Probability in the Engineering and Informational Sciences, **22**, 2008, 231–259. Printed in the U.S.A.
doi: 10.1017/S0269964808000144

ASYMPTOTIC PROPERTIES OF SOJOURN TIMES IN MULTICLASS TIME-SHARED SYSTEMS

ARZAD A. KHERANI
General Motors India Science Lab
Bangalore, India
E-mail: arzad.kherani@gm.com

We consider two multiclass discriminatory process sharing (DPS)-like time-shared $M/G/1$ queuing systems in which the weight assigned to a customer is a function of its class as well as (1) the attained service of the customer in the first system and (2) the residual processing time of the customer in the second system. We study the *asymptotic slowdown*, the ratio of expected sojourn time to the service requirement, of customers with very large service requirements. We also provide various results dealing with ordering of conditional mean sojourn times of any two given classes. We also show that the sojourn time of an arbitrary customer of a particular class in the standard DPS system (static weights) with heavy-tailed service requirements has a tail behavior similar to that of a customer from the same class *that starts a busy period*.

1. INTRODUCTION AND DISCUSSION OF RELATED LITERATURE

The standard egalitarian processor sharing (EPS) system is an example of a time-shared system in which the *weight* given to a customer remains static throughout its sojourn. One can device complex schemes of allocating weights to a customer in a time-shared system. Age-based[1] [resp. residual processing time (RPT) based] scheduling policy forms an important class of time-shared systems in which the weight of a customer at any point in time during its sojourn depends on its age (resp. RPT) at that time instant. There are many examples emphasizing practical importance and applicability of age-based and RPT-based schemes. See [2, 3, 7, 8, 17] for properties of SRPT (shortest remaining processing time first), which minimizes the number of active flows sample-pathwise and, as a consequence, the mean transfer delay. Reference 7 advocates the SRPT scheduler for web servers. References [11

0269-9648/08 $25.00

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

and 16] provide instances of age-based scheduling in data networks in which a (possibly virtual) scheduler implements a weighted round-robin scheduling by associating a weight, say $W(x)$, with a customer (connection) that is to transfer its xth packet (thus, $x-1$ is the age of the connection at that time instant). Of particular interest is the form $W(x) = x^{\alpha}$ for $-\alpha < \infty$. Table 1 shows the standard scheduling policies that various values of α provide. Here TCP-CA and TCP-SS represent respectively the congestion avoidance and slow-start phases of TCP (the congestion control protocol used in the Internet). LAS stands for least assigned service first system and FCFS for the first come–first served system. Reference 11 shows that TCP provides a mechanism of implementing age-based scheduling schemes *in a distributed fashion* using its concept of window evolution. This is similar to using round-robin to achieve EPS. A similar table can be arrived at by assuming that the weight of a customer depends on its residual processing time, RPT, instead of its age. SRPT will then be obtained as one extreme point (just like LAS was obtained for age-based scheduling). In principle, one need not restrict attention to TCP-CA and TCP-SS only; many other window evolution schemes have already been proposed in the computer networks literature (e.g., the HighSpeed TCP [6] and Scalable TCP [9]). Considering a bottleneck link in the Internet with several classes of connections, each identified by its congestion control protocol, we naturally end up with a multiclass age-based scheduling system.

If $T(x)$ denotes the expected sojourn time of a customer of size x in a queuing system, then its *expected slowdown*, as defined in [8], is given by $T(x)/x$. Using this performance metric, [8] argued that the EPS system is *fair* since the expected slowdown of any customer is $1/(1-\rho)$, irrespective of its service requirement (here, ρ is the load on the system). Now, since the LAS (resp. SRPT) policies provide absolute priorities to jobs with smaller attained (resp. remaining) service, one might think that these policies *discriminate against* jobs with large service requirements. This motivated Theorem 1 in [8], which states that if the second moment of service requirement is finite, then for very large service requirements, the expected slowdowns are the same in EPS, LAS, and SRPT systems. This result is important, as it establishes that in LAS and SRPT systems, the large jobs, although temporarily penalized, are ensured the same performance as in the *fair* EPS system.

Yet another important insight that one gets from Theorem 1 or [8] is that in LAS and SRPT, large jobs are effectively served only when there is no other customer in the system. This motivated another result, Theorem 2 of [8], which states that the results of Theorem 1 of [8] can be thought of as being the worst-case performance achievable for large jobs (because the large jobs are effectively served only in isolation, thus getting

Table 1. Specific Values of the Parameter α to Achieve Some of the Standard Scheduling Disciplines

			Policy		
	LAS	EPS	TCP-CA	TCP-SS	FCFS
α	$-\infty$	0	0.5	1	∞

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

least possible attention), and the other work-conserving scheduling policies will do no worse, thus ensuring that the asymptotic slowdown is at most $1/(1-\rho)$.

A recent article [1] studied the asymptotic slowdown for the standard discriminatory process sharing (DPS) system [5] with a finite second moment of service requirement distribution. They found that, as opposed to the EPS system (which is a DPS system with only one class), the expected sojourn time of customers from a given class does not grow in a strictly linear fashion. If $T_j(x)$ denotes the expected sojourn time of class j customers with service requirement of x units, then

$$T_j(x) = \frac{x}{1-\rho} + \theta_j(x),$$

where $\theta_j(x)$ is a monotone nondecreasing function. The quantity $\lim_{x\to\infty} \theta_j(x)$ is called the *asymptotic bias* in the expected sojourn time of class j customers. The asymptotic bias is shown to be a finite quantity that depends on the second moments (of service requirements) and weights *of the other classes*. This finiteness of asymptotic bias implies that the asymptotic slowdown in the case of a DPS system is again $1/(1-\rho)$ *for all classes of customers*. The results of [1] can be seen as a generalization of results of [8] to the case of DPS system (where the weight of a customer does not change over time).

Note that in SRPT (resp. LAS), the server gives *all* its attention to the customer(s) with shortest RPT (resp. age). Instead of working only with these kinds of extremal policies, we assume a general age-based (resp. RPT-based) server sharing mechanism with LAS (resp. SRPT) as its special case. In particular, we assume that in the age-based (resp. RPT-based) scheduling system, the weight of a class j customer that has an age (resp. RPT) of x is given by weight function $\omega_j(x)$; clearly, static but class-dependent weights give us the standard DPS system of [1]. For such systems, we study the asymptotic slowdown and asymptotic bias under the assumption of a finite second moment for service requirement distribution. We have refrained from making a comparison by changing the weight functions $\omega_j(\cdot)$, given only the huge degree of freedom that makes such a comparison an unintelligible bunch of statements (this was not the case in the single-class system and has been beautifully done in [8]). Instead, we stop at providing expressions for the asymptotic bias and slowdown, which can be used at will to make such a comparison.

This article can be thought of as an extension of the results presented in [1] to incorporate dynamically changing weights assigned to active customers. For a DPS system with static weights (g_i for a class i customer, $1 \le i \le \mathcal{K}$), the following results were obtained in [1]. Here, $T_i(x)$ represents the mean sojourn time of a class i customer that requires x units of service.

- [1, Lemma 1] If $\rho < 1$ and $g_k > 0$, then $T_k(x) < \infty$ for all $x < \infty$.
- [1, Thm. 2] If $g_k \le g_l$, then $T_k(x) \le T_l(x)$ for all x.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

- [1, Prop. 2] If $g_k \le g_l$, then

$$\begin{aligned} &(\mathrm{P1}) \quad T_k(g_k x) \le T_l(g_l x), \\ &(\mathrm{P2}) \quad \frac{T_k(g_k x)}{g_k x} \ge \frac{T_l(g_l x)}{g_l x}. \end{aligned}$$

- [1, Thm. 3] Let $E[X_j^2]$, the second moment of service requirement distribution of class j customers, be finite for all j. Then,

$$\lim_{x\to\infty} \frac{T_k(x)}{x} = \frac{1}{1-\rho},$$

$$\lim_{x\to\infty} \left(T_k(x) - \frac{x}{1-\rho} \right) = \frac{\sum_{j\neq k} \lambda_j \left(1 - \frac{g_k}{g_j}\right) E[X_j^2]}{2(1-\rho)^2}.$$

The reader is referred to [1] for an interesting interpretation of properties (P1) and (P2). *Note that the underlying assumption in the above results is that the second moment of the service requirement distribution is finite for all the classes.* We will also be adhering to this assumption in the present piece of work.

As mentioned earlier, our contribution in this article is to extend all of the above results to a multi-class system with dynamically changing weights. Yet another novelty of the present article is a *straightforward approach* that yields the more general results with relatively less effort ([1] derives and uses a conservation law for DPS to establish the above results). This approach also has the advantage of providing significant insights into working of such server-sharing systems. We also provide conservation laws for the system under consideration.

Our results support the intuition that in a time-shared system, the asymptotic slowdown (the ratio of sojourn time and the service requirement) of a tagged customer is essentially determined by the customers arriving after the arrival of the tagged customers.

The main results of this article are presented in Sections 2 and 4. In Section 2 we present analysis for conditional mean sojourn times in a multiclass system with dynamic age-dependent weights. Section 4 deals with conditional mean sojourn times in a DPS system with dynamic RPT-dependent weights. Proofs of RPT-based scheduling are similar to those of age-based scheduling; hence, they are omitted.

As a slight digression from the main theme of this work, for the special case of DPS with static weight and Pareto distributed service requirements, we also show in Section 3 that the sojourn time of an arbitrary customer behaves like the sojourn time of a customer starting a busy period. The implications of this result are also discussed.

