Probability in the Engineering and Informational Sciences, **21**, 2007, 335–360. Printed in the U.S.A.
DOI: 10.1017/S0269964807000010

MONOTONICITY PROPERTIES FOR MULTISERVER QUEUES WITH RENEGING AND FINITE WAITING LINES

OUALID JOUINI AND YVES DALLERY
Laboratoire Génie Industriel
Ecole Centrale Paris
Grande Voie des Vignes
92295 Châtenay-Malabry Cedex, France
E-mail: walid.jouini@ecp.fr; yves.dallery@ecp.fr

We consider a markovian multiserver queue with a finite waiting line in which a customer may decide to leave and give up service if its waiting time in queue exceeds its random deadline. We focus on the performance measure in terms of the probability of being served under both transient and stationary regimes. We investigate monotonicity properties of first and second order of this performance with respect to the buffer size, say k. Under the stationary regime, we prove that our service level is strictly increasing and concave in k, whereas we prove under the transient regime that it is only increasing in k.

1. INTRODUCTION

Monotonicity properties of performance measures are useful for understanding and solving optimization problems of queuing systems. Optimization models are being used increasingly in the design of a variety of systems where queuing phenomena arise. Examples include flexible manufacturing systems, as well as service systems and telecommunications networks. For such problems, it is important to know the convexity properties of the performance measures with respect to the design variables. These properties might enable us to reduce the performance optimization problem to a convex programming problem, which is easier to solve. Using a convexity result, Yao and Shanthikumar [34] accelerated their computation procedure to design a loss queuing system subject to constraints

© 2007 Cambridge University Press 0269-9648/07 $25.00

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

on the loss probability. Koole and Pot [19] considered an optimization problem for an $M/M/s/K+M$ queue. The objective function is a profit function of the number of servers and the buffer size. They derived some monotonicity properties about the defined performance measure. Based on these properties, they developed a fast algorithm that avoids the research of all possible solutions to get the global optimum.

Several convexity properties about various performance measures have been investigated in the queuing literature. The major performance measures for delay systems are the average waiting time, the average queue length, and the probability of delay. Those for pure loss systems include basically the probability for a new arrival to be lost. In general, the loss probability is related to systems involving finite buffers or systems with reneging. In this article, we consider a queuing system with impatient customers and a finite waiting line. The performance measure of interest is the probability for a new arrival customer to enter service; or equivalently, the probability to not be lost. We investigate first- and second-order monotonicity properties of our performance measure as a function of the queue size. Note that the design of the buffer size is an important issue in practice. Koole, Nuyens, and Righter [18] addressed this problem by investigating the maximum queue length during a busy period for an infinite buffer size.

Another central feature in many practical queuing systems is the reneging phenomenon; that is, one customer might decide to leave the queue (abandons) before starting service. For instance, call abandonment is not negligible in call center operations. A major drawback in many call center models is assuming customers to be infinitely patient. Garnett, Mandelbaum, and Reiman [7] showed, using numerical examples, that models with and without abandonment tend to give very different performance measures even if the abandonment rate is small. A further application would be manufacturing systems with perishable items. In this article, we analyze the simplest abandonment model, assuming that the customer's patience is exponentially distributed. However, the model is still of interest in practice, as mentioned by Pierson and Whitt [25]. The authors have shown, using various simulation experiments, that the $M/M/s+M$ model provides a good approximation of the $M/GI/s+GI$ model.

Here is how the rest of the article is organized. In Section 2 we review the literature close to our work. In Section 3 we present the framework of the work: Section 3.1 is devoted to formulate the queuing model, and Section 3.2 gives definitions and some preliminary results. In Section 4 we focus on the first-order monotonicity results. In Section 4.1 we start by proving two helpful lemmas before proceeding to the main result. Next, using coupling arguments, we establish that the transient and stationary probabilities of being served are increasing in the buffer size. In Section 4.2 we prove the result for the stationary performance measure using an analytical approach. In Section 5 we prove that the stationary probability of being served is strictly concave in the buffer size. Some numerical illustrations of the results are also presented. In Section 6 we conclude and propose some directions for future research.

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

2. LITERATURE REVIEW

In this section, we review the literature related to this article. We start by presenting some articles investigating monotonicity results for models without reneging. Second, we focus on those for models incorporating reneging.

We classify the results for models with infinitely patient customers into three classes: pure loss, limited buffer, and infinite buffer models. For pure loss systems, Harel [9] proved that the throughput of an $M/G/s/s$ model is concave in the arrival and service rates. He also characterized the traffic intensity below which the Erlang loss formula is convex in the arrival rate and above which it is concave. Furthermore, he showed that the Erlang loss formula is convex in the service rate. For the same model, Messerli [22] proved that the loss probability is a convex function of the number of servers. Additional properties of the loss probability are also discussed by Jagerman [13].

As for systems with limited buffer, Nagarajan and Towsley [23] investigated the convexity of the loss probability in the $M/M/1/K$ queue with respect to the traffic intensity and the service rate. They showed that the loss probability is convex in the service rate. However, they proved that there is a value of the traffic intensity that exactly delineates the convex and concave regions of the loss probability as a function of the traffic intensity. Pacheco [24] considered for his part a more general model with many servers, namely the $M/M/s/K$ queue. He proved that the loss probability is convex in the queue capacity. Meister and Shanthikumar [21] proved many convexity results for tandem queuing systems. Several interesting stochastic comparisons of various variants of multiserver queues with limited buffer were also derived by Berger and Whitt [4].

In what follows, we review some monotonicity results for models with infinite queue capacity. Tu and Kumin [28] proved that the expected number of customers in a $G/G/1$ queue is convex in the service rate. They also showed that the result does not hold for a $GI/GI/2$ queue. Surprisingly, Harel [10] showed that the expected number of customers in an $M/D/s$ queue is convex in both arrival and service rates. For the $M/M/s$ queue, Lee and Cohen [20] showed that the average queue length and the probability of delay are both convex in the arrival rate. For the same model, Harel and Zipkin [12] established that the average sojourn time as well as its standard deviation are convex in arrival and service rates. Again about the $M/M/s$ queue, Jagers and van Doorn [14] focused on the performance measure in terms of the probability for a customer to wait no longer than a given threshold. Note that this service level is widely used in call centers. The authors show that the probability of interest is concave as a function of the number of servers, if the latter is strictly greater than the offered load. We refer the reader for further convexity properties to Weber [30, 31], Grassmann [8], Shanthikumar [27], Harel and Zipkin [11], Shaked and Shanthikumar [26], and Koole [17].

Let us now turn to the second area of literature related to this article. Queues with impatient customers have received some attention in the queuing literature. The results focus especially on performance evaluation. We refer the reader to Ancker

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

and Gafarian [1], Garnett et al. [7], and references therein for simple models assuming exponential reneging times. Other articles have allowed reneging to follow a general distribution. Related studies include those by Baccelli and Hebuterne [3], Brandt and Brandt [6], Ward and Glynn [29], and Zeltyn and Mandelbaum [35]. Concerning the monotonicity properties, few results were derived. This is due to the mathematical complexities of such problems. Bhattacharya and Ephremides [5] considered multiserver queues with impatient customers. They showed that the transient number of lost customers is a monotone function with respect to the arrival rate, the service rate, as well as the reneging rate. Armony, Plambeck, and Seshadri [2] considered a holding cost in an $M/M/C$ queue with impatient customers. They proved that this function is decreasing and convex in the service rate and the number of customers. Some sensitivity results for the Erlang-A model can also be found in Whitt [32].

3. FRAMEWORK

This section is devoted to formulate the general framework of the article. First, we describe the queuing system and detail the processes assumptions. Second, we define the performance measures of interest, namely the fraction of customers who get service under both transient and stationary regimes. We next develop some preliminary results.

3.1. Model Formulation

Consider a multiserver queuing system with a single class of customers. The model consists of a set of s parallel, identical servers and a finite queue (waiting line). There is a maximum number of customers that might be simultaneously present; we assume that the system can hold at most a total of K customers, including those in service. Clearly, $K \geq s$, and we denote the queue capacity by $k = K - s$, $k \geq 0$. The system is operated in such a way that at any time, any customer can be addressed by any server. So upon arrival, a customer is addressed by one of the available servers, if any. If not, the customer joins the queue if less than K customers are present in the system. If not, the customer is refused entry and departs immediately without service. He is blocked and considered lost. In addition, we assume that customers are impatient. After entering the queue, a customer will wait a random length of time for service to begin. If service has not begun by this time, he will renege (abandon), and again considered to be lost. Finally, retrials are ignored, and reneging is not allowed once a customer starts his service.

The arrival of customers is assumed to follow a Poisson process. Interarrival times are independent and identically distributed (i.i.d.) and exponentially distributed with rate λ. Successive service times are assumed to be i.i.d., independent from the arrival process, and follow an exponential distribution with rate μ. Times before reneging are assumed to be i.i.d. and exponentially distributed with rate γ. Following similar arguments, the system can be modeled as an $M/M/s/K + M$ queue. The symbol M after the $+$ is to indicate the markovian assumption for reneging times. Note that owing

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

to reneging, the system is always ergodic even if the queue has infinite capacity. Also, ergodicity would always be assured for our system because of its limited capacity, even if the customers were assumed to be infinitely patient. In conclusion, the system we consider here is unconditionally ergodic.

