Extremal GI/GI/1 queues given two moments: exploiting Tchebycheff systems

Yan Chen & Ward Whitt

Queueing Systems
Theory and Applications

ISSN 0257-0130

Queueing Syst
DOI 10.1007/s11134-020-09675-7

ONLINE FIRST

Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Springer

Author's personal copy

Queueing Systems
https://doi.org/10.1007/s11134-020-09675-7

Extremal *GI*/*GI*/1 queues given two moments: exploiting Tchebycheff systems

Yan Chen[1] · **Ward Whitt**[1]

Received: 6 February 2020 / Revised: 13 July 2020 / Accepted: 19 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
This paper studies tight upper bounds for the mean and higher moments of the steady-state waiting time in the $GI/GI/1$ queue given the first two moments of the interarrival-time and service-time distributions. We apply the theory of Tchebycheff systems to obtain sufficient conditions for classical two-point distributions to yield the extreme values. These distributions are determined by having one mass at 0 or at the upper limit of support.

Keywords GI/GI/1 queue · Tight bounds · Extremal queues · Bounds for the mean steady-state mean waiting time · Moment problem

Mathematics Subject Classification 60K25 · 65C50 · 90B22

1 Introduction

In this paper, we apply the theory of Tchebycheff (T) systems from [17] to identify the extremal interarrival-time and service-time distributions with given first two moments for the mean and higher moments of the steady-state waiting time in the $GI/GI/1$ queue. Thus, this paper contributes to a long-standing open problem for the classical $GI/GI/1$ queueing model: determining a tight upper bound (UB) for the mean steady-state waiting time, and the distributions that attain them, given the first two moments of the interarrival-time and service-time distributions; see [10,31] and references therein. The seminal paper on extremal queues was by B. A. Rogozin in 1966 [23].

This paper is a sequel to [7] in which we applied T systems to determine interarrival-time and service-time distributions with given moments and other properties that

✉ Ward Whitt
ww2040@columbia.edu

Yan Chen
yc3107@columbia.edu

1 Industrial Engineering and Operations Research, Columbia University, New York, USA

Author's personal copy

maximize or minimize the asymptotic decay rate of the steady-state waiting time in the $GI/GI/k$ queue. The theory of T systems was previously used for queueing problems in [11–16,24,25,29,30].

1.1 The $GI/GI/1$ model

The $GI/GI/1$ single-server queue has unlimited waiting space and the first-come first-served service discipline. There is a sequence of independent and identically distributed (i.i.d.) service times $\{V_n : n \geq 1\}$, each distributed as V with cumulative distribution function (cdf) G, which is independent of a sequence of i.i.d. interarrival times $\{U_n : n \geq 1\}$ each distributed as U with cdf F. With the understanding that the first customer (customer 1) arrives at time 0, V_n is the service time of customer n, while U_n is the interarrival time between customers n and $n + 1$.

Let $\equiv$ denote equality by definition. Let U have mean $E[U] \equiv 1$ and squared coefficient of variation (scv, variance divided by the square of the mean) c_a^2; let a service time V have mean $E[V] \equiv \tau \equiv \rho$ and scv c_s^2, where $\rho < 1$, so that the model is stable.

Let W_n be the waiting time of customer n, i.e., the time from arrival until starting service, assuming that the system starts with an initial workload W_0 having cdf H_0 with a finite mean. The sequence $\{W_n : n \geq 0\}$ is well known to satisfy the Lindley recursion

$$W_n = [W_{n-1} + V_n - U_n]^+, \quad n \geq 1, \tag{1}$$

where $x^+ \equiv \max\{x, 0\}$. Let H_n be the cdf of W_n, which is determined by (1). Let $W \equiv W_\infty$ (both used) be the steady-state waiting time, satisfying $W_n \Rightarrow W_\infty$ as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution; see §§X.1–X.2 of [2]. The cdf H_∞ of $W \equiv W_\infty$ is the unique cdf satisfying the stochastic fixed point equation

$$W_\infty \stackrel{\mathrm{d}}{=} (W_\infty + V - U)^+, \tag{2}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. If $P(W_0 = 0) = 1$, then $W_n \stackrel{\mathrm{d}}{=} \max\{S_k : 0 \leq k \leq n\}$ for $n \leq \infty$, $S_0 \equiv 0$, $S_k \equiv X_1 + \cdots + X_k$ and $X_k \equiv V_k - U_k$, $k \geq 1$. Under the specified finite moment conditions, for $1 \leq n \leq \infty$, W_n is a proper random variable with finite mean given by

$$E[W_n | W_0 = 0] = \sum_{k=1}^{n} \frac{E[S_k^+]}{k} < \infty, \quad 1 \leq n < \infty,$$
$$\text{and } E[W_\infty] = \sum_{k=1}^{\infty} \frac{E[S_k^+]}{k} < \infty. \tag{3}$$

1.2 Classical steady-state results: exact, approximate and bounds

For the $M/GI/1$ special case, when the interarrival time has an exponential distribution, we have the classical Pollaczek–Khintchine formula

Author's personal copy

$$E[W] = \frac{\tau\rho(1 + c_s^2)}{2(1 - \rho)} = \frac{\rho^2(1 + c_s^2)}{2(1 - \rho)}. \quad (4)$$

A natural commonly used approximation for the $GI/GI/1$ model, inspired by (4), which we call the heavy-traffic approximation, because it is motivated by the early heavy-traffic limit in [18], is

$$E[W] \equiv E[W(\rho, c_a^2, c_s^2)] \approx \frac{\rho^2(c_a^2 + c_s^2)}{2(1 - \rho)}. \quad (5)$$

The heavy traffic limit for the mean states that $(1-\rho)E[W(\rho, c_a^2, c_s^2)] \to (c_a^2 + c_s^2)/2$ as $\rho \uparrow 1$.

The most familiar UB on $E[W]$ is the [19] bound

$$E[W] \le \frac{\rho^2([c_a^2/\rho^2] + c_s^2)}{2(1 - \rho)}, \quad (6)$$

which also satisfies the same heavy traffic limit.

A better UB depending on these same parameters was obtained by [9]. In particular, the [9] UB replaces the term c_a^2/ρ^2 by $(2-\rho)c_a^2/\rho$, i.e.,

$$E[W] \le \frac{\rho^2([(2 - \rho)c_a^2/\rho] + c_s^2)}{2(1 - \rho)}. \quad (7)$$

Note that $(2-\rho)/\rho < 1/\rho^2$ because $\rho(2-\rho) < 1$ for all ρ, $0 < \rho < 1$.

In contrast to the tight UB that we study, the tight lower bound (LB) for the steady-state mean has been known for a long time; see §5.4 of [26], §V of [29], Theorem 3.1 of [10] and references there. The LB is

$$E[W] \ge \frac{\rho((1 + c_s^2)\rho - 1)^+}{2(1 - \rho)}. \quad (8)$$

The LB is attained asymptotically at a deterministic interarrival time with specified mean and at any three-point service-time distribution that has all mass on nonnegative-integer multiples of the deterministic interarrival time. The service part follows from [22]. (All service-time distributions satisfying these requirements yield the same mean.)

2 The main results

In this section, we state our main results. These results will be proved in following sections.

Author's personal copy

2.1 Sets of probability distributions with specified moments

Let $\mathcal{P}_n$ be the set of all probability measures on a subset of the positive real line $[0, \infty)$ with specified first n moments. The set $\mathcal{P}_n$ is a convex set, because the convex combination of two probability measures is just the mixture, i.e., for all p, $0 \leq p \leq 1$, $pP_1 + (1-p)P_2 \in \mathcal{P}_n$ if $P_1 \in \mathcal{P}_n$ and $P_2 \in \mathcal{P}_n$, because the n^{th} moment of the mixture is the mixture of the n^{th} moments, which is just the common value of the components. Let $\mathcal{P}_{n,k}$ be the subset of probability measures in $\mathcal{P}_n$ that have support on at most k points.

Let $\mathcal{P}_2(m, c^2)$ be the subset of all cdf's in $\mathcal{P}_2$ with support in the interval $[0, \infty)$ having mean m and second moment $m^2(c^2+1)$. Let $\mathcal{P}_2(m, c^2, M)$ be subset of $\mathcal{P}_2(m, c^2)$ denoting all cdf's with support in the closed interval $[0, M]$, where $1 + c^2 < M < \infty$. (The last property ensures that the set is non-empty.) Let subscripts a and s denote sets for the inter-arrival and service times, respectively. Therefore, $\mathcal{P}_{a,2}(1, c_a^2, M_a)$ is the set of all interarrival-time cdf's F with mean 1, scv c_a^2 and compact support within $[0, M_a]$, while $\mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ is the set of all service-time cdf's G with mean ρ, scv c_s^2 and compact support within $[0, \rho M_s]$.

A special role is played by two-point distributions, which necessarily have finite support. Let $\mathcal{P}_{2,2}(m_1, c^2, M)$ be the set of all two-point distributions with mean m_1 and second moment $m_2 = m_1^2(c^2+1)$ with support in $[0, m_1 M]$. The set $\mathcal{P}_{2,2}(m_1, c^2, M)$ is a one-dimensional parametric family. Any element is determined by specifying one mass point. Let $F_b^{(2)}$ be the cdf that has probability mass $c^2/(c^2+(b-1)^2)$ on $m_1 b$, and mass $(b-1)^2/(c^2+(b-1)^2)$ on $m_1(1-c^2/(b-1))$ for $1+c^2 \leq b \leq M$. The cases $b = 1+c^2$ and $b = M$ constitute the two extremal distributions.