2. DPS WITH AGE-BASED SCHEDULING

Let there be $\mathcal{K}$ customer classes, indexed 1, 2, ..., $\mathcal{K}$. Class i, $1 \le i \le \mathcal{K}$, customers arrive to the system according to a Poisson process with rate λ_i. The arrival processes

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

of different classes are independent of each other. The service requirements of customers from class k have distribution $F_k(\cdot)$; $F_k^c(x) = 1 - F_k(x)$ is the probability that service requirement of a class k customer exceeds x units. We will assume that $F_k(\cdot)$ has infinite support, has mean $E[X_k]$ (or sometimes we only use E_k) and has finite second moment $E[X_k{}^2]$. A customer of class i that has attained an age of x units is assigned a weight $\omega_i(x)$. At any time instant, the active customers receive a share of server in proportion to their respective weights. We will assume that $0 < \omega_i(x) < \infty$ for all i and all $0 \leq x < \infty$. For the classical DPS system of [12], $\omega_i(x) \equiv g_i$. We use the following notation:

$N_j(x)$ is the mean number of class j customers with attained service of at most x present in the steady state (hence, also at customer arrival instants).

$T_j(x)$ is the mean sojourn time of a class j customer requiring x amount of service.

$U_j(x)$ denotes the mean sojourn time of a class j customer that starts a busy period and requires x amount of service. This is also the amount of service imparted to the customer and its descendants.[2]

$C_{k,j}(y, z, x)$ is the contribution of a class k customer (and its descendants) to the sojourn time of a class j customer until the instant when a customer of class j has attained an age of x units and such that at the initial time a customer of class k (resp. j) had age of y (resp. z). Clearly, $z \leq x$.

$\hat{\omega}_i(x) = \int_{u=0}^{x} (1/\omega_i(u))\, du$ so that $d(\hat{\omega}_i(x))/dx = 1/\omega_i(x)$.

Note that $C_{k,j}(y, z, x)$ is defined without reference to other customers in the system; this helps us in studying the evolution of age of any customer with respect to one *tagged* customer. The next lemma reports this formally.

Lemma 1: *If at an instant in time there is a customer of class j having an age of y and also present in the system is a customer of class i having age of z, then the age of customer of class j at the instant when the customer of class i has attained an age of x units (with x ≥ z) is (assuming both of the customers have enough service requirement for this to happen)*

$$\Omega_{j,i}(y, z, x) = \hat{\omega}_j^{-1}(\hat{\omega}_j(y) + \hat{\omega}_i(x) - \hat{\omega}_i(z)),$$

irrespective of the dynamics of other customers.

Proof: See Appendix A. ■

Theorem 1: *For an M/G/1 queue with age-based scheduling, the following hold:*

$$N_k(y) = \lambda_k \left[\int_{x=0}^{y} T_k(x) dF_k(x) + T_k(y) F_k^c(y) \right]. \qquad (\mathbf{1})$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

2.

$$C_{k,j}(y,z,x) = \int_{v=y}^{\infty}\int_{u=y}^{v\wedge\Omega_{k,j}(y,z,x)}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,j}(0,\Omega_{j,k}(z,y,u),x)\right]du\,\frac{dF_k(v)}{F_k^c(y)}$$
$$= \int_{u=y}^{\Omega_{k,j}(y,z,x)}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,j}(0,\Omega_{j,k}(z,y,u),x)\right]\frac{F_k^c(u)}{F_k^c(y)}du.$$

3.

$$U_j(x) = x + \int_{u=0}^{x}\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,j}(0,u,x)\,du.$$

4.

$$T_j(x) = U_j(x) + \int_{u=0}^{\infty}\sum_{k=1}^{\mathcal{K}} C_{k,j}(u,0,x)\,dN_k(u).$$

Proof: See Appendix B. ■

Remarks:

1. Since $N_k(\cdot)$ is expressed in terms of $T_k(\cdot)$, we get, using the expression for $U_j(x)$,

$$T_j(x) = x + \int_{u=0}^{x}\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,j}(0,\ u,\ x)\,du + \int_{u=0}^{\infty}\sum_{k=1}^{\mathcal{K}} C_{k,j}(u,\ 0,\ x)\lambda_k F_k^c(u)\,dT_k(u).$$

   We can use the method of repeated substitution to get an infinite series expansion for $T_j(x)$ in terms of the functions $C_{\cdot,\cdot}(\cdot,\cdot,\cdot)$. However, such an expansion does not furnish significant information directly.

2. An important point to be observed is that in Theorem 1, we have effectively decomposed $T_j(x)$ into two components: (1) the contribution of the existing customers and their descendants and (2) the contribution of the customer itself and its descendants. The second component is the term $U_j(x)$. It turns out that, for file size distributions with a finite second moment, the asymptotic properties of the function $T_j(\cdot)$ is determined by the second component only (i.e., $U_j(\cdot)$); the first component remains bounded by a finite quantity.

Theorem 2: *If $\rho < 1$, then the following hold*:

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

1.

$$C_{k,j}(y,z,x) \le \frac{1}{1-\rho}\int_{u=0}^{\infty} u\frac{dF_k(u+y)}{F_k^c(y)}, \quad \textit{independent of } x \textit{ and } j,$$

$$C_{k,j}(y,z,x) \nearrow_{x\to\infty} \frac{1}{1-\rho}\int_{u=0}^{\infty} u\frac{dF_k(u+y)}{F_k^c(y)}.$$

2. $U_j\ (x) \le x/(1-\rho)$.

3.

(a) $T_j(x)$ *exists and is bounded for each class* j *and* $x<\infty$.

(b) *If* $\bar{R}$ *is the expected total remaining work in steady state, then* $T_j(x) - U_j(x) \le \bar{R}/(1-\rho)$, $\forall x$.

(c) $T_j(x) - U_j(x) \nearrow_{x\to\infty} \overline{R}/(1-\rho) = \left(\sum_l \lambda_l E[X_l^2]\right)/2(1-\rho)^2$.

PROOF: See Appendix C. ■

The results of Theorem 2 correspond to [1, Lemma 1] and [1, Theorem 3]. Slight difference here is that [1, Theorem 3] deals with $T_j(x) - x/1-\rho$ while we look at $T_j(x) - U_j(x)$. Thus, we see that the (average) contribution of the existing customers to the sojourn time of a newly arriving customer is bounded above by a finite quantity.

2.1. Main Result and Discussions

THEOREM 3: *Under conditions of Theorem 2, we have the following*:

1.

$$\frac{x}{1-\rho} - U_j(x) \to_{x\to\infty} \sum_l \lambda_l \Phi_{l,j},$$

where

$$\Phi_{l,j} = \frac{1}{1-\rho}\lim_{x\to\infty}\int_{u=0}^{x}\left[E_l - \int_{y=0}^{\Omega_{l,j}(0,u,x)} F_l^c(y)\,dy\right]du.$$

2.

$$\lim_{x\to\infty} T_j(x) - \frac{x}{1-\rho} = \frac{\overline{R}}{1-\rho} - \sum_l \lambda_l \Phi_{l,j}.$$

Proof: See Appendix D. ■

Remarks:

1. We have obtained a general expression for the asymptotic behavior of sojourn times. Other known results (for standard DPS queue, etc.) can be easily obtained using our expression.
2. Finiteness of $\Phi_{l,j}$ is not required for the validity of Theorem 3 (the result remains valid trivially if $\Phi_{l,j} = \infty$ for some l). In fact, one can easily see that $\Phi_{l,j} = \infty$ cannot be ruled out or cannot be considered as an undesirable scenario, as it is a possibility for systems of practical importance. A trivial example is when the weight function $\omega_j(x)$ is of the form $(x + 1)^{\alpha}$ with $\alpha \to \infty$, and $\omega_l(x) = 1$, approaching FCFS scheduling for the class j customer (for which $U_j(x)/x \to 1$ for any service requirement distribution). A nontrivial example is when $\omega_j(x) = x + 1$ and $\omega_l(x) = 1$, like in the Slow-Start phase of TCP, when the bandwidth sharing gets *locked on* to the initial state. Indications of these observations are already in [10]. In the case $\Phi_{l,j} = \infty$ for some l, it is seen that the asymptotic slowdown is
$$\lim_{x\to\infty} \frac{U_j(x)}{x} = \frac{1}{1-\rho} - \sum\nolimits_l \frac{\lambda_l}{1-\rho} \lim_{x\to\infty} \frac{1}{x} \int_{u=0}^{x} \left[E_l - \int_{y=0}^{\Omega_{l,j}(0,u,x)} F_l^c(y)\, dy \right] du$$
$$> \frac{1}{1-\rho},$$
implying a larger share of server being allocated to class j customers in the presence of class l jobs.
3. Consider the busy period in which the tagged customer (with service requirement x) arrives. Let the customers in this busy period be identified by their arrival sequence in the busy period. For the ith arriving customer in this busy period, let $\mathbf{D_i}$ denote the (random) set of its descendants; we will follow the convention that $i \in \mathbf{D_i}$. Let $\mathbf{C}_\mathrm{e}$ denote the set of customer index that the tagged arrival found in the system on its arrival. Let the index of the tagged customer in the busy period be the integer-valued random variable $\mathbf{T}$. Clearly, $\mathbf{D_T} \subset \cup_{i\in \mathbf{C}_\mathrm{e}} \mathbf{D_i}$; in particular, $\mathbf{T} \in \cup_{i\in\mathbf{Ce}} \mathbf{D_i}$. Let $\mathbf{C_s}$ denote the set of *all* the customers that the tagged customer sees in the system during its sojourn; note that $\mathbf{C_s} = (\cup_{i\in\mathbf{C}_\mathrm{e}} \mathbf{D_i})$. Clearly, $\mathbf{C_e} \subset \mathbf{C_s}$ and $\mathbf{C_s} \subset \cup_{i\in\mathbf{C}_\mathrm{e}} \mathbf{D_i}$. We also need to keep in mind that $\mathbf{D_T}$ need not be contained in $\mathbf{C_s}$, as the family of descendants of the tagged customer might have new arrivals after its departure as well.

