, from (1.4),

$$P(W^* = 0) = P\left([C_0 W^* + X_0]^+ = 0\right) = P(W^* = 0) P(X_0 \leqslant 0), \tag{2.14}$$

where C_0 and X_0 are independent of W^*. Since $P(X_0 > 0) > 0$, we must have $P(W^* = 0) = 0$.

(e) Stochastic monotonicity when $W_0 = 0$ follows from lemma 2. The possibility of different cases follows from the standard case with $P(C_0 = 1) = 1$ (see p. 14 of Borovkov [2]). □

Remarks

(2.3) As we should expect, theorem 1 concludes that in (1.4) the multiplicative adjustment C_n is stronger than the additive adjustment X_n. Moreover, the distribution of C_n for positive values does not cause instability provided that $P(C_0 \leqslant 0) > 0$. Given that C_n assumes only positive values, the critical quantity for stability is $E[\log C_0]$, which is understandable when we see that in successive iterations of (1.4) we encounter the products $(C_n, C_{n-1} \ldots C_0)$ (see lemma 2).

(2.4) Under extra independence conditions, the results of Laslett, Pollard and Tweedie [13] are applicable to this problem, but the familiar mean-drift criterion yields only a weaker sufficient condition for stability, namely, $P(C_0 > 0) = 1$ and $E[C_0] < 1$. (By Jensen's inequality, $E[C_0] < 1$ implies that $E[\log C_0] < 0$, but not conversely.) However, an appropriate Liapounov function to obtain the same results for the i.i.d. case is $h(x) = \log x$.

(2.5) A standard random walk argument can be used to establish divergence in the final statement of theorem 1(d) under stronger conditions, namely, $\{(C_n, X_n)\}$ being a sequence of independent vectors with $EC_0 \geqslant 1$ and $P(X_0 > 0, C_0 \geqslant 1) > 0$. Since $P(X_0 > 0, C_0 \geqslant 1) > 0$, arbitrarily large values of W_n will occur infinitely often w.p.1. For sufficiently high w, $\{W_n\}$ starting above w is greater than a random walk with mean step size $E[(C-1)w] + E[X] > 0$. By theorems 8.2.4, 8.3.4 and 8.4.4 of Chung [7], this random walk will diverge to $+\infty$ before going below w with positive probability. Since the process W_n necessarily exceeds w infinitely often, $W_n \to \infty$ w.p.1. (If $P(C_0 > 1 + \epsilon) = 1$ and $P(X_0 \geqslant -m) = 1$, then the argument is much more direct.)

(2.6) To see that independence is needed in the final statement of theorem 1(d) and remark 2.5, note that without it arbitrarily large values of w need not be reached by W_n. When $P(X_0 \leqslant m) = 1$ and $E[X_0] < 0$, it is possible to have $X_i + \ldots + X_n \leqslant m$ for all i and n. Indeed this occurs in example 2.5 below. □

Examples

(2.1) Suppose that we only modify the service times via (1.3); i.e., let $P(A_0 = 0) = 1$ and $P(B_0 > 0) > 0$. Theorem 1(d) implies that if $\{(U_n, V_n, B_n)\}$ is a sequence of independent vectors with $P(V_0 > U_0) > 0$, then W_n *diverges* to $+\infty$ w.p.1 for all initial workloads W_0 and all nominal traffic intensities ρ.

(2.2) Suppose that we only modify the interarrival times via (1.2); i.e., let $P(B_0 = 0) = 1$ and $P(A_0 > 0) > 0$. Theorem 1(b) and (c) imply (without any independence assumptions) that if $P(0 \leqslant A_0 \leqslant 1) = 1$, then W_n *converges* in distribution to a proper limit for all initial workloads W_0 and all nominal traffic intensities ρ.

(2.3) Suppose that we consider a symmetric case involving both (1.2) and (1.3); i.e., let A_0 be distributed the same as B_0 with $P(0 \leqslant A_0 \leqslant 1) = 1$ and $P(A_0 = 0) < 1$. As a consequence, $E[C_0] = 1$, so that in a certain mean-value sense (1.4) is like the standard $G/G/1$ queue. However, theorem 1(c) implies (without any independence assumptions) that, if $P(C_0 = 1) < 1$, then W_n *converges* in distribution to a proper limit for all initial workloads W_0 and all nominal traffic intensities ρ. The situation is not all good here, however, because $E[W_n] \to \infty$ as $n \to \infty$ when $\rho \geqslant 1$ (see the corollary to theorem 2).