Since we are only interested in the extremal cdf's here, we will use different notation. We let $F_0 \equiv F_{1+c^2}^{(2)}$, because it is the unique element that has lower mass point 0 and we let $F_u \equiv F_M^{(2)}$, because it is the unique element that has upper mass point $m_1 M$. We use this definition for both the cdf's we consider: F of U and G of V, but recall that our parameter specification with $E[U] = 1$ makes the support of F_u be $[0, M_a]$, while the support of G_u is $[0, \rho M_s]$. Therefore, with $M_a \geq 1 + c_a^2$ for F and $M_s \geq 1 + c_s^2$ for G, we have:

- $F_0 : c_a^2/(1+c_a^2)$ on 0 and $1/(1+c_a^2)$ on $1+c_a^2$;
- F_u: $(M_a-1)^2/(c_a^2+(M_a-1)^2)$ on $1-c_a^2/(M_a-1)$ and $c_a^2/(c_a^2+(M_a-1)^2)$ on M_a;
- $G_0 : c_s^2/(1+c_s^2)$ on 0 and $1/(1+c_s^2)$ on $\rho(1+c_s^2)$;
- G_u: $(M_s-1)^2/(c_s^2+(M_s-1)^2)$ on $\rho(1-c_s^2/(M_s-1))$ and $c_s^2/(c_s^2+(M_s-1)^2)$ on ρM_s.

2.2 Extremal distributions for higher moments of *W*

Ever since [3] (see p. 97 of [26]), it is known that the extremal theory is quite orderly for higher moments (and cumulants) even though it is challenging for the mean. Thus, we start by applying the T system theory to the higher moments. To treat higher moments, we require that the service time V has a finite moment generating function (mgf) and

Author's personal copy

that implies the same is true for the transient and steady-state waiting time; see §3 of [7] and references there. For a nonnegative random variable Z, we say that it has a finite mgf if there exists $t^* > 0$ such that

$$E[e^{tZ}] < \infty \quad \text{for} \quad t < t^*. \tag{9}$$

That implies that all moments of Z are finite. We remark that condition (9) can be relaxed. In order for $E[W^k]$ to be finite for $k \geq 1$, it suffices to have $E[V^{(k+1)}] < \infty$; for example, see §10.2 of [2].

Theorem 1 (higher steady-state moments) *Consider the $GI/GI/1$ model where $F \in \mathcal{P}_{a,2}(1, c_a^2)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2)$.*

(a) *Let the service-time cdf G be fixed satisfying (9). Then,*

$$E[W(F_u, G)^k] \leq E[W(F, G)^k] \leq E[W(F_0, G)^k] \tag{10}$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ and $k \geq 2$. For each F and $k \geq 2$, these extrema are unique.

(b) *Let the interarrival-time cdf F be fixed. Then,*

$$E[W(F, G_0)^k] \leq E[W(F, G)^k] \leq E[W(F, G_u)^k] \tag{11}$$

for all $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ and $k \geq 2$. For each G and $k \geq 2$, these extrema are unique.

(c) *Suppose that neither F nor G is fixed. Then,*

$$E[W(F_u, G_0)^k] \leq E[W(F, G)^k] \leq E[W(F_0, G_u)^k] \quad \textit{for all} \quad k \geq 2 \tag{12}$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ with $M_s < \infty$. For each $k \geq 2$, these extrema are unique.

We prove Theorem 1 in Sect. 4 by first establishing results for the transient mean and then taking limits. We apply stochastic comparison results from [25] and [11], which are intimately related to T systems. We apply a variant of Theorem 1 in Sect. 6 to establish a natural condition for the continuity of the mean steady-state waiting time in the $GI/GI/1$ queue. This provides an extension of the continuity theorem in §X.6 of [2]. We do so by applying the bounds to establish uniform integrability.

Given Theorem 1, it is natural to expect that corresponding results also hold for the steady mean $E[W]$. However, the proof does not apply to that case. Moreover, counterexamples to the natural analogs of Theorem 1 (a) and (b) above (without additional conditions) were provided, respectively, in §8 of [31] and in Sect. V of [29]. Indeed, counterexamples for (b) are provided by Theorem 2 (b) in Sect. 2.3. However, we conjecture that the analog of Theorem 1 (c) is valid. Accordingly, we directly studied the distribution of $W(F_0, G_u)$ and its limiting behavior as $M_s \to \infty$ in [5]. Theorem 2 there provides a tractable bound for the limit of $E[W(F_0, G_u)]$ as $M_s \to \infty$, which serves as an excellent approximation of the conjectured tight upper bound.

Author's personal copy

2.3 Extremal distributions for the steady-state mean

We now turn to the more challenging problem of the mean $E[W]$. To obtain corresponding comparison results for the steady-state mean approaching Theorem 1, we will exploit stochastic-order properties for cdf's of nonnegative random variables; for example, see §8 of [20] and Ch. 1 of [28]. Recall that a hyperexponential (H_k, mixtures of k exponentials) distribution is completely monotone (CM), which in turn has strictly decreasing failure rate (DFR), which has a strictly decreasing pdf, which has a strictly concave cdf, i.e., we have the implications

$$H_k \mapsto \text{CM} \mapsto \text{DFR} \mapsto \quad \text{strictly concave cdf}. \tag{13}$$

To show the dependence of random variables on the cdf assigned to them, we will include the cdf in parentheses, so we write $U(F)$ $(V(G))$ for an interarrival time U with cdf F (service time V with cdf G). Let $W(F, G)$ denote the steady-state waiting time when the pair (U, V) have the pair of cdf's (F, G). When we write sums of random variables as occurs in the Lindley recursion (1), we assume that the random variables are independent.

We prove part (a) of the following theorem in Sect. 5 and then apply the same methods to prove parts (b) and (c) in later sections.

Theorem 2 (Extremal distributions for the steady-state mean) *Consider the class of $GI/GI/1$ queues with $F \in \mathcal{P}_{a,2}(1, c_a^2)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2)$, $0 < \rho < 1$, where $\mathcal{P}_{a,2}$ and $\mathcal{P}_{s,2}$ are non-empty.*

(a) If the service-time cdf $G \in \mathcal{P}_{s,2}$ is completely monotone and $1 + c_a^2 \le M_a \le \infty$, then

$$W(F_u, G) \le_{icx} W(F, G) \le_{icx} W(F_0, G) \quad \textit{for all} \quad F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \tag{14}$$

so that

$$E[W(F_u, G)] \le E[W(F, G)] \le E[W(F_0, G)] \quad \textit{for all} \quad F \in \mathcal{P}_{a,2}(1, c_a^2, M_a). \tag{15}$$

The extrema in (14) and (15) are uniquely attained.

(b) If the interarrival-time cdf $F \in \mathcal{P}_{a,2}$ is strictly concave and $1 + c_s^2 \le M_s < \infty$, then

$$W(F, G_u) \le_{icx} W(F, G) \le_{icx} W(F, G_0) \quad \textit{for all} \quad F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \tag{16}$$

so that

$$E[W(F, G_u)] \le E[W(F, G)] \le E[W(F, G_0)] \quad \textit{for all} \quad F \in \mathcal{P}_{a,2}(1, c_a^2, M_a). \tag{17}$$

Author's personal copy

If the cdf F has support in $[0, M_a]$ and is strictly convex, then (16) and (17) hold with the roles of G_0 and G_u switched.

(c) *If $M_s < \infty$ and*

$$E[(W(F_0, G_u) + V(G) - U(F) - t)^+] \\ \leq E[(W(F_0, G_u) + V(G_u) - U(F_0) - t)^+] \quad \text{for all} \quad t, \tag{18}$$

then $W(F, G) \leq_{icx} W(F_0, G_u)$ and $E[W(F, G)] \leq E[W(F_0, G_u)]$. If (18) holds for all $F \in \mathcal{P}_{a,2}(1, c_a^2)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$, then

$$\sup\{E[(F, G)] : (F, G) \in \mathcal{P}_{a,2}(1, c_a^2) \times \mathcal{P}_{s,2}(\rho, c_s^2, M_s)\} = E[W(F_0, G_u)]. \tag{19}$$

It is worthwhile to mention the $E[W(F_u, G_0)]$ is not a lower bound (see numerical study in [6]). We regard Theorem 2 (c) as a promising tool to do further analysis together with the algorithms and properties of $W(F_0, G_u)$ developed in [5]. From those results, it suggests a tractable way to justify the optimum by solving an approximate version of the stochastic optimization in (18), i.e., we solve, for all $t \geq 0$,

$$\sup\left\{\frac{1}{n}\sum_{i=1}^{n}(W_i(F_0, G_u) + V_i(G) - U_i(F) - t)^+ : F \in \mathcal{F}, G \in \mathcal{G}\right\} \\ = E[W(F_0, G_u)], \tag{20}$$

where $\mathcal{F}$ is a proper finite support over $[0, M_a]$ and $\mathcal{G}$ is also a proper finite support over $[0, \rho M_s]$.