   The idea now is decompose the sojourn time of the tagged customer by looking at the *contribution* to the sojourn time of the tagged customer by the following:
   (a) Descendants of customers who were present in the system at the arrival instant of the tagged customer, but are not descendants of the tagged customer. This is the set $\mathbf{C_s}\backslash\mathbf{D_T}$.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

(b) Descendants of the tagged customer itself (i.e., $\mathbf{C_s} \cap \mathbf{D_T}$).

By *contribution* here we mean the amount of service given to the customers of these sets *during the sojourn* of the tagged customer.

The expected contribution of these two sets are then $T_j(x) - U_j(x)$ and $U_j(x)$, respectively.

4. Part 3 of Theorem 2 says that the expected contribution from the (descendants of the) existing customers remains bounded above by the mean residual busy period length. Further, this bound is *achieved* as the service requirement of the tagged customer increases to very large value. This is intuitive because we are assuming that the weight given to any customer is strictly positive during its sojourn; this ensures that the descendants of the existing customers get served at a strictly positive rate and, hence, the stability of the system implies that any such family dies out eventually.

   It thus follows that the asymptotic bias in the sojourn time of a customer can come only from the contribution of the descendants of the tagged customer itself (i.e., the $U_j(x)$ term). This is confirmed by Theorem 3.
5. Part 1 of Theorem 3 is clearly intuitive because as $x \to \infty$, $C_{l,j}(0, u, x)$ is essentially the expected busy period length with the exceptional first service [18] (the first customer is from class l).
6. Recall the way we had to write Eq. (D.1) of Appendix D in order to study the asymptotic property of $x/(1-\rho) - U_j(x)$. We were not able to apply the monotone or dominated convergence theorems to the expression preceding Eq. (D.1). The problem essentially was that even though for any *given* u, the quantity $C_{l,j}(0, u, x) \to_{x\to\infty} E_l/(1-\rho)$, one cannot claim the same for $C_{l,j}(0, x-u, x)$. The reason for asymptotic bias is the incomplete *busy period* started by a class l customer that arrives to the system when the tagged customer has u amount of remaining service. It is then clear that the contribution of such a class l customer will be less than the busy period with exceptional first service. Further, the relative weights and the service requirement distributions of the various classes will now play an important role, the reason being that if the weight of the tagged customer is decreasing with its age, then the other customers will grab most of the server share from it and, hence, their contribution will be closer to the mean busy period with exceptional first service. Whereas if the weight of tagged customer increases with age, the contribution from other customers will be less. This effect is observed in a simulation study not reported here.
7. It was observed in [13] that large sojourn time in the EPS system *with the light-tailed service requirement* is essentially due to "work brought along by customers arriving *during* the sojourn time of the tagged customer" (i.e., large cardinality of the set $(\cup_{i\in\mathbf{C_e}} \mathbf{D_i})\backslash\mathbf{C_e}$). The result of Theorem 2 *suggests* (although a formal proof is not yet available to us) that essentially it is because of the large cardinality of $\mathbf{D_T}$. Note that both of these sets of customers arrive *during* the sojourn of the tagged customer.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

8. The breakup of the sojourn time that we have used helps us in understanding the asymptotic property of the sojourn times by only looking at the contribution of the descendants of the tagged customer, forgetting about the contribution of the existing customers. This is a significant simplification since most of the difficulty in such asymptotic analysis is caused when considering the contribution of descendants of the existing customers. In a recent article [14], we used this approach to prove tail equivalence between the sojourn time of a customer starting a busy period and the service requirement distribution for a single-class age-based scheduling system.

2.2. Ordering of Sojourn Times

The function $\Omega_{j,i}(y, z, x)$ plays a central role in obtaining several interclass orderings of $T_k(\cdot)$ to be presented in the article. This function allows us to study the *relative* evolution of age of two customers while *forgetting about the other customers in the system*. We now present some structural properties of this function, most of which are under the following condition:

$$\mathbf{C1}: \qquad \Omega_{j,i}(y,z,x) > \Omega_{j,j}(y,z,x) \quad \text{for all } x, y, \text{ and } z \leq x.$$

Condition **C1** amounts to requiring that a *class j customer does better against a class i customer compared to its standing against a class j customer.*

Let $\Omega_{j,i}(x) \triangleq \Omega_{j,i}(0,0,x)$, then we have the following:

Lemma 2: *For a multiclass age-based scheduling system, we have the following*:

Property 1:

$$\boldsymbol{C1} \Leftrightarrow \hat{\omega}_i(x) - \hat{\omega}_i(z) > \hat{\omega}_j(x) - \hat{\omega}_j(z).$$

Property 2:

$$\boldsymbol{C1} \Leftrightarrow \frac{d}{dx}\hat{\omega}_i(x) > \frac{d}{dx}\hat{\omega}_j(x) \quad (i.e., \omega_i(x) < \omega_j(x) \textit{ for all } x).$$

Property 3:

$$\boldsymbol{C1} \Rightarrow \Omega_{k,i}(y,z,x) > \Omega_{k,j}(y,z,x) \quad \textit{for all } k, x, y, \textit{ and } z \leq x.$$

Property 4:

$$\boldsymbol{C1} \Rightarrow \Omega_{i,k}(y,z,x) < \Omega_{j,k}(y,z,x), \quad \textit{for all } k, x, y, \textit{ and } z \leq x.$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

Property 5:

$$\frac{d\Omega_{j,i}(x)}{dx} = \frac{\omega_j(\Omega_{j,i}(x))}{\omega_i(x)}.$$

Property 6: *$C_{l,k}(y, z, x)$ is a decreasing function of z.*

PROOF: See Appendix E. ■

We list an explanation of the results of Lemma 2.

Property 2: This is intuitive, as it says that **C1** is possible iff the weight given to a class j customer with age x is larger than that given to a class i customer.

Property 3 and 4: In case $\omega_k(\cdot)$ is a nondecreasing function, it is clear that, in view of property 2, a class k customer would do better against a class i customer in comparison to its performance against a class j customer. However, properties 3 and 4 are valid in an even more general setting. Hence, these two properties *cannot* be seen as intuitive corollary to property 2, except for the case where $\omega_k(\cdot)$ is a nondecreasing function.

Property 6: This property uses the fact that $d\,\Omega_{j,k}(y, z, x)/dz < 0$, which, again, is not intuitive for general functions $\omega_j(\cdot)$ and $\omega_k(\cdot)$. For example, one might expect an increase in $\Omega_{j,k}(y, z, x)$ with an increase of $\delta > 0$ in z if both $\omega_k(\cdot)$ and $\omega_j(\cdot)$ are decreasing functions so that class j customer might gain more service with an increase in z. However, it turns out that such a gain is nullified by the δ decrease in the class k customer timeline.

THEOREM 4: *If **C1** holds, then the following hold*:

1. *For all k, x, y, and z, $C_{k,i}(y, z, x) > C_{k,j}(y, z, x)$.*
2. *For all x, $T_i(x) \geq T_j(x)$.*
3. *If it is also true that $d\Omega_{j,i}(x)/dx \geq 1$ for all x, then for all x, $T_i(x) \leq T_j(\Omega_{j,i}(x))$.*

PROOF: See Appendix F. ■

Clearly, results of Theorem 4 are the equivalents of Theorem 2 of and [1] (P1) of [1, Prop. 2]. Note the additional condition $d\Omega_{j,i}(x)/dx \geq 1$ required to obtain the equivalent of (P1) of [1, Prop. 2]. In the case of [1], this condition is not required because for classical DPS, $\Omega_{j,i}(x) = g_j x/g_i$ so that

$$\frac{d\Omega_{j,i}(x)}{dx} = \frac{g_j}{g_i} = \frac{\omega_j(x)}{\omega_i(x)} \geq 1$$

holds trivially under condition **C1** in view of the results of Lemma 2.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

An attempt to generalize (P2) of [1, Prop. 2] requires another definition: The *discounted weight* of a customer of class j at an age of u with respect to a later age $y \geq u$ is

$$D_j(y, u) \triangleq \frac{\omega_j(u)}{y}.$$

THEOREM 5: *If **C1** is true and if it also holds that* $D_j(\Omega_{j,i}(x), \Omega_{j,i}\,(u)) \leq D_i(x, u)$ for all x and $u \leq x$, *then, for all* x,

$$\frac{T_i(x)}{x} \geq \frac{T_j(\Omega_{j,i}(x))}{\Omega_{j,i}(x)}.$$

PROOF: See Appendix G. ■

Again, for classical DPS, it is easily shown that $D_j(\Omega_{j,i}(x), \Omega_{j,i}\,(u)) = D_i(x, u)$ for any i and j; hence, this condition was not required in [1].

3. DPS WITH STATIC WEIGHT

We now consider a DPS system where class i customers have a static weight $g_i > 0$. We will be assuming that the service requirement distribution of class i is Pareto distributed with mean $E[X_i]$ and some shape parameter $\zeta_i \in (1, 2)$. For the DPS system *with static weights*, it is easily shown that

$$\lim_{x\to\infty} \frac{U_i(x)}{x} = \frac{1}{1-\rho}.$$

Note that this result has been obtained using the expression for $\Phi_{l,j}$ from the previous section, not from the results of [1], as now we are dealing with an infinite second moment of the service requirement distribution.