As we will establish later, we need not specify here the scheduling policy, except the assumption that it is work-conserving. A scheduling policy is defined to be work-conserving (nonidling) if the waiting customers get into service as soon as a server is free. No server stays idle as long as there are customers to serve. Several familiar disciplines of service as the first come–first served (FCFS), the last come–first served (LCFS), the random selection (RS) disciplines are characterized to be work-conserving. A further possible classification for scheduling policies is the preemption/nonpreemption schemes. In preemptive cases, we can interrupt the service of a customer to let a new arrival start service. However, a discipline of service is said to be nonpreemptive if there is no possibility for service interruption.

3.2. Preliminaries

In this subsection, we focus on characterizing the performance measure of interest. It is defined in terms of the fraction of customers who get service (i.e., the fraction of customers who are not blocked and who do not renege).

Let us consider an interval of time $[0, t]$, $t > 0$. We initially assume that the system starts empty. Given that t units of time have elapsed, let $n(t)$ and $s(t)$ be the total number of arrivals (including blocked customers) and the number of those who enter service, respectively. We define the transient fraction of customers who enter service, $Q(t)$, during $[0, t]$ as $Q(t) = s(t)/n(t)$. Taking the limit as $t \to \infty$ of $Q(t)$ leads to the stationary fraction Q of successful departures: $Q = \lim_{t\to\infty} Q(t)$. Let $b(t)$ and $r(t)$ be the number of blocked customers and that of those who renege, respectively. Since the quantities $s(t)$ and $n(t) - b(t) - r(t)$ coincide in the long run, then Q can be rewritten as

$$Q = \lim_{t\to\infty} \frac{n(t) - b(t) - r(t)}{n(t)}. \tag{3.1}$$

In what follows, we derive a closed-form expression for Q. We denote the system state by a random variable taking nonnegative integer values representing the total number of customers in system (including those in service). We should first distinguish two sequences of probabilities that could describe the limiting state of our system. The first sequence is denoted by $p(i)$ and represents the steady-state probability that the system is in state i, $0 \le i \le K$, seen by an outside random observer (at a random instant). The second is denoted by $q(i)$ and represents the probability that an arriving customer finds the system in state i, which is, in general, different from $p(i)$, $0 \le i \le K$. The existence of both sequences is assured due to the ergodicity property of our system.

Clearly, the quantity Q represents the probability in the infinite horizon for a new arrival customer to enter service, which involves, as a consequence, the sequence $q(i)$.

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

The latter is too difficult to be computed; however, the sequence $p(i)$ is easy to derive, as we will detail below. Fortunately, we state for our system that the quantities $p(i)$ and $q(i)$ coincide for any state i. This is due to the well-known PASTA property (Poisson Arrivals See Time Averages) also referred to as ROP (Random Observer Property), which holds here given that the arrival process is Poisson. The PASTA property is based on the memoryless property of the Poisson process, which allows one to generate a sequence of arrivals that take a random look at the system. We refer the reader to Kleinrock [16] for a further explanation and to Wolff [33] for a rigorous proof.

Let us now come back to Eq. (3.1) by dividing both the numerator and the denominator on the right-hand side over t. Computing Q can be reduced thereafter to computing separately the ratios $n(t)/t$, $b(t)/t$, and $r(t)/t$ as t goes to ∞. Recall that the mean number of customers per unit of time is λ. Hence, in the long run (as $t \to \infty$), the ratio $n(t)/t$ converges by construction to λ. As for the limit of $b(t)/t$ as $t \to \infty$, we might recognize it as the probability for a new arrival to be blocked times the mean arrival rate λ. So, it is the probability that a new arrival finds a full system times λ, namely the quantity $\lambda q(K)$, or equivalently $\lambda p(K)$. Let us now focus on the limit of $r(t)/t$ as t goes to infinity. One might recognize this quantity as the mean number of reneging per unit of time seen by a random outside observer. It takes, on average, $1/\gamma$ units of time for one customer waiting in queue to renege. Thus, as $t \to \infty$, $r(t)/t$ converges to the mean number of customers in queue (in the distant future) times γ. Based on the previous analysis, Q can be rewritten as

$$Q = 1 - p(K) - \frac{\gamma}{\lambda} \sum_{i=s+1}^{K} (i - s)p(i). \tag{3.2}$$

To get explicitly the expression of Q, we move on to compute the stationary probabilities $p(i)$, for $0 \leq i \leq K$. In the usual way, we model our system as a finite continuous-time birth–death process with discrete state space taking nonnegative integer values ranging from 0 to K and defined on a probability space. The birth rates are constant and equal to λ. The death rates are state dependent; when moving from state i to state $i - 1$, the death rate is $i\mu$ for $1 \leq i \leq s$ and it is $s\mu + (i - s)\gamma$ for $s < i \leq K$. Under the stationary regime, we easily get a set of K recursive equations relating $p(i)$ and $p(i + 1)$ for $0 \leq i \leq K - 1$. Proceeding to solve by iteration leads to

$$p(i) = \frac{\lambda^i}{i!\mu^i} p(0) \quad \text{for } 0 \leq i \leq s$$

and (3.3)

$$p(i) = \frac{\lambda^i}{s!\mu^s \prod_{j=1}^{i-s} (s\mu + j\gamma)} p(0) \quad \text{for } s < i \leq K,$$

where $p(0)$ is the steady-state probability to have no customers in system and, obviously, $p(i) = 0$ for $i > K$. Then we couple the last set of equations with the

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

probability conservation relation (i.e., $\sum_{i=0}^{\infty} p(i) = 1$) to get

$$p(0) = \left(\sum_{i=0}^{s} \frac{\lambda^i}{i!\,\mu^i} + \frac{\lambda^s}{s!\,\mu^s} \sum_{i=s+1}^{K} \frac{\lambda^{i-s}}{\prod_{j=1}^{i-s} (s\mu + j\gamma)} \right)^{-1}, \tag{3.4}$$

which determines all stationary probabilities. We still have to substitute them into Eq. (3.2) to obtain Q.

4. PROOF OF FIRST-ORDER MONOTONICITY PROPERTY

One can intuitively state that the performance measures $Q(t)$ and Q increase with respect to the queue capacity k, keeping the parameters λ, μ, γ, and s constant. The idea is that although adding more places in the waiting line might increase abandonments, it is clear that it could not deteriorate the performances we consider here. On the contrary, it allows for more customers to enter service. If not, it will, at worst, achieve an equal fraction of successful departures compared to a system with less queue capacity. In this section, we rigorously prove these results using two different approaches. In Section 4.1 we prove, using coupling arguments, that $Q(t)$ and Q increase in k for a more general case, namely for a $GI/M/s/K + M$ queue. In Section 4.2 we consider our original system (the $M/M/s/K + M$ queue) and prove, using an analytical approach, that Q increases in k.

4.1. Sample Path Approach

We start with a tangential development that will be of a great help to prove our main result. Let us relax some assumptions in our original system by considering a $GI/GI/s/K + M$ queue. We assume that interarrivals and service times are i.i.d., but we allow them to follow a general distribution. In Lemma 1, we present an interesting result about the relation between the performance measures of interest and the scheduling policy under which the system is working. For the rest of the article, we denote by Π the set of work-conserving nonpreemptive scheduling policies.

The following result in Lemma 1 can be seen as an extension of that in Theorem 2 of Jouini, Pot, Dallery, and Koole [15]. In the latter, the author proved a conservation result for the probability of being lost in a queuing system with an infinite buffer size. Here we prove the conservation result for the probability of being served by adding blocking (whenever the system is full).

LEMMA 1: *Consider a $GI/GI/s/K + M$ queue. Times before reneging are assumed to be i.i.d. and exponentially distributed. Then the probability of being served Q is constant for any work-conserving nonpreemptive scheduling policy.*

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

PROOF: The result is trivial for a queue with no capacity or with capacity 1. In such cases, it is clear that the system behaves identically for any policy $\pi \in \Pi$, and as a consequence, Q remains constant. Otherwise, for $k \geq 2$, we prove the result by coupling arguments. Consider two $GI/GI/s/K + M$ models, say Model 1 and Model 2. We assume that Model 1 and Model 2 have identical parameters except for the scheduling policies. Model 1 and Model 2 are working under policies π_1 and π_2, respectively, such that $\pi_1 \in \Pi$, $\pi_2 \in \Pi$, and $\pi_1 \neq \pi_2$. Our approach is based on a single sample path. In both models, we create identical successive arrival epochs as well as identical successive service times. However, since times before reneging are exponentially distributed, the decision for one customer to abandon the queue is not affected by his elapsed waiting time. This is stochastically equivalent to create randomly, for our sample path, a new maximum time of patience for each customer in queue after each selection for service epoch (or equivalently after each successful departure epoch). Assume that at time $t = 0$, both systems are empty, and let work begins.