3 Connecting to basic T system theory

As indicated above, we apply the theory of T systems, as reviewed in §2 of [7], which draws on [17]. In particular, we apply Lemma 2.1 in §2.3 of [7], which is a consequence of the tractable Wronskian condition for a T system.

Definition 1 (T System) Consider a set of $n + 1$ continuous real-valued functions $\{u_i(t) : 0 \leq i \leq n\}$ on the closed interval $[a, b]$. This set of functions constitutes a T system if the $(n+1)^{\text{st}}$-order determinant of the $(n+1) \times (n+1)$ matrix formed by $u_i(t_j)$, $0 \leq i \leq n$ and $0 \leq j \leq n$, is strictly positive for all $a \leq t_0 < t_1 < \cdots < t_n \leq b$.

Equivalently, except for an appropriate choice of sign, we could instead require that every non-trivial real linear combination $\sum_{i=0}^{n} a_i u_i(t)$ of the $n + 1$ functions (called a u-polynomial; see §I.4 of [17]) possesses at most n distinct zeros in $[a, b]$. (Non-trivial means that $\sum_{i=0}^{n} a_i^2 > 0$.)

We next state a consequence of Lemma 2.1 in §2.3 of [7]. Let $\phi^{(n)}$ denote the nth derivative of the function ϕ.

Author's personal copy

Lemma 1 (*From the* $(n+1)$st *derivative to a* T *system*) *Consider the real-valued functions* $u_i(t) \equiv t^i$, $0 \le i \le n$, *and* ϕ *on the interval* $[a, b]$ *for* $0 \le a < b < \infty$. *Suppose that* ϕ *has* $n+1$ *continuous derivatives. If* $\phi^{(n+1)}(t) > 0$ *for* $a \le t \le b$, *then* $\{u_0(t), u_1(t), \dots, u_n(t), \phi(t)\}$ *is a* T *system of functions on* $[a, b]$. *If* $(-1)^{n+1}\phi^{(n+1)}(t) > 0$ *for* $a \le t \le b$, *then* $\{u_0(t), u_1(t), \dots, u_n(t), -\phi(t)\}$ *is a* T *system of functions on* $[a, b]$.

As reviewed in §2 of [7], Lemma 1 applies to our setting when $n = 2$. For Theorem 2 (a), we want the UB and LB of the integral

$$\int_0^{M_a} \phi(u)\, \mathrm{d}F(u), \tag{21}$$

so that we will be applying Lemma 1 over the interval $[0, M_a]$. In part (a) of our queueing extremal problem we work with the integral form in (21) with integrand

$$\phi(u) \equiv \int_0^{\infty} h((y-u)^+)\, \mathrm{d}\Gamma(y) = h(0)\Gamma(u) + \int_{u+}^{\infty} h(y-u)\, \mathrm{d}\Gamma(y), \quad 0 \le u \le M_a, \tag{22}$$

where Γ is a cdf of a nonnegative real-valued random variable Y with a finite moment generating function (mgf), i.e., satisfying (9).

The following lemma combines Lemma 1 with the known extremal distributions in a T system, as given in Theorem 2.4 of [25].

Lemma 2 *If the condition of Lemma 1 is satisfied with* $n = 2$ *and* $(-1)^3\phi^{(3)}(u) > 0$ *for* $0 \le u \le M_a$, *then*

$$\sup\left\{\int_0^{M_a} \phi(u)\, \mathrm{d}F(u) : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)\right\} = \int_0^{M_a} \phi(u)\, \mathrm{d}F_0(u) \tag{23}$$

and

$$\inf\left\{\int_0^{M_a} \phi(u)\, \mathrm{d}F(u) : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)\right\} = \int_0^{M_a} \phi(u)\, \mathrm{d}F_u(u). \tag{24}$$

If the condition of Lemma 1 is satisfied with $n = 2$ *and* $(-1)^3\phi^{(3)}(u) < 0$ *for* $0 \le u \le M_a$, *then the roles of* F_0 *and* F_u *are switched in (23) and (24).*

We now give sufficient conditions on h and the cdf Γ in (22) for the system $\{1, u, u^2, -\phi(u)\}$ to be a T system on $[0, M_a]$. For a real-valued function h of a real variable that has at least k continuous derivatives, let $h^{(k)}$ denote its k^{th} derivative; let $h^{(0)} \equiv h$. Let 1_A be the indicator function of the set A, which equals 1 on A and 0 on its complement. For part of this result, we will be assuming that the cdf Γ has a smooth pdf γ, but we will relax that assumption in Sect. 7.1.

Author's personal copy

Lemma 3 (*Condition for the third derivative to be negative*) *Consider a nonnegative real-valued random variable Y with a finite mgf (satisfying (9)) and the cdf Γ with support in $[a, b]$ or $[a, b)$ such that*

$$0 \le a < M_a \le b \le \infty. \tag{25}$$

For ϕ in (22), in order to have

$$(-1)^3 \phi^{(3)}(u) > 0 \quad for \quad 0 \le u \le M_a, \tag{26}$$

so that $\{1, u, u^2, -\phi(u)\}$ is a T system on $[0, M_a]$, implying that F_0 attains the UB in (23), while F_u attains the LB (24), each of the following is a sufficient condition:

(i) $h(x) \equiv x$ *and Γ has a positive pdf γ that is differentiable with $\gamma^{(1)}(x) < 0$ for $a \le x \le M_a$,*
(ii) $h(x) \equiv x^2$ *and Γ has a positive pdf γ for $a \le x \le M_a$,*
(iii) $h(x) \equiv h(x; p) \equiv x^p \quad for \quad p \ge 3$,
(iv) $h(x) \equiv h(x; t) \equiv e^{tx} - tx - \frac{(tx)^2}{2} - \frac{(tx)^3}{6} = 1 + \sum_{k=4}^{\infty} \frac{(tx)^k}{k!} \quad for \quad t > 0$,
(v) $h^{(k)}(x) > 0, \quad a < x \le M_a, \quad 0 \le k \le 3 \quad and \quad h^{(k)}(a) = 0, \quad 1 \le k \le 2.$

For the function $h(x) \equiv x$ in condition (i), the condition on γ is necessary as well as sufficient, given that γ has a continuous positive derivative. In condition (i), if instead $\gamma^{(1)}(x) > 0$ for $0 \le x \le M_a$, then the roles of F_0 and F_u are switched in (23) and (24).

Proof First, observe that condition (9) implies that all integrals are finite. Next, we consider what happens if $0 \le u \le a$ with $a > 0$ or $u > b$. First, if $u \le a$, then $\phi(u) = E[h(Y - u)]$, $0 \le u \le a$, so that the desired property of ϕ holds over $[0, a]$. In particular,

$$\phi^{(3)}(u) = -\int_a^{\infty} h^{(3)}(y - u)\, \mathrm{d}\Gamma(y) < 0, \quad 0 \le u \le a. \tag{27}$$

On the other hand, if $u \ge b$, then $\phi(u) = h(0)$, so that the desired property cannot hold for $u > b$. However, we have ruled that case out by assuming that $M_a \le b$. It suffices for Γ to have a pdf over $[a, M_a]$.

In each case, we can apply Lemmas 1 and 2 with (22). To do so, we apply the Leibniz rule for differentiation of an integral with (22). Using that condition with $a \le u \le M_a$, we have

$$\begin{aligned}\phi(u) &= \int_a^{\infty} h((y-u)^+)\, \mathrm{d}\Gamma(y) = \int_u^{\infty} h(y-u)\, \mathrm{d}\Gamma(y) + h(0)\Gamma(u) \quad \text{and} \\ \phi^{(1)}(u) &= -\int_u^{\infty} h^{(1)}(y-u)\, \mathrm{d}\Gamma(y) - h(0)\gamma(u) + h(0)\gamma(u) \\ &= -\int_u^{\infty} h^{(1)}(y-u)\, \mathrm{d}\Gamma(y).\end{aligned} \tag{28}$$

Author's personal copy

For $h(x) \equiv x$ in condition (i), we have $h^{(1)}(x) = 1$ for all x, so that

$$\phi^{(1)}(u) = -\int_u^\infty h^{(1)}(y-u)\,\mathrm{d}\Gamma(y) = -\int_u^\infty \mathrm{d}\Gamma(y) = -(1-\Gamma(u)), \quad (29)$$

so that, by the condition on Γ,

$$\phi^{(2)}(u) = \gamma(u) > 0 \quad \text{and} \quad \phi^{(3)}(u) = \gamma^{(1)}(u) < 0 \quad \text{for} \quad u \geq a. \quad (30)$$

From the form of $\phi^{(3)}(u)$ in (30), we see that the condition on γ is necessary as well as sufficient. We also see that the UB and LB are switched if instead $\gamma^{(1)}(u) > 0$.

Turning to $h(x) = x^2$ in condition (ii), we use $h^{(1)}(0) = 0$ and $h^{(2)}(x) = 2$ for all x with the second line of (28) to get

$$\phi^{(2)}(u) = \int_u^\infty h^{(2)}(y-u)\,\mathrm{d}\Gamma(y) = 2\int_u^\infty \mathrm{d}\Gamma(y) = 2(1-\Gamma(u)) > 0, \quad (31)$$

so that $\phi^{(3)}(u) = -2\gamma(u) < 0$ for $a \leq u \leq M_a$.