(2.4) To see that there need not be either convergence to a proper limit or divergence to $+\infty$, and to see the reason for having $P(A_0 \leqslant 1)$ in examples 2.2 and 2.3, consider the purely deterministic model with $U_n = 1$, $V_n = 2$, $A_n = 2$ and $B_n = 0$ for all n, and $W_0 = 0$. Then $W_{2n+1} = 1$ and $W_{2n} = 0$ for all n. Since $C_n = -1$, this example is in the setting of theorem 1(a).

(2.5) To see that W_n can be stochastically bounded when $E[\log C_0] > 0$ in theorem 1(d), let

$$P\big(C_{2n} = e^3 \text{ and } C_{2n+1} = e^{-2} \text{ for all } n\big)$$
$$= P\big(C_{2n} = e^{-2} \text{ and } C_{2n+1} = e^3 \text{ for all } n\big) = 1/2$$

and

$$P(X_{2n} = -100 \text{ and } X_{2n+1} = 1 \text{ for all } n)$$
$$= P(X_{2n} = 1 \text{ and } X_{2n+1} = -100 \text{ for all } n) = 1/2,$$

where $\{C_n\}$ is independent of $\{X_n\}$. If $W_0 = 0$, then

$$P(W_{2n} = 1 \text{ and } W_{2n+1} = 0 \text{ for all } n)$$
$$= P(W_{2n+1} = 1 \text{ and } W_{2n} = 0 \text{ for all } n) = 1/2;$$

i.e., $W_n \Rightarrow W$ where $P(W = 0) = P(W = 1) = 1/2$. On the other hand, if $W_0 = 1000$, then $W_n \to \infty$ w.p.1., so that whether or not $\{W_n\}$ is stochastically bounded depends on the initial workload W_0.

(2.6) To clearly see that convergence to a proper limit can occur for all initial workloads W_0 and all nominal traffic intensities ρ, as claimed in theorem 1(c), we give an example for which it is easy to calculate the limiting distribution explicitly. Let $P(X_0 = x) = 1$ for $x > 0$ (so that $\rho > 1$) and let

$$P(C_{2n+1} = c_1 \text{ and } C_{2n} = c_2 \text{ for all } n)$$
$$= P(C_{2n+1} = c_2 \text{ and } C_{2n} = c_1 \text{ for all } n) = 1/2,$$

where $c_1 > c_2 > 0$ and $c_1 c_2 < 1$. Then the two-step transition operator applied to any state w is either

$$T_1(w) = c_1[c_2 w + x] + x = c_1 c_2 w + (1 + c_1)x$$

or

$$T_2(w) = c_2[c_1 w + x] + x = c_1 c_2 w + (1 + c_2)x,$$

each with probability 1/2. Operators T_1 and T_2 are contractions with unique fixed points

$$w_1^* = \frac{(1 + c_1)x}{1 - c_1 c_2} \quad \text{and} \quad w_2^* = \frac{(1 + c_2)x}{1 - c_2 c_2}.$$

Moreover, $c_2 w_1^* + x = w_2^*$ and $c_1 w_2^* + x = w_1^*$. Hence, $W_n \Rightarrow W$ as $n \to \infty$, where

$$P(W = w_1^*) = P(W = w_2^*) = 1/2,$$

independent of W_0 and x. Unlike the standard $GI/GI/1$ queue, this example is stable without 0 being visited infinitely often. □

3. Stochastic comparisons

We now make some stochastic comparisons in the spirit of Stoyan [19], which for example allow us to compare our model to the standard $GI/GI/1$ queue. Let $\leqslant_{\mathrm{d}}$ denote stochastic order and $\leqslant_{\mathrm{c}}$ increasing convex order, as on pp. 4, 8, 26 of [19]. The following is the analog of the external monotonicity result in Stoyan's theorem 5.2.1. Let a second subscript index the system.