Both models behave identically until a busy period starts and the following situation occurs: A server becomes idle (after a service completion) and more than one customer are waiting in queue. Let D_i be the epoch of that service completion (which occurs simultaneously in Model 1 and Model 2). For both models, let n be the number of customers in queue just before D_i, $2 \leq n \leq k$. At D_i, the idle server in Model 1 selects one customer from the queue who can be different from that selected by the same idle server in Model 2. However, the number of customers in queue goes down by one for both models, it becomes $n - 1$. Recall that up to now, the number of blocked customers as well as that of those who abandoned the queue are identical for both models.

At epoch D_i, we create for each customer, waiting in the queue of Model 1, a new patience time. Without altering distributions, since times before reneging are identically distributed, we create the same set of $n - 1$ maximum patience times as in Model 1, and we assign them arbitrarily to the customers waiting in Model 2. After D_i, three events are possible: One customer reneges, or a new arrival occurs, or a server becomes idle (service completion). Recall that by construction of our single sample path, these events occur simultaneously in both models. Assume now that the first event occurs, then the number of customers who abandon the queue goes up by one in both models and, as a consequence, remains identical for them. It is still also identical if another customer abandons the queue. In general, it is the case as long as there are customers waiting in queue. If not, both models will behave identically, anyway. Assume now that a new arrival occurs. Note that the number of customers in queue is the same in both models. If the queues are currently full (i.e., k are waiting for service), then the new arrival will be blocked, and systems states remain unchanged. However, if at least there is one available space, the number of customers in queue goes up by one in both models. Note that if another arrival occurs or that one customer abandons the queue, then, the number of customers in queue will increase by one (or remains unchanged if the system is full) or will decrease by one, respectively. The main conclusion is that the number of blocked customers as well as that of those who abandon the queue will vary identically in both models.

Assume now that one server becomes idle. If the number of customers in queue is currently less or equal to 1, it is obvious to see that policies π_1 and π_2 will select at the same

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

time the unique available customer, if any. Otherwise, the busy period ends in both models, so both policies will again select identically new arrivals for service until the beginning of the next busy period. However, if the number of customers in queue is greater or equal to 2, the selected customer for service might be different from one model to another. As earlier, we create for the remaining set of waiting customers, the same set of patience times. Again, we can state that the number of blocked customers as well as that of those who abandon the queue remain the same for both models.

Carrying on using the same arguments, we state that in a distant future, the number of blocked customers and that of those who renege in Model 1 coincide with those in Model 2. Since the number of arrivals are also equal for both models, the service level in terms of the fraction of successful departures is unchanged: $Q_{\pi 1} = Q_{\pi 2}$. This completes the proof. ■

Although the probability of being served is independent of the scheduling policy, the mean waiting time in queue for the served customers does depend on the scheduling policy. Jouini et al. [15] have proved the latter result when considering the particular case of a $GI/GI/s + M$ queue. They have also characterized the policies under which upper and lower bounds of the mean waiting time are achieved.

We should note, however, that the result in Lemma 1 does not hold if times before reneging are not i.i.d. and exponentially distributed, or if service times at any point during an arbitrary busy period are order of service dependent, we need to assume that no service needs are created or destroyed within the system: No renege in the midst of service, no forced idleness of servers, and so on.

In Lemma 2, we show that Q is still unchanged for any work-conserving scheduling policy (with preemption or not) if we further assume that service times are i.i.d. and exponentially distributed.

LEMMA 2: *Consider a $GI/M/s/K + M$ queue. Times before reneging are assumed to be i.i.d. and exponentially distributed. Then, the probability of being served Q is constant for any workconserving scheduling policy.*

PROOF: We again show the result using coupling arguments. Based on a single sample path, we compare the quantity Q for two identical $GI/M/s/K + M$ models, say Model 1 and Model 2, working under two different scheduling policies π_1 and π_2, respectively. We assume that π_1 and π_2 are work-conserving and we do not restrict them to be nonpreemptive. We use a similar approach to that for Lemma 1. The only difference is only when an interruption of service occurs in one of the models. Note that just before the epoch of that event, both models are identical: All servers are busy, the same number n of customers in queue, the same remaining service times, and the same set of remaining times before reneging for waiting customers in queue. Without loss of generality, assume that in Model 1, a new arrival interrupts the service of a customer currently in service. Since service times are assumed to be exponentially distributed, the remaining time for a service completion is not affected by the elapsed time in service. This allows us to create randomly, for our sample path, a new set of remaining service

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

times for the customers currently in service (s customers in both models) and a new set of patience time for waiting customers in queue (n customers in both models). Continuing the sample path comparison in the long run will subsequently show that Q coincides for both models. This completes the proof of the lemma. ■

In Theorem 1, we show the main result of first-order monotonicity for a $GI/M/s/K+M$ queue. The analysis resorts in part to the previous preliminary results of this section.

THEOREM 1: *Consider a $GI/M/s/K+M$ queue. Times before reneging are assumed to be i.i.d. and exponentially distributed. Then the probability of being served Q is strictly increasing in the buffer size k.*

PROOF: To prove the result of the theorem, it suffices to compare the achieved Q for the two following systems. The first, say Model 1, is a $GI/M/s/K+M$ queue with k waiting spaces. From Lemma 1, it does not restrict generality to assume that Model 1 works under the FCFS discipline of service. The second model is identical to the first in all parameters except that it has $k+1$ waiting spaces. From Lemma 2, the latter is equivalent, in terms of the achieved Q, to a $GI/M/s/K+M$ queue, say Model 2, with $k+1$ waiting spaces and working under any preemptive work-conserving policy. In summary, it is left to establish that the stationary probability of being served in Model 2, say Q_2, is strictly greater than that in Model 1, say Q_1.

The proof follows the sample path approach. Before proceeding to the details, let us characterize a specific preemptive work-conserving policy, say π, under which Model 2 is operated. We divide the queue in Model 2 (with capacity $k+1$) into two virtual queues. The first, say queue 1, has capacity k. The second, say queue 2, has the remaining capacity (i.e., 1). Upon arrival, a customer is addressed by one of the available servers, if any. If not, the customer must join one of the queues. We will specify the queue joining policy later. Customers in queue 1 have priority over customers in queue 2 in the sense that servers are handling customers belonging to queue 1 first. The priority rule is preemptive, which simply means that a server currently serving a customer pulled from queue 2, while a new arrival customer joins queue 1, will interrupt this service and turn to the queue 1 customer. Within each queue, customers are served in order of their arrival (i.e., under the FCFS discipline).

Let us now couple Model 1 and Model 2 and let work begins. Both models behave identically until the situation where in Model 1 all servers are busy, k customers are waiting in queue, and a new arrival occurs. Let us stop our clock temporarily. We denote that customer as the "low customer." Clearly, the "low customer" is blocked in Model 1 because the system is currently full; however, he joins the waiting line in Model 2. We assign him to queue 2 (with lower priority). Recall that up to now the number of customers served is identical in both models. Let our clock resume ticking: Arrivals, blocking, abandonments, as well as service completions will occur at the same epochs in both models until the busy period in system 1 ends (which occurs with probability 1 due to the ergodicity condition). We distinguish two possible

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

cases for the "low customer": either he has meanwhile abandoned or he is still waiting in queue 2. In the first case, both systems states become identical again. In the second case (i.e., if the "low customer" is still waiting), then we assign him to the server currently idle in Model 2. As long as the current idle period in Model 1 does not finish, we let the "low customer" stay in service. If the "low customer" finishes his service before a new arrival occurs (at the same epoch in both models), Model 2 will have one more service completion compared to Model 1, and all events in both models become identical again. If not (i.e., if the idle period in Model 1 ends and the "low customer" has not successfully left Model 2), then we interrupt his service and we put him back in queue 2. Here, the idea of choosing policy π is to ensure an identical behavior, in both models, with regard to all customers except for the "low customers." Such customers are blocked (lost) in Model 1; however, they join queue 2 in Model 2.

From the previous arguments, one might easily deduce that $Q_1 \leq Q_2$. Let us now proceed to establish that $Q_1 < Q_2$. It is clear that one "low customer" at most might be present in Model 2 at a given observation moment. Let us further define a particular cycle duration referred to as the "low cycle." The "low cycle" starts when a "low customer" enters Model 2 and it terminates upon the arrival of the next "low customer." The latter allows the following "low cycle" to start, and so on. The duration of a "low cycle" is given by the time it takes so that the "low customer" who starts the cycle either reneges or successfully finishes his service plus the time it takes starting from that epoch until the next "low customer" arrival epoch. Since the systems we consider here are stable, any busy period in Model 1 ends with probability 1 (i.e., its duration is finite ($<\infty$)). In addition, knowing that times before reneging are finite, we state that the "low cycle" duration is also finite. Furthermore, since interarrival times, times before reneging, as well as service times are i.i.d. and also independent of each other, it then follows that "low cycle" durations are also i.i.d. Next, assuming the stationary regime and observing that there is a nonzero probability that a "low customer" finishes successfully his service within its corresponding "low cycle," it yields from the Law of Large Numbers that there is a nonzero proportion of "low customers" that will finish successfully their service. Thus, we state that the number of customers being served in Model 2 is strictly greater than that in Model 1. Finally, it is implied that the stationary probability of being served Q is strictly increasing in the buffer size k. This completes the proof. ■

In a parallel to the proof of Theorem 1, we also state that the probability of being served under the transient regime, $Q(t)$, is an increasing function of k. Note that it is not necessarily strictly increasing in k as it is the case for the quantity Q.