Conditions (iii) and (iv) are both special cases of condition (v), which implies that

$$\phi^{(3)}(u) = -\int_u^\infty h^{(3)}(y-u)\,\mathrm{d}\Gamma(y) < 0. \quad (32)$$

□

4 Proof of Theorem 1

We prove Theorem 1 by establishing results for the transient higher moments. We do so by applying stochastic comparison results from [11,25], which are intimately connected to the theory of T systems. From Theorem 2.1 of [25], the stochastic partial order $X_1 \leq_{2,n} X_2$ holds for any n with $n \geq 2$ if and only if

$$E[((X_1 - t)^+)^n] \leq E[((X_2 - t)^+)^n] \quad \text{for all} \quad t \in \mathbb{R}. \quad (33)$$

From Theorem 3.2 of [11], the stochastic partial order $X_1 \leq_{3-cx} X_2$ holds if and only if both $X_1 \leq_{2,2} X_2$ and $E[(X_1)^j] = E[(X_2)^j]$ for $j = 1, 2$. Hence, $X_1 \leq_{3-cx} X_2$ implies that $X_1 \leq_{2,2} X_2$. Moreover, $X_1 \leq_{2,2} X_2$ implies $X_1 \leq_{2,n} X_2$ for all $n > 2$, as shown in Corollary 1 to Theorem 2.1 in [25]. As shown in §5 of [11], for random variables $X(F)$ with cdf F on the bounded interval $[0, M]$,

$$X(F_0) \leq_{3-cx} X(F) \leq_{3-cx} X(F_u). \quad (34)$$

We need two lemmas:

Author's personal copy

Lemma 4 (*Order for differences of random variables*) *If $U_1 \leq_{3-cx} U_2$ and $V_1 \leq_{3-cx} V_2$, where U_i and V_i are independent real-valued random variables for each i, then*

$$V_1 - U_2 \leq_{3-cx} V_2 - U_1. \tag{35}$$

Proof Combine Propositions 3.10 and 3.11 (vi) of [11]. □

Lemma 5 (*Preservation of order for positive-part function*) *If $U_1 \leq_{2,n} U_2$, then*

$$(U_1 - t)^+ \leq_{2,n} (U_2 - t)^+ \quad \textit{for all} \quad t \in \mathbb{R}. \tag{36}$$

Proof This is an easy consequence of the definition in (33). □

We can apply the above to prove an ordering of all the transient waiting times in the $GI/GI/1$ queue. The proof of Theorem 1 follows directly from the following theorem.

Theorem 3 (Order for transient and steady-state waiting times) *Let $W_{i,n}$ be the waiting time of customer n in two $GI/GI/1$ queues with pairs of interarrival-time and service-time distributions (U_i, V_i), $i = 1, 2$. Let the systems start empty or with ordered initial waiting times $W_{1,0} \leq_{2,2} W_{2,0}$. If $U_1 \geq_{3-cx} U_2$ and $V_1 \leq_{3-cx} V_2$, then*

$$W_{1,n} \leq_{2,2} W_{2,n} \quad \textit{for all} \quad n \geq 1 \tag{37}$$

and

$$W_1 \leq_{2,2} W_2 \tag{38}$$

for the associated steady-state waiting times.

Proof We carry out the proof by mathematical induction. If the systems start empty, we get $W_{1,1} \leq_{2,2} W_{2,1}$ by combining Lemmas 4 and 5. Given $W_{1,n-1} \leq_{2,2} W_{2,n-1}$, we get $W_{1,n} \leq_{2,2} W_{2,n}$ by applying the preservation of order under convolution by (C) in §2.1 of [25] and the positive part function in Lemma 5. We get the final order for the steady-state waiting times from the preservation under convergence $W_{i,n} \Rightarrow W_i$ as $n \to \infty$ for each i using §2.1 of [25]. □

We conclude this section by providing a simulation illustration and sanity check for Theorems 3 and 1. Figure 1 plots simulation estimates of the transient second moments $E[W_n(F, G)^2]$ for seven values of n for $(F, G) = (F_u, E_2)$, (E_2, E_2) and (F_0, E_2) (left) and $(F, G) = (E_2, G_0)$, (E_2, E_2) and (E_2, G_u) (right) for the case $c_a^2 = c_s^2 = 0.5$, $\rho = 0.5$ and $M_a = M_s = 10$.

5 Proof of Theorem 2 (a)

We next prove Theorem 2 (a): finding the extremal interarrival-time cdf F on $[0, M_a]$ for the mean $E[W(F, G)]$ for any given service-time cdf G. We again apply the theory

Author's personal copy

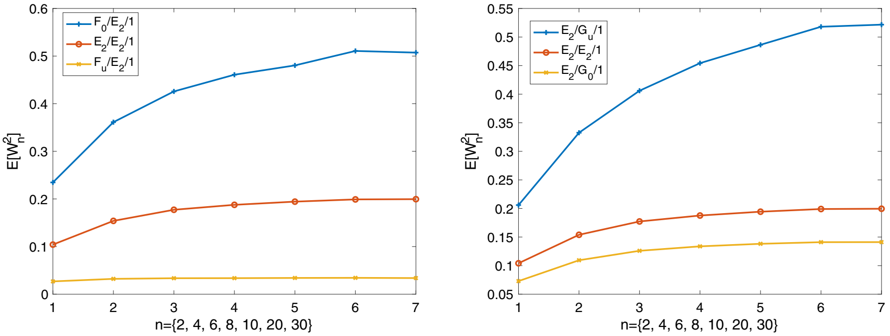

Fig. 1 Simulation estimates of $E[W_n(F, G)^2]$ for various (F, G) with $c_a^2 = c_s^2 = 0.5$, $M_a = M_s = 10$, $\rho = 0.5$

of T systems. After treating part (a), we apply the same methods to treat Theorem 2 (b) and (c). The remaining details in the proof of Theorem 2 (b) and (c) are given in Sects. 7.2 and 7.3.

There are several steps in the proof of Theorem 2 (a). First, we apply increasing convex stochastic order to show that (39) in Theorem 4 is a sufficient condition for all the desired conclusions. Then, Theorems 5 and 6 are devoted to providing sufficient conditions to establish (39), as stated in Theorem 2 (a). The T-system theory enters in the proof of Theorem 5.

5.1 Exploiting increasing convex stochastic order

For Theorem 2 (a), we apply Lemma 3 (i) with the random variable Y being the sojourn time, i.e., the time spent by the arrival in the system, also called the response time. It is the sum of two independent nonnegative random variables, one being a service time V and the other steady-state waiting time W or the transient waiting time W_n for $n \geq 0$. Let $Y \equiv W + V$ and let $Y_n \equiv W_n + V$. Let H and H_n be the cdf of W and W_n, respectively. Let $Y_n(H_0, F, G)$ and $Y(F, G)$ denote the dependence of Y_n and Y on the underlying cdf triple (H_0, F, G) of (W_0, U, V) and similarly for other random variables. (Recall that the steady-state distributions are independent of the initial conditions, assuming a finite mean $E[W_0]$). We can apply the previous results to deduce the following two theorems.

For the first theorem, we exploit increasing convex stochastic order, denoted by $\leq_{icx}$, and state results in that form. For real-valued random variables, $Z_1 \leq_{icx} Z_2$ if $E[f(Z_1)] \leq E[f(Z_2)]$ for all non-decreasing convex functions f for which the expectations are well defined; for example, see §1.5 of [20].

Theorem 4 (A one-step condition for ordered steady-state means) *Consider the $GI/GI/1$ model with given service-time cdf G satisfying (9). Let $W_0 \stackrel{d}{=} W(F_1, G)$ and $Y_0 \stackrel{d}{=} Y(F_1, G)$, where $W(F, G)$ and $Y(F, G)$ are the steady-state waiting time and sojourn time using $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$. For any $F_2 \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$, if*

Author's personal copy

$$\int_0^{M_a} E[(Y_0(F_1, G) - t - u)^+] \, \mathrm{d}F_2(u) \\ \leq \int_0^{M_a} E[Y_0(F_1, G) - t - u)^+] \, dF_1(u) \quad \textit{for all} \quad t \geq -M_a, \tag{39}$$

then

$$W_n(F_2, G) \leq_{icx} W_{n-1}(F_2, G) \quad \textit{and} \quad Y_n(F_2, G) \leq_{icx} Y_{n-1}(F_2, G) \quad \textit{for} \quad n \geq 1 \tag{40}$$

for

$$\begin{aligned} Y_n(F_i, G) &\equiv W_n(F_i, G) + V(G) \quad \textit{for} \quad n \geq 2, \\ Y_1(F_i, G) &\equiv W_0 + V(G) = W(F_1, G) - U(F_i))^+, \\ W_n(F_i, G) &\equiv (Y_{n-1}(F_i, G) - U(F_i))^+ \quad \textit{for} \quad n \geq 2 \\ \textit{and} \quad W_1(F_i, G) &\equiv (Y_0 - U(F_i))^+. \end{aligned} \tag{41}$$

Hence,

$$W(F_1, G) \geq_{icx} W(F_2, G) \quad \textit{and} \quad Y(F_1, G) \geq_{icx} Y(F_2, G) \tag{42}$$

and thus

$$E[W(F_1, G)] \geq E[W(F_2, G)] \quad \textit{and} \quad E[Y(F_1, G)] \geq E[Y(F_2, G)]. \tag{43}$$

Proof We start by observing that the increasing convex stochastic ordering

$$Y(F_1, G) - U(F_1) \geq_{icx} Y(F_1, G) - U(F_2), \tag{44}$$

where the random variables Y and U are independent, is equivalent to the expectation orderings

$$E[(Y(F_1, G) - U(F_1) - t)^+] \geq E[(Y(F_1, G) - U(F_2) - t)^+] \quad \text{for all} \quad t, \tag{45}$$

by virtue of Theorem 1.5.7 of [20]. We can then rewrite (45) equivalently as

$$\int_0^{M_a} E[(Y(F_1, G) - u - t)^+] \, \mathrm{d}F_1(u) \\ \geq \int_0^{M_a} E[(Y(F_1, G) - u - t)^+] \, \mathrm{d}F_2(u) \quad \text{for all} \quad t. \tag{46}$$

Since U has support in $[0, M_a]$, we only need to consider $t \geq -M_a$. Thus, the condition in (39) is equivalent to each of the expressions in (44)–(46).

