Probability in the Engineering and Informational Sciences, **27**, 2013, 99–114.
doi:10.1017/S0269964812000320

CONDITIONAL SOJOURN TIMES OF PROCESSOR-SHARING QUEUES

WEI-YI LEE AND CHIA-LI WANG
Department of Applied Mathematics
National Dong Hwa University, Hualien, Taiwan, ROC
E-mail: cwang@mail.ndhu.edu.tw

Queues operated by a processor-sharing mode have important applications in many modern systems. However, because of the simultaneous sharing of service capacity by all customers, the distribution function and moments of the sojourn time are difficult to derive, even with a given initial condition. In addition, when a limit on the number of customers in the system is enforced to ensure the quality of service, the sojourn time becomes more complicated. In recent literature, the distribution function is obtained via the Laplace–Stieltjes transform. In this paper, we take a pure algebraic approach to derive the moments of the sojourn time. We obtain an iterative formula and use it to investigate properties of the conditional sojourn time. The approach is simple and intuitive, and applies to queues with multiple class customers as well.

1. INTRODUCTION

In the classical paper of queueing control, Naor [9] considered a single-server system with Poisson arrivals, exponential service times, and a finite system capacity limit, the $M/M/1/N$ queue. He showed that the self-optimization and the social-optimization of a given utility function for the system under first-in-first-out (FIFO) service discipline are in general different. He demonstrated the difference by explicitly calculating the expected conditional waiting time of an entering customer given the number of customers in the system upon arrival.

When one wants to apply Naor's analysis to a modern queueing model, the *processor-sharing* (PS) queue, he may have found no handy formula for the corresponding distribution or expectation as those of the FIFO queue. Indeed, despite the great demand, there was no such closed form in the literature. The difficulty in deriving the distribution function of the conditional sojourn time is because it depends on not only the present number of customers in system, but also the future arrivals.

The earliest work on this problem, to our best knowledge, is by Coffman, Muntz Jr., and Trotter [5]; in which they obtain the conditional sojourn time distribution on both the number seen by the arrival and his service time. Their result is useful from the customer's point of view.

On the other hand, customers are usually assumed to be i.i.d. by the system administrator. So, the conditional sojourn time distribution that only depends on the number seen is

 0269-9648/13 $25.00

more practical for system management. But, it is difficult to obtain directly from [5] a simple expression for the moment of the conditional sojourn time for only the number in the system. Thus, Sengupta and Jagerman [10] took a different approach to get the rth moment of the conditional sojourn time as a polynomial of degree r, and the Laplace–Stieltjes transform of the distribution.

To be specific, they considered an $M/M/1/\infty$ PS queue with arrival rate λ and service rate μ. Define the traffic intensity $\rho = \lambda/\mu$, and let $W(n)$ denote the conditional sojourn time of an entering (tagged) customer who sees $n-1$ customers upon his arrival. They obtained the first moment as

$$E^{\mathrm{PS}}[W(n)] = \frac{n+1}{\mu(2-\rho)}. \tag{1}$$

Clearly, (1) is a linear function of n, whereas for the same queue under FIFO, $E^{\mathrm{FIFO}}[W(n)] = n/\mu$, is also linear in n. Moreover, $E^{\mathrm{FIFO}}[W(n)]$ is finite no matter whether the system is stable or not, yet $E^{\mathrm{PS}}[W(n)]$, depends on λ, is shown (later by [11]) to be finite iff $\rho < 2$.

As mentioned in [10], it is interesting to note that

$$E^{\mathrm{PS}}[W(n)] < E^{\mathrm{FIFO}}[W(n)] \text{ iff } n-1 > \frac{\rho}{1-\rho},$$

where $\rho/(1-\rho)$ is the expected number of customers in system under either FIFO or PS. That is to say, when the number of customers seen by an arrival is greater than the mean stationary number in system, joining the PS queue has a smaller mean conditional sojourn time than the FIFO queue. This is true for $\rho < 1$.

The recent progress on this topic made by Zhen and Knessl [11], where a new formula, as well as an approximation, for the conditional sojourn time distribution is obtained. It then derives various asymptotic limits as n and/or ρ are large. Also, an up-to-date reference of this topic can be found therein.

In many applications, the system administration will enforce a limit on the number of customers in system to ensure the quality of service as discussed in [2], or potential customers will set up a stop-loss threshold for his utility as assumed in [9]. When the number reaches the limit, arrivals will be rejected or leave without service. While the limit lets the conditional sojourn time to become stochastically smaller, it makes the system more complicated to analyze.

The first results of the PS queue with finite system limit have appeared very recently. Borst, Boxma, and Hegde [2] obtain a Laplace–Stieltjes transform of the distribution of the conditional sojourn time, denoted it as $W_{\mathrm{N}}(n)$, for a $M/M/1/N$ PS queue with state-dependent service rates.

Having a finite number of recursive equations, they use a matrix representation for the transforms, and then invert the transform to get a phase-type distribution of $W_{\mathrm{N}}(n)$ as

$$P\{W_{\mathrm{N}}(n) > t\} = \sum_{j=1}^{N} \frac{A_{n,j}}{-\omega_j} e^{\omega_j t}, \tag{2}$$

where $\omega_j, j = 1, \ldots, N$, are the N roots of the determinant of matrix $\boldsymbol{M}$ that is, composed of coefficients of a set of linear equations, and $A_{n,j}$ are the residues of the partial fraction

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

expansion of

$$\psi_n(\omega) = \frac{\det M_i}{\det M}$$

for the given roots. The matrix M_i is equal to M with the nth column replaced by a particular vector.

To reduce the computational effort involved in finding the appropriate roots, [2] also provides two approximations for the distribution.

In a sequel, Boxma, Hegde, and Nunez-Queija [3] extends the approach to multi-class *discriminatory* processor-sharing (DPS) queue, that is, customers of distinct classes have different service rates.

To compute the moments of $W_{\mathrm{N}}(n)$ by (2), the only available closed form in the literature, one has to find the roots of the determinant of an $N \times N$ matrix, and deals with possible round-off errors. Hence, the need of accessible closed forms for the moments remains to be met which motivates our present study.

To that end, we start with the same conditional argument as [2], [3], and [10] that is, both intuitive and simple, then take a different approach from there. We show that the recursive nature of the conditional argument leads to an iterative formula for moments of $W_{\mathrm{N}}(n)$.

This formula not only unifies the expression for the moments of the conditional sojourn time for standard PS queue, PS queue with state-dependent service rate considered in [2] and multi-class DPS queue in [3], it also takes less computational effort in deriving moments than from the matrix-form distribution. With the formula, we further obtain an approximation of $E[W_{\mathrm{N}}(n)]$ for the standard PS queue complementary to those proposed in [2] and [3].