4.2. Analytical Approach

In this section, we again consider our original $M/M/s/K + M$ queue described in Section 3.1. As shown in Section 3.2, a closed-form expression of the quantity Q can be derived. This allows us to again prove the result of Theorem 1 using an analytical approach. The analysis we address in this subsection is, in particular, useful for the proof of the convexity result in Section 5. Before giving the details of the proof of the

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

monotonicity property, we begin with some preliminary results by means of Properties 1 and 2. For the rest of the article, an empty sum is being interpreted as zero and an empty product is being interpreted as one.

Our objective is to show that Q is strictly increasing in k for an $M/M/s/K+M$ queue. To do so, we consider two models. The first is an $M/M/s/K+M$ queue with parameters λ, γ, μ, and s and k waiting spaces, $k \geq 0$. The second model is identical to the first; however, it has a larger buffer with $k+1$ waiting spaces. Recall that for our analysis, we do not need to specify the scheduling policy except that it is work-conserving. Next, it suffices to show that the stationary probability of being served in the first model, say Q_k, is strictly lower than that in the second model, say Q_{k+1}; or, equivalently, if we introduce the sequence $\{U_k, k \geq 0\}$ defined as $U_k = Q_{k+1} - Q_k$, it remains for us to establish that $U_k > 0$ for all $k \geq 0$. From Eq. (3.2), U_k can be rewritten as

$$U_k = p_k(k+s) - p_{k+1}(k+s+1) + \frac{\gamma}{\lambda}\left(\sum_{i=s+1}^{k+s}(i-s)p_k(i) - \sum_{i=s+1}^{k+s+1}(i-s)p_{k+1}(i)\right). \tag{4.1}$$

The stationary probabilities are given by Eqs. (3.3) and (3.4). The subscripts are to indicate to which system the stationary probabilities correspond, either for the one with queue capacity k or for that with queue capacity $k+1$. In Property 1, we state a useful relation between U_k and U_{k+1} for any nonnegative integer k.

PROPERTY 1: *For all $k \geq 0$, the following holds:*

$$U_{k+1} = \frac{s!\mu^s \sum_{i=0}^{s}\phi_i + \sum_{i=s+1}^{k+s}\rho_i}{s!\mu^s \sum_{i=0}^{s}\phi_i + \sum_{i=s+1}^{k+s+2}\rho_i} \frac{\lambda}{s\mu + (k+2)\gamma} U_k, \tag{4.2}$$

where

$$\phi_i = \frac{\lambda^i}{i!\mu^i} \quad \text{for } i \geq 0,$$

and (4.3)

$$\rho_i = \frac{\lambda^i}{\prod_{j=1}^{i-s}(s\mu + j\gamma)} \quad \text{for } i \geq s+1.$$

PROOF: The proof is provided in Appendix A. ■

Note that proving Property 1 represents the "hard" part of the proof of the monotonicity result as well as that of the convexity result. One can verify that Eq. (4.2)

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

holds for different special cases. For instance, let us consider an infinite-server queue $M/M/s/K+M$ ($s \to \infty$). Taking the limit in Eq. (4.2) as s goes to ∞ implies that $U_k = 0$ for all nonnegative integers $k \geq 1$ (in addition, from Eq. (A.4), for example, we have $U_0 = 0$), which obviously agrees with the classical queuing results. The result also holds for an $M/M/s/K+M$ queue with infinitely impatient customers ($\gamma = \infty$). In that case, the $M/M/s/K+M$ queue is equivalent to a loss system (without waiting space). Thus, it is easy to see that the quantity Q_k does not depend on the buffer size k. So, $U_k = 0$ for any $k \geq 0$, which agrees with Eq. (4.2).

Although we present in Property 2 an inequality that directly seems to be of independent interest, it is useful for forthcoming proofs of our results.

PROPERTY 2: *Let λ and μ be strictly positive reals and let $\{N_s, s \geq 1\}$ be a sequence defined as*

$$N_s = s\mu \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} - \lambda \sum_{i=0}^{s-1} \frac{\lambda^i}{i!\mu^i}. \tag{4.4}$$

Then, $N_s > 0$ for all $s \geq 1$.

PROOF: The inequality holds by induction. We have $N_1 = \mu > 0$; then Property 2 holds for $s = 1$. Assume now that $N_s > 0$ for a given $s \geq 1$ and let us show that $N_{s+1} > 0$. From Eq. (4.4), N_{s+1} can be written as

$$\begin{aligned}
N_{s+1} &= (s+1)\mu \sum_{i=0}^{s+1} \frac{\lambda^i}{i!\mu^i} - \lambda \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} \\
&= s\mu \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + s\mu \frac{\lambda^{s+1}}{(s+1)!\mu^{s+1}} + \mu \sum_{i=0}^{s+1} \frac{\lambda^i}{i!\mu^i} - \lambda \sum_{i=0}^{s-1} \frac{\lambda^i}{i!\mu^i} - \frac{\lambda^{s+1}}{s!\mu^s} \\
&= N_s + \frac{s\lambda^{s+1}}{(s+1)!\mu^s} - \frac{\lambda^{s+1}}{s!\mu^s} + \mu \sum_{i=0}^{s+1} \frac{\lambda^i}{i!\mu^i} \\
&= N_s - \frac{\lambda^{s+1}}{(s+1)!\mu^s} + \mu \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \frac{\lambda^{s+1}}{(s+1)!\mu^s} \\
&= N_s + \mu \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i}.
\end{aligned} \tag{4.5}$$

Using the induction assumption, it follows that $N_{s+1} > 0$. Finally, we conclude that $N_s > 0$ for all $s \geq 1$. This completes the proof. ■

In Theorem 2, we state the main result of this section. Having Properties 1 and 2, we are now ready to establish the first-order monotonicity property of the probability of being served, Q, with respect to the buffer size k.

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

THEOREM 2: *Consider an $M/M/s/K+M$ queue. Times before reneging are assumed to be i.i.d. and exponentially distributed. Then Q is strictly increasing in the buffer size k.*

PROOF: As explained in the beginning of this section, proving the theorem is equivalent to proving that U_k is strictly positive for $k \geq 0$. Keeping the parameters λ, μ, s, and γ constant, the result holds by induction on k.

Let us establish our claim for the first rank, $k = 0$. The quantity U_0 is given by $U_0 = Q_1 - Q_0$, where Q_0 and Q_1 are the probabilities of being served for the $M/M/s/s+M$ (no waiting space) and $M/M/s/s+1+M$ (single waiting space) systems, respectively. Using Eq. (3.2), the probability Q_0 is given by $Q_0 = 1 - p_0(s)$, where $p_0(s)$ is the stationary probability of having s customers in the $M/M/s/s+M$ system. As for Q_1, it is given by $Q_1 = 1 - [(\lambda+\gamma)/\lambda]\, p_1(s+1)$, where $p_1(s+1)$ is the stationary probability of having $s+1$ customers in the $M/M/s/s+1+M$ system.

From Eqs. (3.3) and (3.4), we get

$$Q_0 = 1 - \frac{\dfrac{\lambda^s}{s!\mu^s}}{\displaystyle\sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i}} \tag{4.6}$$

and

$$Q_1 = 1 - \frac{\lambda+\gamma}{s\mu+\gamma} \frac{\dfrac{\lambda^s}{s!\mu^s}}{\displaystyle\sum_{i=0}^{s} \frac{\lambda^i}{i!\,\mu^i} + \frac{\lambda^s}{s!\,\mu^s}\frac{\lambda}{s\mu} + \gamma}. \tag{4.7}$$

Therefore,

$$U_0 = \frac{\lambda^s}{s!\,\mu^s \displaystyle\sum_{i=0}^{s} \frac{\lambda^i}{i!\,\mu^i}} - \frac{\lambda^{s+1} + \lambda^s\,\gamma}{s!\,\mu^s\,(s\mu+\gamma) \displaystyle\sum_{i=0}^{s} \frac{\lambda^i}{i!\,\mu^i} + \lambda^{s+1}}. \tag{4.8}$$

To prove that $U_0 > 0$, we consider U_0 as a real function of γ, for $\gamma \geq 0$, and we study the sign of $U_0(\gamma)$. It is clear that U_0 has the property to be continuous and derivable in γ. Taking the derivative, U_0', of U_0 in γ leads to

$$U_0'(\gamma) = -\frac{s!\mu^s s\mu \displaystyle\sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + s!\mu^s\gamma \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \lambda^{s+1} - (\lambda+\gamma)s!\mu^s \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i}}{\left(s!\mu^s(s\mu+\gamma)\displaystyle\sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \lambda^{s+1}\right)^2} \lambda^s$$