Author's personal copy

Now, given (44), because $(x)^+$ is a non-decreasing convex function, we have

$$W_0 \equiv W(F_1, G) \stackrel{d}{=} (Y(F_1, G) - U(F_1))^+ \geq_{icx} (Y(F_1, G) - U(F_2))^+ \equiv W_1, \tag{47}$$

where $W_1 \equiv W_1(F_1, F_2, G) \equiv [W(F_1, G) + V(G) - U(F_2)]^+$. Then, by Theorem 1.5.5 (b) of [20], we see that the order is maintained if we add the same independent random variable from both sides. That gives

$$Y_0 \equiv W_0 + V(G) \geq_{icx} W_1 + V(G) \equiv Y_1, \tag{48}$$

where independence is assumed in the sums, as usual. Then, by Theorem 1.5.5 (b) of [20] again, we see that the order is maintained if we subtract the same independent random variable from both sides. Hence, from (48) we deduce that

$$Y_0 - U(F_2) \geq_{icx} Y_1 - U(F_2). \tag{49}$$

Then, because $(x)^+$ is a non-decreasing convex functions of x, we have

$$W_1 = (Y_0 - U(F_2))^+ \geq_{icx} (Y_1 - U(F_2))^+ \equiv W_2, \tag{50}$$

By the same reasoning, we deduce recursively, and using mathematical induction, that

$$W_{n-1} \geq_{icx} W_n \quad \text{and} \quad Y_{n-1} \geq_{icx} Y_n \quad \text{for all} \quad n \geq 1. \tag{51}$$

But then observe that $(W_n, Y_n) \Rightarrow (W(F_2, G), Y(F_2, G))$ as $n \to \infty$, so that we can apply Theorem 1.5.9 of [20] to deduce (42), which of course implies (43). □

It now remains to provide a sufficient condition for condition (39) in Theorem 4 in terms of the steady-state sojourn time $Y(F_0, G)$. We remark that it is known that the steady-state waiting-time cdf is always new worse than used (NWU), is concave if the service-time cdf is has decreasing failure rate (is DFR), and is completely monotone if the service-time cdf is completely monotone; see §§1.7–1.9 of [28]. However, these properties are not preserved under convolution in general.

Theorem 5 (Strict concavity condition for F_0) *Consider the $GI/GI/1$ model with given service-time V having cdf G with support in $[a, \infty)$ for $0 \leq a < \rho = E[V]$ and a finite mgf (satisfying (9)). If the sojourn-time cdf $\Gamma(x) \equiv P(Y(F_0, G) \leq x)$ is strictly concave in x over $[a, \infty)$, then condition (39) in Theorem 4 is satisfied for $F_1 = F_0$ and for all $F_2 \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$, so that*

$$Y(F, G) \leq_{icx} Y(F_0, G) \quad \textit{for all} \quad F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \tag{52}$$

and

$$\sup \left\{ E[W(F, G)] : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \right\} = E[W(F_0, G)]. \tag{53}$$

Author's personal copy

If the sojourn-time cdf $\Gamma(x) \equiv P(Y(F,G) \le x)$ *is strictly concave in* x *over* $[a,\infty)$ *for all* $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$*, then condition (39) in Theorem 4 is satisfied for all* $F_2 = F_u$ *and* $F_1 = F$ *for all* F *in* $\mathcal{P}_{a,2}(1, c_a^2, M_a)$*, so that*

$$Y(F_u, G) \le_{icx} Y(F, G) \quad \textit{for all} \quad F \in \mathcal{P}_{a,2}(1, c_a^2, M_a), \tag{54}$$

and

$$\inf \left\{ E[W(F,G)] : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \right\} = E[W(F_u, G)]. \tag{55}$$

Proof The condition on the cdf Γ in Theorem 5 implies condition (39) in Theorem 4. That implication follows by applying Lemma 3 (i) when the service-time cdf has a strictly decreasing pdf, which we have not yet assumed. However, it is possible to treat the more general case by an additional asymptotic argument, as we indicate in Sect. 7.1. In particular, we apply Lemmas 6 and 7. The concavity of the cdf Γ over the entire interval $[a,\infty)$ is important for covering the subtraction by t in condition (39). □

We now provide a sufficient condition for the strict concavity conditions on the sojourn-time distribution in Theorem 5. Recall that a cdf G on $[0,\infty)$ is completely monotone if it is a mixture of exponential cdf's, i.e., if

$$G(x) = \int_0^\infty (1 - e^{-\lambda x})\, \mathrm{d}P(\lambda)$$

for some probability measure P.

Theorem 6 (Hyperexponential sojourn-time distribution) *In the* $GI/GI/1$ *queue, if the service-time distribution is* H_k*, then so is the sojourn-time distribution. Hence, the concavity conditions on the sojourn-time cdf* Γ *in Theorem 5 are satisfied for all* $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ *if the service-time cdf* G *is hyperexponential or completely monotone.*

Proof For the $GI/PH/1$ queue with any interarrival-time cdf, it is known that the waiting time and sojourn time cdf's inherit the phase-type (PH) matrix structure of the service-time cdf; see pp. 46 and 151 of [21] or Corollary 2.2 of [1]. From the diagonal structure of the substochastic matrix in the PH representation when the PH distribution is H_k, we see that the distributions of W and Y are also H_k when the cdf G of V is H_k.

We continue to provide a detailed proof, referring to [21] and using notation there. To do so, we use the known fact that the waiting-time cdf is H_k with an atom at the origin; see §II.5.10 of [8] or Theorem 1.9 D (i) of [28]. Thus, from equation (4.1.41) of [21], we deduce that the matrix function $\Theta(x)$ that appears there in the cdf of W and is characterized in previous equations, is a diagonal matrix with exponential functions $e^{-\mu_i x}$ appearing on the diagonal. (The constant matrices appearing there are all strictly positive, so that there is no cancellation.) Hence, this must also be true for the matrix $\Psi(x)$ in equation (4.1.40) of [21]. Thus, the ODE for the function $\Psi_1(x)$ appearing

Author's personal copy

there involves diagonal matrices S and $\Psi(x)$. Hence, $\Psi_1(x)$ has a diagonal matrix function solution. Thus, the distribution of Y given in equation (4.1.44) is a mixture of exponentials. Given that H_k structure, the cdf Γ has a strictly decreasing pdf, so that the concavity condition is satisfied. The general completely monotone case can be represented as the limit of H_k cdf's, using §7.1. The asymptotic argument was used for the $GI/GI/k$ waiting time by [27] □

We now observe that the concavity condition for the sojourn-time cdf cannot be satisfied for any $GI/E_k/1$ queue for $k \geq 2$ or any queue where the service time is a finite mixture of E_k distributions all of which have $k \geq 2$.

Proposition 1 (*Negative results for Erlang service*) *For any service-time cdf G that has a pdf g that is differentiable and strictly increasing over* $[0, x]$ *for some* $x > 0$, *the pdf γ of the associated steady-state sojourn-time $Y \equiv W + V$ must be strictly increasing over* $[0, x]$, *so that the cdf Γ cannot be strictly concave.*

Proof Let H be the cdf of W, which has an atom at 0. Given that V has a continuous pdf, $Y \equiv V + W$ has a pdf γ, where

$$\gamma(t) = H(0)g(t) + \int_{0+}^{t} g(t-u)\, \mathrm{d}H(u), \quad t \geq 0. \tag{56}$$

Consequently, the derivative satisfies

$$\gamma^{(1)}(t) = H(0)g^{(1)}(t) + \int_{0+}^{t} g^{(1)}(t-u)\, \mathrm{d}H(u), \tag{57}$$

which is strictly positive for $0 \leq t \leq x$ for sufficiently small x by the assumption on g. □

6 Continuity of the mean steady-state waiting time

In this section, we use a variant of Theorem 1 to show that the mean steady-state waiting time is continuous as a function of the underlying pair of cdf's (F, G) under a natural condition. This section thus provides an extension to Corollary X.6.4 in [2] by establishing uniform integrability (UI) of the sequence of waiting times; for example, see §5 of [4].