Now, if we use $V_i(x)$ to denote the second moment of the sojourn time of a class i customer with service requirement x starting a busy period, we can write

$$V_i(x) - U_i(x)^2 = \sum_j \lambda_j \int_{u=0}^{x} V_c^{(j,i)}(u, x)\, du,$$

where

$$V_c^{(j,i)}(u, x) = G_2^{(j,i)}(u, x) + \int_{y=0}^{\infty} \sum_k \lambda_k \int_{z=0}^{y\wedge(g_j(x-u)/g_i)} V_c^{(k,i)}(u + g_i z/g_j, x)\, dz dF_j(y)$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

and

$$G_2^{(j,i)}(u,x) = \int_{y=0}^{\infty} \left[\int_{z=0}^{y \wedge (g_j(x-u)/g_i)} \left(1 + \sum_k \lambda_j C_{k,i}(0, u + g_i z/g_j, x) \right) dz \right]^2 \times dF_j(y).$$

These relations can be interpreted (and have been obtained) in same way as those of Theorem 1.

The main result of this section relates the tail of the probability distribution of the random variable $\mathbf{U}_i$ (the sojourn time of a class i customer starting a busy period) to the probability distribution of the service requirement distribution of class i customer (denoted by a generic random variable $\mathbf{X}_i$)

Theorem 6: *For a DPS queue with static weights and Pareto distributed service requirements*,

$$\lim_{x\to\infty} \frac{P\left(\mathbf{U}_i > \dfrac{x}{1-\rho}\right)}{P(\mathbf{X}_i > x)} = 1.$$

Proof: The proof follows the approach in [14] (which gives the tail equivalence for a single class of customer and is not easily extended to a general multiclass age-based scheduling system). In particular, we show that $G_2^{(j,i)}(u, x) \le V_c^{(j,i)}(u, x) \le G_2^{(j,i)}(u, x)/(1-\rho)$. For the special case of DPS with static weight, it is easy to see that $V_c^{(j,i)}(u, x)$ is a function only of $x-u$; similarly for $G_2^{(j,i)}(u, x)$. This implies that for the DPS system *with static weights*, $V_i(x) - U_i(x)^2 = \sum_j \lambda_j \int_{u=0}^{x} V_c^{(j,i)}(u)\, du$, with $G_2^{(j,i)}(u) \le V_c^{(j,i)}(u) \le G_2^{(j,i)}(u)/(1-\rho)$. Now, since $C_{k,i}(\cdot,\cdot,\cdot)$ is bounded above by $E[X_k]/(1-\rho)$, the mean busy period with first customer from class k, we see that $G_2^{(j,i)}(u) = \int_{y=0}^{\infty} [\rho/(1-\rho)]^2\, [y \wedge (g_j\, u/g_i)]^2\, dF_j(y)$, implying that the asymptotic behavior of $G_2^{(j,i)}(u)$ is like that of $u^{2-\zeta_j}$. This implies that the asymptotic behavior of $V_i(x) - U_i(x)^2$ is like $x^{3-\zeta i}$ or, written in a different manner, is like $x^{2-(\zeta_i-1)}$. Since we are assuming that for all i, $\zeta_i \in (1, 2)$, we see that $\zeta_i - 1 > 0$. Thus, we can invoke Theorem 2.3 of [15] to get the desired result. ■

This result can be seen as a refinement of the result of [4], in which the authors show the tail equivalence for the sojourn time of an arbitrary class j customer and its service requirement distribution. Our result says that the sojourn time of a class j customer that *starts a busy period* also has same behavior as that of an *arbitrary* class j customer. This again supports our approach of looking at the sojourn time of a customer starting a busy period when working with a time-shared system.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

4. DPS WITH RPT-BASED SCHEDULING

Let there be $\mathcal{K}$ customer classes. Classes are indexed $1, 2, \ldots, \mathcal{K}$. Class i, $1 \leq i \leq \mathcal{K}$, customers arrive to the system according to a Poisson process with rate λ_i. The arrival processes of different classes are independent of each other. A customer of class i that has an RPT of x units is assigned a weight $\omega_i(x)$. At any time instant, the active customers receive a share of server in proportion to their respective weights. We will assume that $0 < \omega_i(x) < \infty$ for all i, $1 \leq i \leq \mathcal{K}$, and all $0 \leq x < \infty$. Since most of the results are parallels of those of Section 2, we do not provide an explanation of results in this section. The similarity of results in age-based and RPT-based systems is not surprising in view of the fact that the classical DPS system is common to both the families. The proofs of results given in this section are very similar to those for age-based scheduling; hence, we are not providing them. We here the following notation:

$N_j(x)$ is the mean number of class j customers with RPT of *at most* x present in the steady state (hence, also at customer arrival instants).

$T_j(x)$ is the mean sojourn time of a class j customer requiring x amount of service.

$U_j(x)$ is the mean sojourn time of a class j customer that starts a busy period and requires x amount of service. This is also the amount of service imparted to the customer and its descendants.

$C_{k,j}(y, x)$ is the contribution of a class k customer (and its descendants) to the sojourn time of a class j customer until the instant when the customer of class j has attained an age of 0 units and such that at the initial time the customer of class k (resp. j) had age of y (resp. x).

$\hat{\omega}_i(x) = \int_{u=0}^{x} (1/\omega_i(u))\,du$ so that $d\hat{\omega}_i(x)/dx = 1/\omega_i(x)$.

Let $\Omega_{j,i}(y, z, x)$ be the RPT of a customer of class j at the instant when a customer of class i has an RPT of x units and such that at the initial time the customer of class j (resp. i) had a RPT of y (resp. $z \geq x$). Here, we assume that $x \geq \Omega_{i,j}(z, y, 0)$; that is, the class j customer does not finish before the class i customer reduces its RPT from z to x. In the case that $x \leq \Omega_{i,j}(z, y, 0)$, we use the convention that $\Omega_{j,i}(y, z, \mathrm{x}) = 0$. Then, using the proof of Lemma 1, we can show that

$$\Omega_{j,i}(y, z, x) = \hat{\omega}_j^{-1}(\hat{\omega}_j(y) + \hat{\omega}_i(x) - \hat{\omega}_i(z)).$$

Let $\Omega_{j,i}(x) \triangleq \inf(y : \Omega_{j,i}(y, x, 0) = 0)$; that is, $\Omega_{j,i}(x)$ is the RPT of a class j customer such that it finishes at the same instant as a customer of class i that had an RPT of x initially. Thus, $\Omega_{j,i}(x) = \hat{\omega}_j^{-1}(\hat{\omega}_i(x))$.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

THEOREM 7: *For an $M/G/1$ queue with RPT-based scheduling, we have the following*:

1.

$$N_k(y) = \lambda_k\left[T_k(y)F_k^c(y) + \int_{x=0}^{y} T_k(x)\,dF_k(x)\right].$$

2.

$$C_{k,j}(y,x) = \int_{u=\Omega_{k,j}(y,x,0)}^{y}\left[1 + \sum\nolimits_{l=1}^{\mathcal{K}} \lambda_l \int_{z=0}^{\infty} C_{l,j}(z,\Omega_{j,k}(x,y,u))\,dF_l(z)\right]du.$$

3.

$$U_j(x) = x + \int_{u=0}^{x} \sum\nolimits_{l=1}^{\mathcal{K}} \lambda_l \int_{z=0}^{\infty} C_{l,j}(z,u)\,dF_l(z)\,du.$$

4.

$$T_j(x) = U_j(x) + \int_{y=0}^{\infty} \sum\nolimits_{k=1}^{\mathcal{K}} C_{k,j}(y,x)\,dN_k(y).$$

THEOREM 8: *For any given k, j, y, and x, we have the following*:

1. $C_{k,j}(y,x) \le y/(1-\rho)$, *independent of the value of x. Further,* $C_{k,j}(y,x) \nearrow_{x\to\infty} y/(1-\rho)$.
2. If $\rho < 1$, *then* $U_j(x) \le x/(1-\rho)$.
3. *If* $\rho < 1$, *then* $T_j(x)$ *exists and is bounded for each class j and* $x < \infty$. *Further, we have the following*:
   - *(a)* $T_j(x) - U_j(x) \le \bar{R}/(1-\rho)$.
   - *(b)* $T_j(x) - U_j(x) \nearrow_{x\to\infty} \bar{R}/(1-\rho)$.

As of now, we do not have an exact equivalent of Theorem 3; that is, we do not have a closed-form expression for the asymptotic bias for the quantity $U_j(x)$. An indirect expression can always be found in the following manner.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

Theorem 9: *Under conditions of Theorem 8, we have the following*:

1.

$$\frac{x}{1-\rho} - U_j(x) \to_{x\to\infty} C_j = \sum\nolimits_l \lambda_l \int_{u=0}^{\infty} \left[\frac{E_l}{1-\rho} - \int_{z=0}^{\infty} C_{l,j}(z,u)\, dF_l(z)\right] du$$
$$\geq 0.$$

2.

$$\lim\nolimits_{x\to\infty} T_j(x) - \frac{x}{1-\rho} = \frac{\overline{R}}{1-\rho} - C_j.$$

We now present some structural properties of the function $\Omega_{j,i}(y, z, x)$. We will need the following condition, whose interpretation is similar to **C1** of age-based scheduling,

$$\textbf{C2}: \qquad \Omega_{j,i}(y,z,x) > \Omega_{j,j}(y,z,x) \quad \text{for all } x, y, \text{ and } z \geq x.$$

Also, let $\Omega_{j,i}(x) \triangleq \Omega_{j,i}(0,0,x)$. We now have a lemma equivalent of Lemma 2.