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

$$= -\frac{s!\mu^s s\mu \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + s!\mu^s \gamma \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \lambda^{s+1} - (\lambda+\gamma)\left(s!\mu^s \sum_{i=0}^{s-1} \frac{\lambda^i}{i!\mu^i} + \lambda^s\right)}{\left(s!\mu^s(s\mu+\gamma)\sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \lambda^{s+1}\right)^2}\lambda^s$$

$$= -\frac{s!\mu^s s\mu \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + s!\mu^s \gamma \sum_{i=0}^{s-1} \frac{\lambda^i}{i!\mu^i} + \lambda^s\gamma + \lambda^{s+1}}{\left(s!\mu^s(s\mu+\gamma)\sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \lambda^{s+1}\right)^2}\lambda^s$$

$$+ \frac{\lambda^{s+1} + \lambda s!\mu^s \sum_{i=0}^{s-1} \frac{\lambda^i}{i!\mu^i} + \lambda^s\gamma + s!\mu^s\gamma \sum_{i=0}^{s-1} \frac{\lambda^i}{i!\mu^i}}{\left(s!\mu^s(s\mu+\gamma)\sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \lambda^{s+1}\right)^2}\lambda^s$$

$$= -\lambda^s s!\mu^s \frac{s\mu \sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} - \lambda \sum_{i=0}^{s-1} \frac{\lambda^i}{i!\mu^i}}{\left(s!\mu^s(s\mu+\gamma)\sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \lambda^{s+1}\right)^2}. \tag{4.9}$$

Using the notation in Eq. (4.4), $U_0'(\gamma)$ can be rewritten as

$$U_0'(\gamma) = \frac{-\lambda^s s!\mu^s}{\left(s!\mu^s(s\mu+\gamma)\sum_{i=0}^{s} \frac{\lambda^i}{i!\mu^i} + \lambda^{s+1}\right)^2} N_s. \tag{4.10}$$

Applying Property 2 for s strictly positive integer and for λ and μ strictly positive reals, we easily see that $U_0'(\gamma) < 0$. Then U_0 is a strictly decreasing function in γ for $\gamma \geq 0$. Hence, it follows that

$$U_0(\gamma) > \lim_{\gamma \to +\infty} U_0(\gamma) \quad \text{for } \gamma \geq 0. \tag{4.11}$$

Observing that

$$\lim_{\gamma \to +\infty} U_0(\gamma) = \frac{\lambda^s}{s!\,\mu^s \sum_{i=0}^{s} \frac{\lambda^i}{i!\,\mu^i}} - \frac{\lambda^{s+1} + \lambda^s\,\gamma}{s!\,\mu^s\,(s\mu+\gamma)\sum_{i=0}^{s} \frac{\lambda^i}{i!\,\mu^i} + \lambda^{s+1}} = 0, \tag{4.12}$$

we deduce that $U_0(\gamma) > 0$ for $\gamma \geq 0$. Thereafter, our claim is true for the first rank $k = 0$.

Let us consider $k \geq 0$ and assume that our claim is true for the rank k (i.e., $U_k > 0$). Let us now prove that our claim is true for the rank $k+1$. This is a direct consequence of Property 1. For $s \geq 1$, λ, $\mu > 0$, and $\gamma \geq 0$, we state, using Property 1, that U_{k+1} is the

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

product of U_k and a strictly positive real. So, $U_{k+1} > 0$. Finally, we conclude that $U_k > 0$ for $k \geq 0$, which completes the proof of the theorem. ■

5. PROOF OF SECOND-ORDER MONOTONICITY PROPERTY

In this section, we investigate the second-order property of monotonicity (of the probability of being served) in the queue capacity. First, we prove using a simple counterexample that the transient probability of being served, $Q(t)$, is not concave in k. Second, we state our main result in Theorem 3 about the concavity property. Finally, we present some numerical illustrations of that result.

To prove the nonconcavity of $Q(t)$ as a function of k, we consider three $M/M/1/K + M$ queues denoted by Model 1, Model 2, and Model 3. Assume the discipline of service to be FCFS. The models are identical in all parameters except for the buffer size. Specifically, Models 1, 2, and 3 contain one, two, and three waiting spaces, respectively. During an interval of time $[0, t]$, we denote the transient probability of being served for Model 1 by $Q_1(t)$. We denote those for Models 2 and 3 by $Q_2(t)$ and $Q_3(t)$, respectively. In what follows, we construct one possible sample path that shows that the transient probability of being served is not concave in k. In mathematical terms, it consists of finding an instant t such that $Q_3(t) - Q_2(t) > Q_2(t) - Q_1(t)$.

Initially, the models are empty. Now, let work begin. All models behave identically until the situation where, in each model, the unique available server is busy and there is one waiting customer in queue, say A_1. Thereafter, assume that one arrival, say A_2, occurs before a service completion or an abandonment. Note that this event occurs with a nonzero probability. The customer A_2 is blocked in Model 1, whereas he joins the queue in Models 2 and 3. Assume also that the next event is an arrival denoted by A_3. The customer A_3 is blocked in Models 1 and 2, however, he joins the queue in Model 3. Next, assume that A_2 abandons the queue, which occurs simultaneously in Models 2 and 3. Then assume that A_1 in all models and A_3 in Model 3 finish their service and successfully leave the systems. Let t_{A_3} be the epoch of the departure of A_3. So, we state that during $[0, t_{A_3}]$, the number of served customers in Model 1 is equal to that in Model 2. However, there is one served customer more in Model 3 compared to the other models. In other words, $Q_3(t) > Q_2(t)$ and $Q_1(t) = Q_2(t)$, which leads to the inequality $Q_3(t) - Q_2(t) > Q_2(t) - Q_1(t)$ and closes the discussion.

Turning now to the concavity of the stationary quantity Q as a function of k, we present the following theorem.

THEOREM 3: *Consider an $M/M/s/K + M$ queue. Times before reneging are assumed to be i.i.d. and exponentially distributed. Then Q is a strictly concave function in the buffer size k.*

PROOF: Let us again consider three $M/M/s/K + M$ queues denoted by Model 1, 2, and 3. All models are identical in all parameters except in the buffer size. In Model 1, there are k waiting spaces. However, Model 2 and Model 3 have $k + 1$ and $k + 2$ waiting spaces, respectively. We do not need to specify the scheduling policy except that it is

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

work-conserving. For Models 1, 2, and 3, we denote by Q_k, Q_{k+1}, and Q_{k+2} the stationary probabilities of being served, respectively. Following this introduction, one can easily see that proving our theorem is equivalent to proving that $U_k = Q_{k+1} - Q_k$ is strictly greater than $U_{k+1} = Q_{k+2} - Q_{k+1}$ for all $k \geq 0$. In other terms, it remains to prove that the sequence $\{U_k,\ k \geq 0\}$ is strictly decreasing. Knowing from Theorem 2 that $U_k > 0$ for $k \geq 0$, it suffices to show that $U_{k+1}/U_k < 1$ for $k \geq 0$. From Eq. (4.2), we have

$$\frac{U_{k+1}}{U_k} = \frac{s!\mu^s \sum_{i=0}^{s} \phi_i + \sum_{i=s+1}^{k+s} \rho_i}{s!\mu^s \sum_{i=0}^{s} \phi_i + \sum_{i=s+1}^{k+s+2} \rho_i} \frac{\lambda}{s\mu + (k+2)\gamma} \quad \text{for } k \geq 0, \tag{5.1}$$

which can be rewritten, for $k \geq 0$, as

$$\frac{U_{k+1}}{U_k} = \frac{\left(\left(s!\mu^s \sum_{i=0}^{s-1} \phi_i\right) \frac{\lambda}{s\mu + (k+2)\gamma}\right) + \left(\left(\lambda^s + \sum_{i=s+1}^{k+s} \rho_i\right) \frac{\lambda}{s\mu + (k+2)\gamma}\right)}{s!\mu^s \sum_{i=0}^{s} \phi_i + \sum_{i=s+1}^{k+s+2} \rho_i}. \tag{5.2}$$

On the one hand, Property 2 leads to

$$\lambda \sum_{i=0}^{s-1} \phi_i < (s\mu + (k+1)\gamma) \sum_{i=0}^{s} \phi_i. \tag{5.3}$$

Hence,

$$\left(s!\mu^s \sum_{i=0}^{s-1} \phi_i\right) \frac{\lambda}{s\mu + (k+1)\gamma} < s!\mu^s \sum_{i=0}^{s} \phi_i. \tag{5.4}$$

On the other hand, we have for all i, such that $i > k + s + 1$,

$$\frac{\lambda}{s\mu + (k+2)\gamma} < \frac{\lambda}{s\mu + (i-s+1)\gamma}, \tag{5.5}$$

which implies

$$\frac{\lambda}{s\mu + (k+2)\gamma} \rho_i < \frac{\lambda}{s\mu + (i-s+1)\gamma} \rho_i = \rho_{i+1}. \tag{5.6}$$