For this result, we relax condition (9) and instead assume that the third moment of the service time V is specified as well as the parameters $(1, c_a^2, \rho, c_s^2)$. Let

$$\nu_{s,3} \equiv \frac{E[V^3]}{E[V]^3} = \rho^3 E[V^3], \tag{58}$$

which we assume to be finite. Our new continuity result is

Theorem 7 (Continuity of the mean waiting times) *Consider a sequence of $GI/GI/1$ queueing models indexed by k with underlying interarrival-time and service-time*

Author's personal copy

random variables (U_k, V_k) *having the pair of cdf's* (F_k, G_k) *with the fixed model parameters* $(1, c_a^2, \rho, c_s^2, \nu_{s,3})$ *(to be used in each time period* n*). Let* $W_n^{(k)}$ *be the transient waiting time in time period* n *and let* $W^{(k)}$ *be the steady-state waiting time for model* k*. Suppose that* $F_k \Rightarrow F$ *and* $G_k \Rightarrow G$ *as* $k \to \infty$*. Then,* $W_n^{(k)} \Rightarrow W_n \equiv W_n(F, G)$ *as* $k \to \infty$ *for each* $n \geq 1$ *and* $W^{(k)} \Rightarrow W \equiv W(F, G)$ *as* $k \to \infty$*,*

$$\text{both} \quad \{(W_n^{(k)}) : k \geq 1\} \quad \text{for each } n \geq 1 \quad \text{and} \quad \{(W^{(k)}) : k \geq 1\} \quad \text{are UI} \tag{59}$$

or, equivalently,

$$\begin{aligned} &E[(W_n^{(k)})] \to E[W_n] \quad as \quad k \to \infty \quad \text{for each } n \geq 1 \\ &\text{and } E[(W^{(k)})] \to E[W] \quad as \quad k \to \infty. \end{aligned} \tag{60}$$

The proof of the uniform integrability needed for Theorem 7 exploits upper bounds on the mean waiting times, which are provided by the following variant of Theorem 1.

Theorem 8 (Upper bounds for the second moment of the steady-state waiting time) *Consider the set of* $GI/GI/1$ *queueing models with* $F \in \mathcal{P}_{a,2}(1, c_a^2)$ *and* $G \in \mathcal{P}_{s,3}(\rho, c_s^2, \nu_{s,3})$ *for* $\nu_{s,3}$ *in (58).*

(*a*) *Let the service-time cdf* $G \in \mathcal{P}_{s,3}(\rho, c_s^2, \nu_{s,3})$ *be fixed. Then,*

$$E[W(F, G)^2] \leq E[W(F_0, G)^2] < \infty \tag{61}$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2)$*, where* F_0 *is the two-point cdf with one mass at* 0*.*

(*b*) *Let the interarrival-time cdf* $F \in \mathcal{P}_{a,2}(1, c_a^2)$ *be fixed. Then, there exists a cdf* $\hat{G} \in \mathcal{P}_{s,3}(\rho, c_s^2, \nu_{s,3})$ *such that*

$$E[W(F, G)^2] \leq E[W(F, \hat{G})^2] < \infty \tag{62}$$

for all $G \in \mathcal{P}_{s,2}(\rho, c_s^2, \nu_{s,3})$*.*

(*c*) *Suppose that neither* F *nor* G *is fixed. Then, there exists a cdf* $\hat{G} \in \mathcal{P}_{s,3}(\rho, c_s^2, \nu_{s,3})$ *such that*

$$E[W(F, G)^2] \leq E[W(F_0, \hat{G})^2] < \infty \tag{63}$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2)$ *and* $G \in \mathcal{P}_{s,2}(\rho, c_s^2, \nu_{s,3})$*.*

Proof The proof is a variant of the proof in Theorem 1. As before, we apply T systems. Since we only draw conclusions about the second moment of the steady-state waiting time, it suffices to have the bounded third moment of G in (58). For part (a), we initially impose the finite support bound M_a on F, but the extremal cdf F_0 places no mass on the upper limits. Thus, the bound is independent of M_a. For parts (b) and (c), we use the T-system theory again to exploit the specified third moment of G to construct the extremal upper bound cdf of G given the first three moments, which we denote by $\hat{G}$.

Author's personal copy

The extremal cdf $\hat{G}$ assigns one point to 0 and the other two points to x_1 and x_2 with $0 < x_1 < x_2 < M_s$ for M_s suitably large; for example, see the tables on p. 137 of [12]. By Theorem X.2.1 of [2], $E[W^2] < \infty$ given that $\nu_{s,3} < \infty$. □

Proof of Theorem 7. First, the results for the transient waiting times are elementary, given the Lindley recursion in (1). For the steady-state mean, Corollary X.6.4 of [2] implies that $W^{(k)} \Rightarrow W$ as $k \to \infty$. The condition there that $\{X_k^+ : k \geq 1\}$ be UI for $X_k \equiv V_k - U_k$ is satisfied because $X_k^+ \leq V_k$ and $E[V_k^2] = \rho^2(c_s^2 + 1) < \infty$ for all k. To deduce (59), which is equivalent to (60) because the waiting times are nonnegative (see Theorem 5.4 of [4]), we apply the uniform bound on the second moment provided by Theorem 8. □

7 Remaining proofs

We now provide the remaining proofs.

7.1 Relaxing the PDF condition in Lemma 3

We now relax the pdf condition on Γ in Lemma 3 under conditions (i) and (ii) above. Recall that convergence in distribution can be expressed in terms of cdf's, i.e., corresponds to pointwise convergence at all points x that are continuity points of the limiting cdf. Let $\Rightarrow$ denote convergence in distribution.

Lemma 6 (*Preservation of optimality*) *Suppose that $\{Y_n : n \geq 1\}$ is a sequence of real-valued random variables such that the conditions of Lemma 3 are satisfied for each $n \geq 1$ and $Y_n \Rightarrow Y$ as $n \to \infty$. If F_0 (F_u) yields the UB for (23) and F_u (F_0) yields the LB in (24) for all $n \geq 1$, then the same is true for the limit Y.*

Proof We directly compare F_0 to any alternative cdf F for the UB. First, by the continuous mapping theorem, we obtain

$$\phi_n(u) \to \phi(u) \quad \text{as} \quad n \to \infty \tag{64}$$

for each u from (22). Then, by the dominated convergence theorem,

$$\begin{aligned}\int_0^{M_a} \phi(u)\,\mathrm{d}F(u) &= \lim_{n\to\infty} \int_0^{M_a} \phi_n(u)\,\mathrm{d}F(u) \leq \lim_{n\to\infty} \int_0^{M_a} \phi_n(u)\,\mathrm{d}F_0(u) \\ &= \int_0^{M_a} \phi(u)\,\mathrm{d}F_0(u).\end{aligned} \tag{65}$$

Hence, F_0 remains optimal for the limit. Essentially, the same argument applies to the lower bound. □

Lemma 7 (*Preservation of optimality for the first two moments*) *In order to apply Lemma 6 to condition (i) in Lemma 3 with $\gamma^{(1)}(x) < 0$, $a \leq x \leq M_a$, it suffices to*

Author's personal copy

have, in addition to $Y_n \Rightarrow Y$*, the cdf's* Γ *of* Y *be strictly concave over* $[a, M_a]$*, i.e., have* $\Gamma(x + \delta) - \Gamma(x)$ *be strictly decreasing in* x *over* $[a, M_a]$ *for all* $\delta > 0$*. For condition (ii) in Lemma 3, it suffices to have* $\Gamma(x)$ *be strictly decreasing in* x *over* $[a, M_a]$.

Proof If Γ has the stated property, then Γ can be made the limit of cdf's Γ_n with the properties stated in Lemma 3. □

7.2 Proof of Theorem 2 (b)

The proof for part (b) can be short, because we can apply a variant of the proof for part (a). For part (b), we are concerned with

$$\sup\left\{\int_0^\infty E[(W(F,G) + v - U)^+]\mathrm{d}G(v) : G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)\right\}. \tag{66}$$

It is convenient to use a reverse-time formulation and work with the cdf $\tilde{G}$ of $\tilde{V} \equiv \rho M_s - V$, and adjusting the moments consistently. Then we can focus on

$$\sup\left\{\int_0^\infty E[(W(F,G) - U + \rho M_s - v)^+]\mathrm{d}\tilde{G}(v) : \tilde{G} \in \mathcal{P}_{s,2}(\tilde{\rho}, \tilde{c}_s^2, M_s)\right\}. \tag{67}$$

We make structural assumptions about the cdf's F of U and H of W, which can be relaxed by the asymptotic methods of Sect. 7.1.