THEOREM 2

(a) Let $\{(C_{ni}, X_{ni})\}$ be a sequence of i.i.d. vectors characterizing system i with $P(C_{0i} \geqslant 0) = 1$ for each i. If $W_{01} \leqslant W_{02}$ and $(C_{01}, X_{01}) \leqslant (C_{02}, X_{02})$, then $W_{n1} \leqslant W_{n2}$ for all n, where $\leqslant$ denotes either $\leqslant_{\mathrm{d}}$ or $\leqslant_{\mathrm{c}}$ throughout.

(b) If, in addition, $E[W_{02}] < \infty$, $E[C_{02}] < \infty$ and $E[X_{02}]^+) < \infty$, then $E[W_{n1}] \leqslant E[W_{n2}] < \infty$ for all n. If $E[C_{02}] < 1$ as well, then $W_{ni} \Rightarrow W_i$ and $E[W_{ni}] \to E[W_i] < \infty$ as $n \to \infty$ for each i, and $W_1 \leqslant W_2$ (ordering for the steady-state waiting times) in the same sense.

Proof

(a) First, for $\leqslant_{\mathrm{d}}$ note that $[cw + x]^+$ is a nondecreasing function of (c, w, x) provided that $c \geqslant 0$ and $w \geqslant 0$. Then apply induction. For $\leqslant_{\mathrm{c}}$, note that

$$C_1 W_1 + X_1 \leqslant_{\mathrm{c}} C_2 W_1 + X_2 \leqslant_{\mathrm{c}} C_2 W_2 + X_2$$

by first conditioning on W_1 and then conditioning on (C_2, X_2). Then $[C_1W_1 + X_1]^+ \leqslant_c [C_2W_2 + X_2]^+$ because $[x]^+$ is nondecreasing and convex. Complete the proof by applying induction.

(b) Apply lemma 1 to bound W_{n2} by the unrestricted process. By the assumed independence,

$$EY_{n+1,2} = E\left([C_{n2}]^+ Y_{n2} + [X_{n2}]^+\right) = E\left([C_{n2}]^+\right)EY_{n2} + E\left([X_{n2}]^+\right),$$

which is finite by induction. Apply theorem 5.1 of Vervaat [21] to establish convergence of the moments for the unrestricted process. Since $E[\log C_0] < 0$ when $E[C_0] < 1$ (remark 2.4), theorem 1(b) and (c) imply that $W_{ni} \Rightarrow W_i$ where W_i is proper. Apply the stochastic dominance to get the uniform integrability needed to establish $E[W_{n2}] \to E[W_2] < \infty$ (as on p. 32 of Billingsley [1]). Finally, apply proposition 1.3.2 of Stoyan [19] to get preservation of $\leqslant_c$ order under convergence in distribution. □

The following corollary to theorems 1 and 2 supports a conclusion about example 2.3. In particular, it shows that the means diverge even though the model is stable.

COROLLARY

Let $\{C_{ni}\}$ and $\{X_{ni}\}$ be independent sequences of i.i.d. random variables for each i. Let $P(C_{01} = 1) = 1$, as in a standard $GI/GI/1$ queue. If $E[C_{02}] \geqslant 1$, $W_{01} \stackrel{d}{=} W_{02}$ and $X_{01} \stackrel{d}{=} X_{02}$, then $C_{01} \leqslant_c C_{02}$, so that $W_{n1} \leqslant_c W_{n2}$ for all n. If, in addition, $E[X_{01}] \geqslant 0$, then $W_{n1} \stackrel{P}{\to} \infty$ and $E[W_{n1}] \leqslant E[W_{n2}] \to \infty$ as $n \to \infty$. However, if $E[\log C_{02}] < 0$, then $W_{n2} \Rightarrow W_2$ as $n \to \infty$, where W_2 is proper.