Summing Eq. (5.6) on all i, $s + 1 \leq i \leq k + s$, we get

$$\frac{\lambda}{s\mu + (k+2)\gamma} \sum_{i=s+1}^{k+s} \rho_i < \sum_{i=s+1}^{k+s} \rho_{i+1} = \sum_{i=s+2}^{k+s+1} \rho_i. \tag{5.7}$$

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

Next, observing that

$$\frac{\lambda^{s+1}}{s\mu+(k+2)\gamma} < \frac{\lambda^{s+1}}{s\mu+\gamma} = \rho_{s+1} \quad \text{for } k \geq 0, \tag{5.8}$$

the summation of both inequalities (5.7) and (5.8) leads to

$$\left(\lambda^s + \sum_{i=s+1}^{k+s} \rho_i\right) \frac{\lambda}{s\mu+(k+2)\gamma} < \sum_{i=s+1}^{k+s+1} \rho_i < \sum_{i=s+1}^{k+s+2} \rho_i. \tag{5.9}$$

Finally, it remains to apply relations (5.4) and (5.9) back into relation (5.2), to state that $U_{k+1}/U_k < 1$. This completes the proof of the theorem. ■

To get some numerical illustrations of our results, we consider various $M/M/s/K+M$ models by taking a broad range of parameters values. The service rate is unchanged for all chosen examples, $\mu = 1$. The values of the reneging rate are 0.5, 1, and 2. The number of servers are 1, 2, 3, 5, 10, 15, 50, 70, and 100. To vary the "servers utilization" calculated as $\lambda/s\mu$, we consider $\lambda = 1.8$ for $s = 1, 2$, and 3, $\lambda = 8$ for $s = 5, 10$, and 15, and $\lambda = 60$ for $s = 50, 70$, and 100. For each set of the previous values, the buffer size ranges from 0 to 30. The detailed results are presented in Tables B1–B3.

From the numerical results, we underline the following comments. As expected from Theorems 2 and 3, Q is increasing and concave in k, keeping all remaining parameters constant. One might see that there is no need to go beyond a buffer size around 10 to approximately reach the maximum of the probability of being served (reached within an infinite buffer size). Starting from a system with no waiting space, most of the improvements are achieved by adding two places in the buffer. Obviously, we also see that Q is decreasing with respect to the abandonment rate γ. The reason is simply that the probability to abandon the queue is increasing in the abandonment rate. Furthermore, for a fixed server utilization, large systems allow one to achieve higher service levels. This does not seem at odds with known results; it is due to the pooling effect.

6. CONCLUSIONS

In this article, we considered a queuing system with reneging and a finite buffer size. The model is of interest for the modeling in practice of several systems with impatient customers, such as call centers. We investigated monotonicity results of the probability of being served with respect to the buffer size. These results are helpful when addressing optimizations issues. We considered both transient and stationary quantities of the performance of interest. Under the transient regime, we proved that it is an increasing and nonconcave function of the buffer size. Under the stationary regime, we proved that it is strictly increasing and concave in the buffer size.

As a topic for future research, it would be interesting to investigate, in a similar fashion, the convexity properties of the performance measure as a function of other parameters such as the arrival rate, service rate, reneging rate, and number of servers.

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

Acknowledgments

The authors thank Ger Koole and Mohamed Salah Aguir for their helpful discussions to improve this article.

References

1. Ancker, C.J. & Gafarian, A. (1962). Queueing with impatient customers who leave at random. *Journal of Industrial Engineering* 13: 84–90.
2. Armoney, M., Plambeck, E., & Seshadri, S. (2005). Convexity properties and comparative statics for an M/M/S queue with impatient customers: Why you shouldn't shout at the DMV. http://www.hdl.handle.net/2451/14637.
3. Baccelli, F. & Hebuterne, G. (1981). On queues with impatient customers. In *Performance '81*. Amsterdam: North-Holland, pp. 159–179.
4. Berger, A.W. & Whitt, W. (1992). Comparisons of multi-server queues with finite waiting rooms. *Stochastic Models* 8: 719–732.
5. Bhattacharya, P.P. & Ephremides, A. (1991). Stochastic monotonicity properties of multiserver queues with impatient customers. *Journal of Applied Probability* 28: 673–682.
6. Brandt, A. & Brandt, M. (2002). Asymptotic results and a Markovian approximation for the M(n)/M(n)/C + GI system. *Queueing System: Theory and Applications* 41: 73–94.
7. Garnett, O., Mandelbaum, A., & Reiman, M. (2002). Designing a call center with impatient customers. *Manufacturing & Service Operations Management* 4: 208–227.
8. Grassmann, W. (1987). The convexity of the mean queue size of the M/M/C queue with respect to the traffic intensity. *Journal of Applied Probability* 20: 916–919.
9. Harel, A. (1990). Convexity of the Erlang loss formula. *Operations Research* 38: 499–505.
10. Harel, A. (1990). Convexity results for single-server queue and for multiserver queues with constant service times. *Journal of Applied Probability* 27: 465–468.
11. Harel, A. & Zipkin, P. The convexity of a general performance measure for multiserver queues. *Journal of Applied Probability* 24: 725–736.
12. Harel, A. & Zipkin, P.H. (1987) Strong convexity results for queueing systems. *Operations Research* 35: 405–418.
13. Jagerman, D.L. (1974). Some properties of the Erlang loss function. *Bell System Technical Journal* 53: 525–551.
14. Jagers, A.A. & van Doorn, E.A. (1991) Convexity of functions which are generalizations of the Erlang loss function and the Erlang delay function. *SIAM Review* 33: 281–282.
15. Jouini, O., Pot, A., Dallery, Y., & Koole, G. (2006). Real-time dynamic scheduling policies for multiclass call centers with impatient customers. Working paper. Ecole Centrale, Paris, and Vrije Universiteit, Amsterdam.
16. Kleinrock, L. (1975). *Queueing systems, theory*, Vol. I. New York: Wiley-Interscience.
17. Koole, G. (2004). Convexity in tandem queues. *Probability in the Engineering and Informational Sciences* 18: 13–31.
18. Koole, G., Nuyens, M., & Righter, R. (2005). The effect of service of time variability on maximum queue lengths in MX/G/1 queues. *Journal of Applied Probability* 42: 883–891.
19. Koole, G. & Pot, A. (to appear). A note on profit maximization and monotonicity for inbound call centers. http://www.math.Vu.nl/∼sapot/call-centers/profitmax/art.pdf.
20. Lee, H.L. & Cohen, M.A. (1983). A note on the convexity of performance measures of M/M/C queueing systems. *Journal of Applied Probability* 20: 920–923.
21. Meister, L. & Shanthikumar, J.G. (1990). Concavity of the throughput of tandem queueing system with finite buffer storage space. *Advances in Applied Probability* 22: 764–767.
22. Messerli, E.J. (1972). Proof of a convexity property of the Erlang B formula. *Bell System Technical Journal* 51: 951–953.
23. Nagarajan, R. & Towsley, D. (1992). A note on the convexity of the probability of full buffer in the M/M/1/K queue. CMPSCI Technical Report TR 92–85, MIT, Cambridge, MA.
24. Pacheco, A. (1994). Second-order properties of the loss probability in M/M/s/s + c systems. *Queueing Systems: Theory and Applications* 15: 309–324.

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

25. Pierson, M.P. & Whitt, W. (2006). A statistically-fit Markovian approximation of a basic call-center model. Working paper, Columbia University, New York.
26. Shaked, M. & Shanthikumar, J.G. (1988). Stochastic convexity and its applications. *Advances in Applied Probability* 20: 427–446.
27. Shanthikumar, J.G. (1987). Stochastic majorization of random variables with proportional equilibrium rate. *Advances in Applied Probability* 19: 854–872.
28. Tu, H.Y. & Kumin, H. (1983). A convexity result for a class of GI/G/1 queueing systems. *Operations Research* 31: 948–950.
29. Ward, A.R. & Glynn, P.W. (2003). A diffusion approximation for a Markovian queue with reneging. *Queueing Systems: Theory and Applications* 43: 103–128.
30. Weber, R.R. (1980). On the marginal benefit of adding servers to G/GI/m queues. *Management Science* 26: 946–951.
31. Weber, R.R. (1983). A note on waiting times in single server queues. *Operations Research* 31: 950-951.
32. Whitt, W. (2006). Sensitivity of performance in the Erlang-A queueing model to changes in the model parameters. *Operations Research* 54: 247–260.
33. Wolff, R.W. (1982). Poisson arrivals see time averages. *Operations Research* 30: 223–231.
34. Yao, D.D. & Shanthikumar, J.G. (1987). The optimal input rate to a system of manufacturing cells. *Information Systems and Operational Research* 25: 57–65.
35. Zeltyn, S. & Mandelbaum, A. (2005). Call centers with impatient customers: Many-servers asymptotics of the M/M/n+G queue. *Queueing Systems: Theory and Applications* 51: 361–402.