Lemma 8 *If (i) the cdf* F *is differentiable with a strictly positive pdf* f *that can be expressed as*

$$f(u) = \int_0^u f^{(1)}(x)\,\mathrm{d}x, \quad u \geq 0, \tag{68}$$

where $f^{(1)}$ *is integrable, and (ii)* W *has a cdf* H *with* $H(0) > 0$ *and*

$$H(x) = H(0) + \int_0^x h(w)\,\mathrm{d}w, \quad x \geq 0, \tag{69}$$

where h *is strictly positive and integrable over the halfline, then the integrand* ϕ_s *in (67) with* $k = 1$ *can be expressed as*

$$\phi_s(v) = H(0)E[(\rho M_s - v - U)^+] + \int_0^\infty h(w)E[(w + \rho M_s - v - U)^+]\,\mathrm{d}w > 0, \tag{70}$$

so that the first three derivatives of ϕ_s *exist for* $v > 0$ *and satisfy*

$$\phi_s^{(1)}(v) = -P(U - W \leq \rho M_s - v)$$

Author's personal copy

$$
\begin{aligned}
&= -H(0)F(\rho M_s - v) - \int_0^\infty h(w)F(w + \rho M_s - v)\,\mathrm{d}w < 0,\\
\phi_s^{(2)}(v) &= \theta(v) = H(0)f(\rho M_s - v) + \int_0^\infty h(w)f(w + \rho M_s - v)\,\mathrm{d}w > 0,\\
\phi_s^{(3)}(v) &= \dot{\theta}(v)\\
&= -H(0)f^{(1)}(\rho M_s - v) - \int_0^\infty h(w)f^{(1)}(w + \rho M_s - v)\,\mathrm{d}w > 0, \quad v \geq 0,
\end{aligned} \tag{71}
$$

where $\theta(v)$ is the pdf of $U - W$ over $[0, \rho M_s]$ because $v \in [0, \rho M_s]$, so that ϕ_s is strictly positive, strictly decreasing and strictly convex on $[0, \rho M_s]$. Moreover, from (71) we see that if $f^{(1)} > 0$, then $\phi_s^{(3)}(v) < 0$ as well.

Thus, the previous proof applies until we come to Theorem 6, but we now need to replace the steady-state sojourn time $Y = W + V$ by $U - W$, where W is the steady-state waiting time. Fortunately, the analog of Theorem 6 is already covered by Lemma 8.

7.3 Proof of Theorem 2 (c)

The proof here is essentially the same as the proof of Theorem 4. As before, we establish increasing convex stochastic order as we move from one steady-state distribution to another through a sequence of transient distributions, based on the Lindley recursion (1).

8 Supporting simulation results

In this section, we present results of simulation experiments illustrating the shape conditions for parts (a) and (b) of Theorem 2.

8.1 Supporting simulation results for *F*

We now present supporting simulation results for the shape of the steady-state sojourn time $Y \equiv W + V$ asserted in Theorem 6 and thus for the extremal results in Theorems 4 and 5. We consider the supreumum and infimum for F, for example,

$$E[W(F_0, H_k)] = \sup\{E[W(F, H_k)] : F \in \mathcal{P}_{a,2}(M_a)\}, \quad k \geq 2. \tag{72}$$

Figure 2 shows that for the examples $F_0/H_2/1$ and $F_u/H_2/1$ ($M_a = 10$) with balanced means under traffic level $\rho = 0.5$, the pdf γ of Y under the same simulation settings in both cases has a monotone density. In these cases, the estimated mean steady-state values are $E[W(F_u, H_2)] = 2.00$ and $E[W(F_0, H_2)] = 3.28$.

In addition, Fig. 3 shows supporting simulation results for $F_0/H_5/1$ and $F_u/H_5/1$ ($M_a = 10$) where H_5 has the service rates [1.0, 1.5, 2.0, 2.5, 3.0] with the respective

Author's personal copy

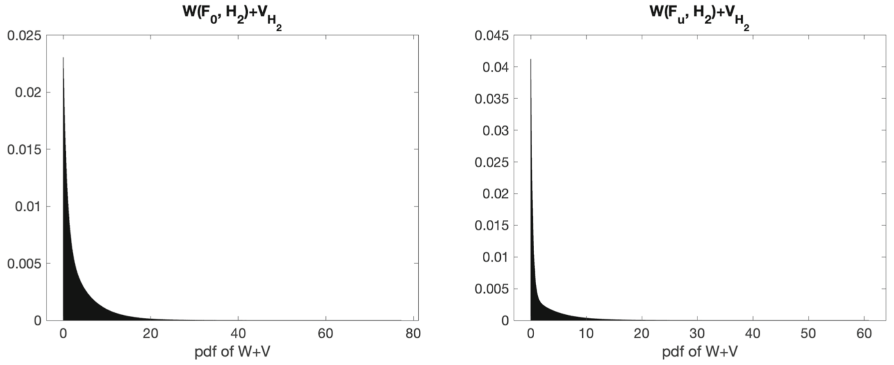

Fig. 2 The shape of the pdf of $W(F_0, H_2) + V_{H_2}$ and $W(F_u, H_2) + V_{H_2}$ for H_2 service with $M_a = 10$, $\rho = 0.5$ and $c_a^2 = c_s^2 = 4$

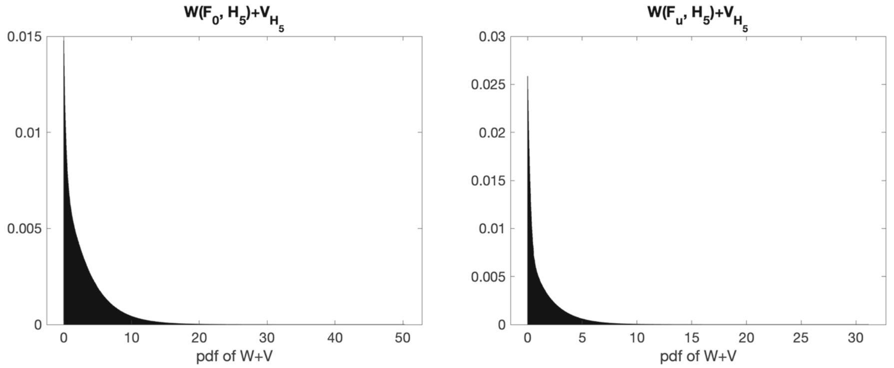

Fig. 3 The shape of the pdf of $W(F_0, H_5) + V_{H_5}$ and $W(F_u, H_5) + V_{H_5}$ for H_5 service (as specified in the text) with $M_a = 10$, $\rho = 0.5$ and $c_a^2 = c_s^2 = 4$

probabilities $[0.1, 0.15, 0.2, 0.25, 0.3]$. In these cases, $E[W(F_u, H_5)] = 1.03$ and $E[W(F_0, H_5)] = 2.75$.

Figure 4 shows corresponding results for E_2 service. Figure 4 shows that the pdf of $W + V$ is nearly monotone, but a careful examination shows that the pdf is increasing over a short interval $[0, x]$. In these cases, $E[W(F_u, E_2)] = 0.528$ and $E[W(F_0, E_2)] = 2.55$.

8.2 Supporting simulation results for *G*

In this concluding subsection, we present simulation results supporting Theorem 2. Figure 5 experimentally confirms the conclusion of Lemma 8 in Sect. 7.2 that the condition of Theorem 2 (b) can be satisfied by presenting simulation estimates of the pdf of $W(H_2, G_0) - U_{H_2}$ and $W(H_2, G_u) - U_{H_2} \leq 0$ with $M_s = 10$ and $\rho = 0.5$ for the case $c_a^2 = c_s^2 = 4$. In these cases, $E[W(H_2, G_0)] = 2.17$ and $E[W(H_2, G_u)] = 2.03$.

Author's personal copy

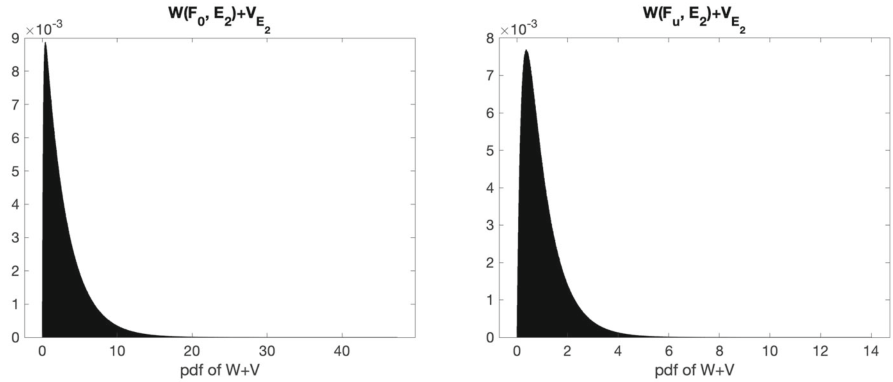

Fig. 4 The shape of the pdf pf $W(F_0, E_2) + V_{E_2}$ and $W(F_u, E_2) + V_{E_2}$ for Erlang E_2 service with $M_a = 10$, $\rho = 0.5$, $c_a^2 = 4$ and $c_s^2 = 0.5$

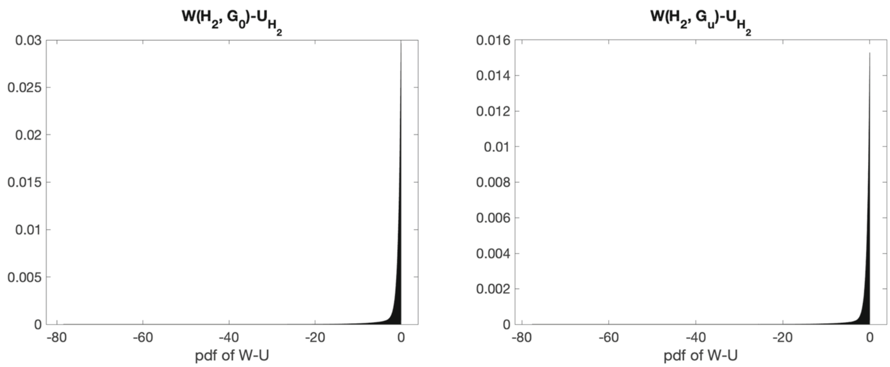

Fig. 5 The shape of the pdf of $W(H_2, G_0) - U_{H_2}$ and $W(H_2, G_u) - U_{H_2} \leq 0$ with H_2 interarrival-time cdf, $M_s = 10$, $\rho = 0.5$ and $c_a^2 = c_s^2 = 4$

Figure 6 presents corresponding simulation results when F has an E_2 distribution. We see that monotonicity fails, just as in Fig. 4. In these cases, $E[W(E_2, G_0)] = 1.01$ and $E[W(E_2, G_u)] = 1.06$. Notice that there is a switch of order of G_0 and G_u going from H_2 arrivals to E_2 arrivals.