4. Normal approximation when $\rho > 1$

In this section we assume that $\{(X_n, C_n)\}$ is a sequence of i.i.d. random vectors with $E[X_0^2] < \infty$, $P(C_0 > 0) = 1$, $E[(\log C_0)^2] < \infty$ and $E[\log C_0] < 0$, so that $W_n \Rightarrow W$ as $n \to \infty$, where W is proper, by theorem 1(c). Drawing on section 6 of Vervaat [21], we show that if $E[X_0] > 0$, which corresponds to $\rho > 1$, and $|E[\log C_0]|$ is suitably small, then W is approximately normally distributed with mean

$$E[W] \approx \frac{E[X_0]}{|E[\log C_0]|} \tag{4.1}$$

and variance

$$Var[W] \approx \frac{(E[X_0])^2 Var[\log C_0]}{2|E[\log C_0]|^3} + \frac{Var[X_0]}{2|E[\log C_0]|} + \frac{E[X_0]Cov[X_0, \log C_0]}{(E[\log C_0])^2}. \tag{4.2}$$

Remarks

(4.1) Since W is nonnegative, one test for the reasonableness of this approximation is that the mean $E[W]$ should be sufficiently far away from 0 in the scale of the standard deviation $(Var[W])^{1/2}$.

(4.2) Assuming that the process $\{W_n\}$ tends not to be near the origin (which is what is happening in this case, asymptotically), we should have

$$E[W] \approx E[C_0 W + X_0] \tag{4.3}$$

(without the positive-part operator) as a reasonable approximation, which yields

$$E[W] \approx \frac{E[X_0]}{1 - E[C_0]}, \tag{4.4}$$

provided that $E[C_0] < 1$. Note that (4.4) is consistent with (4.1) when C_0 tends to be slightly less than 1, i.e, if $C_0 = 1 - \epsilon Z_0$ for some random variable Z_0, because then

$$\log C_0 = \log(1 - \epsilon Z_0) \approx -\epsilon Z_0 = C_0 - 1. \quad \square \tag{4.5}$$

To obtain a proper limit theorem we introduce a sequence of systems indexed by m. Let $W_n(m)$ be the waiting time of the nth customer and $W(m)$ the steady-state wait in system m, which will exist by theorem 1(c) and assumptions below. The sequence of systems is constructed from a single system with sequence $\{(X_n, C_n)\}$ by letting $X_n(m) = X_n$ and $C_n(m) = C_n^{1/m}$ for $n \geqslant 0$ and $m \geqslant 1$. Let $N(a, b)$ represent a normally distributed random variable with mean a and variance b.

THEOREM 3

Let $\{(X_n, C_n)\}$ be a sequence of i.i.d. random vectors with $E[X_0^2] < \infty$, $E[X_0] > 0$, $P(C_0 > 0) = 1$, $E[(\log C_0)^2] < \infty$ and $E[\log C_0] < 0$. If $W_n(m)$ is the wait of customer n determined by the generalized Lindley recursion

$$W_{n+1}(m) = \left[C_n^{1/m} W_n(m) + X_n\right]^+, \quad n \geqslant 0, \tag{4.6}$$

then, for each $t > 0$,

$$m^{-1/2}\left(W_{[mt]}(m) - \mu_t m\right) \Rightarrow Z(t) \quad \text{as} \quad m \to \infty,$$

where $\mu_t \to \mu$, $Z(t) \Rightarrow N(0, \sigma^2)$ and $W_{[mt]}(m) \Rightarrow W(m)$ as $t \to \infty$,

$$\mu = E[X_0] / | E[\log C_0] | \tag{4.7}$$

and

$$\sigma^2 = \frac{Var[\log C_0](E[X_0])^2}{2|E[\log C_0]|^3} + \frac{Var[X_0]}{2|E[\log C_0]|} + \frac{E[X_0]Cov[X_0, \log C_0]}{(E[\log C_0])^2}. \tag{4.8}$$