APPENDIX A
Proof of Property 1

In this appendix, we prove the result in Property 1. Using the notations in Eq. (4.3), the stationary probabilities for an $M/M/s/s+k+M$ queue (k extra waiting lines) given in Eqs. (3.3) and (3.4) can be rewritten as

$$p_k(i) = \phi_i \times p(0) \quad \text{for } 0 \le i \le s, \tag{A.1}$$

$$p_k(i) = \rho_i \times p(0) \quad \text{for } s < i \le s+k, \tag{A.2}$$

and

$$p_k(0) = \frac{s!\mu^s}{s!\mu^s \sum_{i=0}^{s} \phi_i + \sum_{i=s+1}^{s+k} \rho_i}. \tag{A.3}$$

Substituting them into Eq. (4.1) yields, for $k \ge 0$,

$$U_k = \left(\frac{\rho_{k+s}}{s!\mu^s \sum_{i=0}^{s} \phi_i + \sum_{i=s+1}^{k+s} \rho_i} - \frac{\rho_{k+s+1}}{s!\mu^s \sum_{i=0}^{s} \phi_i + \sum_{i=s+1}^{k+s+1} \rho_i} \right) + \frac{\gamma}{\lambda} \left(\frac{\sum_{i=s+1}^{k+s} (i-s)\rho_i}{s!\mu^s \sum_{i=0}^{s} \phi_i + \sum_{i=s+1}^{k+s} \rho_i} - \frac{\sum_{i=s+1}^{k+s+1} (i-s)\rho_i}{s!\mu^s \sum_{i=0}^{s} \phi_i + \sum_{i=s+1}^{k+s+1} \rho_i} \right); \tag{A.4}$$

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

or, equivalently, with some algebra,

$$U_k\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)=\rho_{k+s}\left(+\frac{\rho_{k+s+1}}{s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i}\right)-\rho_{k+s+1}$$
$$+\frac{\gamma}{\lambda}\left(\sum_{i=s+1}^{k+s}(i-s)\rho_i\left(1+\frac{\rho_{k+s+1}}{s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i}\right)-\sum_{i=s+1}^{k+s+1}(i-s)\rho_i\right)$$
$$=\rho_{k+s}+\frac{\rho_{k+s}\ \rho_{k+s+1}}{s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i}-\rho_{k+s+1}$$
$$+\frac{\gamma}{\lambda}\left(-(k+1)\rho_{k+s+1}+\frac{\rho_{k+s+1}}{s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i}\sum_{i=s+1}^{k+s}(i-s)\rho_i\right)$$
$$=\rho_{k+s}+\rho_{k+s+1}\left(-1+\frac{\rho_{k+s}}{s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i}\right.$$
$$\left.-\frac{(k+1)\gamma}{\lambda}+\frac{\gamma}{\lambda}\frac{\sum_{i=s+1}^{k+s}(i-s)\rho_i}{s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i}\right). \qquad \textbf{(A.5)}$$

Calculating further gives

$$U_k\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)$$
$$=\rho_{k+s}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)$$
$$+\rho_{k+s+1}\left(\rho_{k+s}-\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)\right.$$
$$\left.-\frac{(k+1)\gamma}{\lambda}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)+\frac{\gamma}{\lambda}\sum_{i=s+1}^{k+s}(i-s)\rho_i\right)$$
$$=\rho_{k+s}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)$$
$$+\rho_{k+s+1}\left(\rho_{k+s}-\left(1+\frac{(k+1)\gamma}{\lambda}\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)+\frac{\gamma}{\lambda}\sum_{i=s+1}^{k+s}(i-s)\rho_i\right). \qquad \textbf{(A.6)}$$

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

Observing that $\rho_{k+s+1} = [\lambda/(s\mu + (k+1)\gamma)]\rho_{k+s}$, for $k \geq 0$, we can write

$$\begin{aligned}
&U_k\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)\frac{1}{\rho_{k+s}}\\
&=\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i+\rho_{k+s+1}\right)-\frac{\lambda+(k+1)\gamma}{s\mu+(k+1)\gamma}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)\\
&\quad+\frac{\gamma}{s\mu+(k+1)\gamma}\sum_{i=s+1}^{k+s}(i-s)\rho_i\\
&=\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)-\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)\\
&\quad-\frac{\lambda-s\mu}{s\mu+(k+1)\gamma}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)+\frac{\gamma}{s\mu+(k+1)\gamma}\sum_{i=s+1}^{k+s}(i-s)\rho_i. \qquad \textbf{(A.7)}
\end{aligned}$$

Simplifying Eq. (A.7) implies the following relation:

$$\begin{aligned}
&U_k\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)\frac{1}{\rho_{k+s}}\\
&=\rho_{k+s+1}-\frac{\lambda-s\mu}{s\mu+(k+1)\gamma}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)\\
&\quad+\frac{\gamma}{s\mu+(k+1)\gamma}\sum_{i=s+1}^{k+s}(i-s)\rho_i. \qquad \textbf{(A.8)}
\end{aligned}$$

For the rank $k+1$, relation (A.8) becomes

$$\begin{aligned}
&U_{k+1}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+2}\rho_i\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\frac{1}{\rho_{k+s+1}}\\
&=\rho_{k+s+2}-\frac{\lambda-s\mu}{s\mu+(k+2)\gamma}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\\
&\quad+\frac{\gamma}{s\mu+(k+2)\gamma}\sum_{i=s+1}^{k+s+1}(i-s)\rho_i\\
&=\frac{\lambda}{s\mu+(k+2)\gamma}\rho_{k+s+1}\\
&\quad-\frac{\lambda-s\mu}{s\mu+(k+2)\gamma}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i+\rho_{k+s+1}\right)\\
&\quad+\frac{\gamma}{s\mu+(k+2)\gamma}\sum_{i=s+1}^{k+s}(i-s)\rho_i+\frac{(k+1)\gamma}{s\mu+(k+2)\gamma}\rho_{k+s+1}. \qquad \textbf{(A.9)}
\end{aligned}$$

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

Hence,

$$U_{k+1}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+2}\rho_i\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\frac{1}{\rho_{k+s+1}}$$
$$=\frac{s\mu+(k+1)\gamma}{s\mu+(k+2)\gamma}\rho_{k+s+1}-\frac{\lambda-s\mu}{s\mu+(k+2)\gamma}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)$$
$$+\frac{\gamma}{s\mu+(k+2)\gamma}\sum_{i=s+1}^{k+s}(i-s)\rho_i. \qquad \textbf{(A.10)}$$

Multiplying both sides in Eq. (A.10) by $(s\mu + (k + 2)\gamma)/(s\mu + (k + 1)\gamma)$ implies

$$U_{k+1}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+2}\rho_i\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\frac{1}{\rho_{k+s+1}}\quad\frac{s\mu+(k+2)\gamma}{s\mu+(k+1)\gamma}$$
$$=\rho_{k+s+1}-\frac{\lambda-s\mu}{s\mu+(k+1)\gamma}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)+\frac{\gamma}{s\mu+(k+1)\gamma}\sum_{i=s+1}^{k+s}(i-s)\rho_i. \qquad \textbf{(A.11)}$$

From Eqs. (A.8) and (A.11), we next deduce that

$$U_{k+1}\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+2}\rho_i\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\frac{1}{\rho_{k+s+1}}\quad\frac{s\mu+(k+2)\gamma}{s\mu+(k+1)\gamma}$$
$$=U_k\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+1}\rho_i\right)\left(s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i\right)\frac{1}{\rho_{k+s}}. \qquad \textbf{(A.12)}$$

Finally, simplifying Eq. (A.12) and again observing that $\rho_{k+s+1}/\rho_{k+s} = \lambda/(s\mu + (k + 1)\gamma)$, we get, for all $k \geq 0$,

$$U_{k+1}=\frac{s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s}\rho_i}{s!\mu^s\sum_{i=0}^{s}\phi_i+\sum_{i=s+1}^{k+s+2}\rho_i}\quad\frac{\lambda}{s\mu+(k+2)\gamma}U_k, \qquad \textbf{(A.13)}$$

which completes the proof of the property.

APPENDIX B
Numerical Illustrations

In this appendix, we present numerical examples to illustrate the convexity results. We compute the probability of being served as a function of the queue capacity for several systems chosen so as to cover a broad range of parameters values. Systems parameters are presented in Section 5.