9 Conclusions

This paper applies the theory of Tchebycheff (T) systems to identify interarrival-time and service-time cdf's that maximize or minimize the transient and steady-state moments of the waiting time in the $GI/GI/1$ queue, given the first two moments of the underlying interarrival-time and service-time cdf's. The extremal cdf's are classical two-point distributions that are determined by either having one mass point at 0 or the upper limit of the support.

Author's personal copy

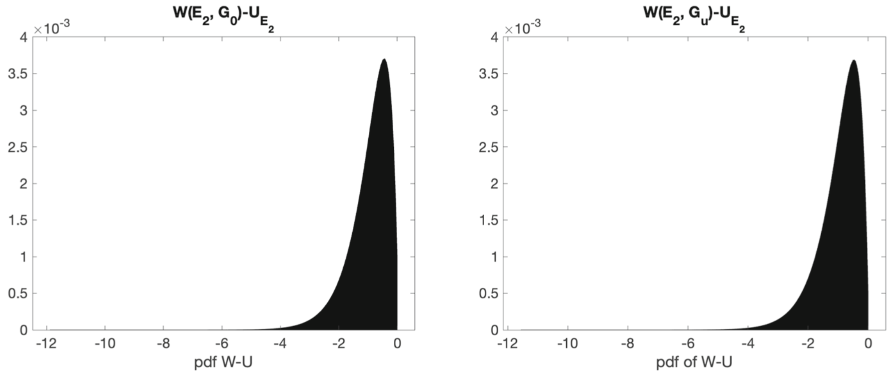

Fig. 6 The shape of the pdf of $W(E_2, G_0) - U_{E_2}$ and $W(E_2, G_u) - U_{E_2} \leq 0$ with E_2 interarrival-time cdf, $M_s = 10$, $\rho = 0.5$ and $c_a^2 = 0.5$, $c_s^2 = 4$

Theorem 1 establishes that higher moments of the steady-state waiting times are maximized (minimized) over all by interarrival-time cdf's F with given first two moments in the $GI/GI/1$ model by the classical extremal two-point cdf F_0 with one mass on 0 (F_u with one mass on the upper limit of support). Corresponding results are also obtained for the service times and for the two cdf's jointly. We prove Theorem 1 in Sect. 4 by combining stochastic comparison results in [25] and [11], which are intimately related to T systems. In Sect. 6, a variant of Theorem 1 is established to produce a continuity result for the mean steady-state waiting time in the $GI/GI/1$ queue, extending Corollary X.6.4 of [2].

Theorem 2 establishes sufficient conditions for corresponding results to hold for the steady state mean $E[W]$. The proofs rely on the tractable characterization of T systems in Sect. 3 in terms of Wronskians, which was used for the asymptotic decay rate in [7]. For given service-time cdf G, F_0 yields the upper bound if G is completely monotone. For optimizing over F, a key supporting result was Theorem 5, establishing the concavity of the sojourn-time cdf when the service-time cdf is completely monotone. For given interarrival-time cdf F, G_0 yields the upper bound if the cdf F is strictly concave, as occurs when G has a strictly decreasing pdf. Increasing convex order is used to give a sufficient condition for the overall upper bound to be $E[W(F_0, G_u)]$, as widely conjectured, but the main extremal problem for the $GI/GI/1$ queue remains unresolved.

Acknowledgements This research was supported by NSF CMMI 1634133. We thank Tomasz Rolski and an anonymous referee for exposing a gap in our first proof of Theorem 1.

References

1. Asmussen, S.: Phase-type representations in random walk and queueing problems. Ann. Probab. **20**(2), 772–789 (1992)
2. Asmussen, S.: Applied Probability and Queues, 2nd edn. Springer, New York (2003)

Author's personal copy

3. Bergmann, R., Daley, D.J., Rolski, T., Stoyan, D.: Bounds for cumulants of waiting times in $GI/GI/1$ queues. Math. Op. Schung Stat. Ser. Optim. **10**(2), 257–263 (1979)
4. Billingsley, P.: Convergence of Probability Measures, 1st edn. Wiley, New York (1968)
5. Chen, Y., Whitt, W.: Algorithms for the upper bound mean waiting time in the $GI/GI/1$ queue. Queue. Syst. **94**, 327–356 (2020a)
6. Chen, Y., Whitt, W.: Extremal $GI/GI/1$ queues given two moments: numerical results exploiting three-point distrdibutions. Columbia University, http://www.columbia.edu/~ww2040/allpapers.html(2020b)
7. Chen, Y., Whitt.W.: Extremal models for the $GI/GI/K$ waiting-time tail-probability decay rate. Oper. Res. Letters **48**, 770–776 (2020c)
8. Cohen, J.W.: The Single Server Queue, 2nd edn. North-Holland, Amsterdam (1982)
9. Daley, D.J.: Inequalities for moments of tails of random variables with queueing applications. Zeitschrift fur Wahrscheinlichkeitsetheorie Verw. Gebiete **41**, 139–143 (1977)
10. Daley, D.J., Kreinin, A.Ya., Trengove, C.D.: Inequalities concerning the waiting-time in single-server queues: a survey. In: Bhat, U.N., Basawa, I.V. (eds.) Queueing and Related Models. pp. 177–223. Clarendon Press, Oxford (1992)
11. Denuit, M., Lefevre, C., Shaked, M.: The s-convex orders among real random variables with applications. Math. Inequal. Appl. **1**(4), 585–613 (1998)
12. Eckberg, A.E.: Sharp bounds on Laplace-Stieltjes transforms, with applications to various queueing problems. Math. Op. Res. **2**(2), 135–142 (1977)
13. Gupta, V., Osogami, T.: On Markov-Krein characterization of the mean waiting time in $M/G/K$ and other queueing systems. Queue. Syst. **68**, 339–352 (2011)
14. Holtzman, J.M.: The accuracy of the equivalent random method with renewal in outs. Bell Syst. Tech. Ical J. **52**(9), 1673–1679 (1973)
15. Johnson, M.A., Taaffe, M.R.: An investigation of phase-distribution moment-matching algorithms for use in queueing models. Queie. Syst. **8**(1–2), 129–148 (1991)
16. Johnson, M.A., Taaffe, M.R.: Tchebycheff systems for probability analysis. Am. J. Math. Manag. Sci. **13**(1–2), 83–111 (1993)
17. Karlin, S., Studden, W.J.: Tchebycheff Systems; With Applications in Analysis and Statistics, vol. 137. Wiley, New York (1966)
18. Kingman, J.F.C.: The single server queue in heavy traffic. Proc. Camb. Phil. Soc. **77**, 902–904 (1961)
19. Kingman, J.F.C.: Inequalities for the queue $GI/G/1$. Biometrika **49**(3/4), 315–324 (1962)
20. Muller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risks. John Wiley and Sons, New York (2002)
21. Neuts, M.F.: Matrix-Geometric Solutions of Stochastic Models. The Johns Hopkins University Press, Baltimore (1981)
22. Ott, T.J.: Simple inequalties for the $D/G/1$ queue. Op. Res. **35**(4), 589–597 (1987)
23. Rogozin, B.A.: Some extremal problems in the theory of mass service. Theor. Probab. Appl. **11**(1), 144–151 (1966)
24. Rolski, T.: Some inequalities for $GI/M/n$ queues. Zast. Math. **13**(1), 43–47 (1972)
25. Rolski, T.: Order relations in the set of probability distribution functions and their applications in queueing theory. Diss. Math. Pol. Acad. Sci. **132**(1), 3–47 (1976)
26. Stoyan, D.: Comparison Methods for Queues and Other Stochastic Models. John Wiley and Sons, New York (1983). Translated and edited from 1977 German Edition by D. J. Daley
27. Szekli, R.: On the complete monotonicity of the waiting time density in some $GI/G/k$ systems. Queue Syst. **1**, 401–406 (1987)
28. Szekli, R.: Stochastic Ordering and Dependence in Applied Probability. Springer, Berlin (1995)
29. Whitt, W.: On approximations for queues, I: extremal distributions. AT&T Bell Lab. Tech. J. **63**(1), 115–137 (1984a)
30. Whitt, W.: On approximations for queues, III: mixtures of exponential distributions. AT&T Bell Lab. Tech. J. **63**(1), 163–175 (1984b)
31. Wolff, R.W., Wang, C.: Idle period approximations and bounds for the $GI/G/1$ queue. Adv. Appl. Probab. **35**(3), 773–792 (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.