Proof

As in remark 4.2, in this case we should expect that the limiting behavior for $W_n(m)$ would be the same as if the positive-part operator were deleted, and indeed this is what occurs. The key result, therefore, is the limit for $Y_n(m)$ satisfying $Y_n(m) = C_n^{1/m} Y_m(m) + X_n$, $n \geqslant 0$, established in section 6 of Vervaat [21]. What remains to be shown is that the limit for $W_n(m)$ satisfying (4.6) must be the same. To do this, we work with the stronger FCLT (functional central limit theorem) version, which is also established in [21]; see remark 6.2 and the proof of theorem 6.1 there. Abstractly this FCLT gives $Z_m \Rightarrow Z$ as $m \to \infty$ in function space, where Z is displayed on line 2 of p. 778 of [21],

$$Z_m(t) = m^{-1/2} \sum_{k=1}^{[mt]} T_k - am^{1/2}(1 - \mathrm{e}^{-bt}), \quad t \geqslant 0, \tag{4.9}$$

with T_k being random variables based on $\{(C_n, X_n)\}$ and a and b are constants with $a > 0$ (because $E[X_0] > 0$) and $b > 0$. Hence, $\hat{Z}_m \Rightarrow Z$ too, where

$$\hat{Z}_m(t) = m^{-1/2} \sup_{0 \leqslant s \leqslant t} \left\{ \sum_{k=1}^{[ms]} T_k \right\} - am^{1/2}(1 - \mathrm{e}^{-bt}), \quad t \geqslant 0, \tag{4.10}$$

by a slight modification of theorem 6.2(ii) of [23]. A modification is required for two reasons: first, the translation term is not of the form $am^{1/2}$ but includes $1 - \mathrm{e}^{-bt}$ and, second, the space of functions and the topology are different (as in [22]). However, essentially the same proof as in theorem 6(ii) of [23] applies. By the continuous mapping theorem, $\hat{Z}_m(t) \Rightarrow Z(t)$ in $\mathbb{R}$ as $m \to \infty$ for each t. Finally, by lemma 2, $\hat{Z}_m(t)$ is asymptotically equivalent to the normalized version of $W_{[mt]}(m)$. (Apply theorem 4.1 of Billingsley [1].) □

Remarks

(4.3) We conjecture that the corresponding limit theorem holds directly for the steady-state wait $W(m)$, i.e.,

$$m^{-1/2}(W(m) - m\mu) \Rightarrow N(0, \sigma^2) \quad \text{as} \quad m \to \infty, \tag{4.11}$$

but we have not yet established it. This defect exists for the standard heavy-traffic limits for queues too.

(4.4) We obtain the normal approximation with (4.1) and (4.2) from theorem 4 by choosing a large t and setting

$$m^{-1/2}(W(m) - \mu m) \approx m^{-1/2}\left(W_{[mt]}(m) - \mu_t m\right) \approx Z(t) \approx N(0, \sigma^2), \tag{4.12}$$

or, equivalently,

$$W(m) \approx N(m\mu, m\sigma^2). \tag{4.13}$$

The m in (4.12) and (4.13) disappears when we go to (4.1) and (4.2) because (4.1) and (4.2) are based on the random variable $C_0^{1/m}$ instead of C_0, and

$$E\left[\left(\log C_0^{1/m}\right)^k\right] = \frac{E\left[(\log C_0)^k\right]}{m^k}; \tag{4.14}$$

i.e., in a given system to be approximated we act as if the actual C_0 is $\overline{C}_0^{1/m}$ for some $\overline{C}_0$ and some suitably large m. (Of course, we obtain $\overline{C}_0$ by setting $\overline{C}_0 = C_0^m$.)

(4.5) The full independence assumed in theorem 3 is not needed. It suffices to have an appropriate FCLT for the partial sums of $(\log C_k, X_k)$, as in (6.4) and (6.5) of [21].

(4.6) In the case $E[X_0] = 0$ $(\rho = 1)$ a limit for $m^{-1/2}W_{[mt]}(m)$ also follows from section 6 of [21]. In this case, since there is no translation term, we can apply the continuous mapping theorem; the limit is $\sup_{0 \leqslant s \leqslant t} Z(s)$ for Z displayed in [21], which seems to be rather complicated.