The conditional sojourn time on the service time of the tagged customer for the PS queue has a relatively earlier development. For the $M/M/1$ PS queue, Kleinrock [7] obtained the nice and simple closed form for the conditional expectation. Within a short time, Coffman, Muntz Jr., and Trotter [5] extended the results to conditioning on both the service time and the number in system. It is worth mentioning that a corresponding result with a finite system limit has not yet been obtained.

The structure of this paper is as follows. In Section 2, we consider the finite-capacity PS queue with homogeneous customers and state-dependent service rate. We derive an iterative formula for any moment of the conditional sojourn time, and establish the increase and concavity of the first moment both in the number seen and in the system limit by the coupling argument. In Section 3, we consider the service rate being state independent, and simplify the formula for $E[W_{\mathrm{N}}(n)]$. By the formula, we recover (1) and show the necessary and sufficient condition of convergence. The monotonic property helps us to obtain simple approximations of $E[W_{\mathrm{N}}(n)]$. The quality of the approximation is demonstrated by numerical experiments. In the final section, we extend the approach to queues with multiple-class customers, where each class has its own arrival rate, service rate, and system limit.

2. AN ITERATIVE FORMULA

Consider an $M/M/1/N$ PS queue with homogeneous customers, arrival rate λ, and state-dependent service rate μ_i when there are i customers in the system.

A special case of state-dependent service rate is for $\mu_i = \mu \times \min\{i, c\}$, which is the service rate of an $M/M/c/N$ PS queue. It operates as FIFO when $i \leq c$ and allocates full service capacity equally to all present customers in the system when $i > c$.

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

Let $W_{\mathrm{N}}(n)$ denote the conditional sojourn time of an entering (tagged) customer who sees $n-1<N$ customers in the queue upon arrival. In the associated queueing process, there are three events that will trigger the change of state, which are an arrival joining the queue, the departure by other customers, and by the tagged customer. By conditioning on the next occurring event, we have the following recursive relation for $1 \le n \le N-1$:

$$W_{\mathrm{N}}(n) = \begin{cases} \exp(\lambda+\mu_n) + W_N(n+1), & \text{w.p.}\lambda/(\lambda+\mu_n); \\ \exp(\lambda+\mu_n) + W_N(n-1), & \text{w.p. } (n-1)\mu_n/n(\lambda+\mu_n); \\ \exp(\lambda+\mu_n), & \text{w.p. } \mu_n/n(\lambda+\mu_n), \end{cases}$$

where "exp(a)" denotes an exponential random variable with rate a and "w.p." stands for "with probability", and for $n=N$:

$$W_N(N) = \begin{cases} \exp(\mu_N), & \text{w.p. } 1/N; \\ \exp(\mu_N) + W_N(N-1), & \text{w.p. } (N-1)/N. \end{cases}$$

With the memoryless property of the exponential distribution, $W_N(n)$ can be represented as a convolution of a random number of independent exponential random variables. That will lead to an expression of the distribution function in an exponential-matrix form as derived in [2]. We take a different approach from here to deriving formulae for the moments of $W_N(n)$.

We begin by showing a recursive relation of exponential moments. Let X be an exponential random variable with rate a, independent of another random variable Y. For any $m \ge 1$,

$$\begin{aligned} E[(X+Y)^m] &= \sum_{n=0}^{m} \binom{m}{n} E(X^n)E(Y^{m-n}) \\ &= \sum_{n=0}^{m-1} \binom{m}{n+1} E(X^{n+1})E(Y^{m-1-n}) + E(Y^m) \\ &= \frac{m}{a} \sum_{n=0}^{m-1} \binom{m-1}{n} E(X^n)E(Y^{m-1-n}) + E(Y^m) \\ &= \frac{m}{a} E[(X+Y)^{m-1}] + E(Y^m), \end{aligned} \tag{3}$$

where the third equality is due to the fact that $E(X^{n+1}) = (n+1)E(X^n)/a$.

Now, let $E[W_N^m(n)]$ be the mth moment of the conditional sojourn time, $m \ge 1$. By conditioning on the next event occurred, we obtain from (3) that

$$\begin{aligned} E[W_N^m(n)] = \frac{m}{\lambda+\mu_n} E[W_N^{m-1}(n)] + \frac{\lambda}{\lambda+\mu_n} E[W_N^m(n+1)] \\ + \frac{n-1}{n} \frac{\mu_n}{\lambda+\mu_n} E[W_N^m(n-1)] \end{aligned} \tag{4}$$

for $1 \le n \le N-1$, and

$$E[W_N^m(N)] = \frac{m}{\mu_N} E[W_N^{m-1}(N)] + \frac{N-1}{N} E[W_N^m(N-1)]. \tag{5}$$

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

Let $\rho_i = \lambda/\mu_i$, $a_1 = \rho_1/(1+\rho_1)$, $b_{m,1} = a_1 E[W_N^{m-1}(1)]/(\mu_1\rho_1)$, and

$$a_n = \frac{\rho_n}{1+\rho_n-(n-1)a_{n-1}/n}, b_{m,n} = \frac{a_n}{\rho_n}\left[\frac{E[W_N^{m-1}(n)]}{\mu_n} + \frac{n-1}{n} b_{m,n-1}\right].$$

for $2 \le n \le N-1$. We have a recursive formula below which is useful in proving the main result.

LEMMA 1: *For* $1 \le n \le N-1$ *and* $m \ge 1$,

$$E[W_N^m(n)] = m b_{m,n} + a_n E[W_N^m(n+1)]. \tag{6}$$

PROOF: The proof is by induction. For $n = 1$, we have from (4),

$$\begin{aligned} E[W_N^m(1)] &= \frac{m}{\mu_1(1+\rho_1)} E[W_N^{m-1}(1)] + \frac{\rho_1}{1+\rho_1} E[W_N^m(2)] \\ &= m b_{m,1} + a_1 E[W_N^m(2)]. \end{aligned}$$

Assume that it holds for $n = k-1 \ge 1$. For $n = k$, it follows from (4) and the induction step that

$$\begin{aligned} E[W_N^m(k)] &= \frac{m}{1+\rho_k}\left[\frac{E[W_N^{m-1}(k)]}{\mu_k} + \frac{k-1}{k} b_{m,k-1}\right] + \frac{\rho_k}{1+\rho_k} E[W_N^m(k+1)] \\ &\quad + \frac{k-1}{k}\frac{1}{1+\rho_k} a_{k-1} E[W_N^m(k)] \\ &= m b_{m,k} + a_k E[W_N^m(k+1)]. \end{aligned}$$