Lemma 3:

1. ***C2*** $\Leftrightarrow \hat{\omega}_i(x) - \hat{\omega}_i(z) > \hat{\omega}_j(x) - \hat{\omega}_j(z)$.
2. ***C2*** $\Leftrightarrow \frac{d}{dx}\hat{\omega}_i(x) < \frac{d}{dx}\hat{\omega}_j(x)$ *(i.e.,* $\omega_i(x) > \omega_j(x)$*).*
3. ***C2***$\Rightarrow \Omega_{k,i}(y, z, x) > \Omega_{k,j}(y, z, x)$ *for all k, x, y, and* $z \geq x$.
4. ***C2***$\Rightarrow \Omega_{i,k}(y, z, x) < \Omega_{j,k}(y, z, x)$ *for all k, x, y, and* $z \geq x$.
5. $C_{l,k}(y, z)$ *is an increasing function of z.*
6. ***C2*** $\Leftrightarrow \Omega_{j,i}(x) \geq x$.

Theorem 10:

1. ***C2*** $\Rightarrow C_{k,i}(y, x) < C_{k,j}(y, x)$ *for all k, y, and x.*
2. ***C2*** $\Rightarrow T_i(x) < T_j(x)$ *for all x.*
3. If $\Omega_{j,i}(x) > x$ *for all x, and if* $(d\Omega_{j,i}(u)/du) \geq 1$ *for all u, then, for all x,*

$$T_i(x) < T_j(\Omega_{j,i}(x)).$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

Define the *discounted weight* of a customer of class j at an age of u with respect to a later age $y \geq u$ as

$$D_j(y,u) \triangleq \frac{\omega_j(u)}{y}.$$

LEMMA 4:

$$\frac{1}{\Omega_{j,i}(x)} \frac{d\Omega_{j,i}(u)}{du} < \frac{1}{x} \quad \textit{iff } D_j(\Omega_{j,i}(x), \Omega_{j,i}(u)) < D_i(x, u).$$

PROOF: The proof follows from the observation that $(d\Omega_{j,i}(u)/du) = (\omega_j(\Omega_{j,i}(u))/\omega_i(u))$. ■

THEOREM 11: *If $\Omega_{j,i}(x)>x$ for all x, y, and $z \leq x$ and if it holds that $D_j(\Omega_{j,i}(x), \Omega_{j,i}(u)) \leq D_i(x, u)$ for all x and $u \leq x$, then, for all x,*

$$\frac{T_i(x)}{x} \geq \frac{T_j(\Omega_{j,i}(x))}{\Omega_{j,i}(x)}.$$

5. CONSERVATION LAWS

For any single-server queue with $\mathcal{K}$ job classes, let $R_k(t)$ be the unfinished work at time t of class k jobs and let $R(t) = \sum_{j=1}^{\mathcal{K}} R_j(t)$ denote the total unfinished work in the system. For any work-conserving discipline, the unfinished work in the system is the same, say $\bar{R}$, regardless of the scheduling discipline employed. In the particular case of age-based or RPT-based scheduling, this implies that the total unfinished work in the system $(R(t)=\sum_{j=1}^{\mathcal{K}} R_j(t))$ is independent of the particular details of the weight functions $(\omega_k(\cdot); k = 1, \ldots, \mathcal{K})$. In the following, we obtain conservation laws for age-based and RPT-based scheduling.

THEOREM 12: *For the system under consideration with age-based scheduling,*

$$\sum_{k=1}^{\mathcal{K}} \lambda_k \int_{y=0}^{\infty} T_k(y) F_k^c(y)\, dy = \overline{R}.$$

PROOF: See Appendix H. ■

THEOREM 13: *For the system under consideration with RPT-based scheduling*,

$$\sum_{k=1}^{\mathcal{K}} \lambda_k \int_{y=0}^{\infty} T_k(y) F_k^c(y)\, dy = \overline{R}.$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

Notice that the conservation law is the same for both age-based and RPT-based scheduling disciplines. This can be explained by the fact that the DPS queue with class-dependent static weights is common to both of these scheduling disciplines.

Notes

[1] Age of a customer at any point in time is the amount of service received by the customer (i.e., its attained Service.)

[2] By descendants of a customer we mean those jobs that arrive during the time the customer or any of its descendants are getting served. Thus, a customer is also a descendant of itself.

References

1. Avrachenkov, K., Ayesta, U., Brown, P., & Nunez-Queija, R. (2005). Discriminatory processor sharing revisited. In *IEEE INFOCOM 2005*.
2. Bansal, N. & Harchol-Balter, M. (2001). Analysis of SRPT scheduling: Investigating unfairness. In *Proceedings of ACM Sigmetrics 2001 Conference*.
3. Borst, S.C., Boxma, O.J., Nunez-Queija, R., & Zwart, A.P. (2003). The impact of the service discipline on delay asymptotics. *Performance Evaluation* 54: 175–206.
4. Borst, S.C., van Ooteghem, D., & Zwart, B. (2005). Tail asymptotics for discriminatory processor sharing queues with heavy-tailed service sequirements. *Performance Evaluation* 61: 281–298.
5. Fayolle, G., Mitrani, I., & Iasnogorodski, R. (1980). Sharing a processor among many job classes. *Journal of the ACM* 27: 519–532.
6. Floyd, S. (2003). RFC 3649: HighSpeed TCP for large congestion windows. Internet Request for Comments, RFC 3649, Experimental. December 2003. Available at http://www.ietf.org/rfc.html
7. Harchol-Balter, M., Schroeder, B., Bansal, N., & Agrawal, M. (2003). Size-based scheduling to improve Web performance. *ACM Transactions Compon Systems* 21: 207–233.
8. Harchol-Balter, M., Sigman, K., & Wierman, A. (2002). Asymptotic convergence of scheduling policies with respect to slowdown. *Performance Evaluation* 49: 241–256.
9. Kelly, T. (2003). Scalable TCP: Improving performance in highspeed wide area networks. *Computer Communication Review* 33: 83–91.
10. Kherani, A.A. & Kumar, A. (2002). Stochastic models for throughput analysis of randomly arriving elastic flows in the internet. In *IEEE Infocom 2002*.
11. Kherani, A.A. & Nunez-Queija, R. (2006). TCP as an implementation of age-based scheduling: fairness and performance. In *IEEE Infocom 2006*.
12. Kleinrock, L. (1967). Time-shared systems: A theoretical treatment. *Journal of the ACM* 14: 242–261.
13. Mandjes, M. & Zwart, B. (2006). Large deviations for sojourn times in processor sharing queues. *Queueing Systems* 52: 237–250.
14. Padhy, S. & Kherani, A.A. (2006). Tail equivalence for some time-shared systems. In *Proceedings of Valuetools 2006 Conference*.
15. Queija, R.N. (2002). Queues with equally heavy sojourn time and service requirement distributions. *Annals of Operations Research* 113: 101–117.
16. Rai, I.A., Urvoy-Keller, G., Vernon, M., & Biersack, E.W. (2004). Performance analysis of LAS-based scheduling disciplines in a packet switched network. In *Sigmetrics 2004*.
17. Schrage, L.E. (1968). A proof of the optimality of the shortest remaining processing time discipline. *Operations Research* 16: 687–690.
18. Wolff, R.W. (1989). *Stochastic modeling and the theory of queues*. Englewood Cliffs, NJ: Prentice-Hall.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

APPENDIX A
Proof of Lemma 1

Let $x_i(t)$ (resp. $x_j(t)$) be the age of the class i (resp. j) customer under consideration at time instant t. Since the choice of time is arbitrary, we can let the initial time be zero so that $x_j(0) = y$ and $x_i(0) = x$. It is then easily seen that

$$\frac{dx_j(t)}{dt} = \frac{\omega_j(x_j(t))}{C(t)} \quad \text{and} \quad \frac{dx_i(t)}{dt} = \frac{\omega_i(x_i(t))}{C(t)},$$

where $C(t)$ is the cumulative weight of all the customers present at time t, so that

$$\frac{dx_j(t)}{dx_i(t)} = \frac{\omega_j(x_j(t))}{\omega_i(x_i(t))}.$$

Solving this differential equation, we see that

$$\hat{\omega}_j(x_j(t)) - \hat{\omega}_j(x_j(0)) = \hat{\omega}_i(x_i(t)) - \hat{\omega}_i(x_i(0)).$$

Substituting the initial values, we get the desired result.

APPENDIX B
Proof of Theorem 1

1. Associate with the jth arriving customer of class k a time function $f_j(u)$ that takes the value 1 when this customer is in the system and has an attained service at most y units; $f_j(u) = 0$ otherwise. Then

$$N_k(y) = \lim_{t\to\infty} \frac{1}{t} \int_{u=0}^{t} \sum_{j=1}^{\infty} f_j(u)\, du = \lambda_k \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \left[\int_{u=0}^{\infty} f_j(u)\, du \right],$$

where the second equality follows from Brumelle's theorem [18]. Let X_j be the total service requirement of the jth arriving customer of class k. Then

$$\begin{aligned} \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \left[\int_{u=0}^{\infty} f_j(u)\, du \right] &= \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \int_{x=0}^{\infty} I_{\{X_j = x\}} \left[\int_{u=0}^{\infty} f_j(u)\, du \right] dx \\ &= \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \int_{x=y}^{\infty} I_{\{X_j = x\}} \left[\int_{u=0}^{\infty} f_j(u)\, du \right] dx \end{aligned}$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

$$+ \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \int_{x=0}^{y} I_{\{X_j = x\}} \left[\int_{u=0}^{\infty} f_j(u)\, du \right] dx$$

$$= \left[F_k^c(y) T_k(y) + \int_{x=0}^{y} T_k(x)\, dF_k(x) \right].$$

The last equality follows because if $X_j \geq y$, then the random variable $\int_{u=0}^{\infty} f_j(u)\, du$ is the same in distribution as the sojourn time of a customer requiring y units of service, otherwise (i.e., if $X_j = x < y$), the random variable $\int_{u=0}^{\infty} f_j(u)\, du$ is the same in distribution as the sojourn time of a customer requiring x units of service.