(4.7) In our normal approximation, we could just as well use the *exact* first two moments of Y, the steady-state distribution of (1.6), as given in theorem 5.1 of Vervaat [21]. Instead of (4.1) and (4.2), we can use the moments $E[Y^k]$ as approximations for the desired moments $E[W^k]$. Thus, provided that $E[C_0] \leqslant 1$, we can use (4.4) instead of (4.1). Similarly, if $E[C_0^k] < 1$ for $k = 1, 2$, then (from 5.2.2 of [21])

$$\begin{aligned} Var[W] \approx Var[Y] = \big\{&2E[X_0]E[X_0C_0](1 - E[C_0]) \\ &+ E[X_0^2](1 - E[C_0])^2 - (1 - E[C_0^2])(E[X_0])^2\big\} \\ &/(1 - E[C_0^2])(1 - E[C_0])^2. \end{aligned} \tag{4.15}$$

Using (4.4) and (4.15) instead of (4.1) and (4.2) obviously can be important if it is easier to calculate $E[C_0^k]$ than $E[(\log C_0)^k]$. □

5. Scheduling arrivals

In this final section we consider the problem of scheduling the interarrival times, as might occur in a production system. For our model, we assume that the service times come from a given sequence $\{V_n\}$ of i.i.d. random variables not subject to control. At each arrival epoch, we must select the next interarrival time, given the history up to that time, i.e., give the service times of all arrivals up to that time and all previous interarrival times. We assume that the interarrival time between customers n and $n+1$ is at least U_n, where $\{U_n\}$ is i.i.d. and independent of $\{V_n\}$, so that we are to determine $\overline{U}_n$, where

$$\overline{U}_n = U_n + D_n \tag{5.1}$$

with

$$D_n = f(U_{i-1}, V_i\colon i \leqslant n). \tag{5.2}$$

If $D_n = A_n W_n$, which is one possibility in (5.2), then (5.2) reduces to (1.2).

The general idea here is that something like what we have previously considered should be a reasonable policy in some circumstances. We mention three natural general criteria for choosing D_n. First, we probably do not want the waiting times to be too large, so that we might want to control the expectation of some increasing function of the waiting time, such as the mean $E[W]$ or a tail probability $P[W > t]$. Second, we may want the throughput to be as high as possible (near the upper limit $1/E[V_0]$), so that we might want to make the probability of emptiness upon arrival, $P(W=0)$, small. Third, we may want the sequence of interarrival times to be relatively smooth, as in production smoothing problems (see pp. 400–413, 438 of Heyman and Sobel [12]), so that we may want to control $|\overline{U}_{n+1} - \overline{U}_n|$ or its distribution. This last objective clearly is important for a policy of the form (1.2) to be reasonable.

From the perspective of applied relevance, there is a difficulty with the scheduling problem as we have formulated it, because the waiting time before beginning service (W_n) often would not be known at the decision point. However, in practice it may be possible to obtain a quick rough estimate of W_n. The analysis below for the idealized case in which W_n is known should be useful to understand the more complex problem in which W_n is estimated.

ZERO LOWER BOUND

First consider the case in which $P(U_0 = 0) = 1$. In this case, we can obviously have the server continuously busy and $W_n = 0$ for all n (and thus clearly satisfy the first two criteria above) by setting

$$\overline{U}_n = D_n = W_n + V_n, \quad n \geqslant 0, \tag{5.3}$$

which becomes simply $\overline{U}_n = V_n$ for $n \geqslant 1$. However, with this policy the successive interarrival times can fluctuate substantially. In particular, for $n \geqslant 1$, this policy yields

$$\overline{U}_{n+1} - \overline{U}_n = V_{n+1} - V_n, \tag{5.4}$$

so that $E[\overline{U}_{n+1} - \overline{U}_n] = 0$ and

$$Var[\overline{U}_{n+1} - \overline{U}_n] = 2Var[V_n]. \tag{5.5}$$

A natural alternative to (5.3) if smoothing $\{\overline{U}_n\}$ is of concern is the smoothed response

$$\overline{U}_n = D_n = \epsilon(W_n + V_n), \quad n \geqslant 0, \tag{5.6}$$

for small positive ϵ. Then W_n can grow but $|\bar{U}_{n+1} - \bar{U}_n|$ is better controlled. Then

$$W_{n+1} = [W_n + V_n - \bar{U}_n]^+ = [(1-\epsilon)W_n + (1-\epsilon)V_n]^+$$
$$= (1-\epsilon)W_n + (1-\epsilon)V_n, \quad n \geq 1, \tag{5.7}$$