■

By further letting

$$a_N = \frac{\rho_N}{1-(N-1)a_{N-1}/N}, \quad b_{m,N} = \frac{a_N}{\rho_N}\left[\frac{E[W_N^{m-1}(N)]}{\mu_N} + \frac{N-1}{N} b_{m,N-1}\right],$$

and $\prod_{i=n}^{n-1} a_i = 1$ for any $n \ge 0$, we have the main result below:

THEOREM 1: *For* $1 \le n \le N$ *and* $1 \le m$,

$$E[W_N^m(n)] = m \sum_{j=n}^{N} b_{m,j} \prod_{k=n}^{j-1} a_k. \tag{7}$$

PROOF: It is easy to see that (7) holds when $N = 1$. For $N > 1$ and $n = N$, substituting (6) into (5) yields

$$\begin{aligned} E[W_N^m(N)] &= m\left[\frac{E[W_N^{m-1}(N)]}{\mu_N} + \frac{N-1}{N} b_{m,N-1}\right] + \frac{N-1}{N} a_{N-1} E[W_N^m(N)] \\ &= m b_{m,N}. \end{aligned}$$

We then substitute the above back into (6) to get $E[W_N^m(N-1)]$. Repeating the substitution, we get

$$E[W_N^m(n)] = m\left(b_{m,n} + b_{m,n+1}a_n + \cdots + b_{m,N}\prod_{j=n}^{N-1} a_j\right),$$

for any $n \in \{1, \ldots, N-1\}$, and the desired result. ■

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

We note that to compute $E[W_N^m(n)]$ by (7) one needs to first compute $E[W_N^k(i)]$ for $k = 1, \ldots, m-1$ and $i = 1, \ldots, N$, the price to pay for deriving higher moments without using the distribution. In particular, the computing procedure is:

1. Compute recursive sequences $\{a_i\}$ and $\{b_{1,i}\}$.
2. Input $\{a_i\}$ and $\{b_{1,i}\}$ into (7) to get $\{E[W_N(i)]\}$.
3. If $m > 1$, derive $\{b_{2,i}\}$ via $\{a_i\}$ and $\{E[W_N(i)]\}$.
4. Input $\{a_i\}$ and $\{b_{2,i}\}$ into (7) again to get $\{E[W_N^2(i)]\}$, and so on, until $E[W_N^m(n)]$ is obtained.

Remark: While (2) is the distribution of W_n and the only available closed form in the literature, if one resorts to it for deriving moments, he has to find the roots of the determinant of an $N \times N$ matrix, and deals with possible round-off errors. Alternatively, the calculation needed by (7) is for $2m$ recursive sequences, namely, $\{a_i\}, \{b_{k,i} : k = 1, \ldots, m\}$ and $\{E[W_N^k(i)] : k = 1, \ldots, m-1\}$, with complexity $O(mN)$ in contrast to $O(N^3)$ for finding the roots of the determinant.

2.1. Monotonicity and Concavity

In this subsection, we will study stochastic properties of $W_N(n)$ via the coupling argument. These properties are useful when one concerns the queueing control as in [9].

For a sequence of positive numbers $\{c_n, n \geq 1\}$, we say that it is *proportional decreasing and convex* if for all $n \geq 1$, $nc_{n+1} - (n+1)c_n$ is negative and increasing in n.

The assumption of proportional decreasing and convex on μ_n is not restrictive and is satisfied by most practical systems. For examples, $\mu_n = \mu, \mu_n = n\mu$ and $\mu_n = \mu \times \min\{i, c\}$ all meet the condition.

Consider a $G/M/1/N$ PS queue with state-dependent service rate μ_n that is, proportional decreasing and convex. Suppose that the tagged customer joins the queue at time 0 and finds n in system, $n \geq 1$. Because the exponential service time is Markovian, the number in system determines the state of the system at any time.

We take $W_N(n)$ as a function of n and let $N_n(t)$ denote the number of customers in system at time t, starting with n at time 0. Then, the tagged customer receives service at rate $\mu_{N_{n+1}(t)}/N_{n+1}(t)$ at t, provided $N_{n+1}(t) > 0$, and has sojourn time $W_N(n+1)$.

Now, suppose that we randomly choose a customer among the n seen by the tagged customer and freeze his service from time 0 until the tagged customer departs. With the freezing, the service rate received by the tagged customer changes to $\mu_{N_n(t)}/N_n(t)$ at t, and the sojourn time becomes $W_N(n)$.

Let $T_1 = \inf_{t>0}\{t : N_n(t) = N_{n+1}(t)\}$. At T_1, we let remaining service times associated to $N_n(T_1)$ be the same as those of $N_{n+1}(T_1)$ such that $N_n(t)$ and $N_{n+1}(t)$ couple from T_1. Let

$$N_n'(t) = \begin{cases} N_n(t), & t < T_1; \\ N_{n+1}(t), & t \geq T_1. \end{cases}$$

Then, $N_{n+1}(t) \geq N_n'(t)$ for all t a.s., and, because $n\mu_{n+1} \leq (n+1)\mu_n$,

$$\frac{\mu_{N_n'(t)}}{N_n'(t)} \geq \frac{\mu_{N_{n+1}(t)}}{N_{n+1}(t)}.$$

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

Consequently, for

$$W'_N(n) = \begin{cases} W_N(n), & \text{if } \min\{W_N(n+1), W_N(n)\} < T_1; \\ W_N(n+1), & \text{otherwise,} \end{cases}$$

$W'_N(n) \le W_N(n+1)$ for all t a.s. Combining this inequality with the facts of $N'_n(t) \stackrel{st}{=} N_n(t)$ and $W'_N(n) \stackrel{st}{=} W_N(n)$, we have shown that $W_N(n)$ is stochastically increasing in n.

Next, for $n \ge 2$, suppose that we freeze one more customer among the n at time 0 so that the service rate becomes $\mu_{N_{n-1}(t)}/N_{n-1}(t)$. Similarly, let $T_2 = \inf_{t>0}\{t : N_{n-1}(t) = N'_n(t)\}$ and couple $N_{n-1}(t)$ and $N'_n(t)$ at time T_2 like before. Define

$$N'_{n-1}(t) = \begin{cases} N_{n-1}(t), & t < T_2; \\ N'_n(t), & t \ge T_2, \end{cases}$$

and corresponding copy $W'_N(n-1)$ of $W_N(n-1)$. We get

$$N_{n+1}(t) \ge N'_n(t) \ge N'_{n-1}(t) \text{ for all } t \text{ a.s.},$$

and, with $n\mu_{n+1} - (n+1)\mu_n$ increasing in n,

$$\frac{\mu_{N'_n(t)}}{N'_n(t)} - \frac{\mu_{N_{n+1}(t)}}{N_{n+1}(t)} \le \frac{\mu_{N'_{n-1}(t)}}{N'_{n-1}(t)} - \frac{\mu_{N'_n(t)}}{N'_n(t)}.$$

Therefore, $W_N(n+1) - W'_N(n) \le W'_N(n) - W'_N(n-1)$, which implies

$$W_N(n+1) - W_N(n) \le_{st} W_N(n) - W_N(n-1),$$

that is, $E[W_N(n)]$ is concave in n.