2. This part follows by first conditioning on the remaining service requirement of the class k customer using the variable v and then summing up the contribution of any arrivals when the class k customer increases its age from u to $u+du$ [so that the age of the class j customer now is $\Omega_{j,k}(z, y, u)$ and, of course, the age of the newly arriving customer is zero].
3. This part follows from a reasoning similar to the one presented in step 2. Here, u conditions on the age of the class j customer and $C_{l,j}(0, u, x)$ accounts for the contribution of any arrival of class l customer when the class j customer increases its age from u to $u + du$ (such an arrival has probability $\lambda_l\, du$).
4. This part follows in a way similar to that used for $U_j(x)$; the only difference is that we also need to consider the contribution of the customers that were present at the instant of arrival of the class j customer. Using the PASTA property, a class j customer will see an average of $dN_k(u)$ customers of class k that have their age in the interval $[u, u+du]$.

APPENDIX C
Proof of Theorem 2

1.

$$C_{k,j}(y, z, x) = \int_{u=y}^{\Omega_{k,j}(y,z,x)} \left[1 + \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,j}(0, \Omega_{j,k}(z, y, u), x) \right] \frac{F_k^c(u)}{F_k^c(y)}\, du$$

$$\leq \int_{u=y}^{\infty} \left[1 + \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,j}(0, \Omega_{j,k}(z, y, u), x) \right] \frac{F_k^c(u)}{F_k^c(y)}\, du$$

$$= \int_{u=y}^{\infty} \frac{F_k^c(u)}{F_k^c(y)}\, du + \sum_{l_1=1}^{\mathcal{K}} \lambda_{l_1} \int_{u_1=y}^{\infty} C_{l_1,j}(0, \Omega_{j,k}(z, y, u_1), x) \frac{F_k^c(u_1)}{F_k^c(y)}\, du_1$$

$$\leq \int_{u=y}^{\infty} \frac{F_k^c(u)}{F_k^c(y)}\, du + \sum_{l_1=1}^{\mathcal{K}} \lambda_{l_1} \int_{u_1=y}^{\infty} \left[\int_{u_2=0}^{\infty} \frac{F_k^c(u_2)}{F_k^c(0)}\, du_2. \right.$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

$$+\sum_{l_2=1}^{\mathcal{K}}\lambda_{l_2}\int_{u_2=0}^{\infty}C_{l_2,j}(0,\Omega_{j,l_1}(\Omega_{j,k}(z,y,u_1),0,u_2),x)\frac{F_{l_1}^c(u_2)}{F_{l_1}^c(0)}\,du_2\Bigg]$$
$$\times\frac{F_k^c(u_1)}{F_k^c(y)}\,du_1.$$

Proceeding similarly by repeatedly substituting the upper bound on $C_{p,q}(\cdot,\cdot,\cdot)$, we get

$$C_{k,j}(y,z,x)\le\theta_k(y)+\sum_{l_1=1}^{\mathcal{K}}\lambda_{l_1}\theta_k(y)\theta_{l_1}(0)+\sum_{l_1=1}^{\mathcal{K}}\sum_{l_2=1}^{\mathcal{K}}\lambda_{l_1}\lambda_{l_2}\theta_k(y)\theta_{l_1}(0)\theta_{l_2}(0)$$
$$+\cdots+\left[\theta_k(y)\sum_{l_1=1}^{\mathcal{K}}\cdots\sum_{l_n=1}^{\mathcal{K}}\lambda_{l_1}\cdots\lambda_{l_n}\theta_{l_1}(0)\theta_{l_n}(0)\right]+\cdots,$$

where for any l and y, $\theta_l(y)=\int_{u=y}^{\infty}(F_l^c(u)/F_l^c(y))\,du$. Clearly, $\sum_{l=1}^{\mathcal{K}}\lambda_l\,\theta_l(0)=\sum_{l=1}^{\mathcal{K}}\lambda_l\int_{u=y}^{\infty}F_l^c(u)\,du=\rho$. Hence, $C_{k,j}(y,\ z,\ x)\le\theta_k(y)/(1-\rho)$. It is also easily shown that $\theta_l(y)=\int_{z=0}^{\infty}z\,(dF_l(z+y)/F_l^c(y))$.

Further, since $0<\omega_k(x)<\infty$ for all k and x, it follows that $\Omega_{k,j}(y,z,x)\nearrow_{x\to\infty}\infty$ so that the upper bounds used above are achieved *in each level of substitution* because the last argument of $C_{p,q}(\cdot,\cdot,\cdot)$ is always x. The second part thus follows. Other way to see this is that the quantity $C_{l,j}\ (0,\cdot,\cdot)$ is bounded above by a finite quantity [i.e., $(\int_{u=0}^{\infty}F_l^c(u)\,du)/(1-\rho)$], independent of value of x, so that one can apply the dominated convergence theorem to

$$C_{k,j}(y,z,x)=\int_{u=y}^{\Omega_{k,j}(y,z,x)}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_lC_{l,j}(0,\Omega_{j,k}(z,y,u),x)\right]\frac{F_k^c(u)}{F_k^c(y)}\,du$$

and let limit $x\to\infty$ to get

$$C_{k,j}(y,z,\infty)=\int_{u=y}^{\infty}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_lC_{l,j}(0,\Omega_{j,k}(z,y,u),\infty)\right]\frac{F_k^c(u)}{F_k^c(y)}\,du. \qquad \textbf{(C.1)}$$

This implies, in particular,

$$C_{k,j}(0,z,\infty)=\int_{u=0}^{\infty}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_lC_{l,j}(0,\Omega_{j,k}(z,0,u),\infty)\right]F_k^c(u)\,du.$$

Repeated substitution of the expression for $C_{p,q}(0,\cdot,\infty)$ above implies that

$$C_{k,j}(0,z,\infty)=\frac{\int_{u=0}^{\infty}F_k^c(u)\,du}{1-\rho}.$$

Note that the above limit is independent of z. Substituting this expression in Eq. (C.1),

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

we get

$$C_{k,j}(y,z,\infty) = \int_{u=y}^{\infty} \frac{F_k^c(u)}{(1-\rho)F_k^c(y)}\,du = \int_{u=y}^{\infty} \frac{u\,dF_k^c(u)}{(1-\rho)F_k^c(y)}.$$

2.

$$\begin{aligned} U_j(x) &= x + \int_{u=0}^{x} \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,j}(0,u,x)\,du \\ &\le x + \int_{u=0}^{x} \sum_{l=1}^{\mathcal{K}} \lambda_l \frac{1}{1-\rho} \int_{z=0}^{\infty} z\,dF_l(z)\,du \\ &= x + \frac{\rho x}{1-\rho} \\ &= \frac{x}{1-\rho}. \end{aligned}$$

3.

$$\begin{aligned} T_j(x) &= U_j(x) + \int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} C_{k,j}(y,0,x)\,dN_k(y) \\ &\le U_j(x) + \frac{1}{1-\rho} \int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} \int_{u=0}^{\infty} u \frac{dF_k(u+y)}{F_k^c(y)}\,dN_k(y). \end{aligned}$$

Noting that the integral $\int_{u=0}^{\infty} u\,(dF_k(u+y)/F_k^c(y))$ is just the expected remaining work in the system of a class k customer with an attained service of y units, it follows that

$$\int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} \int_{u=0}^{\infty} u \frac{dF_k(u+y)}{F_k^c(y)}\,dN_k(y)$$

is the total expected remaining work in steady state (denoted $\bar{R}$), which is finite and independent of the scheduling policy for $\rho < 1$, and we have

$$T_j(x) = U_j(x) + \frac{\overline{R}}{1-\rho} \le \frac{x+\overline{R}}{1-\rho} < \infty.$$

The last part follows similarly using the previous result that the upper bound used above is asymptotically achieved.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

APPENDIX D
Proof of Theorem 3

1. Recall that $U_j(x) \leq x/(1-\rho)$, implying that if there is an asymptotic bias (limit for $U_j(x) - x/(1-\rho)$), it has to be negative. Now,

$$\frac{x}{1-\rho} - U_j(x) = \int_{u=0}^{x} \sum_{l=1}^{\mathcal{K}} \lambda_l \left[\frac{E[X_l]}{1-\rho} - C_{l,j}(0, x-u, x) \right] du, \tag{D.1}$$

where $E[X_j]$ is the mean service requirement of class-j customers. Now,

$$\begin{aligned}
&\int_{u=0}^{x} \left[\frac{E(X_l)}{1-\rho} - C_{l,j}(0,u,x) \right] du \\
&= \int_{u=0}^{x} \left[\frac{E(X_l)}{1-\rho} - \int_{y=0}^{\Omega_{l,j}(0,u,x)} \left[1 + \sum_{l=1}^{\mathcal{K}} \lambda_k C_{k,j}(0, \Omega_{j,l}(u,0,y), x) \right] F_l^c(y)\, dy \right] du \\
&= \frac{1}{1-\rho} \int_{u=0}^{x} \left[E(X_l) - \int_{y=0}^{\Omega_{l,j}(0,u,x)} F_l^c(y)\, dy \right] du \\
&\quad - \int_{u=0}^{x} \int_{y=0}^{\Omega_{l,j}(0,u,x)} \sum_{k=1}^{\mathcal{K}} \lambda_k \left[C_{k,j}(0, \Omega_{j,l}(u,0,y), x) - \frac{E[X_k]}{1-\rho} \right] F_l^c(y)\, dy\, du.
\end{aligned} \tag{D.2}$$