$$E[W] = \frac{(1-\epsilon)E[V_0]}{\epsilon}, \quad Var[W] = \frac{(1-\epsilon)^2 Var[V_0]}{\epsilon(2-\epsilon)}, \tag{5.8}$$

and

$$\bar{U}_{n+1} - \bar{U}_n = \epsilon(W_{n+1} + V_{n+1}) - \epsilon(W_n + V_n)$$
$$= \epsilon V_{n+1} - \epsilon^2 V_n - \epsilon^2 W_n, \quad n \geq 1, \tag{5.9}$$

so that

$$E[\bar{U}_{n+1} - \bar{U}_n] \to \epsilon(1-\epsilon)E[V_0] - \epsilon^2 E[W] = 0 \tag{5.10}$$

and

$$Var[\bar{U}_{n+1} - \bar{U}_n] \to \epsilon^2(1+\epsilon^2)Var[V_0] + \epsilon^4 Var[W]$$
$$= \frac{2\epsilon^2}{2-\epsilon} Var[V_0]. \tag{5.11}$$

Note that (5.11) is of order $O(\epsilon^2)$, so that (5.6) enables us to reduce $Var[\bar{U}_{n+1} - \bar{U}_n]$ dramatically, but at the expense of increasing W.

RANDOM LOWER BOUND

We now return to (5.1) without assuming that $P(U_n = 0) = 1$. We shall find a policy of the form $D_n = d + \epsilon(W_n + V_n)$ that tends to keep the process W_n in a prescribed interval $[a, b]$. To do this, we apply the normal approximation in section 4 to produce control parameters d and ϵ so that

$$P(W \leq a) \approx P(W > b) \approx \pi \tag{5.12}$$

for any specified probability π. Our solution will require that $E[V_0] > E[U_0]$, i.e., $\rho > 1$, and the other assumptions of theorem 3.

Since $\bar{U}_n = U_n + d + \epsilon(W_n + V_n)$, in this case we have $C_n = 1 - \epsilon$, $X_n = (1-\epsilon)V_n - U_n - d$ and, by (4.4) and (4.15),

$$E[W] \approx \frac{(1-\epsilon)E[V_0] - (E[U_0] + d)}{\epsilon} \tag{5.13}$$

and

$$Var[W] \approx \frac{Var X_0}{\epsilon(2-\epsilon)} = \frac{(1-\epsilon)^2 Var[V_0] + Var[U_0]}{\epsilon(2-\epsilon)}. \tag{5.14}$$

We first use the desired range $r \equiv b - a$ to specify ϵ. Since

$$r \equiv b - a = 2\beta\sqrt{Var[W]}, \tag{5.15}$$

where $P(N(0, 1) > \beta) = \pi$, we can apply (5.14) to obtain

$$\epsilon = 1 - \left[1 - \frac{\left(Var[V_0] + Var[U_0]\right)^2}{\left(Var[V_0] + [r/2\beta]^2\right)^2}\right]^{1/2}, \tag{5.16}$$

which has a solution provided that $Var[U_0] < (r/2\beta)^2$.

Next we use the intended mean $E[W] \approx (a+b)/2$ to solve for d. We apply (5.13) to get

$$E[W] = \frac{a+b}{2} = \frac{(1-\epsilon)E[V_0] - E[U_0] - d}{\epsilon}, \tag{5.17}$$

so that

$$d = (1-\epsilon)E[V_0] - E[U_0] - \frac{\epsilon(a+b)}{2}. \tag{5.18}$$

Of course, a feasible solution requires that $d > 0$ in (5.18). A necessary condition is $E[V_0] > E[U_0]$, but ϵ determined by (5.16) must also be sufficiently small.