To conclude, we have shown:

THEOREM 2: *Consider the $G/M/1/N$ PS queue with state-dependent service rate μ_n. If μ_n is proportional decreasing and convex, then $W_N(n)$ is stochastically increasing and $E[W_N(n)]$ is concave in n for all ρ.*

We now study the property of $W_N(n)$ in N.

Let $L_N(t)$ denote the number of customers in the system at time t with limit N such that the tagged customer receives service at rate $\mu_{L_N(t)}/L_N(t)$ at t. Note that the system limit will affect W_n only if there are rejected arrivals due to the limit in the sojourn of the tagged customer.

Suppose $N \ge 2$ and we do not serve any customer who enters at $t > 0$ with $L_N(t^-) = N - 1$ when the tagged customer is in system. As a consequence, the service rate received by the tagged customer changes to $\mu_{L_{N-1}(t)}/L_{N-1}(t)$ at time t.

By the same coupling argument, we have $L_N(t) \ge L'_{N-1}(t)$, which, together with $n\mu_{n+1} \le (n+1)\mu_n$, imply $W'_{N-1}(n) \le W_N(n)$ for all t a.s. Because $W'_{N-1}(n) \stackrel{st}{=} W_{N-1}(n)$, $W_N(n)$ increases stochastically in N.

Furthermore, for $N \ge 3$, if any customer who enters at $t > 0$ with $L_N(t^-) = N - 2$ will neither be served, then by the coupling

$$L_N(t) \ge L'_{N-1}(t) \ge L'_{N-2}(t)$$

for all t a.s. This order and $n\mu_{n+1} - (n+1)\mu_n$ increasing in n yield $W_N(t) - W'_{N-1}(t) \le W'_{N-1}(t) - W'_{N-2}(t)$, and the concavity of $E[W_N(n)]$ in N follows.

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

THEOREM 3: *Consider the $GI/M/1/N$ PS queue with state-dependent service rate μ_n. If μ_n is proportional decreasing and convex, then $W_N(n)$ is stochastically increasing and $E[W_N(n)]$ is concave in N for all ρ.*

3. STATE-INDEPENDENT SERVICE RATE

When the service rate is state independent, that is, $\mu_i = \mu$ for all $i = 1, \ldots, N$, (7) can be substantially simplified. In this section, we will demonstrate that for $E[W_N(n)]$, and discuss its properties.

First of all, with

$$\begin{aligned} a_n &= \frac{\rho}{1+\rho-(n-1)a_{n-1}/n}, \\ b_{1,n} &= \frac{a_n}{\rho}\left[\frac{1}{\mu}+\frac{n-1}{n}b_{1,n-1}\right] = \frac{1}{n\mu}\,n\frac{a_n}{\rho}+\frac{1}{n}(n-1)\frac{a_n}{\rho}b_{1,n-1} \\ &\vdots \\ &= \frac{1}{n\mu}\sum_{k=1}^{n} k\prod_{i=k}^{n}\frac{a_i}{\rho}. \end{aligned} \tag{8}$$

Alternatively, we can rewrite $b_{1,1}$ as $(2-3a_1)/\mu(2-\rho)$, and, inductively,

$$b_{1,n} = \frac{n+1-(n+2)a_n}{\mu(2-\rho)}, 2 \le n \le N-1, \text{ and } b_{1,N} = \frac{N+1-a_N}{\mu(2-\rho)}. \tag{9}$$

Then, for $1 \le n \le N-1$, (6) becomes

$$E[W_N(n)] = \frac{n+1-(n+2)a_n}{\mu(2-\rho)} + a_n E[W_N(n+1)]. \tag{10}$$

We note that, from (8), $2-\rho$ in the denominator will not cause any problem to (10), nor to the second main result below.

THEOREM 4: *For all $N \ge 1$ and $1 \le n \le N$,*

$$E[W_N(n)] = \frac{1}{\mu(2-\rho)}\left(n+1-\prod_{i=n}^{N} a_i\right). \tag{11}$$

PROOF: It is easy to see that (11) holds when $N = 1$. For $N > 1$ and $n = N$, substituting (10) into (5) yields

$$\begin{aligned} E[W_N(N)] &= \frac{1}{\mu}\left(1+\frac{(N-1)[N-(N+1)a_{N-1}]}{N(2-\rho)}\right)+\frac{N-1}{N}\,a_{N-1}\,E[W_N(N)] \\ &= \frac{1}{\mu(2-\rho)}\,\frac{N+1-(N+1)(N-1)a_{N-1}/N-\rho}{1-(N-1)a_{N-1}/N} \\ &= \frac{1}{\mu(2-\rho)}\,[N+1-a_N]. \end{aligned}$$

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

We then substitute the above back into (10) to get $E[W_N(N-1)]$. Repeating the substitution, we obtain

$$\begin{aligned} E[W_N(n)] &= \frac{1}{\mu(2-\rho)}\{[n+1-(n+2)a_n] + a_n[n+2-(n+3)a_{n+1}] + \cdots \\ &\quad + a_n \cdots a_{N-1}[N+1-a_N]\} \\ &= \frac{1}{\mu(2-\rho)}\left[\sum_{j=n}^{N-1}[j+1-(j+2)a_j]\prod_{i=n}^{j-1} a_i + (N+1-a_N)\prod_{i=n}^{N-1} a_i\right]. \end{aligned}$$

Consequently, the desired result follows by rearranging the terms. ■

The simple expression of (11) makes it computationally efficient: to get the first moment of $W_N(n)$ one only needs to compute the recursive sequence of $\{a_n\}$.

Furthermore, its resemblance to (1), like the mean waiting time of an $M/M/1/k$ FIFO queue to that of one without the capacity limit, allows us to gain insights into its asymptotic property. In particular, by letting $N \to \infty$, we can recover (1) and its stability condition via (11) with the following result (the proof is in the Appendix):

LEMMA 2: *For the queue with state-independent service rate,*

$$\lim_{N\to\infty} \prod_{i=n}^{N} a_i = 0 \text{ iff } \rho < 2.$$

Remark: The increasing property in Theorem 3 can also be shown by (A.1): For $\rho \le 2$, $n+1-\prod_{i=n}^{N}$ is positive and increasing in N; for $\rho > 2$, $n+1-\prod_{i=n}^{N}$ is negative and decreasing in N. So, we can see from (11) that $E[W_N(n)]$ is increasing in N.