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

Table B1. Values of Q_k (in %) for $\gamma = 0.5$

	$\lambda = 1.8$			$\lambda = 8$			$\lambda = 60$		
k	$s = 1$	$s = 2$	$s = 3$	$s = 5$	$s = 10$	$s = 15$	$s = 50$	$s = 70$	$s = 100$
0	35.7143	63.3484	81.9733	52.0992	87.8339	99.0899	78.3881	97.6256	99.9999
1	44.3548	73.3209	89.1589	56.3701	90.9867	99.5033	79.3985	98.0027	100.0000
2	47.5087	77.5281	92.0093	58.5389	92.9632	99.7085	80.1610	98.3100	100.0000
3	48.8651	79.3467	93.0916	59.7405	94.2080	99.8076	80.7453	98.5601	100.0000
4	49.4796	80.1024	93.4736	60.4509	94.9854	99.8542	81.1989	98.7634	100.0000
5	49.7536	80.3947	93.5977	60.8924	95.4627	99.8755	81.5551	98.9283	100.0000
6	49.8691	80.4986	93.6349	61.1773	95.7490	99.8850	81.8377	99.0615	100.0000
7	49.9140	80.5324	93.6452	61.3664	95.9160	99.8890	82.0639	99.1688	100.0000
8	49.9299	80.5425	93.6478	61.4942	96.0106	99.8908	82.2466	99.2548	100.0000
9	49.9352	80.5454	93.6485	61.5815	96.0625	99.8915	82.3952	99.3235	100.0000
10	49.9367	80.5461	93.6486	61.6411	96.0901	99.8918	82.5169	99.3780	100.0000
11	49.9372	80.5463	93.6487	61.6815	96.1044	99.8919	82.6174	99.4211	100.0000
12	49.9373	80.5463	93.6487	61.7086	96.1115	99.8919	82.7008	99.4550	100.0000
13	49.9373	80.5463	93.6487	61.7264	96.1149	99.8919	82.7704	99.4815	100.0000
14	49.9373	80.5463	93.6487	61.7379	96.1165	99.8919	82.8289	99.5021	100.0000
15	49.9373	80.5463	93.6487	61.7450	96.1173	99.8919	82.8782	99.5180	100.0000
16	49.9373	80.5463	93.6487	61.7494	96.1176	99.8919	82.9200	99.5302	100.0000
17	49.9373	80.5463	93.6487	61.7519	96.1178	99.8919	82.9556	99.5395	100.0000
18	49.9373	80.5463	93.6487	61.7534	96.1178	99.8919	82.9860	99.5466	100.0000
19	49.9373	80.5463	93.6487	61.7542	96.1178	99.8919	83.0121	99.5519	100.0000
20	49.9373	80.5463	93.6487	61.7546	96.1179	99.8919	83.0345	99.5559	100.0000
21	49.9373	80.5463	93.6487	61.7548	96.1179	99.8919	83.0539	99.5589	100.0000
22	49.9373	80.5463	93.6487	61.7549	96.1179	99.8919	83.0706	99.5611	100.0000
23	49.9373	80.5463	93.6487	61.7550	96.1179	99.8919	83.0851	99.5627	100.0000
24	49.9373	80.5463	93.6487	61.7550	96.1179	99.8919	83.0977	99.5639	100.0000
25	49.9373	80.5463	93.6487	61.7550	96.1179	99.8919	83.1087	99.5648	100.0000
26	49.9373	80.5463	93.6487	61.7550	96.1179	99.8919	83.1182	99.5654	100.0000
27	49.9373	80.5463	93.6487	61.7550	96.1179	99.8919	83.1265	99.5658	100.0000
28	49.9373	80.5463	93.6487	61.7550	96.1179	99.8919	83.1337	99.5662	100.0000
29	49.9373	80.5463	93.6487	61.7550	96.1179	99.8919	83.1399	99.5664	100.0000
30	49.9373	80.5463	93.6487	61.7550	96.1179	99.8919	83.1454	99.5665	100.0000

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

TABLE B2. Values of Q_k (in %) for $\gamma = 1$

	$\lambda = 1.8$			$\lambda = 8$			$\lambda = 60$		
k	$s = 1$	$s = 2$	$s = 3$	$s = 5$	$s = 10$	$s = 15$	$s = 50$	$s = 70$	$s = 100$
0	35.7143	63.3484	81.9733	52.0992	87.8339	99.0899	78.3881	97.6256	99.9999
1	42.9864	71.9585	88.3281	56.1529	90.8551	99.4904	79.3906	98.0001	100.0000
2	45.2522	74.8962	90.3886	58.1089	92.6103	99.6776	80.1380	98.3012	100.0000
3	46.0253	75.8487	90.9857	59.1433	93.6035	99.7606	80.7026	98.5414	100.0000
4	46.2760	76.1248	91.1378	59.7245	94.1442	99.7954	81.1343	98.7317	100.0000
5	46.3486	76.1951	91.1719	60.0621	94.4250	99.8094	81.4682	98.8810	100.0000
6	46.3671	76.2108	91.1787	60.2603	94.5635	99.8147	81.7291	98.9972	100.0000
7	46.3713	76.2140	91.1799	60.3754	94.6283	99.8166	81.9351	99.0867	100.0000
8	46.3721	76.2146	91.1801	60.4406	94.6570	99.8173	82.0990	99.1549	100.0000
9	46.3723	76.2147	91.1802	60.4761	94.6690	99.8175	82.2305	99.2063	100.0000
10	46.3723	76.2147	91.1802	60.4945	94.6739	99.8176	82.3367	99.2447	100.0000
11	46.3723	76.2147	91.1802	60.5036	94.6757	99.8176	82.4229	99.2730	100.0000
12	46.3723	76.2147	91.1802	60.5078	94.6764	99.8176	82.4932	99.2937	100.0000
13	46.3723	76.2147	91.1802	60.5097	94.6766	99.8176	82.5508	99.3086	100.0000
14	46.3723	76.2147	91.1802	60.5105	94.6767	99.8176	82.5980	99.3192	100.0000
15	46.3723	76.2147	91.1802	60.5108	94.6767	99.8176	82.6368	99.3267	100.0000
16	46.3723	76.2147	91.1802	60.5109	94.6767	99.8176	82.6687	99.3320	100.0000
17	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.6949	99.3356	100.0000
18	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7163	99.3380	100.0000
19	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7338	99.3397	100.0000
20	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7481	99.3408	100.0000
21	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7596	99.3415	100.0000
22	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7688	99.3420	100.0000
23	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7762	99.3423	100.0000
24	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7821	99.3425	100.0000
25	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7867	99.3426	100.0000
26	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7902	99.3427	100.0000
27	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7930	99.3427	100.0000
28	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7951	99.3428	100.0000
29	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7966	99.3428	100.0000
30	46.3723	76.2147	91.1802	60.5110	94.6767	99.8176	82.7978	99.3428	100.0000

https://doi.org/10.1017/S0269964807000010 Published online by Cambridge University Press

Table B3. Values of Q_k (in %) for $\gamma = 2$

	$\lambda = 1.8$			$\lambda = 8$			$\lambda = 60$		
k	$s = 1$	$s = 2$	$s = 3$	$s = 5$	$s = 10$	$s = 15$	$s = 50$	$s = 70$	$s = 100$
0	35.7143	63.3484	81.9733	52.0992	87.8339	99.0899	78.3881	97.6256	99.9999
1	41.2371	70.1107	87.1346	55.7787	90.6222	99.4669	79.3752	97.9950	100.0000
2	42.5412	71.7812	88.3617	57.3866	92.0354	99.6247	80.0936	98.2842	100.0000
3	42.8391	72.1385	88.6026	58.1445	92.6993	99.6847	80.6211	98.5063	100.0000
4	42.8970	72.2021	88.6419	58.5085	92.9858	99.7055	81.0121	98.6735	100.0000
5	42.9064	72.2116	88.6473	58.6788	93.0991	99.7122	81.3045	98.7969	100.0000
6	42.9077	72.2129	88.6480	58.7540	93.1400	99.7142	81.5250	98.8861	100.0000
7	42.9079	72.2130	88.6480	58.7848	93.1537	99.7147	81.6922	98.9493	100.0000
8	42.9079	72.2130	88.6480	58.7964	93.1578	99.7148	81.8195	98.9931	100.0000
9	42.9079	72.2130	88.6480	58.8004	93.1590	99.7149	81.9164	99.0228	100.0000
10	42.9079	72.2130	88.6480	58.8017	93.1594	99.7149	81.9901	99.0426	100.0000
11	42.9079	72.2130	88.6480	58.8021	93.1594	99.7149	82.0460	99.0555	100.0000
12	42.9079	72.2130	88.6480	58.8022	93.1595	99.7149	82.0879	99.0637	100.0000
13	42.9079	72.2130	88.6480	58.8022	93.1595	99.7149	82.1192	99.0688	100.0000
14	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1422	99.0720	100.0000
15	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1588	99.0739	100.0000
16	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1708	99.0750	100.0000
17	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1791	99.0756	100.0000
18	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1849	99.0760	100.0000
19	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1888	99.0762	100.0000
20	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1914	99.0763	100.0000
21	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1930	99.0763	100.0000
22	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1941	99.0764	100.0000
23	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1948	99.0764	100.0000
24	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1952	99.0764	100.0000
25	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1954	99.0764	100.0000
26	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1956	99.0764	100.0000
27	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1956	99.0764	100.0000
28	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1957	99.0764	100.0000
29	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1957	99.0764	100.0000
30	42.9079	72.2130	88.6480	58.8023	93.1595	99.7149	82.1957	99.0764	100.0000