As noted above, a primary motivation for considering policies of this form is to control the fluctuations in the interarrival times. We have done this in two ways. First, given that $a \leqslant W_n \leqslant b$, we have overall bounds on the final interarrival times, i.e.,

$$U_n + d + \epsilon(a + V_n) \leqslant \overline{U}_n \leqslant U_n + d + \epsilon(b + V_n). \tag{5.19}$$

Second, we have controlled the short-run fluctuations in $\{\overline{U}_n\}$, i.e.,

$$\overline{U}_{n+1} - \overline{U}_n = U_{n+1} - (1+\epsilon)U_n + \epsilon V_{n+1} - \epsilon^2 W_n - \epsilon^2 V_n, \tag{5.20}$$

so that $E[\overline{U}_n] \approx E[V_n]$ and $E[\overline{U}_{n+1} - \overline{U}_n] \approx 0$ for large n, and

$$\begin{aligned} Var[\overline{U}_{n+1} - \overline{U}_n] &\to \left(1 + (1+\epsilon)^2\right)Var[U_0] + \epsilon^2(1+\epsilon^2)Var[V_0] + \epsilon^4 Var[W] \\ &= \left(\frac{2(2+\epsilon)}{2-\epsilon}\right)Var[U_0] + \left(\frac{2\epsilon^2}{2-\epsilon}\right)Var[V_0]. \end{aligned} \tag{5.21}$$

Acknowledgements

I am grateful to Sidney Browne of Columbia University for discussions that motivated this work, Peter Glynn of Stanford University for pointing out the appropriate Liapounov function in remark 2.4, and the referees for several helpful comments.

References

[1] P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968).

[2] A.A. Borovkov, *Stochastic Processes in Queueing Theory* (Springer, New York, 1976).

[3] A.A. Borovkov, *Asymptotic Methods in Queueing Theory* (Wiley, Chichester, 1984).

[4] A. Brandt, The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, Adv. Appl. Prob. 18 (1986) 211–220.

[5] P.H. Brill, Single-server queues with delay-dependent arrival streams, Prob. Eng. Inf. Sci. 2 (1988) 231–247.

[6] J.R. Callahan, A queue with waiting time dependent service times, Naval Res. Logist. Qtrly. 20 (1973) 321–324.

[7] K.L. Chung, *A Course in Probability Theory*, 2nd ed. (Academic Press, New York, 1974).

[8] E.V. Denardo and C.S. Tang, Pulling a Markov production line I: introduction and the full restoration rule, Yale University (1988).

[9] E.V. Denardo and C.S. Tang, Pulling a Markov production line II: the proportional restoration rule and operating characteristics, Yale University (1988).

[10] C.M. Harris, Queues with state dependent service rates, Oper. Res. 15 (1967) 117–130.

[11] C.M. Harris, Some results for bulk arrival queues with state dependent service times, Management Sci., Theory 16 (1970) 313–326.

[12] D.P. Heyman and M.J. Sobel, *Stochastic Models in Operations Research*, vol. 2 (McGraw–Hill, New York, 1984).

[13] G.M. Laslett, D.B. Pollard and R.L. Tweedie, Techniques for establishing ergodic and recurrence properties of continuous-valued Markov chains, Naval Res. Logist. Qtrly. (1978) 455–472.

[14] B. Lisek, A method for solving a class of recursive stochastic equations, Z. Wahrscheinlichkeitstheorie verw. Gebiete 60 (1982) 151–161.

[15] R.M. Loynes, The stability of a queue with non-independent interarrival and service times, Proc. Camb. Phil. Soc. 58 (1962) 497–520.

[16] V.I. Mudrov, Queueing with impatient customers and variable service times, linearly dependent on the queueing time of the customer, Problemy Kibernetiki 5 (1961) 283–285.

[17] M.J.M. Posner, Single-server queues with service time depending on waiting time, Oper. Res. 21 (1973) 610–616.

[18] M. Rosenshine, Queues with state dependent service times, Transportation Res. 1 (1967) 97–104.

[19] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models* (Wiley, Chichester, 1983).

[20] S. Sugawara and M. Takahashi, On some queues occurring in an integrated iron and steel works, J. Oper. Res. Soc. Japan (1965) 16–23.

[21] W. Vervaat, On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables, Adv. Appl. Prob. 11 (1979) 750–783.

[22] W. Whitt, Stochastic Abelian and Tauberian theorems, Z. Wahrscheinlichkeitstheorie verw. Gebiete 22 (1972) 251–267.

[23] W. Whitt, Some useful functions for functional limit theorems, Math. Oper. Res. 5 (1980) 67–85.

[24] W. Whitt, On the heavy-traffic limit theorem for $GI/G/\infty$ queues, Adv. Appl. Prob. 14 (1982) 171–190.