3.1. Bounds and Approximation

Although (11) is more computationally accessible, it still involves an iterative formula and its calculation relies on a programmed procedure. Hence, it would be nice to have a good approximation of $E[W_N(n)]$ that can be computed quickly by hand or, at most, a calculator. For that purpose, we start with the monotonicity and bounds of $\{a_n\}$ (the proof is in the Appendix).

LEMMA 3: *The sequence $\{a_n\}$ is strictly increasing in n. In addition,*

$$a_n \le \frac{2\rho}{1+\rho+\sqrt{(1-\rho)^2+4\rho/n}}, 1 \le n \le N-1, \text{ and}$$

$$a_N \le \rho\left(1-\frac{N-1}{N}\frac{2\rho}{1+\rho+\sqrt{(1-\rho)^2+4\rho/(N-1)}}\right)^{-1}.$$

Denote the bound of a_N by a_N^*. We substitute it in (11) to form the approximation

$$E[W_N(N)] \approx \frac{1}{\mu(2-\rho)}(n+1-a_N^*). \tag{12}$$

The advantage of having (12) is that one does not need to compute from a_1 all the way to a_n just for an estimation of $E[W_N(N)]$. Besides, it is very accurate. Table 1 shows the relative error of (12) to the exact value.

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

TABLE 1. Absolute Relative Errors of (12) for $\mu = 1$

N	$\rho = 0.3$	$\rho = 0.7$	$\rho = 1.5$	$\rho = 2$	$\rho = 10$
3	2.17E − 4	3.63E − 3	4.22E − 2	3.53E − 2	3.94E − 3
10	2.30E − 5	2.11E − 3	5.61E − 2	4.24E − 2	1.32E − 3
50	2.51E − 7	1.08E − 4	2.10E − 2	2.25E − 2	2.53E − 4
100	3.25E − 8	1.94E − 5	1.17E − 2	1.29E − 2	1.26E − 4

TABLE 2. Absolute Relative Errors of (13) for $N = 30$ and $\mu = 1$

n	$\rho = 0.3$	$\rho = 0.7$	$\rho = 1.5$	$\rho = 3$	$\rho = 10$
5	0.000	0.000	0.094	0.392	0.100
15	0.000	0.000	0.018	0.013	0.005
25	0.000	0.009	0.038	0.012	0.000

Clearly, $E[W(n)]$ in (1) is an upper bound of $E[W_N(n)]$ for $\rho \le 2$. It is very tight when $\rho \le 1$, and can also serve as an approximation of $E[W_N(n)]$.

Another upper bound of $E[W_N(N)]$ that is, both simple and tight for $\rho > 2$ is N/μ. To perceive it, let $W_N^*(N)$ denote the sojourn time in the queue with the infinite arrival rate. Then, it is not hard to see that

$$W_N(N) \le_{st} W_N^*(N) = \sum_{i=1}^{M} E_i,$$

where random variables $M \sim$ Geo$(1/N)$ and E_i's are i.i.d. and $\sim \exp(\mu)$. Respective expectations are ordered in the same direction.

For $\rho > 1$, we use the increasing property of $\{a_n\}$ and approximate

$$\prod_{i=n}^{N-1} a_i \approx \left(\frac{\sum_{i=n}^{N-1} a_i}{N-n} \right)^{N-n} \approx \left(\frac{a_n + a_{N-1}}{2} \right)^{N-n}.$$

Substituting the above into (11), we obtain the following approximation for $\rho > 1$.

LEMMA 4: *For* $1 \le n \le N-1$,

$$E[W_N(n)] \approx \begin{cases} \dfrac{1}{\mu(2-\rho)}(n+1), & \text{for } \rho \le 1; \\ \dfrac{1}{\mu(2-\rho)} \left\{ n + 1 - a_N^* \left(\dfrac{a_n^* + a_{N-1}^*}{2} \right)^{N-n} \right\}, & \text{for } \rho \ge 1. \end{cases} \tag{13}$$

We provide numerical comparisons between the approximation and the exact values in Table 2.

In Table 2, one can see that the approximation is quite accurate, except when ρ is slightly larger than 2 and n is small. That is because when the curvature of $E[W_N(n)]$ is sharp and $N - n$ is large, using $(a_n + a_{N-1})/2$ to approximate the geometric mean of $\{a_n, \ldots, a_{N-1}\}$ is a long shot. It does better as either ρ or n gets larger.

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

One can see from Figure 4 in [2] that (13) would outperform the approximations proposed there for approximating $E[W_N(n)]$ of PS queue with state-independent service rate. But, one has to bear in mind that those approximations are for the distribution, whereas (13) is only for the moments.

4. MULTIPLE-CLASS QUEUES

For the PS queue with state-independent service rate, we now consider that customers are heterogeneous.

Suppose there are k classes of customers joining the queue for service. A class-i customer arrives by a Poisson process with rate λ_i, and is assigned a service weight r_i so that when there are n_j of class-j customers in the system, $j = 1, 2, \ldots, k$, he is served at rate

$$\frac{r_i}{\sum_{j=1}^{k} n_j r_j}\,\mu = \frac{r_i}{\mathbf{n}\cdot\mathbf{r}}\,\mu,$$

where $\mathbf{n}\cdot\mathbf{r}$ is the dot product of $\mathbf{n} = (n_1, n_2, \ldots, n_k)$ and $\mathbf{r} = (r_1, r_2, \ldots, r_k)$.

Suppose further that the queue has a system limit N_i on class-i customers, and let $\mathbf{N} = (N_1, N_2, \ldots, N_k)$.

Heterogeneous customers with different service allocation of this queue is called DPS. The setup of the multiple class DPS queue has applications in admission control and pricing of the Internet system when users of the system have various purposes of using the Internet. We refer interested readers to Altman, Avrachenkov, and Ayesta [1] for a complete survey and existing results on the DPS queue.

Under this consideration, using vectors for states and matrices for transitions are unavoidable.

Let $W_{\mathbf{N}}(i, \mathbf{n})$ denote the conditional sojourn time of a class-i customer in a multiple class $M/M/1/\mathbf{N}$ PS queue given the state of the system $\mathbf{n} - \delta_i$ upon arrival, where δ_{ji} is the Kronecker delta and $\delta_i = (\delta_{1i}, \delta_{2i}, \ldots, \delta_{ki})$.