Now,

$$\begin{aligned}
&\int_{u=0}^{x} \int_{y=0}^{\Omega_{l,j}(0,u,x)} \left[C_{k,j}(0, \Omega_{j,l}(u,0,y), x) - \frac{E[X_k]}{1-\rho} \right] F_l^c(y)\, dy\, du \\
&= \int_{y=0}^{\Omega_{l,j}(0,0,x)} \int_{u=0}^{\hat{\omega}_j^{-1}(\hat{\omega}_j(x) - \hat{\omega}_l(y))} \left[C_{k,j}(0, \Omega_{j,l}(u,0,y), x) - \frac{E[X_k]}{1-\rho} \right] F_l^c(y)\, dy\, du \\
&= \int_{y=0}^{\Omega_{l,j}(0,0,x)} \int_{u=\Omega_{l,j}(0,0,y)}^{x} \left[C_{k,j}(0,u,x) - \frac{E[X_k]}{1-\rho} \right] F_l^c(y)\, dy\, du \\
&= \int_{u=0}^{x} \int_{y=0}^{\Omega_{l,j}(0,0,u)} \left[C_{k,j}(0,u,x) - \frac{E[X_k]}{1-\rho} \right] F_l^c(y)\, dy\, du \\
&= \int_{u=0}^{x} E_l \left[C_{k,j}(0,u,x) - \frac{E[X_k]}{1-\rho} \right] \tilde{F}_l(\Omega_{l,j}(0,0,u))\, du \\
&= \int_{u=0}^{x} E_l \left[C_{k,j}(0, x-u, x) - \frac{E[X_k]}{1-\rho} \right] \tilde{F}_l(\Omega_{l,j}(0,0,x-u))\, du,
\end{aligned} \tag{D.3}$$

where $\tilde{F}_l(\Omega_{l,j}(0, 0, u))$ is the equilibrium distribution of $F_l(\cdot)$ (see [18]). The reason for writing the last expression in terms of $x-u$ instead of the second to last expression is that we wish to let $x \to \infty$, which is not justified for the second to last term (cannot use either the dominated convergence theorem or the monotone convergence theorem).

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

However, when working with the last expression, we can use the dominated convergence theorem in the following manner: Observe that for any $\epsilon > 0$,

(a)

$$C_{k,j}(0, x+\epsilon-(u+\epsilon), x+\epsilon)\tilde{F}_l(\Omega_{l,j}(0,0,x+\epsilon-(u+\epsilon))) \\ \geq C_{k,j}(0, x-u, x)\tilde{F}_l(\Omega_{l,j}(0,0,x-u)),$$

(b)

$$\left(\frac{E[X_k]}{1-\rho} - C_{k,j}(0, x+\epsilon, x+\epsilon)\right)\tilde{F}_l(\Omega_{l,j}(0,0,x+\epsilon)) \\ \geq \left(\frac{E[X_k]}{1-\rho} - C_{k,j}(0,x,x)\right)\tilde{F}_l(\Omega_{l,j}(0,0,x)),$$

and

(c)

$$\left(\frac{E[X_k]}{1-\rho} - C_{k,j}(0,0,x)\right)\tilde{F}_l(\Omega_{l,j}(0,0,0)) = 0.$$

The first inequality implies that the *directional derivative* of the function $f(u, x) = C_{k,j}(0, x-u, x)\tilde{F}_l(\Omega_{l,j}(0, 0, x-u))$ is positive along the direction (1, 1). The second and third expressions give the ordering (with changing x) of the function $f(x, u)$ evaluated at the extreme points (i.e., $u=0$ and $u=x$). This three expressions imply that the function $(E[X_k]/(1-\rho) - C_{k,j}(0, x-u, x))\, \tilde{F}_l(\Omega_{l,j}(0, 0, x-u))$ is monotone in x for any given u. Thus, we can invoke the monotone convergence theorem to interchange the limit ($x\to\infty$) and integral. This interchange implies, in particular, that

$$\lim_{x\to\infty}\int_{u=0}^{x} E_l\left[C_{k,j}(0,x-u,x) - \frac{E[X_k]}{1-\rho}\right]\tilde{F}_l(\Omega_{l,j}(0,0,x-u))\,du \\ = \lim_{x\to\infty}\int_{u=0}^{x} E_l\left[C_{k,j}(0,x-u,x) - \frac{E[X_k]}{1-\rho}\right]du,$$

since the function $\Omega_{l,j}(0, 0, (x-u))$ is assumed to be strictly increasing in x. This implies that, with the notation

$$\zeta_{l,j} = \lim_{x\to\infty}\int_{u=0}^{x}\left[\frac{E[X_l]}{1-\rho} - C_{l,j}(0,u,x)\right]du,$$

taking the limit $x\to\infty$ in Eq. (D.2)

$$\zeta_{l,j} = \Phi_{l,j} + E_l\sum_k \lambda_k \zeta_{k,j},$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

where

$$\Phi_{l,j} = \frac{1}{1-\rho} \lim_{x\to\infty} \int_{u=0}^{x} \left[E[X_l] - \int_{y=0}^{\Omega_{l,j}(0,u,x)} F_l^c(y)\, dy \right] du$$
$$= \frac{E_l}{1-\rho} \lim_{x\to\infty} \int_{u=0}^{x} \tilde{F}_l^c(\Omega_{l,j}(0,u,x))\, du.$$

Summing the above expression for $\Phi_{l,j}$ over all values of l and taking limit $x\to\infty$ in Eq. (D.1), we see that the asymptotic bias of $x/(1-\rho)-U_j(x)$ is

$$\sum_l \lambda_l \zeta_{l,j} = \sum_l \lambda_l \left(\Phi_{l,j} + E_l \sum_k \lambda_k \zeta_{k,j} \right),$$

implying that

$$\lim_{x\to\infty} \frac{x}{1-\rho} - U_j(x) = \sum_l \lambda_l \zeta_{l,j} = \sum_l \frac{\lambda_l \Phi_{l,j}}{1-\rho}.$$

2. This part follows from Theorem 2 and the limiting value of $U_j(x)-x/(1-\rho)$ obtained above.

APPENDIX E
Proof of Lemma 2

1. Observe that $\hat{\omega}_k(x) = \int_{u=0}^{x}(1/\omega_k(u))\, du$ is a strictly increasing function of x for all k. Hence, $\Omega_{j,i}(y,z,x) > \Omega_{j,j}(y,z,x) \Leftrightarrow \hat{\omega}_j^{-1}(\hat{\omega}_j(y) + \hat{\omega}_i(x) - \hat{\omega}_i(z)) > \hat{\omega}_j^{-1}(\hat{\omega}_j(y)+$
$\hat{\omega}_j(y) + \hat{\omega}_i(x) - \hat{\omega}_i(z) > \hat{\omega}_j(y) + \hat{\omega}_j(x) - \hat{\omega}_j(z) \Leftrightarrow \hat{\omega}_i(x) - \hat{\omega}_i(z) > \hat{\omega}_j(x) - \hat{\omega}_j(z)$
$\hat{\omega}_j(y) + \hat{\omega}_i(x) - \hat{\omega}_i(z) > \hat{\omega}_j(y) + \hat{\omega}_j(x) - \hat{\omega}_j(z) \Leftrightarrow \hat{\omega}_i(x) - \hat{\omega}_i(z) > \hat{\omega}_j(x) - \hat{\omega}_j(z)$.
2. This is a simple corollary to the first assertion.
3. $\Omega_{k,i}(y,z,x) = \hat{\omega}_k^{-1}(\hat{\omega}_k(y) + \hat{\omega}_i(z)) > \hat{\omega}_k^{-1}(\hat{\omega}_k(y) + \hat{\omega}_j(x) - \hat{\omega}_j(z)) = \Omega_{kj}(y,z,x)$, where the inequality follows from Lemma 2 and the fact that $\hat{\omega}_k(x)$ is a strictly increasing function of x for all k.
4. $\hat{\omega}_i(\Omega_{i,k}(y,z,x)) = \hat{\omega}_i(y) + \hat{\omega}_k(x) - \hat{\omega}_k(z)$ and $\hat{\omega}_j(\Omega_{i,k}(y,z,x)\hat{\omega}_k = \hat{\omega}_j(y) + \hat{\omega}_k(x) - \hat{\omega}_k(z)$. Thus, $\hat{\omega}_i(\Omega_{i,k}(y,z,x)) - \hat{\omega}_i(y) = \hat{\omega}_j(\Omega_{j,k}(y,z,x)) - \hat{\omega}_j(y)$. In view of result of Lemma 2, this is possible only if $\Omega_{i,k}(y,z,x) < \Omega_{j,k}(y,z,x)$.
5. This part is straightforward.
6. Differentiating with respect to z the expression for $C_{k,j}(y,z,x)$ from Theorem 1,

$$\frac{dC_{k,j}(y,z,x)}{dz} = \int_{u=y}^{\Omega_{k,j}(y,z,x)} \sum_{l=1}^{\mathcal{K}} \lambda_l \left. \frac{C_{l,j}(0,v,x)}{dv} \right|_{v=\Omega_{j,k}(z,y,u)} \frac{d\Omega_{j,k}(z,y,u)}{dz} \frac{F_k^c(u)}{F_k^c(y)}\, du$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

$$+\frac{d\Omega_{k,j}(y,z,x)}{dz}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,j}(0,\Omega_{j,k}(z,y,\Omega_{k,j}(y,z,x)),x)\right]$$
$$\times\frac{F_k^c(\Omega_{k,j}(y,z,x))}{F_k^c(y)}.$$

It is easily shown that $\Omega_{j,k}(z,\ y,\ \Omega_{k,j}(y,\ z,\ x))=x$ so that $C_{l,j}\ (0,\ \Omega_{j,k}\ (z,\ y,\ \Omega_{k,j}\ (y,\ z,\ x)),\ x)=0$. Thus, we get

$$\frac{dC_{k,j}(y,z,x)}{dz}=\int_{u=y}^{\Omega_{k,j}(y,z,x)}\sum_{l=1}^{\mathcal{K}}\lambda_l\frac{C_{l,j}(0,v,x)}{dv}\bigg|_{v=\Omega_{j,k}(z,\,y,\,u)}\frac{d\Omega_{j,k}(z,y,u)}{dz}\frac{F_k^c(u)}{F_k^c(y)}du$$
$$+\frac{d\Omega_{k,j}(y,z,x)}{dz}.$$

Now it can also be seen that $d\Omega_{k,j}(y,\ z,\ x)/dz\le 0$ and $d\Omega_{j,k}(z,\ y,\ u)/dz\ge 0$. Hence, repeated substitution of the above expression for $dC_{\cdot,\cdot}(\cdot,\ z,\ \cdot)/dz$ yields a solution for $dC_{\cdot,\cdot}(\cdot,z,\cdot)/dz$ that is non-negative. The result follows.