In [3], an exponential matrix expression is derived for the exact distribution of $W_N(1, \mathbf{n})$ that is, similar to (2). The main function of the expression is for investigating the appropriateness of the approximated distribution obtained therein.

By the same reasoning for the derivation of (4), we first let $\rho_i = \lambda_i/\mu$ and $\rho = \rho_1 + \cdots + \rho_k$, then, for $0 \le n_j < N_j, j \ne i$, we obtain that when $1 \le n_i < N_i$,

$$E[W_{\mathbf{N}}^m(i, \mathbf{n})] = \frac{1}{1+\rho}\left\{\frac{mE[W_{\mathbf{N}}^{m-1}(i, \mathbf{n})]}{\mu} + \sum_{j=1}^{k} \rho_j E[W_{\mathbf{N}}^m(i, \mathbf{n} + \delta_j)] + \sum_{j=1}^{k} \frac{(n_j - \delta_{ji}) r_j}{\mathbf{n}\cdot\mathbf{r}} E[W_{\mathbf{N}}^m(i, \mathbf{n} - \delta_j)]\right\}, \tag{14}$$

and when $n_i = N_i$,

$$E[W_{\mathbf{N}}^m(i, \mathbf{n})] = \frac{1}{1+\sum_{j\ne i}\rho_j}\left\{\frac{mE[W_{\mathbf{N}}^{m-1}(i, \mathbf{n})]}{\mu} + \sum_{j\ne i} \rho_j E[W_{\mathbf{N}}^m(i, \mathbf{n} + \delta_j)] + \sum_{j=1}^{k} \frac{(n_j - \delta_{ji}) r_j}{\mathbf{n}\cdot\mathbf{r}} E[W_{\mathbf{N}}^m(i, \mathbf{n} - \delta_j)]\right\}.$$

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

Consequently, by letting

$$a_{\mathbf{n}}(i) = \begin{cases} \dfrac{\rho_i}{1+\rho - a_{\mathbf{n}-\delta_i}(i)(n_i-1)r_i/(\mathbf{n}\cdot\mathbf{r})}, & 1 \le n_i < N_i; \\ \dfrac{\rho_i}{1+\sum_{j\neq i}\rho_j - a_{\mathbf{n}-\delta_i}(i)(n_i-1)r_i/(\mathbf{n}\cdot\mathbf{r})}, & n_i = N_i, \end{cases} \tag{15}$$

and for $1 \le n_i \le N_i$

$$b_{m,\mathbf{n}}(i) = \frac{a_{\mathbf{n}}(i)}{\rho_i}\left\{ m\frac{E[W_{\mathbf{N}}^{m-1}(i,\mathbf{n})]}{\mu} + \sum_{j\neq i}\rho_j E[W_{\mathbf{N}}^m(i,\mathbf{n}+\delta_j)] \right.$$
$$\left. + \sum_{j\neq i}\frac{n_j r_j}{\mathbf{n}\cdot\mathbf{r}}E[W_{\mathbf{N}}^m(i,\mathbf{n}-\delta_j)] + \frac{(n_i-1)r_i}{\mathbf{n}\cdot\mathbf{r}}\, b_{m,\mathbf{n}-\delta_i}(i)\right\}, \tag{16}$$

we obtain the corresponding recursive formula

$$E[W_{\mathbf{N}}(i,\mathbf{n})] = b_{m,\mathbf{n}}(i) + a_{\mathbf{n}}(i)E[W_{\mathbf{N}}(i,\mathbf{n}+\delta_i)].$$

Regarding the boundary condition, that is, for some $n_l, l \neq i$, in $\mathbf{n}$ such that $n_l = N_l$, we just replace every ρ_l in the above equations by 0. For $\mathbf{n}$ lying outside the feasible region, that is, $n_i \notin \{1,\ldots,N_i\}$, and/or some $n_j \notin \{0,1,\ldots,N_j\}$, define $E[W_{\mathbf{N}}(i,\mathbf{n})] = 0$.

Consequently, we derive the iterative formula below:

THEOREM 5: *For the multiple-class DPS queue with service weight vector* $\mathbf{r}$ *and system limit* $\mathbf{N}$*, the expected conditional sojourn time of a class-i customer given system state* $\mathbf{n}-\delta_i$ *upon arrival is*

$$E[W_{\mathbf{N}}(i,\mathbf{n})] = \sum_{j=0}^{N_i-n_i} b_{m,\mathbf{n}+j\delta_i}(i)\prod_{l=0}^{j-1} a_{\mathbf{n}+l\delta_i}(i), \tag{17}$$

where $a_{\mathbf{n}}(i)$ *and* $b_{m,\mathbf{n}}(i)$ *are defined as (15) and (16), respectively.*

We note that $a_{\mathbf{n}}(i)$, for some fixed i, is independent of $E[W_N^m(i,\mathbf{n})]$ for all m and can be treated as coefficients when solving the set of linear equations (17). Furthermore, by the recursive formula (16), we obtain

$$b_{m,\mathbf{n}}(i) = \frac{a_{\mathbf{n}}(i)}{\rho_i}\sum_{l=0}^{n_1-1}\left\{ \frac{mE[W_{\mathbf{N}}^{m-1}(i,\mathbf{n}-l\delta_i)]}{\mu} + \sum_{j\neq i}\rho_j E[W_N^m(i,\mathbf{n}-l\delta_i+\delta_j)] \right.$$
$$\left. + \sum_{j\neq i}\frac{n_j r_j}{(\mathbf{n}-l\delta_i)\cdot\mathbf{r}}E[W_N^m(i,\mathbf{n}-l\delta_i-\delta_j)]\right\}\prod_{n=1}^{l}\frac{a_{\mathbf{n}-n\delta_i}(i)(n_i-n)r_i}{\rho_i(\mathbf{n}-(n-1)\delta_i)\cdot\mathbf{r}} \tag{18}$$

From (17) and (18), one can see that $E[W_N^m(i,\mathbf{n})]$, with the $(m-1)$th moment known, depends on $E[W_N^m(i,\mathbf{n}-(n_i-n)\delta_i \pm \delta_j)], n = 1,2,\ldots,N_i$, for any $1 \le n_i \le N_i$. That means the vector $(E[W_N^m(i,\mathbf{n})]; n_i = 1,\ldots,N_i)$ is a linear combination of vectors $(E[W_N^m(i,\mathbf{n}-\delta_j)]; n_i = 1,\ldots,N_i)$ and $(E[W_N^m(i,\mathbf{n}+\delta_j)]; n_i = 1,\ldots,N_i)$ for $j \neq i$, which allows us to transform the linear equations (14) into vector equations.

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

In the rest of the section, we focus on the 2-class DPS queue, and demonstrate the derivation of exact $E[W_N(1,\mathbf{n})]$ via (17), whereas $E[W_N(2,\mathbf{n})]$ can be similarly derived.

We first recall that for matrix $\mathbf{M} \in R^{m\times n}$, write $\mathbf{M} = [\mathbf{M}_{*1}\ \mathbf{M}_{*2}\cdots\mathbf{M}_{*n}]$, where $\mathbf{M}_{*j} \in R^{m\times 1}, j = 1,2,\ldots,n$. Then

$$\begin{bmatrix}\mathbf{M}_{*1}\\ \mathbf{M}_{*2}\\ \vdots\\ \mathbf{M}_{*n}\end{bmatrix} \in R^{mn\times 1}$$

is said to be the vec-function of $\mathbf{M}$, and is written vec($\mathbf{M}$).

Let $\mathbf{W} = [E[W_N(1,\mathbf{n})]]$ and $\mathbf{B} = [b_{1,\mathbf{n}}(1)]$ be two $N_1 \times (N_2+1)$ matrices, and $\mathbf{A} = \text{diag}[\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{N_2}]$ be an $N_1(N_2+1) \times N_1(N_2+1)$ matrix with

$$\mathbf{A}_j = \begin{bmatrix} 1 & a_{(1,j)}(1) & a_{(1,j)}(1)a_{(2,j)}(1) & \cdots & \prod_{l=1}^{N_1-1} a_{(l,j)}(1)\\ 0 & 1 & a_{(2,j)}(1) & \cdots & \prod_{l=2}^{N_1-1} a_{(l,j)}(1)\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & \cdots & \cdots & \cdots & 1 \end{bmatrix}_{N_1\times N_1}.$$

We can express (17) in a matrix form as

$$\text{vec}(\mathbf{W}) = \mathbf{A}\text{vec}(\mathbf{B}). \tag{19}$$

Moreover, we have from (18) that

$$b_{1,\mathbf{n}}(1) = \frac{a_{\mathbf{n}}(1)}{\rho_1}\sum_{l=0}^{n_1-1}\left\{\frac{1}{\mu} + \rho_2 E[W_N(1,(n_1-l,n_2+1))] + \frac{n_2 r_2}{(n_1-l,n_2)\cdot\mathbf{r}} E[W_N(1,(n_1-l,n_2-1))]\right\}\prod_{n=1}^{l}\frac{a_{\mathbf{n}-n\delta_1}(1)(n_1-n)r_1}{\rho_1(n_1+1-n,n_2)\cdot\mathbf{r}} \tag{20}$$

Thus, we can let

$$\mathbf{C}_j = \begin{bmatrix} a_{(1,j)}(1) & 0 & \cdots & 0\\ a_{(2,j)}(1)\dfrac{a_1(1,j)r_1}{\rho_1(2,j)\cdot\mathbf{r}} & a_{(2,j)}(1) & \cdots & 0\\ a_{(3,j)}(1)\prod_{n=1}^{2}\dfrac{a_1(n,j)nr_1}{\rho_1(n+1,j)\cdot\mathbf{r}} & a_{(3,j)}(1)\dfrac{a_1(2,j)2r_1}{\rho_1(3,j)\cdot\mathbf{r}} & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ a_{(N_1,j)}(1)\prod_{n=1}^{N_1-1}\dfrac{a_1(n,j)nr_1}{\rho_1(n+1,j)\cdot\mathbf{r}} & a_{(N_1,j)}(1)\prod_{n=2}^{N_1-1}\dfrac{a_1(n,j)nr_1}{\rho_1(n+1,j)\cdot\mathbf{r}} & \cdots & a_{(N_1,j)}(1)\end{bmatrix},$$

$$\mathbf{D}_j = \text{diag}\left[\frac{1}{(1,j)\cdot\mathbf{r}}, \frac{1}{(2,j)\cdot\mathbf{r}}, \cdots, \frac{1}{(N_1,j)\cdot\mathbf{r}}\right],$$

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

both be of size $N_1 \times N_1$, and $\mathbf{E}$ be an $N_1 \times (N_2+1)$ matrix with all entries being 1. With $\mathbf{C} = \text{diag}[\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{N_2}]$ and

$$\mathbf{D} = \begin{bmatrix} \mathbf{0} & \rho_2\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ r_2\mathbf{D}_1 & \mathbf{0} & \rho_2\mathbf{I} & \cdots & \mathbf{0} \\ \mathbf{0} & 2r_2\mathbf{D}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & N_2 r_2 \mathbf{D}_{N_2} & \mathbf{0} \end{bmatrix},$$

where $\mathbf{I}$ is a square identity matrix, we then express (20) as

$$\text{vec}(\mathbf{B}) = \frac{1}{\rho_1}\mathbf{C}\left(\frac{1}{\mu}\text{vec}(\mathbf{E}) + \mathbf{D}\text{vec}(\mathbf{W})\right). \tag{21}$$

Plugging (21) into (19), we get

$$\left(\mathbf{I} - \frac{1}{\rho_1}\mathbf{ACD}\right)\text{vec}(\mathbf{W}) = \frac{1}{\rho_1\mu}\mathbf{AC}\text{vec}(\mathbf{E}).$$

Finally, we conclude the derivation by Cramer's rule and state the last main result below:

THEOREM 6: *For the 2-class DPS queue, the solution of* $E[W_N(1,\mathbf{n})]$ *is*

$$E[W_N(1,(i,j))] = \frac{det(\mathbf{F}_{i,j})}{det(\mathbf{F})} \tag{22}$$

for $i = 1, \ldots, N_1, j = 0, 1, \ldots, N_2$, *where*

$$\mathbf{F} = \frac{1}{\rho_1}\begin{bmatrix} \rho_1\mathbf{I} & -\rho_2\mathbf{A}_0\mathbf{C}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ -r_2\mathbf{A}_1\mathbf{C}_1\mathbf{D}_1 & \rho_1\mathbf{I} & -\rho_2\mathbf{A}_1\mathbf{C}_1 & \cdots & \mathbf{0} \\ \mathbf{0} & -2r_2\mathbf{A}_2\mathbf{C}_2\mathbf{D}_2 & \rho_1\mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -N_2 r_2 \mathbf{A}_{N_2}\mathbf{C}_{N_2}\mathbf{D}_{N_2} & \rho_1\mathbf{I} \end{bmatrix}$$

and $\mathbf{F}_{i,j}$ *is the matrix obtained by replacing the* (jN_1+i)*th column of* $\mathbf{F}$ *by*

$$\frac{1}{\rho_1\mu}\ diag[\mathbf{A}_0\mathbf{C}_0\ \mathbf{A}_1\mathbf{C}_1\ \cdots\ \mathbf{A}_{N_2}\mathbf{C}_{N_2}]vec(\mathbf{E}).$$

The computing procedure of (22) is:

1. Derive $\{a_{\mathbf{n}}(1)\}$ by (15).
2. Form block matrices $\mathbf{A}$ and $\mathbf{C}$.
3. Construct block matrix $\mathbf{D}$ and, consequently, matrices $\mathbf{F}$ and $\mathbf{F}_{i,j}$.
4. Compute det($\mathbf{F}$) and det($\mathbf{F}_{i,j}$).
5. Input associated determinants into (22) to get $E[W_N(1,(i,j))]$.