APPENDIX F
Proof of Theorem 4

1.

$$C_{k,i}(y,z,x)=\int_{u=y}^{\Omega_{k,i}(y,z,x)}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,i}(0,\Omega_{i,k}(z,y,u),x)\right]\frac{F_k^c(u)}{F_k^c(y)}\,du$$
$$C_{k,j}(y,z,x)=\int_{u=y}^{\Omega_{k,j}(y,z,x)}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,j}(0,\Omega_{j,k}(z,y,u),x)\right]\frac{F_k^c(u)}{F_k^c(y)}\,du.$$

From previous lemmas and results,

$$C_{k,i}(y,z,x)\ge\int_{u=y}^{\Omega_{k,i}(y,z,x)}\left[1+\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,i}(0,\Omega_{j,k}(z,y,u),x)\right]\frac{F_k^c(u)}{F_k^c(y)}\,du$$
$$\ge\int_{u=y}^{\Omega_{k,i}(y,z,x)}\frac{F_k^c(u)}{F_k^c(y)}du$$
$$+\int_{u=y}^{\Omega_{k,j}(y,z,x)}\sum_{l=1}^{\mathcal{K}}\lambda_l C_{l,i}(0,\Omega_{j,k}(z,y,u),x)\frac{F_k^c(u)}{F_k^c(y)}\,du.$$

Thus, if $\Delta_{k,i,j}(y,z,x) \triangleq C_{k,i}(y,z,x) - C_{k,j}(y,z,x)$, then

$$\Delta_{k,i,j}(y,z,x) > \int_{u=y}^{\Omega_{k,i}(y,z,x)} \frac{F_k^c(u)}{F_k^c(y)}\,du \\ + \int_{u=y}^{\Omega_{k,j}(y,z,x)} \sum_{l=1}^{\mathcal{K}} \lambda_l \Delta_{l,i,j}(0,\Omega_{j,k}(z,y,u),x)\frac{F_k^c(u)}{F_k^c(y)}\,du.$$

This equation can be solved for $\Delta_{k,i,j}(y,z,x)$ by the method of repeated substitution (as this is the Fredholm equation), and since the additive term is positive here, it follows that $\Delta_{k,i,j}(y,z,x) > 0$ for all k, x, y, and $z < x$. The proof is complete.

2.

$$T_i(x) = x + \int_{u=0}^{x} \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,i}(0,u,x)\,du + \int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} C_{k,i}(y,0,x)\,dN_k(y),$$

$$T_j(x) = x + \int_{u=0}^{x} \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,j}(0,u,x)\,du + \int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} C_{k,j}(y,0,x)\,dN_k(y).$$

Since $C_{l,i}(y,z,x) > C_{l,j}(y,z,x)$, the theorem follows.

3.

$$C_{l,j}(y,\Omega_{j,i}(z),\Omega_{j,i}(x)) = \int_{u=y}^{\Omega_{l,j}(y,\Omega_{j,i}(z),\Omega_{j,i}(x))} \\ \times \left[1 + \sum_{k} \lambda_l C_{k,j}(0,\Omega_{j,l}(\Omega_{j,i}(z),y,u),\Omega_{j,i}(x))\right] \\ \times \frac{F_l^c(u)}{F_l^c(y)}\,du$$

$$C_{l,i}(y,z,x) = \int_{u=y}^{\Omega_{l,i}(y,z,x)} \left[1 + \sum_{k=1}^{\mathcal{K}} \lambda_k C_{k,i}(0,\Omega_{i,l}(z,y,u),x)\right] \\ \times \frac{F_l^c(u)}{F_l^c(y)}\,du.$$

Now,

$$\Omega_{l,j}(y,\Omega_{j,i}(z),\Omega_{j,i}(x)) = \hat{\omega}_l^{-1}(\hat{\omega}_l(y) + \hat{\omega}_i(x) - \hat{\omega}_i(z)) = \Omega_{l,i}(y,z,x),$$

$$\Omega_{j,l}(\Omega_{j,i}(z),y,u) = \hat{\omega}_j^{-1}(\hat{\omega}_j(\Omega_{j,i}(z)) + \hat{\omega}_l(u) - \hat{\omega}_l(y)) \\ = \hat{\omega}_j^{-1}(\hat{\omega}_i(z) + \hat{\omega}_l(u) - \hat{\omega}_l(y)),$$

$$\Omega_{j,i}(\Omega_{i,l}(z,y,u)) = \hat{\omega}_j^{-1}(\hat{\omega}_i(\hat{\omega}_i^{-1}(\hat{\omega}_i(z) + \hat{\omega}_l(u) - \hat{\omega}_l(y)))) \\ = \Omega_{j,l}(\Omega_{j,i}(z),y,u).$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

Using these relations and subtracting the expression for $C_{l,j}(y, \Omega_{j,i}(z), \Omega_{j,i}(x))$ from that of $C_{l,i}(y, z, x)$ obtained above, we get Fredholm equation in the quantity $C_{l,j}(y, \Omega_{j,i}(z), \Omega_{j,i}(x)) - C_{l,i}(y, z, x)$ with zero additive constant. Repeated substitution then shows that $C_{l,j}(y, \Omega_{j,i}(z), \Omega_{j,i}(x)) - C_{l,i}(y, z, x) - 0$ is the only possible solution; that is, $C_{l,j}(y, \Omega_{j,i}(z), \Omega_{j,i}(x)) = C_{l,i}(y, z, x)$ for all l, x, y, and $z < x$.

Now, it is also easily shown that $\Omega_{j,i}(x) > x$. Also,

$$T_i(x) = x + \int_{u=0}^{x} \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,i}(0,u,x)\,du + \int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} C_{k,i}(y,0,x)\,dN_k(y),$$

$$\begin{aligned} T_j(\Omega_{j,i}(x)) &= \Omega_{j,i}(x) + \int_{u=0}^{\Omega_{j,i}(x)} \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,j}(0,u,\Omega_{j,i}(x))\,du \\ &\quad + \int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} C_{k,j}(y,0,\Omega_{j,i}(x))\,dN_k(y) \\ &= \Omega_{j,i}(x) + \int_{z=0}^{x} \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,j}(0,\Omega_{j,i}(z),\Omega_{j,i}(x))\,d\Omega_{j,i}(z) \\ &\quad + \int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} C_{k,j}(y,\Omega_{j,i}(0),\Omega_{j,i}(x))\,dN_k(y). \end{aligned}$$

Using the fact that $C_{l,j}(y, \Omega_{j,i}(z), \Omega_{j,i}(x)) = C_{l,i}(y, z, x)$ for all l, x, y, and $z < x$, the theorem follows since $d\Omega_{j,i}(x)/dx \geq 1$.

APPENDIX G
Proof of Theorem 5

From the proof of Theorem 4,

$$\begin{aligned} \frac{T_i(x)}{x} - \frac{T_j(\Omega_{j,i}(x))}{\Omega_{j,i}(x)} &= \int_{u=0}^{x} \sum_{l=1}^{\mathcal{K}} \lambda_l C_{l,i}(0,u,x) \left(\frac{1}{x} - \frac{\frac{d\Omega_{j,i}(u)}{du}}{\Omega_{j,i}(x)} \right) du \\ &\quad + \int_{y=0}^{\infty} \sum_{k=1}^{\mathcal{K}} C_{k,i}(y,0,x)\,dN_k(y) \left(\frac{1}{x} - \frac{1}{\Omega_{j,i}(x)} \right). \end{aligned}$$

Now,

$$\begin{aligned} \frac{1}{x} - \frac{d\Omega_{j,i}(u)/du}{\Omega_{j,i}(x)} &= \frac{1}{x} - \frac{\omega_j(\Omega_{j,i}(u))}{\omega_i(u)\Omega_{j,i}(x)} \\ &= \frac{1}{\omega_i(u)} (D_i(x,u) - D_j(\Omega_{j,i}(x), \Omega_{j,i}(u))) \geq 0. \end{aligned}$$

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press

The result follows using the additional fact that $\left(\frac{1}{x} - 1/\Omega_{j,i}(x)\right) \geq 0$.

APPENDIX H
Proof of Theorem 12

From Theorem 1,

$$\frac{dN_k(y)}{dy} = \lambda_k F_k^c(y) \frac{dT_k(y)}{dy}.$$

Now, the expected unfinished work in the system at any instant is $\bar{R}$ independent of the scheduling discipline. This quantity for age-based scheduling is

$$\overline{R} = \sum_k \int_{x=0}^{\infty} \int_{y=x}^{\infty} (y-x) \frac{dF_k(y)}{F_k^c(x)} \, dN_k(x),$$

where y is used to condition on the total service requirement of a customer that has an attained age of x. The theorem follows using expression for $dN_k(y)/dy$ and simple math.

https://doi.org/10.1017/S0269964808000144 Published online by Cambridge University Press