The procedure for corresponding expectation of class-2 customers is the same.

Regarding the implementation, we first note that $\mathbf{A}_j, \mathbf{C}_j$ and $\mathbf{D}_j$ are all triangular matrices whose determinants are just the products of the diagonal entries. Secondly, $\mathbf{F}$ is

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

a block tridiagonal matrix that appears in many scientific and engineering applications. It can be found in the literature the analytic formulae of its inverse and determinant, see, for example, Huang and McColl [6] and Molinari [8], as well as efficient algorithms and even the source code in the *C* language [4] for solving the associated system.

Acknowledgments

The majority of the study was done during the second author's visit at the Department of Mathematical and Computing Sciences, Tokyo Institute of Technology. The author is grateful for the hospitality and support received there.

References

1. Altman, E., Avrachenkov, K., & Ayesta, U. (2006). A survey on discriminatory processor sharing. *Queueing Systems* 53: 53–63.
2. Borst, S.C., Boxma, O.J., & Hegde, N. (2005). Sojourn times in finite-capacity processor-sharing queues. *Proceedings of the 1st Euro-NGI Conference*. Rome, 53–60.
3. Boxma, O.J., Hegde, N., & Nunez-Queija, R. (2006). Exact and approximate analysis of sojourn times in finite discriminatory processor sharing queues. *International Journal of Electronics and Communications* 60: 109–115.
4. http://www.cfdengineer.com/articles/BlockTriDiagonal.shtml
5. Coffman, E.G., Muntz Jr. R.R., & Trotter, H. (1970). Waiting time distributions for processor-sharing systems. *Journal of the ACM* 17: 123–130.
6. Huang, Y. & McColl, W.F. (1997). Analytical Inversion of General Tridiagonal Matrices. *Journal of Physics A: Mathematical and General* 30: 7919–7933.
7. Kleinrock, L. (1967). Time-shared systems: a theoretical treatment. *Journal of the ACM* 14: 242–261.
8. Molinari, L.G. (2008). Determinants of block tridiagonal matrices. *Linear Algebra and its Applications* 429: 2221–2226.
9. Naor, P. (1969). The regulation of queue size by levying tolls. *Econometrica* 37: 15–24.
10. Sengupta, B. & Jagerman, D.L. (1985). A conditional response time of the $M/M/1$ processor-sharing queue. *AT&T Technical Journal* 64: 409–421.
11. Zhen, Q. & Knessl, C. (2009). On Sojourn Times in the $M/M/1-PS$ Model, Conditional on the Number of Other Users. *Applied Mathematics Research eXpress* 2: 142–167.

APPENDIX

Proof of Lemma 2: It suffices to show that $\rho < 2$ is the necessary and sufficient condition for

$$\frac{\prod_{i=n}^{N+1} a_{N+1}(i)}{\prod_{i=n}^{N} a_N(i)} < 1, \tag{A.1}$$

where $a_M(i)$ denote a_i with queue limit M. Since $a_{N+1}(i) = a_N(i)$ for $i = n, \ldots, N-1$, we will omit the queue limit when there is no ambiguity.

With some algebraic manipulation, we have

$$\begin{aligned}\frac{\prod_{i=n}^{N+1} a_{N+1}(i)}{\prod_{i=n}^{N} a_N(i)} &= \frac{a_{N+1}(N+1)a_{N+1}(N)}{a_N(N)} \\ &= \frac{\rho[1-(N-1)a_{N-1}/N]}{1-(N-1)a_{N-1}/N+\rho/(N+1)} \\ &= 1-\frac{2-\rho-(\rho-1)[N-(N+1)a_{N-1}](N-1)/N}{(N+1)[1-(N-1)a_{N-1}/N]+\rho}.\end{aligned} \tag{A.2}$$

Now, from (8) and (9), the above numerator becomes

$$(2-\rho)\left[1-(\rho-1)\sum_{k=1}^{N-1}\frac{k}{N}\prod_{i=k}^{N-1}\frac{a_i}{\rho}\right]. \tag{A.3}$$

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press

The term in the bracket is positive when $\rho \leq 1$. When $\rho > 1$, because $a_n < 1$ for $n = 1, \ldots, N-1$ (which can be seen from the recursion of a_n),

$$(\rho - 1) \sum_{k=1}^{N-1} \frac{k}{N} \prod_{i=k}^{N-1} \frac{a_i}{\rho} \leq (\rho - 1) \frac{a_{N-1}/\rho}{1 - 1/\rho} \leq 1$$

so that the term in the bracket is positive too. Therefore, the sign of (A.3) is determined by $2 - \rho$, and (A.2) is smaller than 1 iff $2 - \rho > 0$. ■

Proof of Lemma 3: Firstly, it is easy to see that $a_2 - a_1 > 0$. Then, the monotonicity can be shown by writing

$$\begin{aligned} a_{n+1} - a_n &= \frac{\rho}{1 + \rho - n a_n/(n+1)} - \frac{\rho}{1 + \rho - (n-1) a_{n-1}/n} \\ &= \frac{\rho [n a_n/(n+1) - (n-1) a_{n-1}/n]}{(1 + \rho - n a_n/(n+1))(1 + \rho - (n-1) a_{n-1}/n)} \end{aligned}$$

and arguing inductively that the numerator is positive.

To derive the upper bound, we use the increasing property of $\{a_n\}$ to get

$$a_n < \frac{\rho}{1 + \rho - (n-1) a_n/n}.$$

This inequality holds if either

$$a_n < \frac{1 + \rho - \sqrt{(1-\rho)^2 + 4\rho/n}}{2(n-1)/n} \text{ or } a_n > \frac{1 + \rho + \sqrt{(1-\rho)^2 + 4\rho/n}}{2(n-1)/n},$$

but the later is impossible because $a_n < 1$. Then, multiplying both the numerator and denominator by $1 + \rho + \sqrt{(1-\rho)^2 + 4\rho/n}$ yields the upper bounds. ■

https://doi.org/10.1017/S0269964812000320 Published online by Cambridge University Press